Referee comment on the paper:

Using NDII patterns to constrain semi-distributed rainfall-runoff models in tropical nested catchments

By Nutchanart Sriwongsitanon, Wasana Jandang, James Williams, Thienchart Suwawong, Ekkarin Maekan, and Hubert H.G. Savenije

I have finished my review of the paper “Using NDII patterns to constrain semi-distributed rainfall-runoff models in tropical nested catchments”, by Sriwongsitanon et al., submitted to HESS. This paper outlines a comparison study of five models on 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 NDII remote sensing metrics to inform the distribution of soil stores (models FLEX-SD-$NDII_{MaxMin}$ and FLEX-SD-$NDII_{Avg}$), 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 paper is very well-written, with interesting and appropriately prepared tables and figures, but it seems to me that there are two critical misinterpretations of some of the results. I consider that it would be acceptable for publication in HESS, subject to technical corrections on the following two approaches:

1) At multiple points in the paper, the authors report that the model “gained realism” (e.g., lines 29, 395, 397, 400), however, it is not until section 5.2 (lines 406-407) that some kind of objective definition of the term “realism” is given, since this is to be tested in the models by comparing the outputs of the models to observations of NDII and the global scale SWI dataset for verification. Furthermore, it is not until the conclusions section (lines 480-481) that the realism of the model parameters is directly associated with characteristics such as their distribution according to catchment characteristics comprising catchment area, reach length, and the NDII.

I believe that the methodological section should include a clear introduction of the definition that will be given in this article to the term "realistic", since the methods and definitions mentioned above are relatively subjective, since they are not supported by a statistical test or similar that allows drawing meaningful conclusions associated with a level of confidence about how true the values of the parameters (or model outputs) are, in relation to what we would scientifically consider their "true value" (which is usually achieved through methods such as statistical inference).

To be clearer, in my opinion, the article does not contain scientific evidence that the model gained realism in terms of its true values (parameters, results), but rather in terms of what the authors define as "realism" in lines 406-407 and 480-481 (which is valid as long as you make a formal definition of how the term will be used).

2) I think that both section 5.4.1 (Su-NDII relationship, this being an induced/forced relationship) and item 1 of the discussion section (exploring the causality between the aforementioned induced/forced relationship and the degree of aridity) should be deleted. The reason why is given in the first four lines of the discussion section. It makes no sense for me to present as a relevant finding that the NDII time series correlates well with Su values, considering that Su values were overtly and systematically constrained by NDII during the modelling exercise, and a marginal gain in the efficiency of the models was rather a trivial, expected result. Therefore, I do not believe that this particular modelling exercise

implies any scientific confirmation that the NDII is a “reasonable index to indicate root zone soil moisture during the dry season”, if the only argument is the already expected higher correlation, which were in fact induced by procedure, between Su and NDII.

Again, concluding about signatures in catchments with various soil moisture capacities also seems methodologically inappropriate to me. If Su values from FLEX-SD-NDIIAvg (or any other NDII-based model) produce relatively higher NSE for sub-catchments with more evergreen forest, it is simply because the model forces these Su values to behave according to their corresponding NDII values, which in turn are directly affected by vegetation densities. Consequently, I would suggest removing all NDII time series from Figure 9 and others, and comparing only the simulated root zone moisture storage (Su) with SWI, which is in fact a relatively more independent spatiotemporal variable.

Despite all of the above, leaving aside the attempt of establishing a supposedly independent correlation between Su and NDII, and later use it to explain natural processes such as aridity or forest cover (which are the very factors underlying the NDII estimates and the forced modelling of Su), it would be interesting to try to explain why the NDII constrains Su so unevenly in sub-catchments with different percentages of evergreen forest. I think this exercise should conclude only on model structures, calibration methods, or uncertainty/sensitivity analysis of model parameters versus Su, rather than trying to provide a causal explanation for natural processes that have not been measured directly or statistically.