Interactive comment on "A Deep-Learning Hybrid-Predictive-Modeling Approach for Estimating Evapotranspiration and Ecosystem Respiration" by Jiancong Chen et al.

Jiancong Chen et al.

nigel_chen1993@berkeley.edu

We thank the reviewer for spending time to review our manuscript and appreciate all the helpful suggestions and comments. Below, we describe how we have addressed the reviewer's comments in the revised manuscript. .

Reviewer: This study tested and validated a hybrid predictive modeling (HPM) approach at eight flux tower sites and three snow measurements in western North America. Modeled predictions of annual evapotranspiration and ecosystem respiration fluxes were significantly correlated with observations but less accurate at sub-annual resolution. These results are very promising but also demonstrate some limitations of the current HPM approach.

General comments: This is a thorough analysis and very promising study for the application of hybrid models to simulate ecosystem fluxes. It's also a well written manuscript. However, minimal effort has been devoted to generating the type of process-based/transferrable information that's expected from a top journal. Instead, the discussion is generally couched in terms of supporting previous work to emphasis model performance. In my opinion, some re-interpretation of results is needed to move this beyond a basic model validation.

A good recent example is Wieder et al. 2017 that also compared CLM (point mode) to flux tower observations in complex terrain. Although their study only considered one site, they explicitly focused on periods of relative agreement versus disagreement between the modeled and observed fluxes to yield broadly testable ecohydrological hypotheses. Given the multi-site focus of the current study, I don't think the same level of detailed inquiry is required, but additional synthetic analysis would increase the scope and the subsequent impact of this work.

Relatedly, Figures 4-9 all show similar long-term time series data with scatterplots that lend themselves to similar interpretations in terms of R2 of MAE. These are useful, but perhaps they could be condensed and/or supplemented with other figure types that were more conductive to process-based interpretation. For example, I found Figures 11e and 11f fascinating insofar as they highlighted seasonal differences between vegetation types, but little explanation was provided to "unpack" these results (grasslands and shrublands not even mentioned). Likewise, Figures 12a and 12b present a rich opportunity to speak to differences between the biophysical controls on ET at the SNOTEL and East River sites. Some of the specific factors I'm left wondering about are differences in snow accumulation and melt between sites, evaporation versus transpiration, and heterotrophic versus autotrophic respiration. I understand that you don't have all these measurements, but you've generated a lot of suggestive data that could be leveraged to push this field of research.

Response: We thank the reviewer for his high-level evaluation and suggestions. We agree with the reviewer that process-based interpretation of watershed dynamics is highly valuable, while such interpretation was present but not adequately described in our initial manuscript. We tried to improve such interpretation in the present version (Also at Q13, Q14 and Q18). Correspondingly, we expanded our discussion on how snowmelt timing influences vegetation dynamics and how different vegetation responds to snowmelt at the East River Watershed with the available NDVI dataset (L481-L491). With the HPM-based estimation results at the East River Watershed, we further unpack how ET and R_{ECO}

dynamics shift correspondingly to snowmelt timing, and vary among different vegetation groups, including shrublands and grasslands. We extracted detailed numbers in 2012 (earlier snowmelt year) and 2015 (normal year) for comparison to support our findings and explored how the interactions between snowmelt and vegetation influence ET and R_{ECO} dynamics (L520-L529). We also emphasized what we have observed for shrublands and meadows throughout the result and discussion sections and expanded the discussion around figure 12. In addition, we further clarified the role of meteorological forcing attributes and vegetation types in ET and R_{ECO} dynamics at the East River Watershed and investigated how input variables control ET and R_{ECO} differences among different sites (L535-L550). Following the reviewer's comments, we also clarified the reasons about feature selection for current HPM development (L334-L339) and recommended to include soil moisture or other important parameters for sites with seasonally dry periods (L575-L587). Considering the reviewer's comment about mountainous watersheds, we clarified the reason guiding Fluxnet, SNOTEL and the East River Watershed sites selection (L165-L178). We also included US-Me2 (Oregon site) in this study as suggested in Q21 (Figure A5).

In the following sections, we have addressed the reviewer's comments point-by-point.

Specific Comments:

Q1: L66-67: Reading back through the manuscript, this seems at odds with the practice of using single flux towers to represent the larger ecoregion (section 4.2). I don't actually have a problem with that research design, but this heterogeneity discussion may not be the best way to set things up.

Response: We have improved the discussion on using flux towers while trying to evaluate the effect of heterogeneity in the revised manuscript (L66-L73). We agree that flux towers are mainly installed and maintained at relatively flat areas and thus provide fluxes and meteorological conditions that are not indicative of the heterogeneity in each ecoregion. Still, we believe that HPM models trained at these sites can capture the interactions among ET, R_{ECO} and meteorological forcing and vegetation data and then reproduce at some extent the heterogeneity in ET and Reco within the ecoregion based on the variability in meteorological forcing and vegetation data. We validated this approach in the Use Case 2. While the results obtained in the various Uses cases show that the developed approach performed well, they also indicate the current limitations of this approach. The main limitation is due to insufficient resolution of meteorological reanalysis product, which did not reflect the corresponding heterogeneity. In addition, snow and soil moisture data could presumably improve HPM estimation especially during seasonally dry periods, however these datasets are difficult to be obtained over space and time. In the discussion section of the revised manuscript, we have also acknowledged ongoing research that focus on resolving data limitation issues caused by complex terrain, such as NASA's Asteroid Redirect Mission and Surface Atmosphere Integrated Field Laboratory Data.

Q2: L 73-75: Also uneven hydrologic distribution due to lateral flow in complex terrain (e.g., Chang et al. 2018) that results in heterogeneous fluxes.

Response: We have now acknowledged this perspective (L83-L86).

Q3: L109: Has NDVI been defined?

Response: We have now added the definition (L116).

Q4: L142-144; L365-367: After reading through the manuscript once, I'm not convinced this objective was met or even really addressed, which was confusing because I kept expecting to come across these results. The small-scale heterogeneity results must be expanded or else it may not be a fatal flaw to just remove this language/objective if the analysis didn't work out (as you intimate on L574-577). In any case, *the current manuscript introduction/objectives/results are inconsistent with respect to the degree of focus on this topic.*

Response: We agree that the stated objective of reconstructing small-scale heterogeneity was not entirely met. While the method has shown promising results for predicting ET at other locations in similar topographic position (i.e., flat area), we intended to investigate how meteorological forcing and vegetation heterogeneity influence ET and R_{ECO} at the East River Watershed and the nearby SNOTEL stations. With comparison of meteorological forcing data between weather station and DAYMET data (Figure S3 and S4), we concluded that the insufficient resolution in meteorological reanalysis products limited the ability to estimate the spatiotemporal variability of ET and R_{ECO} in mountainous watersheds, where slope aspect influence the energy balance. This confirm the importance of improving meteorological reanalysis products. While we recognize the above limitation, the impact of vegetation on the ET and R_{ECO} dynamics can be assessed as NDVI data are obtained at much higher resolution. This is what we focused on in section 4.4. The objective and related discussions have been modified and clarified (L148-L151, L366-L374).

Q5: L143: Replace "CO" with "Colorado, USA" for the global audience.

Response: We have made this change (L150).

Q6: L150: I'm curious how you defined "mountainous watersheds" for this study. I've been to the Walnut Gulch sites and they didn't strike me as the least bit mountainous. Also, with respect to my comment on L142-144, how important is the "mountainous" aspect anyway? I understand the broader impacts for water resources, but you'd reach a wider audience if the results were presented in a more general way. I see advantages and disadvantages to both mountain-specific and general analyses, but details/justification (mountain) or else re-framing (general) is needed in either case.

Response: We define mountainous watersheds as high-elevation watersheds that encompass mountainous ecozones (e.g., montane, subalpine and alpine areas). In this study, we mainly refer to the East River Watershed in Colorado, USA as the representative site for mountainous watersheds (Use Case 4).

The FLUXNET sites were selected to test HPM's capability and limitations under different climate conditions, which may not necessarily locate in mountain regions. For example, US-Ton and the Walnut Gulch site should not be treated as mountainous watershed sites. We have clarified the language and the justification in the revised manuscript (L157-L163, L166-L170).

Q7: L162: How were the eight FLUXNET stations selected? Some justification needed here. Was it to facilitate the paired approach in section 4.2?

Response: The Fluxnet sites were selected to sample different climate type, cover a wide range of meteorological and vegetation conditions, and provide continuous >5 years' time series data. These sites represent different ecoregions from Californian Mediterranean, to Sierra Madre Piedmont to Western Cordillera and Boreal Plain. Using sites in various ecosystems enabled us to evaluate the performance of HPM across different sites located in the same ecoregion and evaluate differences in processes driving the ET and R_{ECO} response in various ecoregions. For example, at ecoregions limited by energy conditions (e.g., CA-Oas), current HPM estimations are good, whereas at ecoregions with seasonally dry periods, additional variables (e.g., soil moisture data) might be needed to improve HPM accuracy. The choices of these sites not only facilitated the paired approach in Use Case 2 and 4, but also enabled us to assess HPM limitations at different ecoregions. In the recent revision of the manuscript, additional sets were considered, including US-Me2 (Q21).

Q8: L164: Table 1 indicates that the Saskatchewan sites are colder than US-NR1.

Response: We have corrected this mismatch (L168).

Q9: Table 1: I assume the periods of records are truncated at 2015 because you used the FLUXNET2015 product? This should be specified. Watch significant figures through-out this table.

Response: Yes. We have clarified this point in the revised manuscript (L190).

Q10: L227: Why was it necessary to treat this site different than the others? Please provide details about this "cleaning" procedure and why it was needed.

Response: We have identified some data gaps and erroneous data for the ET data at US-NR1 from the FLUXNET2015 database. The data cleaning framework provided in Rungee et al., 2019 is well documented. We made this decision after visualizing the raw data at US-NR1, where measurements during winter periods are likely uncertain. (L233-L234)

Q11: L367: The previous text makes it sounds like three (nor four) cases – confusion.

Response: Thank you for your comment. We have clarified the paragraph (L370-L372).

Q12: Table 3: You probably don't need a table just to say that "sn" was included at three of the eight sites. Especially because you already have so many display items.

Response: Thank you for your comments. We have made the recommended change.

Q13: L378-380: I'm very curious as to whether this was also the case at the seasonally dry Walnut Gulch sites? If so, it speaks to systematic bias where the model captures ET dynamics during energy-limited but not water-limited periods. This strikes me as a major result (see general comments) and could be leveraged to make recommendations about the input variables that are necessary for various systems.

Response: Thank you for your comments. We agree that the current HPM models with only meteorological attributes and NDVI as features generally captures ET dynamics during energy-limited but not water-limited periods. Variables (i.e., precipitation and constructed index, *sn*) provide indirect information regarding water stress, and we have observed high prediction accuracy during winter time and early growing season. However prediction accuracy usually decreases during peak growing season (summer time), especially at ecoregions that experience dry periods (e.g., occurrence of drought (Sloat et al., 2015; Wainwright et al., 2020). Based on these observations, HPM performs better during energy-limited periods than water-limited periods. As soil moisture is the variable that directly quantifies subsurface moisture stress, including soil moisture as a key variable at HPM can be effective in improving HPM performance. However, soil moisture data highly depends on depth and other subsurface properties (more details provided in Q19). Due to these reasons, we decided not to include soil moisture at the current stage for HPM development (L334-L339) and emphasized the importance of including necessary variables when data becomes available for various systems (L615-L626).

Q14: L383-395: Wouldn't the model overestimate (not underestimate) Reco if it can't account for soil moisture limitation during this time? Please clarify

Response: LSTM captures the long-term temporal fluctuations over time really well. But less frequent signals from peak growing season can be neglected due to the decreasing statistical significance. The underestimation of R_{ECO} during peak growing season is resulted from LSTM emphasizing on capturing long term dynamics and smoothing the larger values that occur less frequently.

Q15: L408-411: Seemingly contradicts L66-67

Response: We have further clarified this bullet point in the revised manuscript (L66-L73). Also see response for Q1 response.

Q16: L493: Units mismatch

Response: We have made the necessary correction (L519).

Q17: L495-505: Discussion

Response: We have made the necessary changes.

Q18: L516-518: In my mind, this is a missed opportunity to gain processed-based (and thus transferrable) insight. What about these sites could factor into ET differences that are so much greater than the Reco differences? See general comments.

Response: Thank you for your comment. Through further investigating data inputs and model performance, we believe the ET and R_{ECO} estimation at the East River Watershed are limited by the insufficient resolution of input data. There are two major factors that lead to the differences of ET and R_{ECO} among these sites. The first perspective is that HPM ET model is more sensitive to temperature and radiation inputs compared to NDVI whereas NDVI, temperature and radiation are all influential for R_{ECO} estimations. The second perspective is about data resolution and uncertainty. Data provided by SNOTEL weather stations are more accurate than DAYMET reanalysis data. We observed a greater differences in temperature and radiation at the SNOTEL sites whereas there's very small differences at the East River sites (Figure S3). Summer temperature differences among SNOTEL sites can be over 3℃ but there's a barely 0.2℃ differences in DAYMET data used for the East River sites. From Landsat data, we could distinguish the differences in NDVI at the East River sites (Figure 10) and these differences are well captured by HPM R_{ECO} model. Considering both the model and data perspectives, it explained why we observed these differences in ET and R_{ECO} at these sites. With the high prediction accuracy for Use Case I scenarios, we believe HPM is capable to capture the interactions among ET, R_{ECO} and input variables. If high resolution meteorological data becomes available at the East River Watershed such as the Surface Atmosphere Integrated Field Laboratory (SAIL), we believe HPM can better distinguish how meteorological forcing heterogeneity controls ET and R_{ECO} and more process-based interpretation can be learned from HPM estimations. These discussions have been implemented in the revised manuscript (e.g., L535-L550; L597-L604; L627-L637).

Q19: L544-546: How hard would it be to add moisture into the model? Why wasn't it added in the first place? I'm not suggesting that you re-do the analysis, but the readers will be very interested in this information.

Response: For Use Case 1 situations, it is applicable and relatively easy to add soil moisture into the model. However, for the other Use Cases where the model will be transferred over space, it is difficult to directly use soil moisture as an input variable given soil moisture measurements are often made at various depths, and other relevant soil characteristics are likely different at different sites. The dependence on depth and other soil characteristics limits the model transferability at inadequately monitored watersheds (L334-L339). Thus, we have decided not to include soil moisture as an input variable. In L615-L626 of the revised manuscript, we now recommend the inclusion of soil moisture and other important variables if data are available and researchers have information regarding site characteristics. For example, at ecoregions that experience seasonally dry periods, it is useful to add soil moisture in HPM.

Q20: L563-568: It's not clear to me what model results "present similar dynamic trends" to the moisture limitation invoked by Hu et al. 2010 (and a host of larger scale, more recent work). My current understanding is that the model breaks down somewhat in the presence of moisture limitation, which I consider an interesting and valid result/contribution, but you can't have it both ways i.e., the model either does or does not capture fluxes during periods of relative moisture limitation. Perhaps I'm missing something.

Response: Thank you for your comment. We meant to say ET and R_{ECO} estimation from HPM at the East River Watershed are comparable to the other studies. Specifically, Hu et al. (2010) were able to use snow, branch and soil sample data at US-NR1 to conclude that a longer growing season led to less carbon sequestration. At the East River Watershed, HPM estimated smaller R_{ECO} for evergreen forests (592 to 639gCm^{-2}) that have longer growing season, compared to deciduous forests (642 to 698 gCm^{-2}). In the revised manuscript, we also compared 2012 (earlier snowmelt and longer growing season length) to other years and discussed how these dynamics influence seasonal and annual ET and R_{ECO} (L520-529).

The current HPM model does not include soil moisture due to reasons described in response to Q19, however other attributes, such as precipitation and sn, provide indirect information regarding the moisture inputs. Thus, fluxes estimated from HPM during moisture limiting periods are still reasonable. In the revised manuscript, we have demonstrated this limitation of HPM and suggested to incorporate soil moisture data in addition to precipitation and sn when they become available.

Q21: L609: Still need more convincing about how "mountainous" was defined and why these sites were chosen, in particular with respect to other "mountainous" sites in the FLUXNET2015 database. I'm thinking of sites in New Mexico and possibly Oregon off the top of my head.

Response: Thank you for your comment. In our study, the representative site of mountainous watersheds is the East River Watershed, and the surrounding US-NR1 and SNOTEL sites. Other Fluxnet sites were selected to test the capability and limitations of HPM under other climate conditions (e.g., US-Ton, US-Wkg) and are not considered as mountainous sites. We have also included US-Me2 (Oregon) in our study. We did not include US-VCM (New Mexico site) after recognizing the occurrence of fire in 2013. These changes have been clarified in the revised manuscript. We have also added the estimation of ET and R_{ECO} at US-Me2. Results of US-Me2 is attached here and also included as Figure A5 in the revised manuscript.

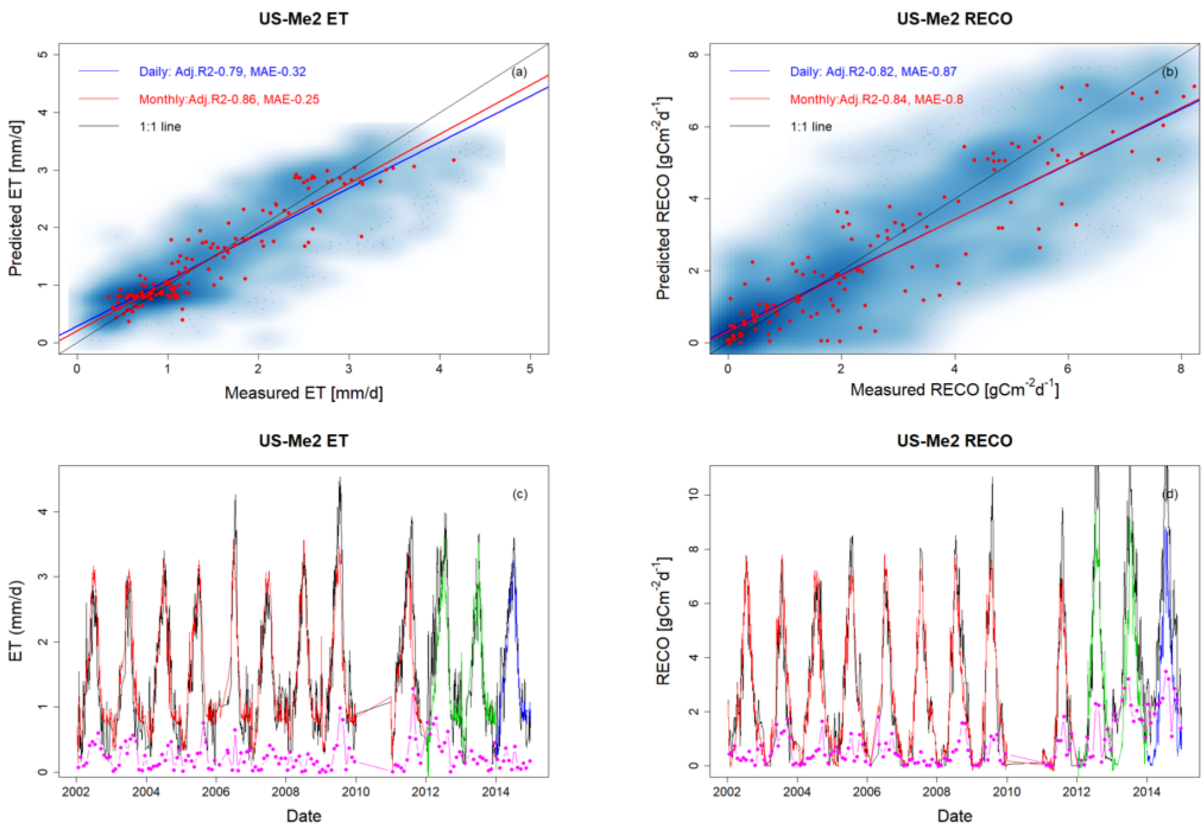

Figure 1. Added HPM estimation of ET and R_{ECO} at US-Me2. Data from 2011 are partially missing, which may decreases LSTM performance at US-Me2.

Reference:

Sloat, L. L., Henderson, A. N., Lamanna, C. and Enquist, B. J.: The Effect of the Foresummer Drought on Carbon Exchange in Subalpine Meadows, Ecosystems, 18(3), 533–545, doi:10.1007/s10021-015-9845-1, 2015.

Wainwright, H. M., Steefel, C., Trutner, S. D., Henderson, A. N., Nikolopoulos, E. I., Wilmer, C. F., Chadwick, K. D., Falco, N., Schaettle, K. B., Brown, J. B., Steltzer, H., Williams, K. H., Hubbard, S. and Enquist, B. J.: Satellite-derived foresummer drought sensitivity of plant productivity in Rocky Mountain headwater catchments: spatial heterogeneity and geological-geomorphological control, Environ. Res. Lett., doi:10.1088/1748-9326/ab8fd0, 2020.