CONTENT

1. Important points
2. Authors' response to the Editor
3. Authors' response to Reviewer #1
4. Authors' response to Reviewer #2
5. Authors' response to Reviewer #3
6. Manuscript with tracked changes

IMPORTANT POINTS

1. All of the reviewers' comments are now addressed.
2. All of the editor's comments are now addressed.
3. Section 2 (Study Area, Data and Methodology), Section 3 (Results), and Section 4 (Discussion) are extensively rewritten in the revised manuscript to take care of the concerns expressed by the reviewers and the editor.
4. Two new figures (Fig. 4 and 5) are now added to the revised manuscript, one explaining the updated model and the other describing the methodology.
5. Previously uploaded review responses are updated according to the revised manuscript, as necessary.

RESPONSE TO THE EDITOR

Editor:

We have now received the reports of three referees and the responses of the authors including a revised manuscript. My general view is that the paper is technically sound and the contents are relevant for the audience of HESS. However, I found the paper hard to read (particularly the methodology part that is very hard to follow) as did reviewers 2 and 3.

Authors:

Dear Editor,

Thank you for your constructive comments, which have helped us to improve and revise our manuscript. We believe that the readability of the manuscript has been significantly improved as a result. Please also see our response to your specific comments in the following. Please also note that we have made some (minor) updates to our previously uploaded review responses according to the revised manuscript.

The review comments really helped us a lot in improving the manuscript and we thank the reviewers for that. Also, many thanks to the commenters for their constructive inputs.

Editor:

There are a number of issues that need consideration, including a better organization and synthesis and a more detailed explanation of methods and physical implications. These issues are listed in the comments by the reviewers, and though I realize that they have been partially addressed by the authors, the revised manuscript still needs further work in this regards.

Authors:

In the revised manuscript, we have made significant changes to the results and discussion sections. Please refer to the version with tracked changes to see the changes made.

Editor:

The study approach section needs considerable reworking. The dot point format is hard to follow and includes few details. Some of the material here should include explanations that are in the results section like, for example, how constrains are implemented. Other aspects that need more explanation are the calibration/optimisation procedure and the determination of parameters of the new formulation.

Authors:

We have now avoided the dot point format in our revised manuscript. We have also transferred a lot of contents from the results section to the methods section. Both sections are now extensively rewritten. We have also included some more details about the calibration/optimization procedures. There is a new figure (Fig. 5) added to the method section demonstrating our study approach.

Editor:

There are also too many subtitles, the authors should make an effort to provide a section with a better flow and easier to read.

Authors:

We have now removed most of the sub-titles and made the sections easier to read by maintaining a flow.

Editor:

There is also not enough discussion on the physical implications of the newly improved model HyMod V2. Figure 4 is a start but more can be said in the text, particularly in the results or discussion sections.

Authors:

We have now expanded the discussion on the newly improved model.

Editor:

Please note that, after the revision stage, the revised manuscript and author responses will be sent to the referees to complete the review process.

Authors:

Thank you for this information.

RESPONSE TO REVIEWER #1

Authors:

We thank the reviewer, who identified himself as Prof. Abdolreza Bahremand, for nicely summarizing the key aspects of our study and pointing out their importance. We have addressed all of his comments and made the suggested changes in our revised manuscript.

NOTE: Page and line numbers mentioned in the response correspond to the revised manuscript.

Reviewer:

The paper and its discussion and conclusion has much more other useful contents than what has been given in the abstract. I guess the limitation of the abstract word numbers (500 words) has been the reason for this. 2. One concluding sentence (like those written in the conclusion) should be added here.

Authors:

As per the reviewer's suggestion, we have now modified the last sentence of the abstract as in the following to summarize our main outlook:

"*Results indicate that while both approaches can provide improved simulations of streamflow, the second approach significantly improves the simulation of actual evapotranspiration, which substantiates the importance of making 'diagnostic structural improvements' to hydrologic models whenever possible*."

Reviewer:

GLEAM and HyMod could be other keywords for this paper? Don't you think so?

Authors:

We agree with the reviewer on this and have now included both GLEAM and HyMod as keywords.

Reviewer:

Here, in such case I am sure the authors know better than me that the strong correlation is not enough :)

Authors:

Yes, we agree that a strong correlation is not enough. Therefore, in our revised manuscript, we have now included a detailed discussion on the evaluation of GLEAM using several other error statistics. We are now citing one book chapter and four papers for this discussion.

<u>Page 2 Line 21 – Page 3 Line 4</u>

"… *Worldwide evaluations suggest that satellite-based ET estimates are strongly correlated (~0.83) with ground-based observations made at flux towers (Demaria and Serrat-Capdevila, 2016).*

For this study, we use the Global Land Evaporation Amsterdam Model (GLEAM) as the source of the satellite-based ET (SET) data. In the GLEAM algorithm, ET is computed using only a small number of satellite-based inputs, which makes it particularly beneficial for application to sparsely gauged basins. Miralles et al. (2011) have shown that GLEAM estimates of evaporation are strongly correlated (0.80) with annual cumulative evaporation estimated via eddy covariance at 43 stations, and have very low (-5%) average bias. The correlations at individual stations are strong (0.83) for all vegetation and climate conditions, and improve to 0.9 for monthly time series (Miralles et al., 2011). McCabe et al. (2016) reported satisfactory statistical performance (R2 = 0.68; Root Mean Square Difference = 64 Wm−2; Nash-Sutcliffe Efficiency = 0.62) of GLEAM when compared against data from 45 globally-distributed eddy-covariance stations. Michel et al. (2016) compared Priestley-Taylor Jet Propulsion Laboratory model (PT-JPL), Moderate Resolution Imaging Spectroradiometer evaporation product (PM-MOD), Surface Energy Balance System (SEBS), and GLEAM simulations against 22 FLUXNET tower-based flux observations and found GLEAM and PT-JPL to more closely match in-situ observations for the selected towers and reference period (2005-2007). Their extended analysis over 85 towers also had a similar overall outcome. Miralles et al. (2016) compared three process-based ET methods (PM-MOD, GLEAM and PT-JPL) against surface water balance from 837 globally distributed catchments, and reported that GLEAM and PT-JPL provide more realistic estimates of ET. They found these two products to provide superior overall performance for most ecosystem and climate regimes, whereas PM-MOD tends to underestimate the flux in tropics and subtropics."

Reviewer:

and whether the model provides improved ....[here, i proposed this little correction to make it in harmony with the previous sentences. So, i think adding the conjunction "whether", like what you did for the previous sentence, is better here.]

Authors:

Thanks for the suggestion. We have now modified the sentence as:

Page 3 Line 16-17

"*Finally, we test whether the use of GLEAM SET can further improve the performance of the structurally modified model, and whether there is any decline in model performance if GLEAM SET data become unavailable*."

Reviewer:

Unnecessary abbreviation makes the text a bit boring, as the text has already too many :) If you wish to reduce them then just start with the name of the rivers...

Authors:

We agree with the reviewer on this. We have now removed the unnecessary abbreviations from our revised manuscript.

Reviewer:

The abbreviation, HAET, has not been introduced in the paper so far. Perhaps you mean HyMod AET. First I thought that H stands for Hargreaves...

Authors:

Thanks for pointing this out. We have now introduced HAET in Section 2.3.

Reviewer:

after the remove of the GAET data.

Authors:

We have now modified this part of the sentence as:

"… *after the removal of the GAET data*."

Reviewer:

This sentence needs a little bit of improvement (grammar correction).

Authors:

Thanks for pointing this out. We changed the sentence as in the following:

“Therefore, our results suggest that ET constraining approach should be implemented only for the simulation periods when SET data are available.”

However, this statement seemed redundant once the constraining results were already discussed. Therefore, we deleted this from the revised manuscript.

RESPONSE TO REVIEWER #2

Authors:

We thank the reviewer for reviewing our manuscript and providing his/her valuable feedbacks. We have now addressed all of his/her comments and discussed them in the following. As the reviewer mentioned, there were some places in the manuscript which created confusions and the concepts seemed circular. We agree with the reviewer on that. These were mainly due to the lack of sufficient care in the use of terminology. We have revised the manuscript to resolve these issues and make our message more clear-cut. Thanks to the reviewer’s feedback, the paper is now much improved.

Reviewer:

This paper presents results from a study examining the use of satellite estimates of actual evapotranspiration (SET) to firstly constrain and secondly modify a HyMod model of Nyangores River Basin in Kenya. Although the ideas presented here are interesting, I found that the reasoning used in the study was circular and I’m not convinced by the results. I think the presentation of the material is too much like a report and the method and results are often mixed up, with the vast majority of the method discussion provided in Section 3 which is nominally the results section. The paper also refers to another publication in preparation by the same authors on this catchment and without seeing this it is difficult to understand the similarity and any potential overlaps between the two publications. It’s not clear why this paper would be presented first. I recommend that the paper is rejected and the authors undertake more extensive validation of the method in a catchment where there is data other than the SET to allow comparisons.

Authors:

The manuscript is designed such that all the analyses steps are clearly stated and their results are thoroughly discussed. This is important since we recommend this approach for similar investigations, due to the fact that it’s inclusive. It takes into account several important issues, including process constraining, use of constraint adjustment, usefulness of model (re)calibration, utilizing new information (from satellite-based sources), diagnostic model structural improvement, and uncertainty analysis. However, as pointed out by the reviewer, we do see that some method discussions could be removed from the results section and put back to the methods section itself. We have now made significant changes in both of these sections to take care of this issue in our revised manuscript.

Regarding the point on the second publication, we do have another manuscript under review, however, we would like to clarify that the objective and scope of that manuscript are quite different as compared to this one. That manuscript reports on the development of a multi-model and multi-product (satellite)-based probabilistic operational streamflow forecasting platform for sparsely-gauged basins and does not in any way address the problem of model structural correction/improvement. We are ready to share the manuscript with the reviewer and the editor personally if necessary to resolve this concern.

Since the other manuscript is under review, we are not citing that anymore in this manuscript.

Regarding the comment on other available data for comparison, note that the dataset (GLEAM) we are using has already been validated in several recent studies. Although we didn't include the detailed discussion on validation in our initial manuscript, we have now included that part in our revised manuscript (Page 2 Line 21 - Page 3 Line 4). GLEAM has already been evaluated both at local (eddy covariance towers) and global scales. There have been projects that have focused on the topic of the evaluation of GLEAM, e.g. WAter Cycle Multi-mission Observation Strategy-EvapoTranspiration (WACMOS-ET), Global Energy and Water Cycle Exchanges (GEWEX) LandFlux Project, etc.

Several studies (also cited in our manuscript) have found GLEAM to be one of the best ET products. Therefore, we don't think it is necessary to carry out an additional evaluation of GLEAM, given the fact that other studies have already focused on that part. This also does not fit well with the main goals of this manuscript. Moreover, an evaluation study of this kind would stand out on its own as an independent paper, which is clearly beyond the scope of this manuscript.

To be clear, the main objective of this study is NOT to validate/compare actual ET products, which is an interesting topic, but appropriate for a different manuscript. In this study, we explore different structure-related methods (including process constraining) to improve the performance of a rainfall-runoff model, and we have an inclusive design to organize all the steps in a systematic manner. We show how the model deficiencies could be overcome by using new sources of information

Reviewer:

If I understand the method properly, in Case 1 HyMod is run and the AET from the model is found to be different from the SET estimates. So the model is run using SET to constrain the AET in the model by setting the requirement that the AET <= SET. However then the model parameters are found to be unrealistic so the SET is bias corrected so that when the model is constrained to have AET <= SET, the model parameters are more realistic. In all of this there is no evaluation of the SET itself

and the bias correction step implies that there are problems with the SET. So you're trying to match a model to a biased quantity and then changing that quantity and then still trying to match it. It just seems very circular to me. Case 2 follows much the same logic except rather that using the constraint that AET <= SET, the model structure is changed with a variety of different equations that factor the evaporative demand ratio. Finally in Figure 9 the model is compared back to the SET which was used to correct the model I just don't understand how you can accept the SET data without having an external validation. I accept that this is unlikely to exist for the catchment you have chosen but I think you then need to test your method in a more instrumented catchment where you do have external validation data and once you have confidence in the method then you can apply to a poorly gauged basin.

Authors:

This is an important point which we unfortunately did not explain well in the original manuscript. We thank the reviewer for pointing this out. Note that we are not doing any bias-correction of GLEAM in this study, because for that we needed the 'ground truth' of the observed ET. Furthermore, GLEAM has been validated in several recent studies (already cited in our manuscript). Therefore, the term 'bias-correction' was wrongly used and we have now changed that. A more appropriate term in this case would be 'constraint adjustment', which is what we are essentially doing. In Stage-I, the model structure is fixed. When GAET is used as a constraint to the ET process within the model, it introduces bias in the streamflows. Therefore, we adjust the constraint such that that bias is removed. Note that this is NOT indicative of the presence of any actual bias within the GAET estimates. The constraint adjustment factor is a model 'parameter' which corresponds to the structural deficiencies within the model. It may or may not be necessary as the structure changes. In Stage-II analysis, we saw that when the structure was improved (deficiencies reduced), ET constraint adjustment became irrelevant.

Regarding the point of external validation, please see the last three paragraphs of our first response (to Reviewer #2).

Reviewer:

Page 2 – paragraph 3 – at this stage its not clear how ET can be a model target – I think you need to make it clearer at this point that PET is forcing data and AET is a model state.

Authors:

We consider precipitation and PET as the forcings. Note that the precipitation is the only input to the water budget of the model, PET is a constraint to set the upper limit of the actual ET in the original HyMod model. The model produces both discharge and actual ET as outputs. Therefore, we don't see why AET needs to be considered as a model state (as conventionally defined). It is a model simulated output.

Reviewer:

Page 2, line 15 – good correlation of the SET does not give me confidence that the property is not biased which is key for this method and even line 23 where the annual bias is low doesn't guarantee that there are not other biases that are cancelling out throughout the year.

Authors:

We agree with the reviewer on the point of the value of correlation. Actually we had a very brief discussion on the comparison/validation of the ET products in our initial manuscript. We have now expanded that discussion in our revised manuscript (Page 2 Line 21 – Page 3 Line 4), where some additional error statistics (apart from correlation coefficient) are also reported.

Note this was also pointed out by Reviewer #1 and in reply to his comments we showed the new paragraph that has been added to the revised manuscript (Page 4 of this document).

Reviewer:

Page 4 – paragraph 12 – TRMM data is no longer available so not clear why you say that it is available to near-present? The study period is not clear from Section 2 in any case.

Authors:

This is an important point. We have now included this information in our revised manuscript. We are using the TRMM Multi-Satellite Precipitation Analysis (TMPA-RT) dataset which is still available. This is a merged dataset. TRMM Microwave Imager (TMI) was a part of it, which is no more operational (since 8 April 2015) because of fuel and battery issues with the satellite. As mentioned by the developers, the absence of TRMM is not crucial to the production of TMPA and TMPA-RT data.

We discuss the time periods in Section 2.5:

"*The model was run continuously for the 7.5-year period Jan 2003 to June 2010, with the first 4 years (2003 to 2006) used for calibration and the remaining 3.5 years (2007 to mid-2010) used to provide an additional assessment of model performance. Results are shown for the "calibration (4-years)", "evaluation (3.5 years)" and "total (7.5 years)" simulation periods*."

Reviewer:

Page 4, line 34 – here you describe Stage 1 as "constraining" and you are at pains to point out that it is not assimilation and yet in the remainder of the manuscript you continue to use the term assimilation – I think you need to be more careful with the terminology e.g. Page 8, line 23; Page 12, Line 24

Authors:

Thank you for pointing this out. We have now removed the term 'assimilation' wherever required.

Reviewer:

Page 5, Step 1-2 – given this is the method section, there are no details here of the actual constraints. These are provided in the results section. I think this makes the presentation quite confused and doesn't provide the reader with much of a sign post or guide as where the research is heading. Similar comments for Step II-1 where the four equations are mentioned.

Authors:

We agree with the reviewer on this, and we want to point out that we have now made significant changes in the methods, results, and discussion sections in our revised manuscript to take care of this issue.

Reviewer:

Page 7, Line 24 – I don't understand why you validate your water balance using satellite precipitation which has its own concerns. Why not use some ground based data as well?

Authors:

Note that the TMPA data used in this study has been bias corrected using rain gauge measurements from the study area. The detailed methodology is discussed in the other manuscript (streamflow forecasting). As mentioned earlier, we are ready to share the manuscript personally with the reviewer or the editor.

Reviewer:

Page 7, Line 27 – "based on our expectation of how it would behave" – this comes to my concern about the validation. We generally expect a more robust validation than just a sense that the soil moisture should be smooth. Why should it be smooth for this catchment? You don't appear to have any soil moisture data to validate this statement.

Authors:

Thanks for pointing this out. We agree that in order to make this statement, the soil moisture data need to be studied first. Therefore, we have now removed this sentence from our revised manuscript.

RESPONSE TO REVIEWER #3

Reviewer:

The paper deals with the use of satellite-based evapotranspiration estimates (GAET) to improve results of a simple hydrological model. The general idea of the paper is sound and potentially useful for the hydrological community.

Unfortunately, I see a number of problems with the paper. The main problem for me is the unclear rationale of the methodology. GAET is used in two ways to improve the hydrological model, and the two procedures have problems.

Authors:

We thank the reviewer for his/her valuable comments and acknowledging the importance and relevance of our paper by stating that 'the general idea of the paper is sound and potentially useful for the hydrological community'. We have now thoroughly addressed all of his/her comments and concerns in our response in the following.

Reviewer:

The first procedure "constrains" the hydrological model estimates of evapotranspiration HAET forcing them to be more similar to HAET. This is done in a very prescriptive way, and to some extent may contradict the whole physical basis of the model. The results of this exercise are not successful, as shown by the poor performance of the model in terms of streamflow. There are other ways of constraining intermediate model results, which are more formal and do not compromise the model physics (for example calibration optimization with side constrains). I believe that the first procedure does not present any novelty in terms of ideas or techniques. A thorough justification of why it should be included in the paper based on similar procedures applied successfully elsewhere is needed here.

Authors:

The reviewer expressed two main concerns in this paragraph. First, the way the method is implemented and second, the performance improvements. We fail to agree with the reviewer on either of them.

Regarding the first point, our constraining scheme is conceptually analogous to any filtering technique, where the main goal is to fix the behavior of the model, not its structure/process parameterization. In filtering, the model state at any time step is adjusted based on the observation from that time step so that the model behaves 'more accurately'. A filtering cannot and is not meant to directly correct the model structure. Likewise, in our constraining approach, we try to fix the model behavior without modifying its structure. We modify the structure diagnostically in the next step (Stage-II).

The constraining approach corrects the model behavior in a physically-consistent manner (using new information from the satellite-based actual ET, GLEAM), which is exactly what we want. The water balance, as expected, is also preserved. Therefore, we don't agree that the constraining approach contradicts the physical basis of the model. To our opinion, it actually corrects the model behavior.

Regarding 'calibration optimization with side constraints', note that we are already performing calibration (using SCE-UA which is a global optimization algorithm) using two different types of constraints, one on the parameters (their ranges) and the other on the ET process. This should result into a more physically-consistent model and not 'contradict the physical basis of the model'.

We think that the constraining is an important part of the paper and should remain in it.

Regarding the second point, we did have performance improvement in constraining, although not as much as the second approach, where we changed the model structure. In Table 5 (revised manuscript) it can be seen that in many cases the error statistics improve from Step-1 to Step-4. In calibration, NMSE changes from 0.56 to 0.43, NBσ changes from -0.12 to -0.06, and R changes from 0.76 to 0.83. In Figure 10 (revised manuscript), we do see improvements in the streamflow simulations (compare the blue and green lines in the transformed space to see the improvements more clearly).

Reviewer:

The second procedure modifies the structure of the model by multiplying the ET equation by a factor. Different formulations are used for the factor, which try to capture more of the physics of the problem. This last point is not clearly explained or justified by the authors. The formulations are tested against GAET estimates and the more complex formulation gives the best results. That formulation is then used to predict discharges, which shows some improvement of the model results. A major problem with the procedure is that the model produces a value of the soil moisture storage capacity H that is totally unrealistic (H=12.8 m). The authors do not report

the value of H for the original model without the "improvements" using GAET, but my impression is that it may have been more physically adequate. I think that the author's claim about the advantage of physically-based over data-driven models is weakened by this outcome.

Authors:

We thank the reviewer for helping us improve our explanations. As suggested, we have now explained the idea (i.e. the formulations 'try to capture more of the physics of the problem') more clearly in our revised manuscript.

Please note that there has been a misunderstanding here. The reviewer pointed out an H value of 12.8 m, however, our final model DID NOT produce an H value of 12.8 m. Please refer to Appendix C where we provided a table with all the calibrated parameters. The column with the final model (Stage-II Step-1 Case-D) shows an H value of 866 mm or 0.866 m, which is conceptually realistic. The value the reviewer mentioned is from another step (Stage-II Step-3) which is not the final step. Please note that the H values from all different models (including the original one) are already reported in the table in Appendix C.

We understand that one of the reasons this confusion arose was because our methods and results sections were not written properly in the initial manuscript. In our revised manuscript, we have made significant changes in these sections to present the content more clearly. Hopefully this will guide the readers much better.

Following is a paragraph from the revised manuscript explaining our methodology:

"*The entire study approach is summarized in Fig. 5. As can be seen, only Step-1 is different for both the stages (Stage-I and Stage-II), while the remaining four steps (Step 2-5) are similar. Thus, in each stage, there are five steps altogether. Stage-I Step-1 is for generating benchmark simulations using the calibrated model but without any ET constraint or structural modifications. On the other hand, Stage-II Step-I has four different cases (A-D) corresponding to different structural modifications in the ET process parameterization. Both the benchmark model from Stage-I Step-I and the best performing model from Stage-II Step-I are used in the following steps. Step-2 is based on imposing the ET constraint but without recalibration, meaning that the same set of calibrated parameters as in the benchmark step is further used in this step. Step-3 is based on recalibrating the model while imposing the ET constraint. Step-4 is conceptually similar to Step-3, however, additionally, some constraint adjustments (Eq. 1 and 2) are applied and the adjustment parameters are calibrated together with model parameters (to match the simulated and observed streamflows). Finally, in Step-5, we remove the ET constraint to see whether the performance of the new model will decline when satellite ET data becomes unavailable (note this is no longer the benchmark model since we recalibrated the parameters in Step 4)*."

Reviewer:

In the middle of all this there are a number of methodological details that are also of concern. For example, model calibration is done using the SCE-UA algorithm, which essentially consists of a global optimization method. Since the formulation of the second procedure involves more calibration parameters, how does that affect the optimization?

Authors:

All the calibration runs were carried out with the same settings of SCE-UA, i.e. with the same number of complex and loops, in order to nullify the effects of the optimization algorithm itself. The calibration runs were successful in all the cases. We have already reported the calibrated parameters from all different calibration runs in the appendix.

Reviewer:

Also, there are ways of optimizing parameters with constrains that could be explored as a more formal way of incorporating the additional information from the GAET.

Authors:

We have already addressed the issue of calibration with constraints in one of our previous paragraphs (response to Reviewer #3). The constraint could be on the parameters or it could be on the processes. We are applying constraints on the parameters by setting their limits. We are also imposing constraints on the ET process within the model using the satellite ET data.

Regarding constraining with ET, we found two very good papers (Winsemius et al., 2008, van Emmerik et al., 2015, already cited in our paper) where the authors constrain the model parameters sensitive to ET using the ET data. Note that in this study our approach (and goal) is different. We impose the constraint on the ET process and also modify the model structure.

Reviewer:

Organization is also an issue. There is material in the results that should be in the methods (for example most of 3.1.2. in the results is about how to implement the "constrain" in the model and should be moved to 2.4. study approach). There is also an excessive use of subtitles and dot point type paragraph, which results in a lack of flow throughout the paper.

Authors:

We thank the reviewer for pointing this out. This was also pointed out by Reviewer #2. We have now made several changes in the sections dealing with methodology, results, and discussion, to take care of this issue.

Reviewer:

One lingering question that I have after reading the paper is why this new methodology was used in a study case with limited data and not on a catchment with extensive data where more verification and checks could be done. After all, the essence of the paper to me is the new formulation to improve an existing hydrological model and from that point of view a better set of data for validation is necessary. I would also add that the application to just one catchment may not be enough to demonstrate that the new formulation is better.

Authors:

Please note that one of the main purposes of this study is to develop models for sparsely-gauged basins. That is why we are using the satellite-based actual ET data in the first place. In a well instrumented catchment we could instead use flux tower data directly. A method/model that has worked well in a highly-instrumented catchment doesn't necessarily guarantee that it will also work well in a sparsely-gauged catchment.

The main focus of our project (NASA SERVIR) has been in solving water resources problems in sparsely-gauged basins using observations from space, and our current study is well aligned with the objective and the scope of the project.

We understand the usefulness of testing a model/method in multiple catchments (see the paper Gupta et al. (2014, HESS) by one of the coauthors of this paper), however, doing that was beyond the scope of this study. A rigorous testing is one of our future plans.

Please refer to the last paragraph in 'Discussion' where we point out this issue:

"*Note that this study is based on testing the model on a single basin using a single satellite-based AET product. While not demonstrating universal applicability, the results are clearly indicative and the methodology illustrates how such data can be used to investigate potential improvements to the structures of simple catchment scale models used for hydrologic studies in data scarce regions. A rigorous analyses of the methodology over multiple basins is a potential avenue for future research scope. For more detailed process-based models, the ET process parameters can be calibrated against some reliable satellite-based AET estimates (e.g. GLEAM), or the process representation itself can be improved by adapting some similar strategies that these AET products follow*."

References

van Emmerik, T., Mulder, G., Eilander, D., Piet, M. and Savenije, H.: Predicting the ungauged basin: model validation and realism assessment, Front. Earth Sci., 3(October), 1–11, doi:10.3389/feart.2015.00062, 2015.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M. and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, Hydrol. Earth Syst. Sci., 18(2), 463–477, doi:10.5194/hess-18-463-2014, 2014.

Winsemius, H. C., Savenije, H. H. G. and Bastiaanssen, W. G. M.: Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins?, Hydrol. Earth Syst. Sci., 5(4), 2293–2318, doi:10.5194/hessd-5-2293-2008, 2008.

MANUSCRIPT WITH TRACKED CHANGES

See next page.

Using Satellite-Based Evapotranspiration Estimates to Improve the Structure of a Simple Conceptual Rainfall-Runoff Model

Tirthankar Roy[1], Hoshin V. Gupta[1], Aleix Serrat-Capdevila[1,2] and Juan B. Valdes[1]

[1] Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, AZ, USA
[2] Water Global Practice, The World Bank, Washington DC, USA

Correspondence to: Tirthankar Roy (royt@email.arizona.edu)

Abstract. Daily, quasi-global (50°N-S and 180°W-E), satellite-based estimates of actual evapotranspiration at 0.25° spatial resolution have recently become available, generated by the Global Land Evaporation Amsterdam Model (GLEAM). We investigate the use of these data to improve the performance of a simple lumped catchment~~-~~scale hydrologic model driven by satellite-based precipitation estimates to generate streamflow simulations for a poorly gauged basin in Africa. In one approach, we use GLEAM to constrain the evapotranspiration estimates generated by the model, thereby modifying ~~the~~ daily water balance and improving model performance. In an alternative approach, we instead change the structure of the model to improve its ability to simulate actual evapotranspiration (as estimated by GLEAM). Finally, we test whether the GLEAM product is able to further improve the performance of the structurally modified model. ~~The results suggest~~Results indicate that ~~the modified model~~while both approaches can provide improved simulations of ~~both~~ streamflow~~ and~~, the second approach significantly improves the simulation of actual evapotranspiration, ~~even if GLEAM satellite-based evapotranspiration data are not available.~~which substantiates the importance of making '*diagnostic structural improvements*' to hydrologic models whenever possible.

Keywords. Satellite-based actual evapotranspiration, Global Land Evaporation Amsterdam Model (GLEAM), satellite-based precipitation, streamflow simulation, catchment~~-~~scale modeling, HyMod, poorly gauged basins, diagnostic model structural improvement, reduction of epistemic uncertainty.

1 Introduction

1.1 ~~1.1~~ Statement of the Problem

~~[1]~~ As a primary mechanism in the surface-to-atmosphere portion of the water cycle, evapotranspiration (ET) plays a crucial role in the water and energy budgets of a hydrologic system. In practice, ET can be estimated either from model simulations or from remotely sensed observations. For example, ET can be estimated as a residual of water balance computations, or via a land-surface energy budget (e.g. Monteith, 1965; Priestley and Taylor, 1972), and simple empirical physically based schemes ~~(Hargreaves and Samani, 1985)~~(Hargreaves and Samani, 1985) can be applied in data-scarce regions. Ultimately, the quality of a model-derived estimate of ET depends on the various sources of uncertainty (inputs, parameters, process representation, structure, etc.) inherent to the model-based scheme used, and common problems include both over- and under-estimation of evaporative fluxes (Trambauer et al., 2014). Recently, methods that use satellite-based remotely sensed climatic and environmental observations provide an alternate approach to the estimation of ET and its different components (e.g. Bastiaanssen et al., 1998; Arboleda et al., 2005).

Formatted: Font: 14 pt
Formatted: Left: 0.45", Right: 0.25", Bottom: 0.17"
Formatted: Font: 14 pt
Formatted: Centered, Space Before: 0 pt
Formatted: Font: 11 pt
Formatted: Centered
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"
Formatted: Font: 11 pt
Formatted: Indent: First line: 0.5", No bullets or numbering
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

~~[2]~~ Several studies have advocated and/or implemented the idea of using physically consistent estimates for the parameters of hydrologic models (Pokhrel et al., 2008, 2012; Savenije, 2010; Schaefli et al., 2011; Kumar et al., 2013; Troch et al., 2015 and references therein). However, in catchment-scale modeling, it is common practice to use parameter estimates that are calibrated by adjusting the simulated streamflows to try and match observed data. If due care is not implemented during the calibration strategy, this approach can result in conceptually unrealistic estimates for the parameters. Such a result defeats an important purpose of using conceptual/physically based models (as opposed to empirical data-based models), which is to help us better understand the dynamical behavior of the system.

~~[3]~~ In principle, the potential of such models can be better realized by incorporating more information about the physical system during model development. Such information can take various forms and be incorporated in different ways. Evapotranspiration (ET) can be used to constrain model parameters that are sensitive to the ET process (Winsemius et al., 2008; van Emmerik et al., 2015). Alternatively, ET can be used as a calibration target along with streamflow within a multi-objective setting (Zhang et al., 2009). There has also been a recent drive towards structurally flexible models that are able to both better characterize the uncertainty associated with model structure and use additional information to help reduce such uncertainty ~~(Wagener et al., 2001; Marshall et al., 2006; Clark et al., 2008; Savenije, 2010; Schaefli et al., 2011; Fenicia et al., 2008a, 2008b, 2011; Bulygina and Gupta, 2009, 2010, 2011; Martinez and Gupta, 2011; Nearing, 2013; Nearing and Gupta, 2015; Clark et al., 2015)~~(Wagener et al., 2001; Marshall et al., 2006; Clark et al., 2008; Savenije, 2010; Schaefli et al., 2011; Fenicia et al., 2008a, 2008b, 2011; Bulygina and Gupta, 2009, 2010, 2011; Martinez and Gupta, 2011; Nearing, 2013; Nearing and Gupta, 2015; Clark et al., 2015).

~~[4]~~ A variety of satellite-based remotely sensed estimates of daily precipitation have been available for some time (e.g. Hsu et al., 1997; Joyce et al., 2004; Huffman et al., 2007; Funk et al., 2014), making it possible to consider the model-based generation of streamflow simulations for ungaged locations. Recently, satellite-based remotely sensed estimates of daily ET have become available, based on a variety of different retrieval algorithms of varying complexity (e.g. Bastiaanssen et al., 1998; Arboleda et al., 2005; Miralles et al., 2011a). ~~Worldwide evaluations suggest that satellite-based ET estimates are strongly correlated (~0.83) with ground-based observations made at flux towers (Miralles et al., 2015; García et al., 2016). Miralles et al. (2015) compared three process-based ET methods (the Moderate Resolution Imaging Spectroradiometer evaporation product PM-MOD, the Global Land Evaporation Amsterdam Model evaporation product GLEAM, and the Priestley-Taylor Jet Propulsion Laboratory model PT-JPL) against surface water balance from 837 globally distributed catchments, and reported that GLEAM and PT-JPL provide more realistic estimates of ET. They found these two products to provide superior overall performance for most ecosystem and climate regimes, while PM-MOD tends to underestimate the flux in tropics and subtropics.~~Worldwide evaluations suggest that satellite-based ET estimates are strongly correlated (~0.83) with ground-based observations made at flux towers (Demaria and Serrat-Capdevila, 2016).

~~[5]~~ ~~Of relevance to~~ For this study~~ is that~~, we use the Global Land Evaporation Amsterdam Model (GLEAM~~ product~~) as the source of the satellite-based ET (SET) data. In the GLEAM algorithm, ET is computed using only a small number of satellite-based inputs, which makes it particularly beneficial for application to sparsely gauged basins. Miralles et al. (2011) have shown that GLEAM estimates of evaporation are strongly correlated (0.80) with annual cumulative evaporation estimated via eddy covariance at 43 stations, and have very low (-5%) average bias. The correlations at individual stations are strong

(0.83) for all vegetation and climate conditions, and improve to 0.9 for monthly time series (Miralles et al., 2011). McCabe et al. (2016) reported satisfactory statistical performance (R^2 = 0.68; Root Mean Square Difference = 64 Wm^{-2}; Nash-Sutcliffe Efficiency = 0.62) of GLEAM when compared against data from 45 globally-distributed eddy-covariance stations. Michel et al. (2016) compared Priestley-Taylor Jet Propulsion Laboratory model (PT-JPL), Moderate Resolution Imaging Spectroradiometer evaporation product (PM-MOD), Surface Energy Balance System (SEBS), and GLEAM simulations against 22 FLUXNET tower-based flux observations and found GLEAM and PT-JPL to more closely match in-situ observations for the selected towers and reference period (2005-2007). Their extended analysis over 85 towers also had a similar overall outcome. Miralles et al. (2016) compared three process-based ET methods (PM-MOD, GLEAM and PT-JPL) against surface water balance from 837 globally distributed catchments, and reported that GLEAM and PT-JPL provide more realistic estimates of ET. They found these two products to provide superior overall performance for most ecosystem and climate regimes, whereas PM-MOD tends to underestimate the flux in tropics and subtropics.

~~[6]~~ ~~These reports suggest that satellite-based ET (SET) estimates have the potential to be useful for hydrologic applications.~~ While previous studies have used SET estimates to *constrain the parameters* of hydrologic models (Winsemius et al., 2008; van Emmerik et al., 2015), the recent interest in diagnostic improvements to model structure ~~(Gupta et al., 2008, 2012; Gupta and Nearing, 2014)~~(Gupta et al., 2008, 2012; Gupta and Nearing, 2014) suggests that it would be potentially more valuable to use the ET data to actually *improve the model structure* when possible. This study attempts to explore this possibility in the context of using satellite-based data to drive a streamflow simulation model for a poorly gauged basin in Africa.

1.2 Objectives and Scope

~~[7]~~ In this study, we explore the use of the GLEAM daily SET product (Miralles et al., 2011; Martens et al., 2016) to improve the performance of a simple lumped catchment scale hydrologic model driven by satellite-based precipitation estimates to generate streamflow simulations for a poorly gauged basin in Africa. We first use the GLEAM product to constrain the evapotranspiration estimates generated by the model, thereby improving the daily water balance. Next, we instead change the structure of the model to make it more physically consistent and improve its ability to simulate actual evapotranspiration (as estimated by GLEAM). Finally, we test whether the use of ~~the~~ GLEAM ~~product~~SET can further improve the performance of the structurally modified model, ~~to see if further improvements are achievable. The modified~~and whether there is any decline in model ~~provides improved simulations of both streamflow and evapotranspiration, even when~~ performance if GLEAM-~~satellite-based evapotranspiration estimates are no longer available.~~ SET data become unavailable.

2 Study Area, Data and Methodology

2.1 ~~2.1~~ Study Area

~~[8]~~ This study is carried out for the Nyangores River ~~Basin (NRB),~~basin, which is a sub-basin of the Mara River basin flowing through Kenya (Fig. 1). ~~NRB~~The Nyangores River basin has an aerial coverage of 697 km^2 and is located at the northeastern side of the Mara River Basin (~~MRB;~~ Location: 33°88'E 35°90'E 0°28'S 1°97'S). The perennial Nyangores River originates from the Mau Escarpment (3000 m ASL) fault scarp passing through the western side of the Great Rift Valley in Kenya. It then merges with the Amala River at the Napuiyapi swamp (2932 m ASL) to form the Mara River, which flows all

Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Indent: First line: 0.5", No bullets or numbering
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"
Formatted: Font: 11 pt
Formatted: Normal, Indent: First line: 0.5", No bullets or numbering
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

the way to Lake Victoria at Musoma Bay, Tanzania (1130 m ASL). ~~MRB (or NRB~~Mara River basin (or Nyangores River basin) has two wet seasons consequent to the yearly oscillations of the inter-tropical convergence zone (ITCZ), the primary wet season occurring during March to May ~~(MAM)~~ and the secondary during October to December ~~(OND).~~. The long-term mean rainfall in the Mau Escarpment is around 1500 mm. The rainfall in the basin is influenced by factors like topography, elevation gradient, regional influence of Lake Victoria, sea-surface temperature (SST) of the Indian Ocean, etc.

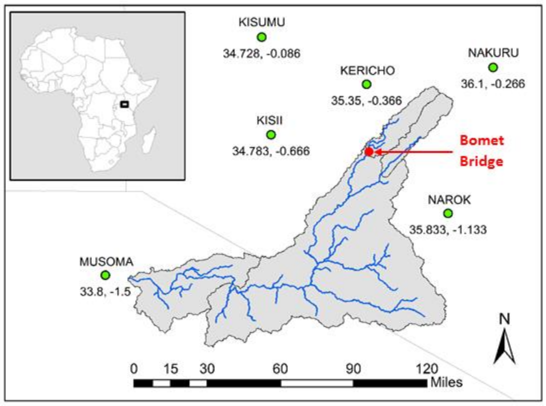


Fig. 1. The Mara River basin and the Nyangores River sub-basin. The discharge station is located at Bomet Bridge (red dot).

~~**[Insert Fig 1]**~~

~~2.2~~ Meteorological stations (green dots) are located in the surrounding regions.

2.2 Data

2.2.1 Estimates of Actual Evapotranspiration

~~*[9]*~~ The source of the SET data used in this study is the Global Land Evaporation Amsterdam Model (GLEAM) Version 3.0. GLEAM comprises a set of algorithms that use remotely sensed climatic and environmental observations to estimate various components of ET. Satellite-based observations of surface net radiation and near-surface air temperature are processed via the Priestley-Taylor Equation (Priestley and Taylor, 1972) to calculate Potential Evapotranspiration (PET), which is then converted to Actual Evapotranspiration (AET) by incorporating an evaporative stress factor obtained from microwave observations of vegetation optical depth (as a proxy for vegetation water content) and root-zone soil moisture (simulations). Interception loss is calculated using the Gash analytical model (Gash, 1979).

~~*[10]*~~ Three different versions of the GLEAM datasets are currently available, depending on the satellite observations used. The version used in this study (GLEAM_v3.0b) is based solely on satellite observations, is quasi-global (50°N-S, 180°W-E), has a spatial resolution of 0.25°, and has a daily temporal coverage of 13 years (2003 to 2015).

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Formatted: Font: 11 pt

Formatted: Normal, No bullets or numbering

Formatted: Font: 11 pt

~~[11]~~ ~~Figure~~ Fig. 2 shows the annual mean of GLEAM AET (GAET) over the entire ~~MRB~~Mara River basin. As can be seen, GAET increases towards the western side of the basin. The annual average GAET varies between 900 mm/year to 1200 mm/year. We computed corresponding estimates of PET via the Hargreaves Equation (HPET) using temperature data collected from the six met stations surrounding the ~~MRB~~Mara River basin (Fig. 1); maximum and minimum temperatures were averaged to obtain time series of maximum and minimum temperatures and processed through the Hargreaves Equation to calculate PET. For a small number (~0.6%) of days, the lumped GAET values were found to be larger than the lumped HPET values; for these few anomalous values, HPET was replaced by GAET. ~~Figure~~Fig. 3 shows the time series of HPET and GAET for ~~NRB~~the Nyangores River basin.

~~[Insert Fig 2]~~

~~[Insert Fig 3]~~

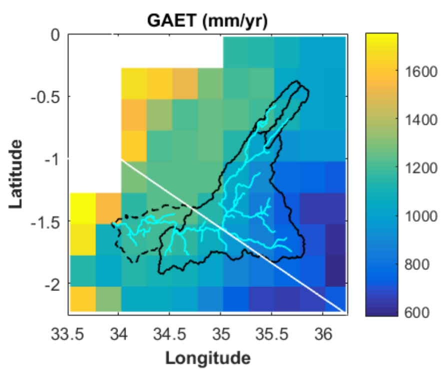


Fig. 2. Annual mean of GAET over the entire Mara River basin.

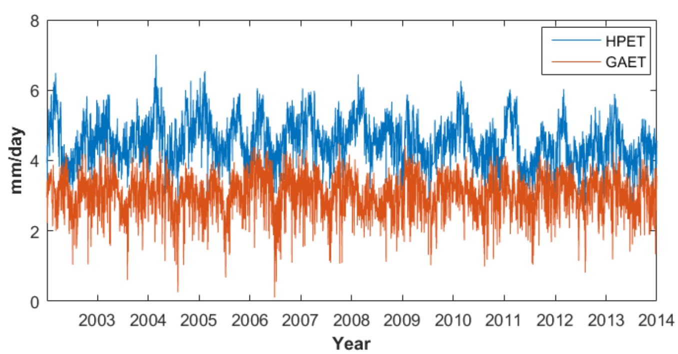


Fig. 3. Time series of HPET and GAET for the Nyangores River basin.

Formatted: Font: 11 pt

Formatted: Space After: 0 pt, No bullets or numbering

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

2.2.2 Estimates of Precipitation

~~*[12]*~~ The Real Time Multi-satellite Precipitation Analysis ~~(TMPA-RT)~~ of the NASA Tropical Rainfall Measuring Mission (TMPA-RT) combines information from multiple satellites to produce a quasi-global (50°N-S, 180°W-E), near-real-time (March 1, 2000 to near-present) precipitation product at 0.25° × 0.25° spatial and 3-hourly temporal resolution (this product is the real-time version of TMPA (Huffman et al., 2007)). ~~Launched~~Until it was shut down on 8 April 2015 due to fuel deficiency and battery issues in ~~1997, TRMM is~~the satellite, TMPA used ~~primarily for research and development, being~~to include the TRMM Microwave Imager (TMI) products. TRMM was the first satellite dedicated ~~to~~for precipitation studies. ~~It is also the first and the only satellite that, in addition to visual, infrared, and passive microwave sensors, also has space borne radar for precipitation measurement using active microwave sensors.~~ The after-real-time TMPA product also incorporates rain gauge information wherever feasible. In this study, we aggregated the 3-hourly TMPA-RT data to daily level, resampled from the coarse resolution (0.25° × 0.25°) to a resolution of 0.05° × 0.05°, and implemented a bias correction using the "Climate Hazards Group InfraRed Precipitation with Station" product ~~(CHIRPS; Funk et al., 2014) and rain gauge measurements (see Roy et al., *in prep*).~~(CHIRPS; Funk et al., 2014, 2015) and rain gauge measurements (details not included).

2.2.3 ~~2.2.3~~ Estimates of Streamflow

~~*[13]*~~ Streamflow data were computed using the calibrated stage-discharge relationship for the Bomet Bridge discharge station (Station ID: 1LA03; Location: 0°47′23.50″S 35°20′47.45″E) on the Nyangores River (drainage area approximately 697 km^2), which is one of the two main tributaries of the Mara River. Data is available ~~data~~ for the period Jan 1, 1996 to Jun 30, 2010, during which time only about ~8% of the records are missing.

2.2.4 Estimates of Temperature

~~*[14]*~~ We computed PET using the Hargreaves Equation ~~(Hargreaves and Samani, 1985)~~(Hargreaves and Samani, 1985), the annual mean of which closely matched the reported PET value for the study area (WREM, 2008). The temperature data used in the Hargreaves Equation were extracted from the Global Surface Summary of the Day (GSOD) product produced by the National Climatic Data Center (NCDC) in Asheville, NC. The daily temperature data includes multiple observations and are available in three forms: maximum, minimum, and average.

2.3 The Hydrologic Model HyMod-V2

~~The spatially lumped HyMod Version 1 (HyMod-V1) conceptual rainfall-runoff model with six parameters has previously been implemented for satellite-rainfall-based simulation and forecasting of streamflow for the study area (Roy et al., *in prep*). The model is driven using mean daily precipitation and PET data to generate daily estimates of AET and streamflow. Nonlinear vertical flow processes are controlled by a two-parameter soil moisture accounting (SMA) module based on the~~ The spatially lumped HyMod Version 1 (HyMod-V1) conceptual rainfall-runoff model with five/six parameters has previously been used in several studies (Boyle et al., 2000; Vrugt et al., 2003, 2009; Moradkhani et al., 2005; Duan et al., 2007; Wang et al., 2009; Razavi and Gupta, 2016). The model is driven using daily precipitation and PET data to generate daily estimates of AET (HAET: HyMod-generated AET) and streamflow. Nonlinear vertical flow processes are controlled by a two-parameter soil moisture accounting module based on the rainfall excess model proposed by Moore (1985) ~~rainfall excess model.~~. Horizontal ~~routing is achieved via~~ flows are

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: No bullets or numbering

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Outline numbered + Level: 3 + Numbering Style: 1, 2, 3, … + Start at: 3 + Alignment: Left + Aligned at: 0" + Indent at: 0.5"

Formatted: Font: 11 pt

Formatted: No bullets or numbering

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: No bullets or numbering

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

simulated in a linear ~~(ROUT)~~routing module that includes a Nash Cascade for quick-flow routing ~~for fast~~ (overland flow) and a linear reservoir for slow-flow routing ~~for~~ (baseflow~~. Details of the model structure and process equations are presented in Appendix A. We will~~).

~~*[15]*~~ In this study, we improved the ET process parameterization within the model by relying on satellite-based AET estimates provided by GLEAM. In this regard, we were careful to ensure that the structural modifications (1) do not over-complicate the model since that defeats the whole purpose of having simpler models such as HyMod, (2) do not require large number of additional model parameters, (3) are more physically consistent, (4) consistently produce improved ET simulations, and finally (5) do not deteriorate the streamflow simulations. We refer to the structurally modified version of the model as HyMod Version 2 (HyMod-V2~~).~~), as shown in Fig. 4. As can be seen, a new ET resistance module is added to the soil moisture accounting module of the model. A detailed description of the ET resistance module is provided in Section 2.4. Details of the overall model structure and process equations are presented in Appendix A.

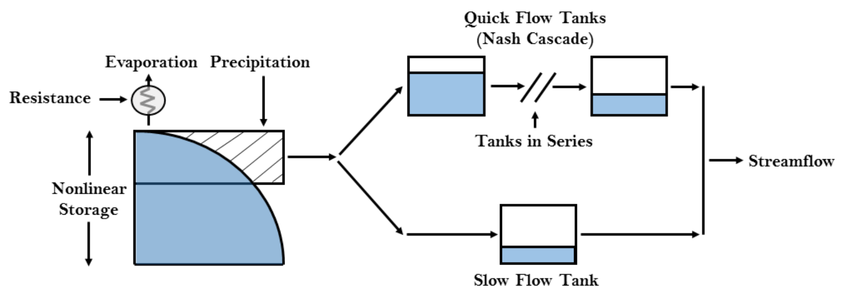


Fig. 4. Schematic diagram of HyMod-V2.

2.4 Study Approach

~~*[16]*~~ We conducted the investigation in two stages. The first stage consists of five steps designed to improve model performance with respect to both streamflow and evapotranspiration (as assessed against data), but ***without*** making structural modifications to the model. The strategy includes using GAET to ***constrain*** simulated evapotranspiration, recalibration of model parameters, and ~~a kind of "bias-correction" of GAET.~~constraint adjustments. In doing so, we specifically do ***not*** directly assimilate GAET into the model (either by a Bayesian data assimilation "nudging" procedure such as the Ensemble Kalman Filter, or by direct insertion), so that the model's representation of overall water balance is not compromised. Accordingly, while we ***are*** extracting information from the GLEAM product, we do so via a process of "*constraining*" rather than "*assimilation*".

~~*[17]*~~ In the second stage, we modify the structure of the model to directly improve its ability to simulate ET (using GAET as the target~~).~~) by trying to capture the physics of the underlying processes more accurately. The steps followed in stage one are repeated so that results of the different strategies can be compared.

Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: No bullets or numbering
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: No bullets or numbering
Formatted: Font: 11 pt
Formatted: Font: 11 pt

~~[18]~~ Conceptually, the main difference between ~~Stages~~ Stage-I and Stage-II is that, in the former, the information provided by GAET is used only to constrain the evapotranspiration fluxes ~~and soil moisture states~~ of the (~~re-calibrated~~recalibrated) model, whereas in the latter the information contained in GAET is used to alter the model structure. While the former provides a temporary improvement to model performance, achieved as long as GLEAM data are available, the latter is expected to provide a lasting improvement to model performance that should persist even when GLEAM data are not available. In the final step of Stage II, we check to see whether the GAET product contains residual information that, not having yet been used to improve the model structure, remains useful for improving model performance via the constraining operation.

~~**2.4.1 Stage One – Constraining HyMod using GAET**~~

~~**Step I-1:** **Benchmark:** HyMod-V1 is forced with TMPA-RT satellite-based precipitation and the parameters are calibrated against observed streamflows. GAET satellite-based evapotranspiration is not used. This is the benchmark step against which model performance improvements are assessed.~~

~~**Step I-2:** **GAET Constrained:** GAET is used to constrain HyMod-V1 HAET. The model parameters remain the same as in Step I.~~

~~**Step I-3:** **Recalibrated GAET Constrained:** GAET is used to constrain HyMod-V1 HAET. The model parameters are re-calibrated to match simulated to observed streamflows.~~

~~**Step I-4:** **Bias-Corrected Recalibrated GAET Constrained:** A "bias-correction" is applied to GAET, and the bias-correction and model parameters are calibrated together to match simulated and observed streamflows. Since no ground-based data are available to bias-correct GAET, we tested two empirical bias correction schemes (shown below):~~

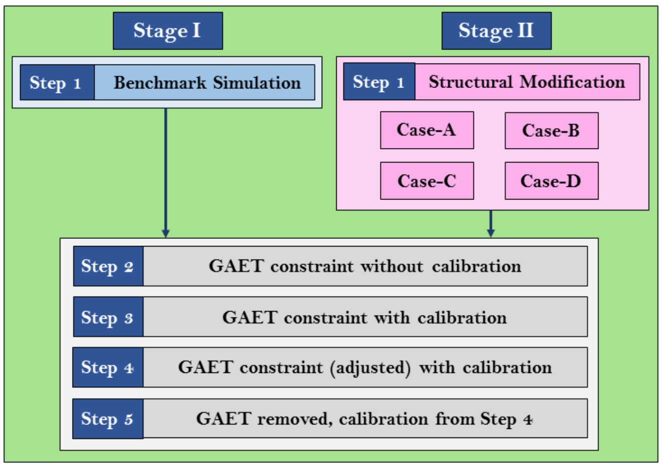
The entire study approach is summarized in Fig. 5. As can be seen, only Step-1 is different for both the stages (Stage-I and Stage-II), while the remaining four steps (Step 2-5) are similar. Thus, in each stage, there are five steps altogether. Stage-I Step-1 is for generating benchmark simulations using the calibrated model but without any ET constraint or structural modifications. On the other hand, Stage-II Step-I has four different cases (A-D) corresponding to different structural modifications in the ET process parameterization. Both the benchmark model from Stage-I Step-I and the best performing model from Stage-II Step-I are used in the following steps. Step-2 is based on imposing the ET constraint but without recalibration, meaning that the same set of calibrated parameters as in the benchmark step is further used in this step. Step-3 is based on recalibrating the model while imposing the ET constraint. Step-4 is conceptually similar to Step-3, however, additionally, some constraint adjustments (Eq. 1 and 2) are applied and the adjustment parameters are calibrated together with model parameters (to match the simulated and observed streamflows). Finally, in Step-5, we remove the ET constraint to see whether the performance of the new model will decline when satellite ET data becomes unavailable (note this is no longer the benchmark model since we recalibrated the parameters in Step 4).

$$y = f(x,\cdots) \text{ and } x = \boldsymbol{a}\cdot \mathrm{GAET} \qquad (1)$$

$$y = f(x,\cdots) \text{ and } x = \boldsymbol{a}\cdot \mathrm{GAET}^{\boldsymbol{b}} \qquad (2)$$

Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Space After: 0 pt
Formatted Table
Formatted: Font: 11 pt
Formatted: Space After: 0 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

~~where~~ In Eq. 1 and 2, ***y*** represents the streamflows after ~~bias correction~~the adjustment of GAET constraint, and ***a*** and ***b*** are the parameters of the ~~bias-correction formulation~~adjustment formulations (***a*** controls the variance and ***b*** controls the degree of non-linearity).



. **Fig. 5.** The approach followed in this study.

ET Constraining

The ET constraint was imposed by modifying the original ET equation of the model (Eq. A5) from $\mathrm{HAET} = \min\{\mathrm{PET}, C_{SMA}\}$ to the new form $\mathrm{HAET} = \min\{\mathrm{PET}, \mathrm{GAET}, C_{SMA}\}$~~**Step I-5: GAET Removed:** The model obtained in Step IV is run without using GAET as a constraint. No re-calibration is performed. This shows what would happen to model performance if real-time GAET data were to become unavailable.~~

~~**2.4.2 Stage Two – Modifying the Structure of HyMod using GAET**~~

~~**Step II-1: Structurally Modified:** The evapotranspiration equation of~~ . Note that this is not a structural modification to the form of the process equation, rather GAET sets the upper limit of HAET in this case, which is more realistic than using PET directly as the upper limit.

Structural Modifications

The ET process parameterization within the original model (HyMod-V1) is modified in Stage-II Step-1 to improve its ability to reproduce GAET ~~(i.e., so that HAET becomes more similar to GAET).~~more accurately without deteriorating the

Formatted: Font: 11 pt, Italic
Formatted: List Paragraph, Indent: Left: 0"
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Indent: Left: 0", First line: 0", Space Before: 0 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

streamflow simulations. Four ~~evapotranspiration~~ET equations of progressive complexity and physical basis are tested. In each case, the model parameters are ~~re-calibrated~~recalibrated to match the simulated streamflows to the observed data. The final result is a structurally modified model called HyMod-V2.

More specifically, the ET equation of HyMod-V1 is multiplied by a function $K(\cdot)$ such that $0 \leq K(\cdot) \leq 1$. This function acts as a resistance to the ET flux of the model. Four different forms for $K(\cdot)$ that represent incremental increases in complexity and physical basis (Table 1) are tested. Writing the main ET equation in the general form:

$$Y_t = K_t \cdot X_t \cdot EDR_t \qquad (3)$$

where Y_t is the AET generated by HyMod ($HAET_t$), X_t is the soil moisture storage (C_{SMA_t}), and EDR_t is the *evaporation demand ratio* computed as $\min\left\{1, \frac{PET_t}{X_t}\right\}$. The most general form for K_t is given by:

$$K_t = K_{min} + [K_{max} - K_{min}] \cdot f(\psi_t) \qquad (4)$$

where, K_{min} and K_{max} are lower and the upper limits for K, and ψ_t is the ratio of actual to maximum storage capacity ($\psi_t = X_t/X_{max}$).

Table 1: K-function in different cases.

Cases	K_{max}	K_{min}	$f(\psi_t)$	Additional Parameters
a	1	0	1	None
b	K_0	0	1	K_0
c	K_0	0	X_t/X_{max}	K_0
d	K_{max}	$\gamma \cdot K_{max}$	$(X_t/X_{max})^{BE}$	K_{max}, γ, BE

~~**Step II-2: GAET Constrained:** Using the "best" structurally modified model (HyMod-V2) from Step II-1, we repeat Step I-2.~~

~~**Step II-3: Recalibrated GAET Constrained:** GAET is used to constrain HAET simulated by HyMod-V2. The model parameters are re-calibrated to match simulated and observed streamflows.~~

~~**Step II-4: Bias-Corrected Recalibrated GAET Constrained:** The empirical "bias-correction" scheme is applied to GAET, and the bias-correction and modified-model (HyMod-V2) parameters are calibrated together to match simulated and observed streamflows.~~

~~**Step II-5: GAET Removed:** HyMod-V2 is run without using GAET as a constraint. No re-calibration is performed. This shows what would happen to performance of HyMod-V2 if real-time GAET data were to become unavailable.~~

Formatted
Formatted: Font: 11 pt
Formatted
Formatted: Indent: Hanging: 0.75"
Formatted: Font: 11 pt
Formatted
Formatted Table
Formatted: Font: 11 pt
Formatted
Formatted: Indent: Hanging: 0.75"
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted Table
Formatted: Indent: Hanging: 0.75"
Formatted: Font: 11 pt
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted: Font: 11 pt
Formatted
Formatted
Formatted
Formatted: Font: 11 pt
Formatted
Formatted
Formatted: Font: 11 pt
Formatted
Formatted
Formatted
Formatted: Font: 11 pt
Formatted
Formatted
Formatted
Formatted
Formatted: Font: 11 pt

2.5 Calibration Methodology and Benchmark Model Calibration

~~[19]~~ Calibration of the model (and ~~bias-correction~~adjustment) parameters was performed ~~by running~~using the SCE-UA algorithm (Duan et al., 1992)~~ with~~. In all cases, the calibration runs were carried out using 10 complexes ~~for~~and 25 loops. ~~Calibration was performed to match~~Model simulated streamflows were matched against the observed streamflows in ~~a~~the λ-transformed space ~~(Box and Cox, 1964; see Appendix B)~~ to minimize the effects of skewness and reduce heteroscedasticity. The λ-transformation was applied after modifying the original equation as proposed by Box and Cox, (1964) (see Appendix B), and the value of the λ parameter was calculated from the observed streamflow records. The performance criterion used was the Mean Squared Error (MSE) of transformed flows. The model was run continuously for the 7.5-year period Jan 2003 to June 2010, with the first 4 years (2003 to 2006) used for calibration and the remaining 3.5 years (2007 to mid-2010) used to provide an additional assessment of model performance. Results are shown for the "*calibration (4-years)*", "*evaluation (3.5 years)*" and "*total (7.5 years)*" simulation periods. ~~All the streamflow errors statistics reported in this study are in the λ-transformed space.~~

2.6 Metrics Used for Performance Evaluation

~~[20]~~ Four metrics are used in this study to assess the model performance (Table ~~1~~2). These metrics measure performance in regards to overall mean squared error, bias, variability, and correlation (see Gupta et al., 2009), are computed in the transformed space where applicable (e.g. for streamflows), and are normalized to be comparable.

Table 2: Performance evaluation metrics used in this study.

Metrics	Equations
Normalized Mean Square Error (NMSE)	$\mathrm{MSE} = \mathrm{mean}((\mathrm{O_i} - \mathrm{S_i})^2)$; $\mathrm{NMSE} = \frac{\mathrm{MSE}}{\mathrm{var(O)}}$
Normalized Bias in Mean (NBμ)	$\mathrm{NB}\,\mu = \frac{\mathrm{mean(S)} - \mathrm{mean(O)}}{\mathrm{mean(O)}}$
Normalized Bias in Standard Deviation (NBσ)	$\mathrm{NB}\,\sigma = \frac{\mathrm{std(S)} - \mathrm{std(O)}}{\mathrm{std(O)}}$
~~[Insert Table 1]~~	
Correlation Coefficient (ρ)	$\rho = \frac{\sum_{i=1}^{N}(\mathrm{O_i} - \mathrm{mean(O)})(\mathrm{S_i} - \mathrm{mean(S)})}{\sqrt{\sum_{i=1}^{N}(\mathrm{O_i} - \mathrm{mean(O)})^2 \sum_{i=1}^{N}(\mathrm{S_i} - \mathrm{mean(S)})^2}}$
O: Observed flows; S: Simulated flows; N: Number of data points.	

3 Results

3.1 Results from Stage ~~-~~I (~~Constraining Simulated AET~~ET Constraints)

~~3.1.1 Benchmark Model (Step I-1)~~

~~[21]~~ The performance of the benchmark model HyMod-V1, driven using TMPA-RT satellite-based precipitation and with parameters calibrated to match simulated streamflow to observed data, is reported in Table ~~4~~5. The NMSE varies between 0.56 (calibration period) and 0.84 (evaluation period), where NMSE = 0.56 means that on average only about 44%

Formatted: Font: 11 pt
Formatted: No bullets or numbering
Formatted ...
Formatted: No bullets or numbering
Formatted ...
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted ...
Formatted ...
Formatted ...
Formatted: Font: 11 pt
Formatted ...
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted ...
Formatted: No bullets or numbering
Formatted ...

(1.0 - 0.56 = 0.44) of the variability in the flows has been explained. This is not surprising given the use of a simple lumped conceptual model driven by satellite-based estimates of precipitation for a poorly gauged basin. The flow biases are small (NBμ < 15%) indicating that long-term water balance is approximately preserved. The calibrated values of the model parameters are reported in Appendix C.

~~[22]~~ Table ~~2~~3 presents a comparison of the model generated HAET with the GAET data (for the total 7.5 year simulation period). HAET tends to be larger on average, varies over a wider range, is considerably more skewed, and is less kurtotic. Some of the reasons for this can be understood from the time-series plot and scatterplot shown in Fig. ~~4~~6. The behavior of HAET tends to be more erratic and, although both HAET and GAET show seasonal patterns, the former regularly drops to zero or near zero (explained by the very simple, threshold-like, ET process representation in the model, which does not contain a resistance term). The result is that HAET and GAET are not well correlated (Fig. ~~4b~~6b) and have different shapes for their empirical probability distributions (Fig. ~~5~~7). Even if we were to ignore the time-steps when HAET drops closer to zero, HAET is strongly positively biased (too large), which results from trying to satisfy the potential evapotranspiration (PET).

~~[23]~~ Table ~~3~~4 reports a water balance estimate WBAET of the mean annual AET for the basin, obtained by subtracting mean annual streamflow (at the discharge station) from mean annual precipitation (estimated from TMPA-RT). WBAET is similar in magnitude to GAET, and we have GAET < WBAET < HAET < HPET, indicating that the AET computed by the model tends to be a little high.

~~[Insert Table 2]~~

~~[Insert Fig 4]~~

~~[Insert Fig 5]~~

~~[Insert~~ Table 3: Descriptive statistics of GAET and HAET.

Statistics	GAET	HAET
Maximum	4.62	6.12
Minimum	0.119	0.00
Mean	3.03	3.52
Median	3.08	4.21
Mode	0.11	0.00
Std. Dev.	0.59	1.72
Skewness	-0.58	-1.03
Kurtosis	3.84	2.62

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: List Paragraph

Formatted: Font: Not Bold, Font color: Auto

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

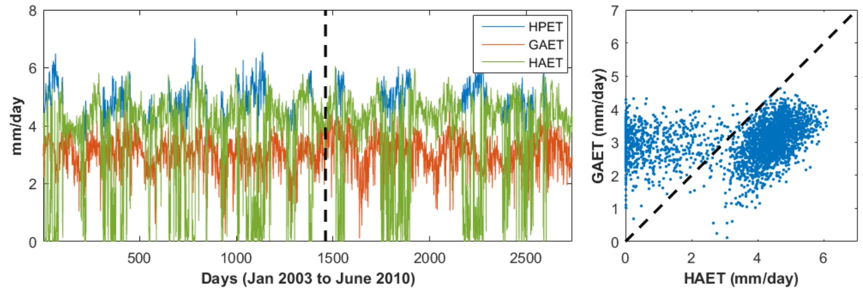


Fig. 6. Time series and scatter plots of HPET, GAET, and HAET.

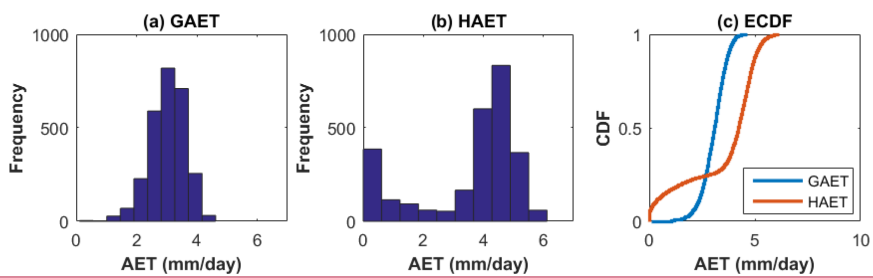


Fig. 7. Histogram and ECDF plots of GAET and HAET.

Table 4: Annual mean of AETs and HPET.

Source	**Annual Mean (mm)**
GAET	1100
HAET	1263
WBAET	1146
HPET	1704

~~3.1.2 Using GLEAM AET to Constrain~~ The benchmark model was constrained using the **~~Model (Step I-2)~~**

~~*[24]* Next, GLEAM satellite-based daily~~GAET estimates ~~of AET (GAET) were used to constrain the HAET estimates generated by HyMod-V1. The constraint is imposed by modifying the original evapotranspiration equation of the model (Eq. A5) from HAET = min{PET, C_{SMA}} to the new form HAET = min{PET, GAET, C_{SMA}}; this is not a 'structural' modification to the form of the process equation – it simply constrains HAET ≤ GAET. The practical effect of this modification is that daily changes in soil moisture storage~~

Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: List Paragraph
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

~~become smoother and thereby closer to our expectation of how they should behave. The parameters of~~in Step-2 but the model parameters were not recalibrated.

~~[25]~~ As can be seen in Table ~~4 indicates that~~5, the model performance has become significantly worse due to streamflow becoming positively biased. Given that GAET < HAET, on average in the previous step, this makes sense because imposing GAET as a constraint alters the water balance of the model.

~~**3.1.3 Recalibration of the Model Constrained Using GLEAM AET (Step I-3)**~~

~~[26]~~ To try and fix ~~the~~this water-balance problem~~ introduced during Step I-2~~, we recalibrated the parameters of the model to improve the match to observed streamflows ~~(~~while continuing to use GAET to constrain HAET in the model (Step-3). Although the large positive bias was reduced (Table ~~4~~5) and the NMSE statistic ~~is~~was improved as compared to Step~~ I~~-2, most of the error statistics deteriorated for all three periods (calibration, evaluation, and total simulation) as compared to Step~~ I~~-1. Importantly, this calibration step resulted in an unrealistically large value of 17.36 meters for the size of the soil moisture storage (previously a more realistic 0.76 meters). This value is clearly conceptually and physically inconsistent (the realistic range is about zero to 2 meters), and while it ~~improves~~improved calibration period performance, the lack of consistency is reflected in the sharp deterioration in performance during evaluation. Unrealistic parameter values, such as this, are indications of either severe errors in the data, or structural errors in the model. Since GAET agrees well with WBAET on average, it is likely that the major cause here is model structural inadequacy (Gupta et al., 2012). For completeness, we next ~~check~~checked (Step~~ I~~-4) to see whether this problem ~~can~~could instead be resolved by implementing an empirical ~~bias-correction~~adjustment to GAET.

~~**3.1.4 Applying a Bias Correction to the GLEAM AET (Step I-4)**~~

~~[27]~~ We tested two empirical ~~bias correction~~constraint adjustment schemes (Eq. 1 and 2) applied to the GAET data~~ to obtain a quantity we call BC-GAET~~, and calibrated the additional parameters (from these equations) along with ~~the~~ HyMod-V1 parameters. Results for both schemes were similar, but Eq. 2 provided slightly better performance for the evaluation and total simulation periods and so we selected Eq. 2. Compared to the ~~Benchmark~~ benchmark (Step~~ I~~-1~~,~~), NMSE and NBμ calibration period statistics reduced from 0.56 to 0.43 and 12% to 6%, respectively (Table ~~4~~5) while ρ increased from (0.76/0.66/0.72; Cal/Eval/Tot) to (0.83/0.74/0.78). Perhaps more important, the calibrated value of parameter H is now 0.65 meters, which is within the conceptually acceptable range.

~~**3.1.5 GLEAM Data Removed (Step I-5)**~~

~~[28]~~ Finally, the model obtained in Step~~ I~~-4 was run without the use of GAET~~ (actually BC-GAET)~~ to see how well the model would perform if GAET data were to become unavailable. The results (Table ~~4~~5) results indicate that model performance does not deteriorate significantly when GAET data become unavailable and, in some cases, the performance is better than the benchmark.

Table 5: Streamflow error statistics for calibration, evaluation, and total simulation (in parenthesis) for all five different cases in **Stage-I** analysis.

Metrics	Step-1	Step-2	Step-3	Step-4	Step-5
Calibration					
NMSE	0.56	1.68	0.64	0.43	0.60
NBμ	0.09	0.32	0.12	0.09	-0.09
NBσ	-0.12	-0.15	-0.19	-0.06	-0.04
ρ	0.76	0.81	0.74	0.83	0.75
Evaluation (Total Simulation)					
NMSE	0.84 (0.77)	2.13 (2.17)	0.92 (1.19)	0.88 (0.75)	0.64 (0.56)
NBμ	0.14 (0.14)	0.38 (0.38)	0.09 (0.22)	0.15 (0.16)	-0.01 (-0.02)
NBσ	-0.04 (-0.08)	-0.03 (-0.10)	0.04 (-0.01)	0.17 (0.02)	0.14 (0.05)
ρ	0.66 (0.72)	0.71 (0.76)	0.61 (0.69)	0.74 (0.78)	0.73 (0.74)

~~[Insert Table 4]~~

3.2 Results from Stage~~-~~II (~~Modifying Model Structure~~Structural Modifications)

~~**3.2.1 Modifying the Model Structure to Improve Simulation of AET (as Estimated by GLEAM) (Step II-1)**~~

~~*[29]*~~ Results from Stage~~-~~I confirm that ~~assimilating~~ GAET ~~into HyMod~~constraining can improve the overall ~~model~~ performance of HyMod. However, for operational implementation, the method requires real-time estimates of SET, which could sometimes pose a challenge for practical ~~application~~applications. To overcome the need for real-time data availability, a simple approach could be to establish a functional relationship between HAET and GAET from the historical records and use that relationship to adjust HAET. In our case, however, HAET and GAET did not show a sufficiently strong relationship (Fig. ~~4~~6). Therefore, we instead investigated whether we could use the historical GAET data to improve the structure of the model itself.

~~*[30]*~~ Our previous results (Fig. ~~4~~6) showed that HAET generated by HyMod-V1 did not match well with GAET. This is likely because the entire soil moisture storage ~~is~~was exposed to the ET process~~ and there is representation of the physical processes that act to inhibit ET when the soil moisture content is low~~. Consequently, it is common for all of the soil moisture to evaporate away during a single time step, leaving no water available for evaporation at the next time step (provided no precipitation is added), so that HAET drops to zero. This tendency can be reduced by modifying the ET process representation so that HAET more closely follows GAET. ~~We do this by multiplying the ET equation by a function $K(\cdot)$ such that $0 \le K(\cdot) \le 1$.~~

~~We tested~~ Step-1 in the Stage-II analysis has four different ~~forms for $K(\cdot)$ that represent incremental increases in complexity (~~cases as shown in Table ~~5). Writing the main ET equation in the general form:~~

$$Y_t = K_t \cdot X_t \cdot EDR_t \qquad \cancel{(3)}$$

~~where Y_t is the model generated AET (HAET$_t$), X_t is the soil moisture storage ($C_{SMA,t}$), and EDR_t is the *evaporation demand ratio* computed as $\min\left\{1, \frac{PET_t}{X_t}\right\}$. The most general form for K_t is given by:~~

$$K_t = K_{min} + [K_{max} - K_{min}] \cdot f(\psi_t) \qquad \text{~~(4)~~}$$

~~where, K_{min} and K_{max} are lower and the upper limits for K, and ψ_t is the ratio of actual to maximum storage capacity ($\psi_t = X_t/X_{max}$).~~

~~**[Insert Table 5]**~~

~~***(a) Results from Step II-1a***~~

~~*[32]*~~ ~~This step~~1. The first case (Case-A) is identical to the benchmark step (~~Sect. 2.5~~Step-1) in the Stage-I analysis, where the calibrated HyMod-V1 is run without GAET estimates. ~~Results are summarized in Table 6.~~

~~***(b) Results from Step II-1b***~~

~~*[33]*~~ In ~~this case~~Case-B, $K_t = K_0$ is applied as a constant multiplier to the ET equation (see Table ~~5~~1), thereby acting as a constant surface resistance to ET. Calibration (of all of the model parameters) resulted in improved error statistics (Table 6). The estimate obtained for the surface resistance was $K_0 = 0.73$. However, we again obtained a conceptually unrealistic value (H = 9.5 meters) for the soil moisture storage parameter.

~~***(c) Results from Step II-1c***~~

~~*[34]*~~ In ~~this case~~Case-C, the more complex form $K_t = K_0 \cdot f(\psi_t)$ was used (see Table ~~5~~1). This produced a model performance (Table 6) comparable to that of the previous ~~Step II-1b~~, but with a more realistic ~~value for the~~ calibrated value of the soil moisture storage (H = 0.90 meters). Interestingly, the calibrated value for K_0 was 1, implying that K_0 becomes irrelevant once $f(\psi_t)$ is introduced to the ET equation.

~~***(d) Results from Step II-1d***~~

~~*[35]*~~ Finally, the most complex form $K_t = K_{min} + [K_{max} - K_{min}] \cdot f(\psi_t)$ was ~~used~~introduced in Case-D. The calibration parameter K_0 was used to represent K_{max}, and K_{min} was defined as $K_{min} = \gamma \cdot K_{max}$ via a second calibration parameter γ (ranging from 0 to 1) (see Table ~~5~~1). Results indicate that although the calibration error statistics (Table 6) are similar to that of ~~Step II-1c~~Case-C, the evaluation and total simulation statistics are better. The calibrated value of parameter 'BE' (derived by transforming the parameter 'be'; see Eq. B1) was 0.86, indicating only a mildly non-linear relationship between ψ_t and K (or HAET). The minimum and the maximum limits of K_0 were close to zero and one, respectively, confirming the findings of ~~Step II-1c~~Case-C that once $f(\psi_t)$ is introduced ~~into~~to the ET equation, the need for ~~the~~ K_{max}, and K_{min} parameters largely disappears.

Table 6: Streamflow error statistics for calibration period, evaluation period, and total simulation period (in parenthesis)

for all four different cases in **Stage-II Step-1** analysis.

Metrics	Case-A	Case-B	Case-C	Case-D
		Calibration		
NMSE	0.56	0.52	0.51	0.51
NBμ	0.09	0.10	0.10	0.10
NBσ	-0.12	-0.19	-0.04	-0.04
ρ	0.76	0.77	0.80	0.80
		Evaluation (Total Simulation)		
NMSE	0.84 (0.77)	0.65 (0.77)	0.88 (0.72)	0.84 (0.70)
NBμ	0.14 (0.14)	0.07 (0.16)	0.14 (0.14)	0.14 (0.13)
NBσ	-0.04 (-0.08)	0.00 (-0.06)	0.16 (0.10)	0.16 (0.10)
~~3.2.2~~ ρ	0.66 (0.72)	0.70 (0.75)	0.73 (0.78)	0.74 (0.78)

~~**[Insert Table 6]**~~

Final Model Selection from Stage-II Step ~~II~~-1

~~[36]~~ In this section, we address two main questions: (a) Does the ~~model~~ structural modification of the model (to the representation of the ET process) improve ET estimation~~, and (b) if~~? If so, then (b) what level of complexity is adequate? Table 7 presents the streamflow and AET performance statistics for the total simulation period for the four cases. Since ~~Step II-1a provided~~Case-A provides very poor error statistics for AET (e.g. NMSE = 8.93 and NBσ = 1.89), we ~~disregarded~~disregard this case. Although ~~Step II-1b provided~~Case-B shows the best NBσ (-0.06) statistics for streamflow, and the best NMSE (1.28) and NBμ (0.12) statistics for AET, the value obtained for the soil moisture storage capacity (H) was unrealistic; we therefore also ~~disregarded Step II-1b.~~disregard this case. Comparison of ~~Step II-1c~~Case-C and ~~Step II-1d~~Case-D shows that while their streamflow and AET simulations ~~were~~are similar (Fig. ~~6), Step II-1d provided~~8), Case-D provides slightly better NMSE (0.70) and NBμ (0.13) statistics for streamflow and slightly better ~~R~~ρ (0.49) statistics for AET (Table 7 and Fig. ~~7~~9). We therefore selected the most complex form $K_t = K_{min} + [K_{max} - K_{min}] \cdot f(\psi_t)$ for the ET function (~~Step II-1d~~Case-C). The corresponding model is hereafter referred to as 'HyMod-V2'. Note that for practical applications, any simpler structural modification (Case-B or Case-C) can be adapted, if proves convenient.

~~**[Insert Table 7]**~~

~~**[Insert Fig 6]**~~

~~**[Insert Fig 7]**~~

~~[37]~~ Comparing the streamflow error statistics of Stage-I Step ~~I~~-4 (Table ~~4~~5) and Stage-II Step ~~II-1d~~-1 Case-D (Table 6), we see that they are quite similar, indicating that the ET constraining (first approach) and diagnostic structural improvement (second approach) strategies produce dynamical behaviors that are similar (as measured by the four performance metrics used).

Table 7: Streamflow and AET error statistics in total simulation for all four cases in **Stage-II Step-1** analysis.

Formatted: Font: 11 pt

Formatted Table

Formatted: Font: 11 pt, Bold

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted Table

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt, Italic

Formatted ...

Formatted ...

Formatted ...

Formatted ...

Metrics	Case-A	Case-B	Case-C	Case-D
		Streamflow		
NMSE	0.77	0.77	0.72	0.70
NBμ	0.14	0.16	0.14	0.13
NBσ	-0.08	-0.06	0.10	0.10
ρ	0.72	0.75	0.78	0.78
		AET		
NMSE	8.93	1.28	1.70	1.71
NBμ	0.18	0.12	0.17	0.17
NBσ	1.89	-0.26	-0.10	-0.13
ρ	0.22	0.43	0.48	0.49

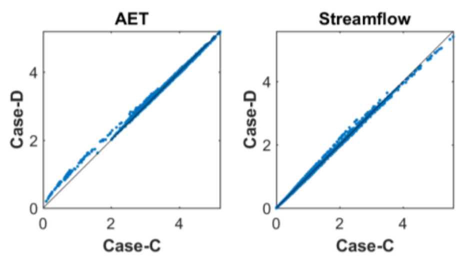


Fig. 8. Scatter plots of streamflow and AET from Case-C and Case-D.

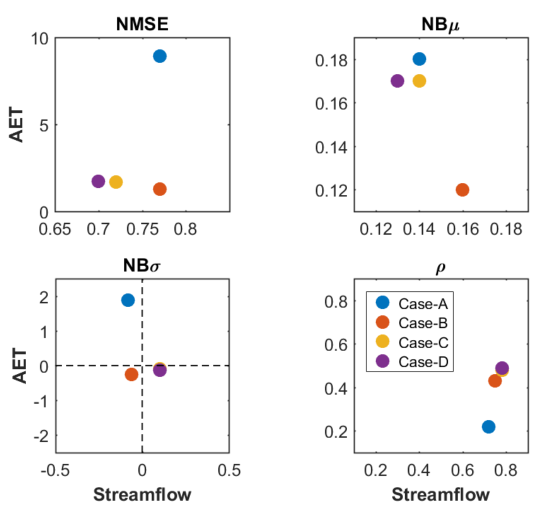


Fig. ~~[38]~~9. Streamflow and AET error statistics in total simulation for all four cases in Stage-II Step-1 analysis.

Is Further Improvement Possible?

The modified model (HyMod-V2) selected in the previous section (Stage-II Step-1 Case-D) was next used with GAET in ~~a similar manner to Steps I~~Step-2 to ~~I~~Step-5~~,~~ (see Fig. 5) to address two questions~~; i) could~~: (1) Could more information from GAET be ~~assimilated~~incorporated (via constraining) into the model, or is the improved model structure (without GAET) already good enough~~, and ii) is~~? (2) Is the ~~bias~~constraint adjustment ~~of~~on GAET (Step-~~IV) necessary~~ 4) still relevant once the model structure has been improved?

~~**3.2.3 Using GLEAM AET to Constrain the Modified Model (Step II-2)**~~

~~*[39]*~~ When GAET was used to constrain the ET process (Step-2) in HyMod-V2~~,~~ without model recalibration (~~the~~ parameters used were from ~~Step II-1d). This introduced~~Case-D), there was significant overestimation bias evident in the ~~simulation~~simulations of ~~streamflows~~streamflow (Table 8). Clearly, recalibration of the modified model ~~is~~was necessary ~~when GAET is imposed as a constraint on ET.~~

~~3.2.4 Recalibration of~~to see if the ~~**Modified Model Constrained Using GLEAM AET (Step II-3)**~~

~~*[40]* Recalibration~~model performance could be improved any further. The recalibration of HyMod-V2 improved the error statistics (Table 8); compare these results with ~~the~~Stage-I Step ~~I~~-3 results in Table 4 derived the same way for HyMod-V1. While a small improvement was obtained for the soil moisture storage capacity parameter H (reduced from 17.4 meters to 12.8 meters), this value remained conceptually inconsistent (too large). Overall, the error statistics deteriorated compared to the best results from ~~Step II-1.~~

~~**3.2.5 Applying a Bias Correction to the GLEAM AET (Step II-4)**~~

~~*[41]*~~ Case-D. The HyMod-V2 parameters were then calibrated along with the parameters of the GAET ~~bias correction~~adjustment equation (using Eq. 2), as in Step I-4. Although the results improved (Table 8), and the value of H parameter became conceptually realistic (0.88 meters), the results were not significantly different from ~~Step II-1~~Case-D in Step-1. Finally, when the GAET data were made unavailable, the model performance remained stable as also seen in Stage-I Step-5 results.

~~**3.2.6 GLEAM Data Removed (Step II-5)**~~

~~*[42]* This step is similar to Step I-5 but with HyMod-V2. The performance of the model remained stable even when GAET was not used (Table 8).~~

~~**[Insert Table 8]**~~

~~*[43]* In~~

Formatted: Font: 11 pt, Italic
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Not Bold
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

Therefore, in regards to the two questions that motivated this section, the results indicate that; (~~a~~1) once information from GAET was ~~assimilated~~incorporated into the model as a modification to the structure, there was no further need for the use of GAET to constrain the simulation of ET (use of GAET even caused some of the results to deteriorate), and (~~b~~2) implementation of a ~~bias correction~~constraint adjustment to GAET (Step ~~II~~-4) ~~did improve the error statistics and resulted~~ in ~~a more conceptually realistic value for the H parameter in the~~the structurally modified model (Stage-II) did not improve the overall results.

Table 8: Streamflow error statistics for calibration, evaluation, and total simulation (in parenthesis) for all five different cases in **Stage-II** analysis.

Metrics	**Step-1**	**Step-2**	**Step-3**	**Step-4**	**Step-5**
Calibration					
NMSE	0.51	1.82	0.64	0.51	0.50
NBμ	0.10	0.32	0.12	0.10	0.10
NBσ	-0.04	0.16	-0.19	-0.04	-0.04
ρ	0.80	0.81	0.74	0.80	0.80
Evaluation (Total Simulation)					
NMSE	0.84 (0.70)	2.46 (2.29)	0.92 (1.19)	0.84 (0.70)	0.84 (0.70)
NBμ	0.14 (0.13)	0.37 (0.36)	0.10 (0.22)	0.14 (0.14)	0.14 (0.13)
NBσ	0.16 (0.10)	0.38 (0.31)	0.05 (-0.01)	0.16 (0.10)	0.16 (0.09)
ρ	0.74 (0.78)	0.70 (0.77)	0.61 (0.69)	0.74 (0.78)	0.74 (0.78)

3.3 Overall Comparison and Analysis of Uncertainty ~~(in Streamflow and AET)~~

[44] ~~Figure 8~~ Fig. 10 compares the streamflow time series obtained from ~~Step~~ Stage-I Step-4 (constraining ET) and ~~Step II-1d~~ (Stage-II Step-1 Case-D (selected model after structural modification) against the benchmark (~~Step~~ Stage-I Step-1) in both actual and λ-transformed space. Simulations from the structurally modified model HyMod-V2 ~~(Step II-1d)~~ follow the observations most closely, followed by the simulations from ~~Step~~ Stage-I Step-4 (ET constraining) and ~~Step I-1 (Benchmark).~~ benchmark. Clearly, while the streamflow simulations are improved by both ET constraining and model structural modification, the latter performs the best.

Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Highlight
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Bold
Formatted Table
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted Table
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

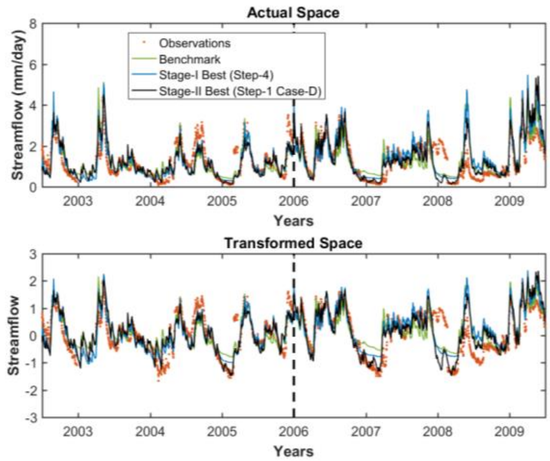


Fig. 10. Time series plots of streamflow for the best simulations in Step I and Step II, the benchmark simulation, ~~**[Insert Fig 8]**~~

[45]

and the observations.

Using the best model from ~~Step~~ Stage-II (HyMod-V2), we next investigate the change in simulation uncertainties for streamflow and AET due to the model structural improvement. ~~Following Roy et al. (*in prep*), we computed the~~The calibration period residual distributions (assumed stationary) were calculated in the λ-transformed space and superimposed ~~them~~ on the daily estimates of the corresponding variables for the total simulation period. Fig. ~~Figure 9~~11 shows the histograms of calibration period residuals for the ~~Benchmark and Final~~benchmark and final (Stage-II Step-1 Case-D) steps ~~(Step II-1d).~~. In both cases (AET and streamflow~~)~~), the residuals become more normally distributed, with the improvement being more prominent for AET. This result is expected, since HyMod-V1 in Step-I showed poor performance in regards to AET. Overall, the structural modification is clearly beneficial.

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt, Not Bold

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

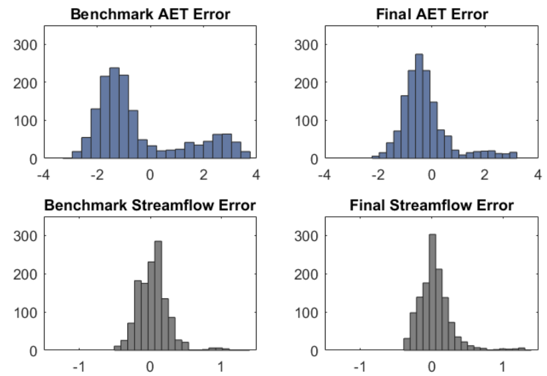


Fig. 11. AET and streamflow error distributions for the benchmark and the final steps.

~~[Insert Fig 9]~~

[46] ~~Figure 10~~ Fig. 12 shows the streamflow and AET time series along with their corresponding 90% confidence intervals for the ~~Benchmark~~benchmark and the ~~Final~~final steps. Both the streamflow and AET simulations improve as a result of the model structural modification. Although the streamflow uncertainty bounds have not narrowed significantly, the flow series is clearly less biased and tracks the recessions better. Meanwhile the AET simulations have improved significantly: (a) the bias has been reduced, (b) the uncertainty bounds are narrower, and (c) the erratic behavior originally seen in the AET simulations (frequent drops to zero) has disappeared. Further, although the improvement in streamflow performance is evident from the statistics in Table ~~4~~5 and 6, the improved behavior is even more apparent in ~~Figure 10~~Fig. 12 where the model can be seen to now track the recessions quite well.

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Space After: 8 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

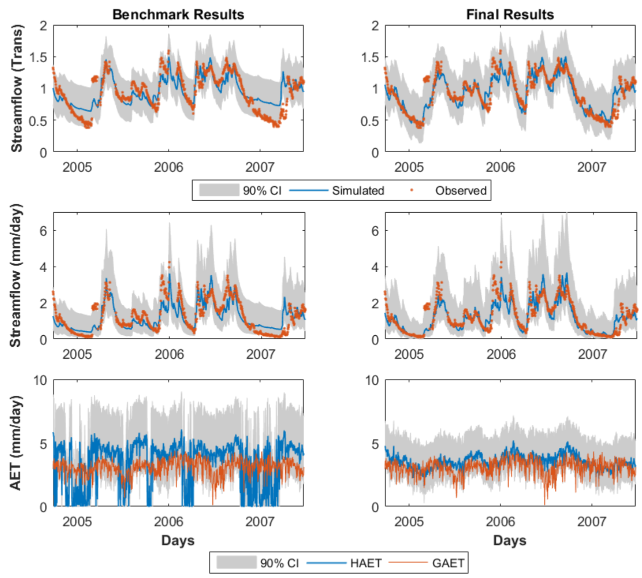


Fig. 12. Time series plots of streamflow (λ-transformed and actual) and AET for the benchmark and the final steps. For clarity, we only show a window of 1000 days.

~~[Insert Fig 10]~~

4 Discussion ~~and Conclusions~~

~~*[47]* This~~ In this study ~~has~~, we have explored two different approaches ~~to~~for using the ~~use of~~ recently available ~~SET data from~~satellite-based AET dataset GLEAM to improve the realism and performance of the conceptual catchment-scale hydrologic model HyMod. In the first approach, ~~SET data were~~ GAET is used as a constraint to ~~constrain~~ the ET ~~estimates~~process equation in the model, while in the second ~~we modified~~approach, the model structure ~~itself. Our study shows~~has been modified so that ~~use of satellite-based information~~ the ET process parameterizations become more physically consistent and realistic. We avoided making the model overly complicated in terms of its structural representations and/or have large number of

Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

parameters, since both of these would defeat its main purpose of being a simple model. Our goal was to increase the realism within the model and improve its performance in simple manners. Furthermore, we also made sure that the improvements in some particular process simulation (e.g. AET) do not deteriorate model's performance for simulating some other process (e.g. streamflow). Our results show that both the approaches (process constraining and structural modification) can ~~clearly~~improve the simulations of streamflow, while the later also significantly improves the AET simulations. Clearly, the satellite-based ET datasets (GLEAM in this case) can significantly benefit the process of hydrologic modeling ~~for~~in poorly gauged basins ~~by providing new sources of information to reduce the epistemic component of model structural uncertainty through improved physical process representation~~.

~~**4.1 Constraining ET**~~

~~*[48]*~~ ~~Use~~ The use of ET data ~~to constrain model simulations~~as a constraint can improve streamflow forecasts, provided some additional processing steps are implemented. ~~Direct insertion of~~ If the GAET ~~into~~data are used directly as a constraint to the ET equation ~~of~~, the ~~HyMod~~ model ~~resulted in~~tends to show bias ~~(Step I-2); the~~ in streamflow simulations. This behavior can be attributed to the fact that once GAET is incorporated, the water balance within the model is altered, the effects of which are reflected in terms of bias in the simulated streamflows. The type of this bias ~~will~~, of course, ~~be~~is subject to change depending on the ~~SET data used.~~dataset. While recalibration of the model ~~improved model~~with the ET constraint improves the performance ~~(Step I-3),~~, it ~~resulted~~can result in ~~a~~ conceptually unrealistic ~~estimate for the storage capacity of the basin. Application of a bias correction to GAET both improved the streamflow forecasts, and resulted in a more realistic estimate for basin storage capacity (Step I-4); use of~~estimates of certain parameters (H in this case). However, the model produces conceptually realistic values of the H parameter if some adjustments are made in the GAET constraints, instead of using them directly. Note that constraint adjustment is not similar to bias correction; for the latter we need the 'ground truth' of actual ET. Therefore, the adjustment process is not necessarily indicative of the presence of any actual bias in GAET estimates. The adjustment factors are model parameters that correspond to the structural deficiencies within the model. They may or may not be necessary as the structure changes. As we have seen in Stage-II Step-4, the constraint adjustment became irrelevant once the structure of the model itself was improved. We also found that a simple adjustment (using a multiplicative ~~and~~ factor) can perform equally well as a more complex alternative (power~~-law type bias correction schemes produced almost similar results, with slightly better error statistics for the latter.~~ function).

~~*[49]* The ET constraining approach can be implemented for real-time forecasting, provided that real-time SET estimates are available. Once the HyMod model structure was appropriately modified to provide good simulations, the model simulations were robust/stable, and the streamflow forecasts did not deteriorate when GAET data was removed. Recalibrating the model (against streamflow) when implemented with GAET assimilation did not improve the model performance when the model was run without GAET (Step I-5). Our results suggest, therefore, the ET constraining approach be implemented only for simulation periods when SET data are available.~~

~~**4.2 Structural Modification**~~

~~*[50]* SET data can prove useful for improving the representation of the ET process in a hydrologic model. Our study was able to derive a simple structural form for HyMod that is robust and enabled the model to produce more accurate estimates of AET. We~~ Improving the model structure provides several other benefits. For example, a model that simulates ET more accurately can be a suitable

Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

candidate for real-time forecasting applications. This type of a model can also prove useful for projecting future water availability. Although ET plays a significant role in the hydrological cycle, traditionally, for conceptual models, the main focus has mainly been drawn towards improving their streamflow simulation performance, while making the ET process overly simplified (e.g. simple water budget). In this study, we show that by incorporating simple but physically consistent structural changes, the ET simulation performance can be improved significantly. In poorly gauged basins, the satellite-based estimates of AET provides useful information to carry out this improvement.

In this study, we tested several conceptually reasonable structural modifications to the model of varying levels of complexity ~~(Step IIa–IId~~and physical basis (Case-A to Case-D), and selected the one that provided the best simulations of both GAET and observed streamflow.

~~[51]~~ We found that relatively simple changes to the ~~HyMod~~model's ET equation significantly improved the ET simulations ~~(~~as assessed by GAET~~). However, while~~. While our goal was to improve the AET and streamflow simulations, we were also careful to ensure that the model parameters remain conceptually realistic. We saw that using a simple multiplicative factor (parameter K) as a resistance to ~~control~~ AET produced excellent streamflow and AET forecasts (~~Step II-1b), it~~Case-B), but resulted in an unrealistic estimate of basin storage capacity~~.~~ (parameter H). In contrast, inclusion of a soil moisture dependent function f(ψ_t), ~~representing the resistance to evaporation,~~ resulted in a more realistic estimate of basin storage capacity without compromising the streamflow and AET simulations. ~~The final model structure (HyMod-V2) establishes a non-linear relationship between AET and evaporative demand.~~

~~[52] Overall, the~~Once the model structure was appropriately modified to provide good simulations, the model simulations were robust/stable, and there was no need to impose the ET constraint (with/without constraint adjustment). The modified model structure provided significantly improved AET forecasts with much narrower uncertainty intervals (see Fig. ~~10~~12), along with reduced bias in streamflow and improved tracking of the streamflow recessions.

~~**4.3 Overall Outlook**~~

~~[53] The validation and total simulation streamflow error statistics from Step I-4 (Table 4) and Step II-1d (Table 6) were quite similar, indicating that both the ET constraining (first approach) and diagnostic structural improvement (second approach) can produce comparable results. However, the latter also resulted in improved AET forecasts, even when SET data were made unavailable. Combining the two schemes did not result in notable additional improvements.~~

~~[54]~~ Overall, by incorporating an additional source of external information in a sensible manner (here by structural modification), the need for calibration can be reduced (note that the model was not calibrated against GAET); see the extensive discussion by Gharari et al., (2014) and Bahremand (2016) on this topic. Nevertheless, given the simplistic nature of the hydrologic models and the large uncertainties that exist therein, some degree of calibration will generally remain important and relevant. We do not mean to imply, therefore, that calibration is not essential, because we will rarely (correctly) know everything we need to know about the system we are modeling. Instead, we should be aware of the strengths and weaknesses involved in the use of calibration and apply it carefully in such a way that useful information is gained about the underlying

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Space After: 6 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

nature of the actual physical system. In this study, we demonstrate the need for both approaches. On the one hand, improving the model structure resulted in improved AET simulations without any need for calibration (to AET). On the other hand, the best streamflow performance was achieved when the structurally modified model~~ structure modification~~ was tuned via a calibration procedure.

~~*[55]*~~ Note that this study is based on testing ~~of a single catchment scale conceptual rainfall-runoff~~the model on a single basin~~,~~ using~~ a single satellite-based precipitation product and~~ a single satellite-based AET product. While not demonstrating universal applicability, the results are clearly indicative and the methodology illustrates how such data can be used to investigate potential improvements to the structures of simple catchment scale models used for hydrologic studies in data scarce regions. A rigorous analyses of the methodology over multiple basins is a potential avenue for future research scope. For more detailed process-based models, the ET process parameters can be calibrated against some reliable ~~SET~~satellite-based AET estimates (e.g. GLEAM), or the process representation itself can be improved by adapting some similar strategies ~~the SET~~that these AET products ~~are based on.~~follow.

4.4 Conclusions

~~*[56]*~~ In conclusion, SET data can be used to improve model performance in different ways. However,~~ data assimilation~~ strategies that result in model structural modifications can generally be expected to provide longer lasting benefits than ones that simply update or constrain the state trajectories of the model. This is because structural modifications can both improve the initial estimates of the state at each time step, and sustain these improvements into future time steps (Bulygina and Gupta, 2009, 2010, 2011; Nearing and Gupta, 2015). In contrast, even though data assimilation to directly improve state estimates can improve model performance, inadequacies in model structure will tend to cause the state estimates to drift away from their ~~ideal~~more appropriate values over time, so that performance deteriorates markedly when the constraining data are not available. Of course, we have only tested a ~~*'constraining'*~~ *“constraining”* strategy to assimilating ET information, which is a relatively simple form of data assimilation (DA) (Houser et al., 1998), and more sophisticated approaches such as the Ensemble Kalman Filter (EnKF) could instead be implemented. However, the efficiency of the EnKF for soil moisture retrieval has been shown to be as low as 30% ~~Nearing et al. (2013a, 2013b)~~(Nearing et al., 2013a, 2013b) and so it is not clear that more sophisticated forms of DA are justified, especially given the large uncertainties associated with both the data and the model structure for this poorly gauged catchment. We leave such investigation for future work.

Code and Data Availability

Data and codes (HyMod-V2 in Matlab) used in this study are available on request from the corresponding author, Tirthankar Roy (royt@email.arizona.edu).

Conflicts of Interests

The authors express no conflicts of interests.

Acknowledgements

This study was supported by the NASA-USAID SERVIR Program through the award 11-SERVIR11-58. The second author acknowledges support by the Australian Centre of Excellence for Climate System Science (CE110001028), and from the EU-funded project ‘‘Sustainable Water Action (SWAN): Building Research Links Between EU and US’’ (INCO-20011-7.6 grant 294947).

Formatted: Space After: 6 pt

APPENDIX

Appendix A: Original HyMod Equations

Formatted: Font: 11 pt

The benchmark version of the spatially lumped conceptual rainfall-runoff model HyMod has six parameters. The model is driven by mean daily precipitation and PET data to generate daily estimates of AET and streamflow. It has two main components, a two-parameter soil moisture accounting (SMA) module based on the Moore (1985) rainfall excess concept, and a linear routing (ROUT) module with parallel quick-flow (fast overland flow) and slow-flow (baseflow) pathways. In the SMA module, the *state variable* (soil moisture storage, C) and the *indicator variable* (storage height, H) are non-linearly related via the following equation (Moore, 1985):

$$C(t) = C_{max}\left(1 - \left(1 - \frac{H(t)}{H_{max}}\right)^{1+b}\right) \qquad \text{(A1)}$$

where the *maximum storage capacity* (C_{max}) and the maximum indicator height (H_{max}) are related as:

$$C_{max} = \frac{H_{max}}{1+b} \qquad \text{(A2)}$$

First, the *initial storage* (C_{beg}) is calculated from the initial indicator height (H_{beg}) using Eq. A1. Next, H_{max} is subtracted from the sum of precipitation (P) and H_{beg} to calculate *overland flow* (OV) as:

$$OV = P + H_{beg} - H_{max} \qquad \text{(A3)}$$

Infiltration (I) is then calculated by subtracting OV from P:

$$I = P - OV \qquad \text{(A4)}$$

and an intermediate indicator height (H_{int}) is computed by adding I to H_{beg}, and used to calculate the *intermediate storage* (C_{int}) via Eq. A2. By subtracting C_{int} from the sum of I and C_{beg} we obtain the *interflow* (IF). Finally, the *total runoff* is obtained by adding together OV and IF.

Finally, the HyMod AET (called HAET) is taken to be the smaller of available water C_{int} and potential demand PET (which is provided as input to the model):

$$HAET = \min\{PET, C_{int}\} \qquad \text{(A5)}$$

and the storage at the end of the time step is computed by subtracting AET from C_{int}:

$$C_{end} = C_{int} - HAET \qquad \text{(A6)}$$

The power coefficients in HyMod ('BE' in Table 5 and 'b' in Eq. A1 & A2) can have values ranging from 0 to infinity. For calibration it is useful to be able to impose finite values to the feasible ranges of the parameters; therefore we applied the following transformation (Eq. A7) which converts the [0,inf) range of parameter BE to the [0,2) range of transformed parameter 'be' so that the search can be conducted on finite range of parameter 'be' instead (similarly for parameter 'b' in Eq. A1 and Eq. A2):

$$BE = \ln(1 - be/2)/\ln(0.5);\ be = [0, 2) \qquad \text{(A7)}$$

Appendix B: The λ-Transformation Used

The λ-transformation on streamflows used in this work is given by the equation:

$$TQ_t = \left(\frac{Q_t}{\mu_{Qobs}}\right)^{\lambda} \qquad (B1)$$

where Q_t and TQ_t represent streamflows in the actual space and the transformed space, μ_{Qobs} is the mean of the observations in the actual space, and λ is the transformation parameter that reduces the skewness. This expression differs slightly from the form $TQ_t = \frac{(Q_t)^{\lambda}-1}{\lambda}$ recommended by (Box and Cox, 1964), in that the flows are normalized by the mean μ_{Qobs} instead of by 1.0 before transformation, and the transformed flows all remain positive. This form works as long as the transformation parameter $\lambda \neq 0$ which is true in our case; if $\lambda = 0$, then one should use $TQ_t = \ln(Q_t)$ as discussed by Box and Cox (1964).

Appendix C: Calibrated HyMod (Actual and Modified) Parameters

The following table provides calibrated parameters of the actual and the modified HyMod models.

Para	**Stage-I Step ~~I~~-1**	**Stage-I Step ~~I~~-3**	**Stage-I Step ~~I~~-4**	**Stage-II Step ~~II-1b~~-1 Case-B**	**Stage-II Step ~~II-1c~~-1 Case-C**	**Stage-II Step ~~II-1d~~-1 Case-D**	**Stage-II Step ~~II~~-3**	**Stage-II Step ~~II~~-4**
H	761.0	17364.0	646.4	9494.0	903.7	866.0	12763.8	878.7
B	1.93	1.95	1.24	1.87	0.29	0.34	1.86	0.32
α	0.48	0.37	0.67	0.38	0.31	0.27	0.45	0.31
Nq	1.44	4.54	1.25	4.22	4.80	4.71	3.51	4.50
Ks	0.00	0.09	0.00	0.10	0.10	0.09	0.10	0.10
Kq	0.10	0.22	0.10	0.19	0.24	0.24	0.18	0.20
Kmax	-	-	-	0.73	1.00	1.00	1.00	1.00
γ	-	-	-	-	-	0.00	1.00	0.00
BE	-	-	-	-	-	0.86	0.06	0.90
a	-	-	1.33	-	-	-	-	1.82
b	-	-	0.93	-	-	-	-	2.00

References

~~Arboleda, A., Ghilain, N. and Gellens-Meulenberghs, F.: The LSA-SAF evapotranspiration product - first results with MSG, in Proceedings of the 2005 EUMETSAT meteorological satellite data user's conference, Dubrovnik, Croatia., 2005.~~

~~Bahremand, A.: HESS Opinions: Advocating process modeling and de-emphasizing parameter estimation, Hydrol. Earth Syst. Sci., 20(4), 1433–1445, doi:10.5194/hess-20-1433-2016, 2016.~~

~~Bastiaanssen, W. G. M., Pelgrum, H., Wang, J., Ma, Y., Moreno, J. F., Roerink, G. J. and van der Wal, T.: A remote sensing surface energy balance algorithm for land (SEBAL)., J. Hydrol., 212-213, 213–229, doi:10.1016/S0022-1694(98)00254-6, 1998.~~

~~Box, G. E. P. and Cox, D. R.: An Analysis of Transformations, J. R. Stat. Soc. Ser. B, 26(2), 211–252, 1964.~~

~~Bulygina, N. and Gupta, H.: Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, Water Resour. Res., 45(12), W00B13, doi:10.1029/2007WR006749, 2009.~~

~~Bulygina, N. and Gupta, H.: How Bayesian data assimilation can be used to estimate the mathematical structure of a model, Stoch. Environ. Res. Risk Assess., 24(6), 925–937, doi:10.1007/s00477-010-0387-y, 2010.~~

~~Bulygina, N. and Gupta, H.: Correcting the mathematical structure of a hydrological model via Bayesian data assimilation, Water Resour. Res., 47(5), W05514, doi:10.1029/2010WR009614, 2011.~~

~~Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. a., Vrugt, J. a., Gupta, H. V., Wagener, T. and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44(12), W00B02, doi:10.1029/2007WR006735, 2008.~~

~~Duan, Q., Sorooshian, S. and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28(4), 1015–1031, doi:10.1029/91WR02985, 1992.~~

~~van Emmerik, T., Mulder, G., Eilander, D., Piet, M. and Savenije, H.: Predicting the ungauged basin: model validation and realism assessment, Front. Earth Sci., 3(October), 1–11, doi:10.3389/feart.2015.00062, 2015.~~

~~Fenicia, F., McDonnell, J. J. and Savenije, H. H. G.: Learning from model improvement: On the contribution of complementary data to process understanding, Water Resour. Res., 44(6), W06419, doi:10.1029/2007WR006386, 2008a.~~

~~Fenicia, F., Savenije, H. H. G., Matgen, P. and Pfister, L.: Understanding catchment behavior through stepwise model concept improvement, Water Resour. Res., 44(1), W01402, doi:10.1029/2006WR005563, 2008b.~~

~~Fenicia, F., Kavetski, D. and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47(11), W11510, doi:10.1029/2010WR010174, 2011.~~

~~Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., Husak, G. J., Michaelsen, J. C. and Verdin, A. P.: A Quasi-Global Precipitation Time Series for Drought Monitoring., 2014.~~

~~García, L., Rodríguez, D., Wijnen, M. and Pakulski, I., Eds.: Earth Observation for Water Resources Management: Current Use and Future Opportunities for the Water Sector, The World Bank., 2016.~~

~~Gash, J. H. C.: An analytical model of rainfall interception by forests, Q. J. R. Meteorol. Soc., 105(443), 43–55, doi:10.1002/qj.49710544304, 1979.~~

Formatted: Font: 11 pt

~~Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H. and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrol. Earth Syst. Sci., 18(12), 4839–4859, doi:10.5194/hess-18-4839-2014, 2014.~~

~~Gupta, H. V and Nearing, G. S.: Debates-the future of hydrological sciences: A (common) path forward? Using models and data to learn: A systems theoretic perspective on the future of hydrological science, Water Resour. Res., 50(6), 5351–5359, doi:10.1002/2013WR015096, 2014.~~

~~Gupta, H. V., Wagener, T. and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22(18), 3802–3813, doi:10.1002/hyp.6989, 2008.~~

~~Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1-2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.~~

~~Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G. and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resour. Res., 48(8), W08301, doi:10.1029/2011WR011044, 2012.~~

~~Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Appl. Eng. Agric., 1(2), 96–99, 1985.~~

~~Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H. and Goodrich, D. C.: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, Water Resour. Res., 34(12), 3405–3420, doi:10.1029/1998WR900001, 1998.~~

~~Hsu, K., Gao, X., Sorooshian, S. and Gupta, H. V.: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks, J. Appl. Meteorol., 36(9), 1176–1190, doi:10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2, 1997.~~

~~Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P. and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, J. Hydrometeorol., 8(1), 38–55, doi:10.1175/JHM560.1, 2007.~~

~~Joyce, R. J., Janowiak, J. E., Arkin, P. a. and Xie, P.: CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution, J. Hydrometeorol., 5(3), 487–503, doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2, 2004.~~

~~Kumar, R., Samaniego, L. and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resour. Res., 49(1), 360–379, doi:10.1029/2012WR012195, 2013.~~

~~Marshall, L., Sharma, A. and Nott, D.: Modeling the catchment via mixtures: Issues of model specification and validation, Water Resour. Res., 42(11), W11409, doi:10.1029/2005WR004613, 2006.~~

~~Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A. and Verhoest, N. E. C.: GLEAM v3.0: satellite-based land evaporation and root-zone soil moisture, 2016.~~

~~Martinez, G. F. and Gupta, H. V.: Hydrologic consistency as a basis for assessing complexity of monthly water balance models for the continental United States, Water Resour. Res., 47(12), W12540, doi:10.1029/2011WR011229, 2011.~~

~~Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A. and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15(2), 453–469, doi:10.5194/hess-15-453-2011, 2011.~~

~~Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F. and Fernaìndez-Prieto, D.: The WACMOS-ET project – Part 2: Evaluation of global terrestrial~~

~~evaporation data sets, Hydrol. Earth Syst. Sci. Discuss., 12(10), 10651–10700, doi:10.5194/hessd-12-10651-2015, 2015.~~

~~Monteith, J. L.: Evaporation and environment, in Proceedings of a symposium of the Society for Experimental Biology, pp. 205–234., 1965.~~

~~Moore, R. J.: The probability-distributed principle and runoff production at point and basin scales, Hydrol. Sci. J., 30(2), 273–297, doi:10.1080/02626668509490989, 1985.~~

~~Nearing, G. S.: Diagnostics And Generalizations Of Parametric State Estimation, The University of Arizona., 2013.~~

~~Nearing, G. S. and Gupta, H. V.: The quantity and quality of information in hydrologic models, Water Resour. Res., 51(1), 524–538, doi:10.1002/2014WR015895, 2015.~~

~~Nearing, G. S., Gupta, H. V., Crow, W. T. and Gong, W.: An approach to quantifying the efficiency of a Bayesian filter, Water Resour. Res., 49(4), 2164–2173, doi:10.1002/wrcr.20177, 2013a.~~

~~Nearing, G. S., Gupta, H. V. and Crow, W. T.: Information loss in approximately Bayesian estimation techniques: A comparison of generative and discriminative approaches to estimating agricultural productivity, J. Hydrol., 507, 163–173, doi:10.1016/j.jhydrol.2013.10.029, 2013b.~~

~~Pokhrel, P., Gupta, H. V. and Wagener, T.: A spatial regularization approach to parameter estimation for a distributed watershed model, Water Resour. Res., 44(12), n/a–n/a, doi:10.1029/2007WR006615, 2008.~~

~~Pokhrel, P., Yilmaz, K. K. and Gupta, H. V.: Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, J. Hydrol., 418-419, 49–60, doi:10.1016/j.jhydrol.2008.12.004, 2012.~~

~~Priestley, C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Mon. Weather Rev., 100(2), 81–92, doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.~~

~~Roy, T., Serrat-Capdevila, A., Gupta, H. and Valdes, J.: A Platform for Probabilistic Multi-model and Multi-product Streamflow Forecasting, *in prep*.~~

~~Savenije, H. H. G.: *HESS Opinions* "Topography driven conceptual modelling (FLEX-Topo)," Hydrol. Earth Syst. Sci., 14(12), 2681–2692, doi:10.5194/hess-14-2681-2010, 2010.~~

~~Schaefli, B., Harman, C. J., Sivapalan, M. and Schymanski, S. J.: *HESS Opinions*: Hydrologic predictions in a changing environment: behavioral modeling, Hydrol. Earth Syst. Sci., 15(2), 635–646, doi:10.5194/hess-15-635-2011, 2011.~~

~~Trambauer, P., Dutra, E., Maskey, S., Werner, M., Pappenberger, F., van Beek, L. P. H. and Uhlenbrook, S.: Comparison of different evaporation estimates over the African continent, Hydrol. Earth Syst. Sci., 18(1), 193–212, doi:10.5194/hess-18-193-2014, 2014.~~

~~Troch, P. A., Lahmers, T., Meira, A., Mukherjee, R., Pedersen, J. W., Roy, T. and Valdés-Pineda, R.: Catchment coevolution: A useful framework for improving predictions of hydrological change?, Water Resour. Res., 51(7), 4903–4922, doi:10.1002/2015WR017032, 2015.~~

~~Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V. and Sorooshian, S.: A framework for development and application of hydrological models, Hydrol. Earth Syst. Sci., 5(1), 13–26, doi:10.5194/hess-5-13-2001, 2001.~~

~~Winsemius, H. C., Savenije, H. H. G. and Bastiaanssen, W. G. M.: Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins?, Hydrol. Earth Syst. Sci., 5(4), 2293–2318, doi:10.5194/hessd-5-2293-2008, 2008.~~

~~WREM: Mara River Basin Monograph. [online] Available from: http://nileis.nilebasin.org/system/files/Mara Monograph.pdf, 2008.~~

~~Zhang, Y., Chiew, F. H. S., Zhang, L. and Li, H.: Use of Remotely Sensed Actual Evapotranspiration to Improve Rainfall–Runoff Modeling in Southeast Australia, J. Hydrometeorol., 10(4), 969–980, doi:10.1175/2009JHM1061.1, 2009.~~

~~TABLES~~

~~Table 1: Performance evaluation metrics used in this study.~~

~~Metrics~~	~~Equations~~
~~Normalized Mean Square Error (NMSE)~~	$\mathrm{MSE} = \mathrm{mean}((\mathrm{O_i} - \mathrm{S_i})^2)$~~; NMSE =~~ $\cancel{\frac{\mathrm{MSE}}{\mathrm{var(O)}}}$
~~Normalized Bias in Mean (NBμ)~~	$\mathrm{NB}\,\mu = \frac{\mathrm{mean(S)} - \mathrm{mean(O)}}{\mathrm{mean(O)}}$

Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt

Normalized Bias in Standard Deviation (NBσ)	$$NB\sigma = \frac{std(S) - std(O)}{std(O)}$$
Correlation Coefficient (ρ)	$$\rho = \frac{\sum_{i=1}^{N}(O_i - mean(O))(S_i - mean(S))}{\sqrt{\sum_{i=1}^{N}(O_i - mean(O))^2 \sum_{i=1}^{N}(S_i - mean(S))^2}}$$
O: Observed flows; S: Simulated flows; N: Number of data points.	

Table 2: Descriptive statistics of GAET and HAET.

Statistics	GAET	HAET
Maximum	4.62	6.12
Minimum	0.119	0.00
Mean	3.03	3.52
Median	3.08	4.21
Mode	0.11	0.00
Std. Dev.	0.59	1.72
Skewness	-0.58	-1.03
Kurtosis	3.84	2.62

Table 3: Annual mean of AETs and HPET.

Source	Annual Mean (mm)
GAET	1100
HAET	1263
WBAET	1146
HPET	1704

Table 4: Streamflow error statistics for calibration, evaluation, and total simulation (in parenthesis) for all five different cases in Stage I analysis.

Calibration					
Metrics	**Step I-1**	**Step I-2**	**Step I-3**	**Step I-4**	**Step I-5**
NMSE	0.56	1.68	0.64	0.43	0.60
NBμ	0.09	0.32	0.12	0.09	-0.09
NBσ	-0.12	-0.15	-0.19	-0.06	-0.04
R	0.76	0.81	0.74	0.83	0.75
Evaluation (Total Simulation)					

Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: List Paragraph
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: List Paragraph
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted Table
Formatted: Font: 11 pt, Bold
Formatted: Font: 11 pt
Formatted Table
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted: Font: 11 pt
Formatted Table

Metrics		Step I-1	Step I-2	Step I-3	Step I-4	Step I-5
NMSE	0.84 (0.77)	2.13 (2.17)	0.92 (1.19)	0.88 (0.75)		0.64 (0.56)
NBμ	0.14 (0.14)	0.38 (0.38)	0.09 (0.22)	0.15 (0.16)		-0.01 (-0.02)
NBσ	-0.04 (-0.08)	-0.03 (-0.10)	0.04 (-0.01)	0.17 (0.02)		0.14 (0.05)
R		0.66 (0.72)	0.71 (0.76)	0.61 (0.69)	0.74 (0.78)	0.73 (0.74)

Table 5~~: K-function in different cases.~~

Cases	K_{max}	K_{min}	$f(\psi_t)$	Additional Parameters
a	1	0	1	None
b	K_0	0	1	K_0
c	K_0	0	X_t/X_{max}	K_0
d	K_{max}	$\gamma \cdot K_{max}$	$(X_t/X_{max})^{BE}$	K_{max}, γ, BE

~~Table 6: Streamflow error statistics for calibration period, evaluation period, and total simulation period (in parenthesis)~~ for all four different cases in Stage II-1 analysis.

Calibration				
Metrics	Step II-1a	Step II-1b	Step II-1c	Step II-1d
NMSE	0.56	0.52	0.51	0.51
NBμ	0.09	0.10	0.10	0.10
NBσ	-0.12	-0.19	-0.04	-0.04
R	0.76	0.77	0.80	0.80
Evaluation (Total Simulation)				
Metrics	Step II-1a	Step II-1b	Step II-1c	Step II-1d
NMSE	0.84 (0.77)	0.65 (0.77)	0.88 (0.72)	0.84 (0.70)
NBμ	0.14 (0.14)	0.07 (0.16)	0.14 (0.14)	0.14 (0.13)
NBσ	-0.04 (-0.08)	0.00 (-0.06)	0.16 (0.10)	0.16 (0.10)
R	0.66 (0.72)	0.70 (0.75)	0.73 (0.78)	0.74 (0.78)

~~Table 7: Streamflow and AET error statistics in total simulation for all four cases in Step II-1.~~

Metric	Step II-1a	Step II-1b	Step II-1c	Step II-1d
	Streamflow			
NMSE	0.77	0.77	0.72	0.70
NBμ	0.14	0.16	0.14	0.13
NBσ	-0.08	-0.06	0.10	0.10
R	0.72	0.75	0.78	0.78
	AET			
NMSE	8.93	1.28	1.70	1.71

- **Formatted:** Font: 11 pt
- **Formatted Table**
- **Formatted:** Font: 11 pt
- **Formatted:** Font: 11 pt
- **Formatted:** Font: 11 pt
- **Formatted** ...
- **Formatted** ...
- **Formatted** ...
- **Formatted:** Font: 11 pt
- **Formatted** ...
- **Formatted** ...
- **Formatted:** Font: 11 pt
- **Formatted** ...
- **Formatted** ...
- **Formatted** ...
- **Formatted:** Font: 11 pt
- **Formatted** ...
- **Formatted** ...
- **Formatted** ...
- **Formatted** ...
- **Formatted:** Font: 11 pt
- **Formatted:** Font: 11 pt
- **Formatted:** Font: 11 pt
- **Formatted Table**
- **Formatted:** Font: 11 pt, Bold
- **Formatted:** Font: 11 pt
- **Formatted Table**
- **Formatted:** Font: 11 pt
- **Formatted:** Font: 11 pt
- **Formatted:** Font: 11 pt
- **Formatted Table**
- **Formatted:** Font: 11 pt
- **Formatted Table**
- **Formatted:** Font: 11 pt
- **Formatted:** Font: 11 pt
- **Formatted:** Font: 11 pt
- **Formatted:** Font: 11 pt
- **Formatted:** Font: 11 pt
- **Formatted:** Font: 11 pt

~~NBμ~~	~~0.18~~	~~0.12~~	~~0.17~~	~~0.17~~
~~NBσ~~	~~1.89~~	~~-0.26~~	~~-0.10~~	~~-0.13~~
~~R~~	~~0.22~~	~~0.43~~	~~0.48~~	~~0.49~~

~~Table 8: Streamflow error statistics for calibration, evaluation, and total simulation (in parenthesis) for all five different cases in Stage II analysis.~~

~~Calibration~~					
~~Metrics~~	~~Step II-1~~	~~Step II-2~~	~~Step II-3~~	~~Step II-4~~	~~Step II-5~~
~~NMSE~~	~~0.51~~	~~1.82~~	~~0.64~~	~~0.51~~	~~0.50~~
~~NBμ~~	~~0.10~~	~~0.32~~	~~0.12~~	~~0.10~~	~~0.10~~
~~NBσ~~	~~-0.04~~	~~0.16~~	~~-0.19~~	~~-0.04~~	~~-0.04~~
~~R~~	~~0.80~~	~~0.81~~	~~0.74~~	~~0.80~~	~~0.80~~
~~Evaluation (Total Simulation)~~					
~~Metrics~~	~~Step II-1~~	~~Step II-2~~	~~Step II-3~~	~~Step II-4~~	~~Step II-5~~
~~NMSE~~	~~0.84 (0.70)~~	~~2.46 (2.29)~~	~~0.92 (1.19)~~	~~0.84 (0.70)~~	~~0.84 (0.70)~~
~~NBμ~~	~~0.14 (0.13)~~	~~0.37 (0.36)~~	~~0.10 (0.22)~~	~~0.14 (0.14)~~	~~0.14 (0.13)~~
~~NBσ~~	~~0.16 (0.10)~~	~~0.38 (0.31)~~	~~0.05 (-0.01)~~	~~0.16 (0.10)~~	~~0.16 (0.09)~~
~~R~~	~~0.74 (0.78)~~	~~0.70 (0.77)~~	~~0.61 (0.69)~~	~~0.74 (0.78)~~	~~0.74 (0.78)~~

~~FIGURES~~

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted Table

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

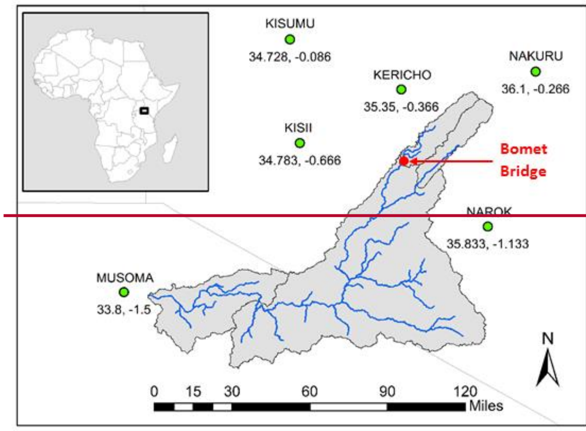
Formatted Table

Formatted: Font: 11 pt

Formatted Table

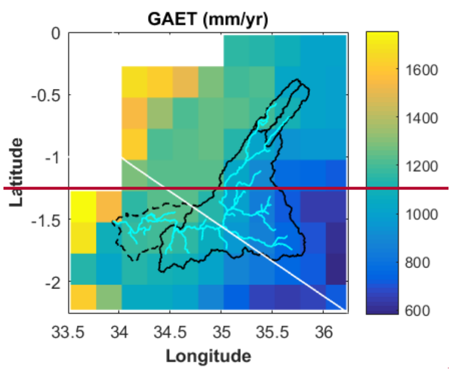
Formatted: Font: 11 pt

Formatted: Font: 11 pt



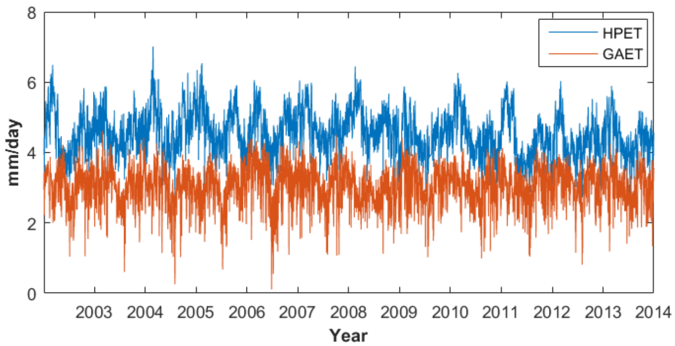
~~**Fig. 1.** The Mara River basin and the Nyangores River sub-basin. The discharge station is located at Bomet Bridge (red dot). Meteorological stations (green dots) are located in the surrounding regions (Adapted from Roy et al., *in prep*).~~

Formatted: Font: 11 pt



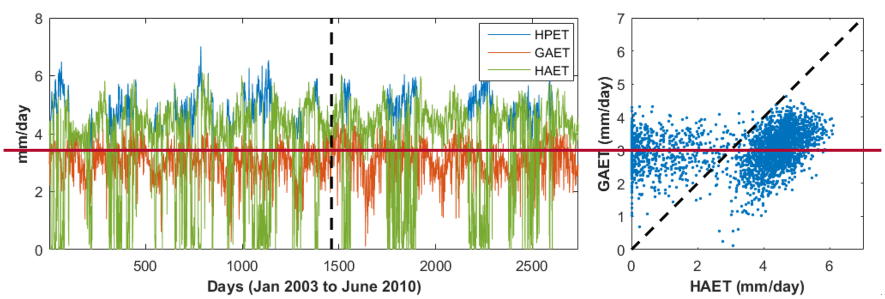
~~**Fig. 2.** Annual mean of GAET over the entire MRB.~~

Formatted: Font: 11 pt



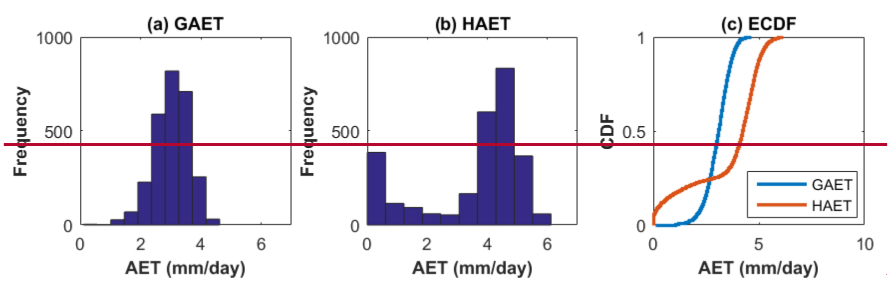
~~Fig. 3. Time series of HPET and GAET for the NRB.~~

Formatted: Font: 11 pt



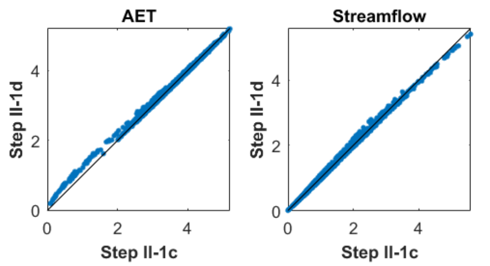
~~Fig. 4. Time series and scatter plots of HPET, GAET, and HAET.~~

Formatted: Font: 11 pt



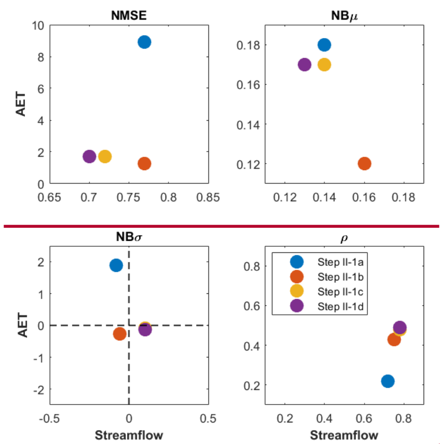
~~Fig. 5. Histogram and ECDF plots of GAET and HAET.~~

Formatted: Font: 11 pt



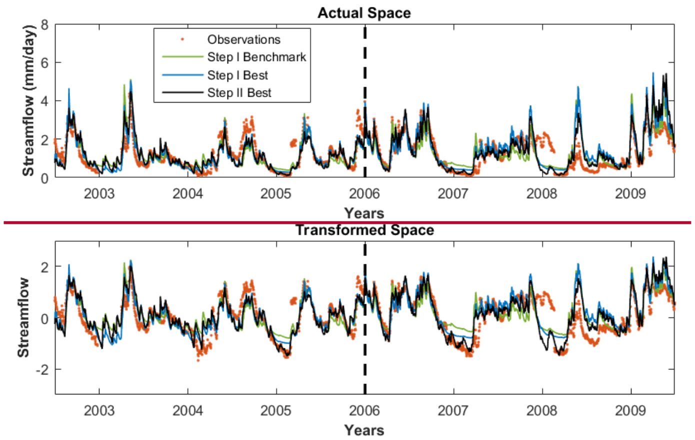
~~Fig.~~ 6. Scatter plots of streamflow and AET from Step II-1a and Step II-1b.

Formatted: Font: 11 pt

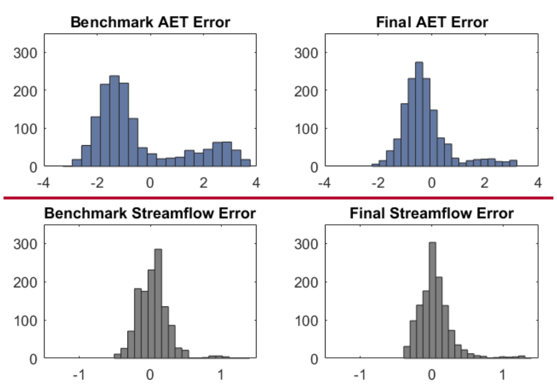


~~Fig.~~ 7. Streamflow and AET error statistics in total simulation for all four cases in Step II-1.

Formatted: Font: 11 pt, Not Bold



~~**Fig. 8.** Time series plots of streamflow for the best simulations in Step I and Step II, the benchmark simulation, and the observations.~~



~~**Fig. 9**~~Arboleda, A., Ghilain, N. and Gellens-Meulenberghs, F.: The LSA-SAF evapotranspiration product - first results with MSG, in Proceedings of the 2005 EUMETSAT meteorological satellite data user’s conference, Dubrovnik, Croatia., 2005.

Bahremand, A.: HESS Opinions: Advocating process modeling and de-emphasizing parameter estimation, Hydrol. Earth Syst. Sci., 20(4), 1433–1445, doi:10.5194/hess-20-1433-2016, 2016.

Bastiaanssen, W. G. M., Pelgrum, H., Wang, J., Ma, Y., Moreno, J. F., Roerink, G. J. and van der Wal, T.: A remote sensing surface energy balance algorithm for land (SEBAL)., J. Hydrol., 212–213, 213–229, doi:10.1016/S0022-1694(98)00254-6, 1998.

Box, G. E. P. and Cox, D. R.: An Analysis of Transformations, J. R. Stat. Soc. Ser. B, 26(2), 211–252, 1964.

Boyle, D. P., Gupta, H. V. and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of

Formatted: Font: 11 pt

Formatted: Font: 11 pt

manual and automatic methods, Water Resour. Res., 36(12), 3663–3674, doi:10.1029/2000WR900207, 2000.

Bulygina, N. and Gupta, H.: Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, Water Resour. Res., 45(12), W00B13, doi:10.1029/2007WR006749, 2009.

Bulygina, N. and Gupta, H.: How Bayesian data assimilation can be used to estimate the mathematical structure of a model, Stoch. Environ. Res. Risk Assess., 24(6), 925–937, doi:10.1007/s00477-010-0387-y, 2010.

Bulygina, N. and Gupta, H.: Correcting the mathematical structure of a hydrological model via Bayesian data assimilation, Water Resour. Res., 47(5), W05514, doi:10.1029/2010WR009614, 2011.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. a., Vrugt, J. a., Gupta, H. V., Wagener, T. and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44(12), W00B02, doi:10.1029/2007WR006735, 2008.

Demaria, E. and Serrat-Capdevila, A.: Validation of remote sensing-estimated hydrometeorological variables, in Earth Observation for Water Resources Management: Current Use and Future Opportunities for the Water Sector, edited by L. E. García, D. J. Rodríguez, M. Wijnen, and I. Pakulski, The World Bank., 2016.

Duan, Q., Sorooshian, S. and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28(4), 1015–1031, doi:10.1029/91WR02985, 1992.

Duan, Q., Ajami, N. K., Gao, X. and Sorooshian, S.: Multi-model ensemble hydrologic prediction using Bayesian model averaging, Adv. Water Resour., 30(5), 1371–1386, doi:10.1016/j.advwatres.2006.11.014, 2007.

van Emmerik, T., Mulder, G., Eilander, D., Piet, M. and Savenije, H.: Predicting the ungauged basin: model validation and realism assessment, Front. Earth Sci., 3(October), 1–11, doi:10.3389/feart.2015.00062, 2015.

Fenicia, F., McDonnell, J. J. and Savenije, H. H. G.: Learning from model improvement: On the contribution of complementary data to process understanding, Water Resour. Res., 44(6), W06419, doi:10.1029/2007WR006386, 2008a.

Fenicia, F., Savenije, H. H. G., Matgen, P. and Pfister, L.: Understanding catchment behavior through stepwise model concept improvement, Water Resour. Res., 44(1), W01402, doi:10.1029/2006WR005563, 2008b.

Fenicia, F., Kavetski, D. and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47(11), W11510, doi:10.1029/2010WR010174, 2011.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Sci. Data, 2, 150066, doi:10.1038/sdata.2015.66, 2015.

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., Husak, G. J., Michaelsen, J. C. and Verdin, A. P.: A Quasi-Global Precipitation Time Series for Drought Monitoring., 2014.

Gash, J. H. C.: An analytical model of rainfall interception by forests, Q. J. R. Meteorol. Soc., 105(443), 43–55, doi:10.1002/qj.49710544304, 1979.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H. and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrol. Earth Syst. Sci., 18(12), 4839–4859, doi:10.5194/hess-18-4839-2014, 2014.

Gupta, H. V and Nearing, G. S.: Debates-the future of hydrological sciences: A (common) path forward? Using models and data to learn: A systems theoretic perspective on the future of hydrological science, Water Resour. Res., 50(6), 5351–5359, doi:10.1002/2013WR015096, 2014.

Gupta, H. V., Wagener, T. and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22(18), 3802–3813, doi:10.1002/hyp.6989, 2008.

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G. and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resour. Res., 48(8), W08301, doi:10.1029/2011WR011044, 2012.

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Appl. Eng. Agric., 1(2), 96–99, 1985.

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H. and Goodrich, D. C.: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, Water Resour. Res., 34(12), 3405–3420, doi:10.1029/1998WR900001, 1998.

Hsu, K., Gao, X., Sorooshian, S. and Gupta, H. V.: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks, J. Appl. Meteorol., 36(9), 1176–1190, doi:10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2, 1997.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P. and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, J. Hydrometeorol., 8(1), 38–55, doi:10.1175/JHM560.1, 2007.

Joyce, R. J., Janowiak, J. E., Arkin, P. a. and Xie, P.: CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution, J. Hydrometeorol., 5(3), 487–503, doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2, 2004.

Kumar, R., Samaniego, L. and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resour. Res., 49(1), 360–379, doi:10.1029/2012WR012195, 2013.

Marshall, L., Sharma, A. and Nott, D.: Modeling the catchment via mixtures: Issues of model specification and validation, Water Resour. Res., 42(11), W11409, doi:10.1029/2005WR004613, 2006.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A. and Verhoest, N. E. C.: GLEAM v3.0: satellite-based land evaporation and root-zone soil moisture, 2016.

Martinez, G. F. and Gupta, H. V.: Hydrologic consistency as a basis for assessing complexity of monthly water balance models for the continental United States, Water Resour. Res., 47(12), W12540, doi:10.1029/2011WR011229, 2011.

McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D. and Wood, E. F.: The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally gridded forcing data, Geosci. Model Dev., 9(1), 283–305, doi:10.5194/gmd-9-283-2016, 2016.

Michel, D., Jiménez, C., Miralles, D. G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., McCabe, M. F., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F. and Fernández-Prieto, D.: The WACMOS-ET project – Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms, Hydrol. Earth Syst. Sci., 20(2), 803–822, doi:10.5194/hess-20-803-2016, 2016.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A. and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15(2), 453–469, doi:10.5194/hess-15-453-2011, 2011.

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F. and Fernández-Prieto, D.: The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sci., 20(2), 823–842, doi:10.5194/hess-20-823-2016, 2016.

Monteith, J. L.: Evaporation and environment, in Proceedings of a symposium of the Society for Experimental Biology, pp. 205–234., 1965.

Moore, R. J.: The probability-distributed principle and runoff production at point and basin scales, Hydrol. Sci. J., 30(2), 273–297, doi:10.1080/02626668509490989, 1985.

Moradkhani, H., Hsu, K.-L., Gupta, H. and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resour. Res., 41(5), n/a-n/a, doi:10.1029/2004WR003604, 2005.

Nearing, G. S.: Diagnostics And Generalizations Of Parametric State Estimation, The University of Arizona., 2013.

Nearing, G. S. and Gupta, H. V.: The quantity and quality of information in hydrologic models, Water Resour. Res., 51(1), 524–538, doi:10.1002/2014WR015895, 2015.

Nearing, G. S., Gupta, H. V., Crow, W. T. and Gong, W.: An approach to quantifying the efficiency of a Bayesian filter, Water Resour. Res., 49(4), 2164–2173, doi:10.1002/wrcr.20177, 2013a.

Nearing, G. S., Gupta, H. V. and Crow, W. T.: Information loss in approximately Bayesian estimation techniques: A comparison of generative and discriminative approaches to estimating agricultural productivity, J. Hydrol., 507, 163–173, doi:10.1016/j.jhydrol.2013.10.029, 2013b.

Pokhrel, P., Gupta, H. V. and Wagener, T.: A spatial regularization approach to parameter estimation for a distributed watershed model, Water Resour. Res., 44(12), n/a-n/a, doi:10.1029/2007WR006615, 2008.

Pokhrel, P., Yilmaz, K. K. and Gupta, H. V.: Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, J. Hydrol., 418–419, 49–60, doi:10.1016/j.jhydrol.2008.12.004, 2012.

Priestley, C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Mon. Weather Rev., 100(2), 81–92, doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

Razavi, S. and Gupta, H. V.: A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application, Water Resour. Res., 52(1), 440–455, doi:10.1002/2015WR017559, 2016.

Savenije, H. H. G.: *HESS Opinions* "Topography driven conceptual modelling (FLEX-Topo)," Hydrol. Earth Syst. Sci., 14(12), 2681–2692, doi:10.5194/hess-14-2681-2010, 2010.

Schaefli, B., Harman, C. J., Sivapalan, M. and Schymanski, S. J.: *HESS Opinions*: Hydrologic predictions in a changing environment: behavioral modeling, Hydrol. Earth Syst. Sci., 15(2), 635–646, doi:10.5194/hess-15-635-2011, 2011.

Trambauer, P., Dutra, E., Maskey, S., Werner, M., Pappenberger, F., van Beek, L. P. H. and Uhlenbrook, S.: Comparison of different evaporation estimates over the African continent, Hydrol. Earth Syst. Sci., 18(1), 193–212, doi:10.5194/hess-18-193-2014, 2014.

Troch, P. A., Lahmers, T., Meira, A., Mukherjee, R., Pedersen, J. W., Roy, T. and Valdés-Pineda, R.: Catchment coevolution: A useful framework for improving predictions of hydrological change?, Water Resour. Res., 51(7), 4903–4922, doi:10.1002/2015WR017032, 2015.

Vrugt, J. A., Gupta, H. V., Bouten, W. and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resour. Res., 39(8), n/a-n/a, doi:10.1029/2002WR001642, 2003.

Vrugt, J. A., Gupta, H. V., Dekker, S. C., Sorooshian, S., Wagener, T. and Bouten, W.: Application of stochastic parameter optimization to the Sacramento Soil Moisture Accounting model, J. Hydrol., 325(1–4), 288–307, doi:10.1016/j.jhydrol.2005.10.041, 2006.

Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V. and Robinson, B. A.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, Stoch. Environ. Res. Risk Assess., 23(7), 1011–1026, doi:10.1007/s00477-008-0274-y, 2009.

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V. and Sorooshian, S.: A framework for development and application of hydrological models, Hydrol. Earth Syst. Sci., 5(1), 13–26, doi:10.5194/hess-5-13-2001, 2001.

Wang, D., Chen, Y. and Cai, X.: State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter, Water Resour. Res., 45(11), n/a-n/a, doi:10.1029/2008WR007401, 2009.

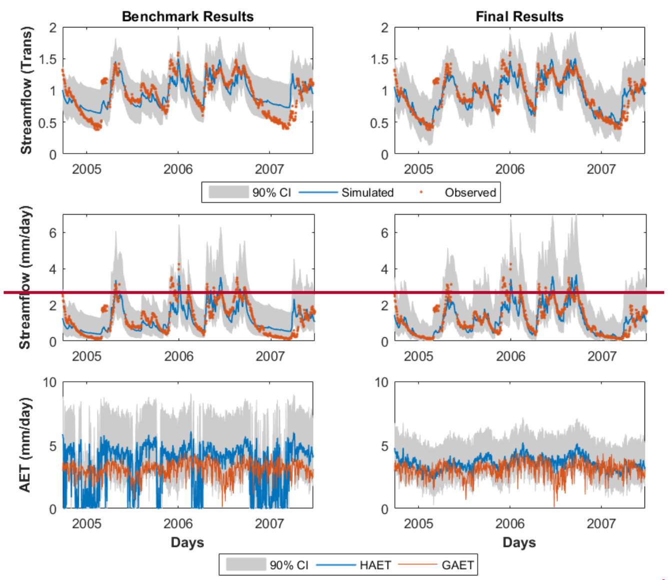
Winsemius, H. C., Savenije, H. H. G. and Bastiaanssen, W. G. M.: Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins?, Hydrol. Earth Syst. Sci., 5(4), 2293–2318, doi:10.5194/hessd-5-2293-2008, 2008.

WREM: Mara River Basin Monograph. [online] Available from: http://nileis.nilebasin.org/system/files/Mara Monograph.pdf, 2008.

Zhang, Y., Chiew, F. H. S., Zhang, L. and Li, H.: Use of Remotely Sensed Actual Evapotranspiration to Improve Rainfall–Runoff Modeling in Southeast Australia, J. Hydrometeorol., 10(4), 969–980, doi:10.1175/2009JHM1061.1, 2009.

~~, AET and streamflow error distributions for the benchmark and the final steps.~~

Formatted: Font: 11 pt



~~Fig. 10.~~ ~~Time series plots of streamflow (λ-transformed and actual) and AET for the benchmark and the final (Stage II-1d) steps. For clarity, we only show a window of 1000 days.~~

Formatted: Font: 11 pt