Hydrology and Earth System Sciences
Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
https://doi.org/10.5194/hess-2021-428-RC2, 2021

Comment on hess-2021-428

Anonymous Referee #2

Referee comment on "Remote sensing-aided rainfall–runoff modeling in the tropics of Costa Rica" by Saúl Arciniega-Esparza et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-428-RC2, 2021

The overarching goal of the manuscript was to use the HPYE model to simulate Costa Rica's catchments with four different model configurations. The authors used remote sensing data as the forcings of the model, in particular, the CHIRPS data for rainfall and MODIS for evaporation. In addition to the usual daily streamflow calibration, the authors incorporated monthly streamflow, PET, and AET data into the model calibration. The use of remote sensing data into hydrological models and their performance is of significant interest to the hydrological community and is important to the local governments with limited access to ground monitoring. The methods are well presented, but the manuscript could benefit from additional information. In particular, a clear evaluation of the model differences and a better time series evaluation would improve it significantly.

1. There is a lack of connection between the supplementary section and the main text. For instance, when the authors introduced the CHIRPS product (~L240), they could link it to Fig. S1 to have a clear picture of the improvement. Another example are tables S1 and S2, which are not mentioned anywhere in the text but would be a useful reference in the discussion section where the authors discuss these hydrological signatures for all models.
2. The abstract needs to be improved by including some of the nice statistics and results from the paper that quantify the improvements. Around L23, the authors talk about the hydrological signatures and that using both daily and monthly streamflow is better than just using the daily flows. However, it is not clear by how much.
3. The authors do not specify which model configuration is the baseline (which I assume is M1). Furthermore, while they present performance statistics, it is unclear if these differences are statistically significant to merit the additional data. Moreover, when they discuss the time-series analysis and the differences between the models, they do so in a descriptive manner to quantify it better. For instance, using a distance metric to evaluate series similarity to the observed data. See DOI: 10.1016/j.rse.2011.06.020 for a summary of some useful metrics. My suggestion would be to make plots of the Mahalanobis distance rather than presenting the original time series (or in addition to Fig. 8).
4. I believe that the first objective should be merged into the other objectives. Running the model (independently of the computer language used) is a trivial objective as it is met from the start of the project.
5. The authors need to explain how they did the catchment extraction in GRASS by

providing additional detail into the used parameters. Also, they need to explain the IDW method in the methods section, define the acronym, and add a reference.
6. The authors need to improve Fig. 3; interpreting it is confusing. Perhaps it would be best to have it with 4 rows rather than arrows, even if there is a degree of repetition.
7. Can the authors modify the presentation of the 86 parameters in L331? It is hard to understand; I would suggest presenting the numbers in parenthesis as the main parameter numbers and then elaborating on how many were linked to soil types, land cover, etc.
8. Can the authors add box plots of the other statistics as supplementary? It is hard to visualize them as isolated numbers. Again, can the authors perform tests of significance on the statistics to determine a significant difference between them?
9. Can the authors mention what the criteria for defining a KGE of 0.5 as acceptable were?
10. Around L605, the authors mention that the corrected temperature improved model performance. The authors need to quantify this performance increase.
11. The authors mention that the streamflow overestimation can be related to a precipitation bias in CHIRPSc. However, from Fig. S1, this does not seem to be the case.
12. When discussing model improvement, please quantify it. The authors mention in L635 that M3 and M4 showed better and more realistic results but failed to quantify the improvement. Moreover, from Fig. 10, it seems that even though KGE was higher for M3 and M4, M1 was able to reproduce the actual spatial distributions of PET and AET better, overlapping more with the observed ranges.
13. In the discussion section, the authors mention that adding PET and AET to the calibration improved model representativeness and link earlier studies. The authors need to also link this assertion to their study, which is one of their objectives.
14. Due to missing tests, I do not see how the authors can conclude that M3 and M4 are better configurations since the statistical significance of the differences has not been evaluated. And in fact, for a lot of the variables, it seemed that M1 performed adequately well compared to M3 and M4. The authors can further support the increased accuracy of M3 and M4 by their link to the FDC information.
15. Finally, I suggest adding ": A case study in Costa Rica." to the title since it was the only region analyzed in the manuscript.
Around L54, the authors mention the opportunities from including additional variables. Please, specify which variables or give a few examples.

Technical corrections:

Around L56, the authors mention that more realistic hydrological partitioning comes at the expense of increased computational cost. Can the authors quantify the time penalties involved?
Around L77, do the authors mean simple bucket models? Any model can be a black-box model.
Around L87, the authors mention that the coarse spatial resolution of the climatological data is an important source of error. Can the authors mention the related uncertainty in the data? (i.e., how much of the model error is associated with the coarse spatial resolution).
L135, the authors mentioned that they merged land covers. Can the authors include how much each merged class contributed to the overall classification?
L159, please remind the reader what both sides are.
L175, please quantify the statement; how well did MODIS compare with the ground data? State r2 or another statistic.
L209, can the authors mention how they chose the soil layer thickness?
L245, the correction factor appears as B in equation one; it appears as BF and BF2 in

equations 2 and 3.
L266, from the text, it is somewhat ambiguous if y refers to each year or the whole period.
L288, the authors should mention that the parameters for correction are part of a monte Carlo simulation and are set to the ranges in Table 3.
L290, sine function.
L320, can the authors justify why only two years were used as a warm-up?
L361, please remind the reader which time series.
L375-379, this information should appear in the introduction.
L405, can the authors normalize the MAE by the mean precipitation? Doing so would help the reader to understand the relative magnitude of the MAE.
L415, please, specify how they affect the performance.
L532, from Fig. 9, it seems that all the models underestimated the real flows to some extent. Is this due to CHIRPSc?
L575, can the authors increase the border thickness of the catchments of Fig. 10? It isn't easy to see them.
L619, is the deviation a positive or negative bias?
L644, what do the authors mean by increased parameter sensitivity?
L650, Can the authors comment why none of the models at the best performing catchment could reproduce the decrease in water content between 2014-2015?