Review of Manuscript

'H2MV (v1.0): Global Physically-Constrained Deep Learning Water Cycle Model with Vegetation'

By Z. Baghirov et al.

Dear Editor,

I have reviewed the manuscript. My conclusions and comments are as follows:

1. Scope

The article is within the scope of GMD.

2. Summary

The authors present the H2MV model, a further development of the H2M model (Kraft et al., 2022). H2MV is a hybrid model for the global terrestrial water cycle. It consists of a conceptual process-based hydrological model including the main terrestrial stocks and fluxes of water, and connected neural networks (partly static, partly dynamic with memory) processing static catchment attributes and dynamic forcing to deliver space-variable, time-static catchment attributes (maximum vegetation-reachable soil water SM_{max}), space-time-variable vegetation states (fraction of absorbed photosynthetically active radiation fAPAR), and various space-time-variable parameters of the process-based hydrological model. Model forcing includes precipitation (P), radiation (Rad) and air temperature (T). Further observations used for model training are terrestrial water storage (TWS), fraction of absorbed photosynthetically active radiation (fAPAR), snow water equivalent (SWE) and runoff (Q). H2MV is trained in a Cross-validation (C-V) approach on 10 spatially mutually exclusive datasets, and validated on an additional spatial holdout set. Model performance is discussed for all predictive variables TWS, fAPAR, SWE and Q on various temporal aggregations (monthly, seasonal, interannual). The authors conclude that generally, model performance is acceptable and shows space-time patterns in agreement with expert expectations and the literature. Further, the authors discuss model equifinality, here expressed as the predictive variability of the target variables among the 10 C-V models. Here, the authors conclude that mainly soil-related parameters are uncertain, and that model errors are dominated by phase shifts.

3. Evaluation

Overall, the work presented by the authors is an interesting and relevant contribution to global land surface modeling. The presentation style is mainly clear and complete, and the conclusions are supported by the results. So there are only minor revisions required to increase clarity and completeness before publication.
"L"=line

L3 "… we explicitly represent vegetation states by the fraction of absorbed photosynthetically active radiation (fAPAR), and by the maximum soil moisture capacity (SM_{max}), …". This is misleading, as it suggests SM_{max} represents a vegetation state. From the rest of the text, I take it that SM_{max} is a spatially variable but temporally invariant representation of the maximum (vegetation-reachable) soil water content, i.e. a purely soil-related property. Therefore I suggest rephrasing with a better distinction between and explanation of the abiotic and biotic controls of soil water capacity.

L8 The authors use the term 'constrain' throughout the manuscript to refer to observables used in an objective function during model training/calibration. As not all readers will be familiar with this use of terms, I suggest adding a related clarification, e.g. in L44.

L37 "Hybrid (or differentiable) modeling aims to address this challenge". The sentence suggests that hybrid modeling is synonymous to differentiable modeling. This is not the case. There are hybrid modeling approaches that do not require differentiability of the process-based part, and not all differentiable model are hybrids. Therefore I suggest rephrasing.

L81 Please add a short information about the length of the available data.

L 165 C-V approach: The authors use 10 validation sets (and one common test data set common) that are mutually exclusive in terms of space, but not time. I understand that a time-exclusive C-V approach for the validation sets may not be possible due to limited data, but, as high spatial correlation may exist between validation and testing sets for the same time, the testing set should be differing from the validation sets in terms of both space and time. This will help to better assess the models space-time generalization capabilities. My suggestion: Train the model on all but the last two years. Use all but the last two years for space-only CV testing in the same way as done now. Use the last two years for space-time independent testing.

L 174 Loss function: Eq. (9) calculates the total loss over all observed targets (TWS, fAPAR, SWE, Q). The targets come with different units, so their influence on total L might be different. How is equal weighting of each target in L assured? Is the loss calculated from the Z-transformed data as in Kraft et al. (2022)?

L201 It is unclear to me what the authors mean by "each estimated process". Please add this information to the text. Also, in L208, the authors refer to "parameters" rather than "processes". Please clarify.

Figs 3, 4, C1, C2, C3. Please add a x-axis label to plots a)

L261 As H2MV is a further development of H2M, a performance comparison between the two is important. The authors provide this comparison in the Appendix in Fig. C5, but do not discuss it. In Fig. C5, it becomes apparent that H2MV performs worse than H2M in terms of at least

- RMSE for TWS, SWE, ET
- SDR for TWS, SWE

While I do not think that a new model generation needs to outperform a previous one for all metrics, the reader will benefit from a more detailed discussion of the performance differences between H2MV and H2M.

Parameter stability: I wonder how time-stable or time-variant the LSTM-predicted parameters of the hydrological model are. Ideally, if the hydrological model would fully contain all relevant processes, the parameters should be static. Time-variations would point at functional deficiencies of the hydrological model, and the time-patterns could point at the nature of these functional deficiencies. See e.g. Fig. 8 in Acuna Espinoza et al. (2024). I do not require that such a discussion is added to the current paper, rather it is a suggestion for further work.

Yours sincerely,

Uwe Ehret

References

Acuña Espinoza, E., Loritz, R., Álvarez Chaves, M., Bäuerle, N., and Ehret, U.: To bucket or not to bucket? Analyzing the performance and interpretability of hybrid hydrological models with dynamic parameterization, Hydrol. Earth Syst. Sci., 28, 2705–2719, https://doi.org/10.5194/hess-28-2705-2024, 2024.

Kraft, B., Jung, M., Körner, M., Koirala, S., and Reichstein, M.: Towards hybrid modeling of the global hydrological cycle, Hydrology and Earth System Sciences, 26, 1579–1614, 2022.