Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-82-RC2, 2020

Hydrology and Earth System Sciences Discussions
Open Access

Interactive comment on "Using NDII pattern for a semi-distributed rainfall-runoff model in tropical nested catchments" *by* Nutchanart Sriwongsitanon et al.

Anonymous Referee #2

Received and published: 25 June 2020

I have finished my review of the paper "Using NDII pattern for a semi-distributed rainfall-runoff model in tropical nested catchments", by Sriwongsitanon et al., submitted to HESS. This paper outlines a comparison study of four models of the same set of nested catchments in Thailand – a lumped model (FLEXL) applied to individual gauges, the semi-distributed version of the same model (FLEX-SD), FLEX-SD modified by using the NDII remote sensing metric to inform the distribution of soil stores (FLEX-SD-NDII), and the independent semi-distributed URBS model. An attempt is made to demonstrate (1) the improved accuracy/realism of using NDII to inform the spatial distribution of soil stores while only calibrating a reference storage quantity and (2) the superiority of FLEX variants over the URBS model. This short paper is generally well-written but

I have found it to include critical experimental design issues, flawed interpretation of results, and procedural omissions. I therefore believe that it is not currently acceptable for publication in HESS, and recommend rejection. I outline the reasons for this below.

Major Comments:

1) The authors only report calibration statistics and performance for their models. There is no effort to validate the models using even standard split-sample validation. Calibration without validation is no longer generally deemed acceptable practice for evaluating hydrological model performance or individual model choices and renders most of the paper results not particularly convincing. In particular, the arguments that the NDII-informed model is "more realistic" due to improved performance in calibration alone are unjustified.

2) At multiple points in the paper, the authors report that the model "gained realism" (e.g., line 15 & 17 of pg 10)– I look at figures 4 and A2 and see only an improvement in baseflow simulation in basins P.20 and P.21 (the only headwater basins unaffected by the Mae Ngad dam overwriting of flows); this is consistent with the quantitative KGE_L metrics which are much more objective assessment of model skill. However, the KGE_L metrics denote a degradation of baseflow in 3 other non-headwater basins – is this still therefore a gain in model realism? Here, the authors can discuss hydrograph fit (in calibration conditions only) but there is no evidence that the mode "gained realism", and I'm rather sure that this is not something that could ever be determined via observation of a hydrograph alone. This is the primary contribution of section 5.1 and it is not defensible from the experimental data. Interpretation of these results seem cherry-picked. Alternate interpretations of the data in table 4 and figures 3,4,5 likewise denote quite inconsistent performance of the FLEX-SD-NDII except for low flow in the 2 headwater basins. Had the authors calibrated the FLEX-SD model using weighted calibration all gauges of interest rather than just P.1 I expect they could get comparable performance without NDII; this likely should have been another test model configuration.

3) The comparison of the correlations of SWI and soil storage characteristics in figure 7 (the other primary piece of evidence supporting the improvement incurred by distributing soil storage information) is likewise unconvincing. All models exhibit the same trend, and the moderate improvements in R2 metrics for a cloud of points about a power law curve (especially when the data relation does not look to have the shape of a power law) is not sufficient to demonstrate model improvement. Again, this is particularly true in light of the fact that all evaluations are done during the calibration period.

4) The authors attribute (at line 25 of pg 9) model discrepancies to observed due to flow regulation at the Mae Ngad Dam upstream of P67. However, they state at line 3 of page 4 that reservoir outflow data was explicitly used as input to model. This flow regulation should therefore be perfectly handled within the model (and in fact would inflate model fit statistics, since the reservoir flows would comprise a rather large portion of the hydrograph, especially under low flow conditions).

5) The reporting of the calibration process is inadequate. What was the calibration period? What was the objective function? How did the authors separate low flow statistics (KGE_L) and high flow statistics (KGE_E)? Was a run-up period used? Why not? Is this hourly NSE or daily NSE? Was the model run at an hourly time step as implied by the use of hourly time lags? The MOSCEM optimization algorithm is uncited. There were a significant number of critical details missing that ensure that these experiments are not replicable.

6) The value of including the URBS model in this comparison is unclear.

All in all, I believe that the paper's approach for distributing soil storage capacities using information gleaned from NDII may have merit, but the experimental design did not clearly demonstrate that this approach actually works for the reasons above. While it was demonstrated via the data that moving from a lumped to a distributed approach somewhat improved model performance (which is not entirely surprising, as additional routing information is included with additional free parameters), this is not new.

Minor Comments:

I have not included the very many minor adjustments I would suggest if I suggested revision, but have touched upon the larger minor issues.

1) In this short 11 page paper, 2.5 pages (plus 2 full page tables and a figure) were dedicated to defining the models already documented elsewhere in Sriwonsitanon et al. 2016 and Carroll 2004. A simple reference would do for most of these details (especially for snow simulation in Thailand!)

2) In presentation of calibrated parameters of table 5, mixing actual calibrated parameters with those calculated from area scaling (Tlag) and the NDII relation (Sumax). The catchment wide Sumax (a calibration parameter) not reported. For some reason the Sumax_i from all of the basins with FLEX-SD-NDII are all less than FLEX-SD, which struck me at odd.

3) observations of improved performance of NDII model @ pg 10 ln 1 not consistent with KGE_L reporting for same basin

4) in table 4, the "best performance is underlined", however, this is not the case, as FLEXL often has best performance.

5) Figure three has little value – a percent bias reporting would be more succinct and equally valuable.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-82, 2020.