

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/hess-2022-62-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2022-62

Anonymous Referee #2

Referee comment on "Leveraging sap flow data in a catchment-scale hybrid model to improve soil moisture and transpiration estimates" by Ralf Loritz et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2022-62-RC2>, 2022

The authors present a catchment-scale hybrid model which is leveraged by sap flow data for more accurate hydrological simulations. The results showed that the hybrid model could lead to more realistic soil moisture estimates than the conventional Jarvis-Stewart equation, especially during drought conditions. The hybrid model predictions could match soil moisture and transpiration equally well as model runs using observed sap flow data and more importantly, hybrid model has good potential extrapolation beyond the study site. Such kind of hybrid model approaches which integrate machine learning methods and physical laws could open promising perspectives for more parsimonious process parametrizations.

These very interesting results have great potential to benefit the scientific community. With some minor clarification, this manuscript will be considered for publication.

I just have several specific questions. First of all, they didn't provide the cross-validation results. Secondly, did you also try the normal neural network, not the GRUs?

Line 129, 130: So, the 32 trees are evenly distributed in the catchment area? Could you

show them on a map?

Line 196, 197: how many predictions time steps? Use 96 hours to predict next hour or next 2 hours? Why not 24h, 48h or 72h?

Line 197~198: How did you prove the network which consists of four layers (input, two hidden, output) with 128 cell/hidden state is the most appropriate structure? Will the different dropout rate affect the results significantly, e.g., 5%, 15%, 20%?

Line 260: So, you're using data from 2014 and 2016 to train the deep learning model, while use the data of 2015 as the test dataset? Did you try cross validation and set 2014 or 2016 as the test dataset to see the results? Are there significant differences between different catchments and years? Could you show the data distribution, e.g., boxplot, of different years and catchments?

Section 2.2.4: It seems that the machine learning model is set to point to sap flow directly? Why not just let machine learning model predict the conductance directly? You could also introduce constrains into the loss function by using equation 1 and 2 to constrain the training process.

I also suggest you should have a flow chart or schematic map for clearly clarifying the hybrid model. This could be more friendly to the readers.

Line 293, 294: Could you further explain why the gcDL under- or overestimates on peaks?

It seems that the model can't not capture the peak value very well? I think if you let the machine learning model predict the conductance directly with constrains from equation 1 and 2 into the loss function, this problem could be mitigated.