Review of "hess-2020-322"
"A Deep-Learning Hybrid-Predictive-Modeling Approach for Estimating Evapotranspiration and Ecosystem Respiration"
Chen et al.

Anonymous

March 21, 2021

Meta

- Disclaimer: I consider myself capable of evaluating the methodological aspects of this study, but I am not an ecologist.
- Note that the list item numbers (e.g., [1-3]) refer to lines in the manuscript.
- Overall great work, thanks for sharing it!

Summary of the Content

The study tests a hybrid modeling approach which integrates different data sources using deep learning to predict daily evapotranspiration (ET) and ecosystem respiration (R_{ECO}). The authors use eddy covariance FLUXNET site-level data and simulations from a physically-based model to train their model on a set of meteorological features (e.g., precipitation, air temperature) and remotely sensed data (NDVI). The model was tuned on different data to test a range of use cases: Training and testing the HPM 1) within FLUXNET site, 2) between FLUXNET site (genrealizeability), 3) emulating a physically-based model, and finally, to 4) study the connection of ET and R_{ECO} and meteorological forcings and vegetation properties. The model was able to learn patterns of ET and R_{ECO} to a satisfying degree, and connections between ET and R_{ECO} and meteorological forcings and vegetation dynamics have been found to be significant at high spatial resolution.

Overall Feedback

Integrating different datasets for ecosystem modeling with neural networks is a hot topic which has the potential to improve the predictability of such complex systems. Much effort has been spent to process different datasets and the proposed approach has been tested in several experiments. I am sure that this study has been conducted with care and I find the results very interesting. I especially like that the authors considered overfitting by using training, validation and prediction sets and tested between-sites genrealizeability.

Recommandation

Minor revisions: I think the manuscript needs improvement (structure, language, consistency, figures, tables), but the study is still interesting and nice overall.

Major Remarks

- I read parts of the manuscript several times to understand how the FLUXNET and CLM data was used (~~combined~~, separately) and how the framawork exactly works, and I am still not sure if I entirely understand it. Also, it took me some time to understand the four experiments ("use cases"), what data was used for training, testing, etc. This is my major critic: I think the manuscript needs a cleaner structure and language.
- For me, the term "hybrid" is a bit confusing here. I assume that you refer to Reichstein (2019), where "(5) Surrogate modelling or emulation" is listed as a hybrid approach, which, once trained, can "achieve simulations orders of magnitude faster than the original physical model without sacrificing much accuracy" and "allows for fast sensitivity analysis, model parameter calibration, and derivation of confidence intervals for the estimates". I think the manuscript would be much easyer to understand if you would make this clearer.

Minor Remarks

General

- I strongly recommend to use colorblind-friendly colors in the plots. The time-series plots with green and red color mixed are particularly problematic. I think that the figures need some more work (general appearance, font size).
- From the HESS guidelines: "Common Latin phrases are not italicized (for example, et al., cf., e.g., a priori, in situ, [...])" (e.g., line 49, in situ).
- From the HESS guidelines: "The abbreviation "Fig." should be used when it appears in running text and should be followed by a number unless it comes at the beginning of a sentence, e.g.: "The results are depicted in Fig. 5. Figure 9 reveals that..."."
- From the HESS guidelines: "Units must be written exponentially (e.g. W m^{-2}).", e.g. line 380 or in axes labels, you use mm/d instead of m d^{-1}.
- You use the notation "Adj.R2-0.94" in some figures (e.g. Fig. 5). This is misleading, please use "Adj.R2: 0.94", "Adj.R2=0.94", or similar.
- Time-series figures: please add a legend for all plots (pink points, red, greeen, blue, black lines).
- Symbol notation: I noticed you use "ET" for evapotranspiration and "R_{ECO}" for ecosystem respiration. I find this is inconsistent, as either you use these as abbreviations, which are not italic ("ET" & "$\mathrm{R_{ECO}}$"), or as mathematical variables ("E" & "R_{ECO}"), where multi-letter symbols are to be avoided due to ambiguity (is "$ET = E \cdot T$?), and subsctipts are only italic if they refere to a variable (such as in x_i, where i is an index), but not if the subscript is a name.
- In general, many small "not so nice" things like units written inconsistently.
- I suggest to not put "learn" in quotes (as in *the model "learns"*) as the term is very commonly used in this context.
- Nice that you split the data in training, validation, and test (prediction) set! This is often not done.

Abstract

- The abstract is too detailed in my opinion, consider to shorten.
- I suggest to state clearly how the approach is hybrid and why you use the approach.

1. Introduction

- Nice review of current methods to estimate ET and $\mathrm{R_{ECO}}$. It *could* be shortened a bit.

2. Site Information, Data Acquisition and Processing

Tab. 1 It is hard to differentiate between the rows visually.

Fig. 1 Consider highlighting the SNOTEL sites visually.

3. Hybrid Predictive Modeling Framework

- I think you don't need to explain the LSTM in detail.

260 Does "deeply connected neural networks" refer to a fully connected neural network?

- For *use case 2*, do you train the model on all sates jointly or on single sites?

282-320 Consider replacing the extensive description of LSTMs with a conceptual high-level description.

326 Would be nice to see if a smaller model does the job (but not essential here).

331 Olah. (2015) → Olah (2015)

340-352 Why did you separate precipitation into rainfall and snowfall and how was the variable *sn* used? If they were used as inputs for the LSTM, why not letting the neural network figure this out, i.e., just inputting the avalable features?

355 I assume you used an LSTM? Then you can just use the term LSTM here, as it has been introduced already instead of "deep-learning recurrent neural networks".

4. Results

- I sugest to move the descriptions of the "use cases" to the methods section, maybe make a table that summarized what data is used for training and testing, the objective of the experiment etc.
- The interpretation would be much easyer if you would show the mean seasonal cycle and the interannual variability!

399-407 This is already discussion of the results.

399-404 I would expect that the LSTM learns SM dynamic, i.e., it represents it (implicitely) in its hidden state. SM would not necessarily be needed as the LSTM learns the ecological memory effects (e.g., Besnard et al. (2019) or Kraft et al. (2019). Adding SM could still help improving the model as it currently does not have much data to learn from compared to the number of parameters. Also, referring to a comment from former Referee #1, I think this should be clarified. This is one of the key advanteges of using models like an LSTM, it can learn ecological memory and thus, variables such as soil moisture may not be needed!

405-407 I agree that LSTMs tend to have issues with extreme values. In my oppinion, this is mostly because extreme values are rare, i.e., the model does not see many anomalous samples, there is less training data for such cases. Maybe you could mention this and provide a source, if you can find one.

Tab. 3 Please write units in exponential form. You could mention that the increase in test performance could be linked to droupout (which I assume is deactivated for inference) in the discussion.

420 I think the representativity of FLUXNET sites for the entire ecoregion is questionable and disputed (?), maybe rephrase.

Fig. 7 The monthly errors used to be pink before, right? I suggest to resuse the same colors.

450 I don't know what an "1-D" model is, consider explaining.

475 The *mechanistic* HPM model?

479 30m → 30 m

479+ Much of it is discussion.

516 17% → 17 %

Fig. 11 Panels (a) and (b) are not very informative, maybe remove?

5. Discussion

559 You refered to "physically-model-based HPM" as "mechanistic HPM" (line 264), you may use the latter one here.

625 Again, I think you need to discuss the "memory aspect". If you have meteorologicla data and site-level variables (e.g. vegetation type, soil properties), and enough training data, an LSTM would learn SM implicitely. This should be added to the discussion, as it is a key selling-point for using deep learning models. I think the message "SM is needed for improving model" is wrong, state variables are *not* needed anymore with DL approaches if the states can be derived from the input data. Of course, it can still be beneficial to add soil moisture, as it would regularize the model and maybe, the complex processes involved (e.g., lateral flux) may not be learnt by the model if the relevant features are missing.

651-660 As an outlook: the model could be trained on FLUXNET and process-based simulations *jointly*.

669 I cite reviewer #1: "Replace *CO* with *Colorado, USA* for the global audience." ;)