Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC1
https://doi.org/10.5194/bg-2022-186-RC1, 2022

Comment on bg-2022-186

Andrew Feldman (Referee)

Referee comment on "Upscaling dryland carbon and water fluxes with artificial neural networks of optical, thermal, and microwave satellite remote sensing" by Matthew P. Dannenberg et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2022-186-RC1, 2022

Dannenberg et al. develop a neural network to predict GPP, ET, and NEE at FLUXNET sites using satellite retrievals of several environmental drivers from several different observing frequencies (optical/thermal from MODIS and microwave from SMAP). DrylANNd is able to predict the GPP and ET seasonal cycle, spatial variability, and, to a lesser degree, their interannual variability. The predictions of NEE are weaker due to satellites not being able to observe respiration. Overall, I find this to be a nice advance and hope this lays the foundation for follow-up studies. The study is very thorough and well-motivated. I support its publication with consideration of points below. With future applications in mind, I encourage the authors to consider several points below as well as some methodological clarifications. Nice work!

-Andrew Feldman

Overall/Major Comments

1) What are the desired use cases of DrylANNd? It is a named model, which indicates a future application as the authors briefly mention for a global study in line 117. Machine learning approaches like this require careful calibration and validation, which the authors have done well here. However, if the conditions change to a different region or globally, what needs to change about the inputs as the predictors and predicted variables? Can we rely on the few dryland locations in the Western US to predict other regions when there may be different rainfall seasonality and vegetation types (i.e. African and Australian drylands) or do we need to train the model in each different defined region? Are we

restricted to certain datasets to serve as the GPP and ET independent variables?

I recommend laying a framework for applications in the discussion by providing more concrete recommendations on how to apply DrylANNd and points about pitfalls that may come about applying DrylANNd at larger spatial scales or other, related to the questions here. I know Section 4.3 may have been an attempt to do this in trying to improve the model overall with new datasets, but I think the authors can expand on that section with regard to these questions and maybe put a more positive outlook on it. Specifically, I recommend being clearer about how DrylANNd can be applied. Next, maybe give a big picture roadmap such as discussing how we may not have as reliable of observation-only data from satellites as we have from FLUXNET to use to train the model on ET or GPP. Therefore, we are constrained to using the model regionally where FLUXNET is available. Perhaps a SIF product (or other) can be used as a predicted variable elsewhere (like Australia) where there are not widespread, publicly available flux tower observations.

2) I want to caution that there may be a drawback in using soil temperature from SMAP L4 as predictor here, especially with regard to the desire to use remote sensing observations to train DrylANNd. The SMAP L4 retrievals are outputs from a land surface model assimilation (see the Reichle et al. 2019 study referenced in the submitted manuscript). While the soil moisture output is highly a function of remote sensing from SMAP's brightness temperatures (especially the 0-5cm product), the soil temperature is likely not as highly influenced by the SMAP observations. Historically, we input soil temperature data from a GMAO model in the process of retrieving L3 SMAP soil moisture – we don't go the other way around to estimate soil temperature. Microwave brightness temperature is a function of physical soil temperature, but more strongly associated with moisture on the surface, and is thus (at least not to my knowledge) not necessarily influencing the soil temperature outputs as heavily in the assimilation. Perhaps the L4 soil temperature output is less of an "effective" remote sensing parameter. I don't think we have good evidence otherwise, though I would be happy for this claim to be refuted which may require a closer look through the literature on assimilating L-band brightness temperature into land surface models. As a consequence, I think SMAP could be overestimated in its ability to explain GPP, NEE, and ET in Figure 8. Since the study's goal is to explain these variables with different observation frequencies from remote sensing instruments, I am not sure the SMAP soil temperature is as appropriate here as the other variables and recommend the MODIS LST or raw infrared data instead.

3) Given that the Western US has seen some unprecedented climatic behavior in the past two decades and especially in the past two years, does this create an issue training DrylANNd on stronger dry response anomalies over 2015-2021? It certainly will be a limitation in applications of predicting future ET and GPP (with regard to my point #1 above).

4) I think some mention of how spatial scale mismatch between datasets has an influence on results is important. For example, the flux towers have a fetch of <1000m. However, some of the remote sensing products have much larger native resolutions here, which could lead to problems with spatial mismatch of data allowing spatial heterogeneity errors to creep in to the prediction performance estimates. This may be motivation to demonstrate the method entirely with flux tower data at FLUXNET sites and see if similar results occur. I leave that up to the authors to try.

Line-specific comments

-L30: Wonderfully written introduction

-L65-90: What about effects of biases from soil color contrast and thus soil contamination on the visible signal?

-L135: Is the gap filling necessary where the NN approach cannot be used on irregularly sampled data? Such gap filling methods could bias a predictive approach if a functional form is used to gap fill (for example, a look up table that may be based on model assumptions). A noisy insertion could eliminate issues of model assumptions becoming imprinted in the prediction model. Maybe gap filling is not very common in the available time series? What percentage of the different time series are gap filled?

-L173: Note that SMAP products do not retrieve soil temperature, though there are some nuances about the assimilation process in L4. See the major point above.

-L225: what is the “holdout model?”

-L228: Are only 7 data points being used in the interannual timescale prediction? It seems 7 data points from all sites are normalized by taking out their mean and aggregated with other sites to increase sample size (as in Fig. 7). Please clarify in the text.

-L233: Are these months averaged in all cases for the warm season or is the max used in the case of visible/NDVI like it was for individual months (as stated in line 185)?

-L234: A word of caution that SMAP went into safe mode in summer 2019 which led to 1-2 months of loss of data. This is unfortunate because leaving this year in the analysis could bias predictions with biased means. Taking 2019 out removes samples from an already short time series. I encourage the authors to assess the consequences of removing 2019.

-L246: By the end of the methods, I have not gotten a picture of precisely what the inputs and outputs are. Are the predictors always from remote sensing and the predicted, independent variables are always from FLUXNET? Table 1 does help, but it may help further to add to table 1 that FLUXNET ET, GPP, and NEE are the independent/predicted variables.

-L293: It might be helpful to mention that time and space are mixed on the left panels of Fig 7 where the spatial patterns might be dominating the good performance there. Only

temporal patterns are shown on the right panels.

-Fig 8: Can the authors indicate in table 1 or elsewhere which variables are grouped into VI only, LST, and SMAP as corresponding to Fig 8?