Author responses to reviewers' comments

We appreciate the reviewers for providing us constructive comments and suggestions. Here we provide a summary of how we addressed each reviewer's questions or comments. All reviewers' comments and questions are in italic, followed by our detailed responses. We used Lx-Ly to represent the lines associated with changes we made in the revised manuscript and Linex-Liney to refer to specific reviewers' comments.

Summary:

1. We improved the manuscript organization and reduced the length significantly (47 pages to 34 pages) as suggested by the reviewers.
2. We reformatted the figures, adopting more friendly color schemes and using consistent color palette as suggested by reviewer #3.
3. We added relevant references as suggested by the reviewers.
4. We checked with HESS guideline and ensured units and Latin words are correctly written.
5. We revised the methodology section as suggested by reviewer #3. We replaced the detailed mathematical demonstration of LSTM with short but high-level explanation. We also added a sub-section explaining the use cases.
6. We addressed reviewer #1's suggestions to consider snow and monsoon's impacts over fluxes. HPM only requires 4 or 5 input variables, so it does not have the capability to explicitly track the movement of snow water and/or monsoon water. To address reviewer's hypothesis, we separated annual ET and R_{eco} into pre-June (January – June) component and post-July (July – December) component. We also included additional years (e.g., 2014) to have different combinations of snow precipitation and monsoon precipitation.
7. We partly addressed reviewer #1's suggestions to perform additional research on delineating fore-summer drought and post-monsoon droughts; identifying differences in snow-dominated watersheds versus monsoon-dominated watershed and quantifying evaporation versus transpiration and autotrophic versus heterotrophic respiration. We performed an independent analysis based upon the Palmer drought sensitivity index (PDSI) and radiation and precipitation data at the East River Watershed to determine major control of watershed dynamics over time. The results indicated that there were no significant differences in meteorological control between US-NR1 and the East River Watershed that occurrences of fore-summer droughts and post-monsoon droughts are highly correlated, and energy limiting conditions may exert more control on watershed dynamics than moisture limitation during late summer and autumn periods. Details of the analysis can be found in our response to reviewer #1's below. We did not consider evaporation versus transpiration and autotrophic versus heterotrophic respiration as suggested by the reviewer because we do not have data to enable this, and collecting them is beyond the context of this study.

Our responses to RC#1.

General comments:

1. The first paragraphs of my previous review (general comments) were not specifically addressed. In my experience, this is not typical for the reviewer response process, and is likely to be unacceptable to some reviewers and/or editors going forward. In particular, I

would still appreciate the authors responding to the third paragraph of my previous review, repeated here:
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 apologize for not having provided a satisfactory response to this reviewer's question in response to our original submission. In L346-L358, we unpacked NDVI dynamics for different vegetation types (including grasslands and shrublands) under various meteorological conditions (e.g., different combinations of snow and monsoon precipitation). We also provided additional analysis that focused on identifying drought conditions between sites as well as fore-summer and post-monsoon droughts (see Line605-Line605 comment). We partly addressed the reviewer's comment about the role of snow precipitation and monsoon precipitation in ET and R_{eco} dynamics using HPM estimations and we also emphasized the occurrence of energy limiting condition (L459-L469, also our response to Line604-Line605 comment). We understand the importance of splitting ET into evaporation/transpiration and R_{eco} into heterotrophic and autotrophic respiration. However, additional datasets and laboratory environments (e.g., isotopes, water use efficiency data) would be needed for this, which is outside the scope of this study.

2. Some of the results and discussion require more nuanced and/or focused interpretation (See detailed comments below). At the same time, the manuscript is long and could be shortened/tightened in many places to more accurately present/highlight key results (details below).

Response: We have shortened the manuscript significantly (47 pages to 34 pages) and also worked to succinctly enhance nuanced interpretation. We also performed additional analysis based on the suggestions and comments. We hope we have struck a reasonable balance in this revision.

L10: The decision to focus on Reco could be set up better. In other words, why Reco instead of NEE or GPP and/or all three? I don't necessarily have a problem with your decision to focus on Reco, but it must be clearly justified.

Response: As is now described on L32-L36, R_{eco} is sensitive to global climate change and plays a vital role in ecosystem carbon cycling (Le Quéré et al., 2009). Increases in R_{eco} may contribute to a global warming acceleration through exerting positive feedback on the climate system (Cox et al., 2000). NEE and GPP are important, however better quantification and estimation of R_{eco} is still needed in order to accurately quantify total carbon emissions from sparsely monitored ecosystems. This is the main reason why we focus on R_{eco}. In our response to reviewer #2's comments, we have developed specific HPM models to estimate NEE at certain FLUXNET sites and the model results are promising (Fig. A6). So we think HPM has the capability to provide NEE and GPP predictions and future studies may consider adopting our framework to better quantify net exchange of carbon and the assimilation component.

L17: Sites within sites?

Response: We improved clarity (L16).

L21: Suggest adding "USA" here for the global audience.

Response: We have modified correspondingly (L18).

L27: Please specify "air, soil, snow", etc. whenever "temperature" is invoked. Lots of room for confusion here because most would expect ET to vary more with air temperature versus Reco that is more sensitive to soil temperature.

Response: While we meant air temperature, we removed that sentence to shorten the abstract. In the revised manuscript, we have specified the use of 'air and soil' temperature to reduce confusion.

L34-L35: Same comment as L10.

Response: We have modified correspondingly

L129-L131: Recent work by Chu et al. 2021 on the representativeness of statistical tower measurement footprints to surrounding areas may be relevant here.

Response: We have described the work of Chu et al. (2021) in the revised manuscript (L50-L52).

L483: *Is it earlier snowmelt triggers the onset of vegetation activity or that higher air temperature trigger both snowmelt and the onset of vegetation activity?*

Response: In our study, we observed earlier increase of NDVI in years with earlier snowmelt (e.g., 2012) and later increase of NDVI in years with later snowmelt (e.g., 2015).
This observation is consistent with Pedersen et al. (2018). The relationship between NDVI, snowmelt timing and air temperature is non-linear in our study and thus we do not

think it is higher air temperature trigger both snowmelt and the onset of vegetation activity. There are studies that reported a positive correlation between NDVI and temperature (Jia et al., 2006) but also no or even weakening relationship between vegetation activity and temperature variability (Piao et al., 2014). We did not intend to imply any causalities among these processes and we have made clarifications in the revised manuscript (L353-L355).

L485-L486: Can you speak to the synoptic meteorological conditions in 2012 versus 2015? Why choose these two years for comparison? Similarly, the comparison of March, April and May between years is interesting, but what about the rest of the year? I'd be very interested in a similar post-monsoon analysis, potentially between years with strong and weak monsoons.

Response: We chose year 2012 as it represents a severe fore-summer drought, and year 2015 because it was a normal/wet year based on the Palmer drought severity index (PDSI). This information has been added in L349-L352. In the revised manuscript, we have selected another year 2014, which was characterized by large snow precipitation but small monsoon precipitation. We added this year to better quantify dynamics for late-summer and autumn months (L390-L396). In addition to monsoon, we want to point out that there was a sharp decline in August (~30%) and September (~40%) radiation compared to June in the three years, indicating the potential of energy limiting condition rather than a monsoon moisture limiting condition (L465-L469). Figure 1 shows the distribution of incident shortwave radiation and similar trends are observed for net radiation that peaks in June (~ 180 $W\ m^{-2}$), and declines significantly in August (~ 90 $W\ m^{-2}$). Please also see our response to Line604-Line605 comment.

L492-L497: Please edit this section to remove/acknowledge differences in NDVI that would be expected due to deciduous versus evergreen physiology. Some of the basic information currently comes across as results. I appreciate the attempt to relate these results back to processes, but this section needs refinement.

Response: We modified the section correspondingly (L347-L349).

L517: What does the "1" syntax correspond to?

Response: For the East River sites, we selected 4 for each vegetation types. "1" is for the first one of each type as shown in Table 2. We clarified this in the manuscript.

L525-L526: Please be specific about the meaning of "drought" in this context. Is it simply meant to connote some limitation to ET and/or Reco? If so, can you justify the underlying expectation that these variables would be affected at the same moisture threshold? I'd also argue that "usually" is the wrong word here. Earlier snowmelt certainly "can" trigger summer drought, but this scenario is subject to modification by monsoon precipitation and other factors as the authors acknowledge in this sentence. See recent work by Knowles et al. 2020, Xu et al. 2020, and many references therein.

Response: We meant that earlier snowmelt is correlated with occurrences of fore-summer drought, and we agree with the reviewer that monsoon may modify drought conditions. We performed additional analysis to look deeper into drought conditions at the East River Watershed, please see our response to Line604-Line605 comment. Due to data availability, soil moisture was never used by HPM at the East River Watershed, so it is not feasible to expect how different soil moisture threshold influence ET and R_{eco} predictions. We also want to point out that energy limiting condition for late-summer and autumn periods may occur as stated in our response to Line485-Line486 comment.

L583-L596: I support this opportunity to discuss physiological differences between evergreen and deciduous vegetation, but simply citing Baldocchi et al. 2010 is insufficient. More thorough and nuanced discussion that incorporates foundational research on this topic is required.

Response: Our original intent is to investigate whether HPM models can incorporate vegetation heterogeneity to quantify ET and R_{eco} differences between different vegetation types with only 4 or 5 input features. We cited Baldocchi et al. 2010 to confirm that HPM estimation for deciduous forest and evergreen forest are reasonable and seek for physical explanation from their studies. This is mainly from a modeling perspective to explore limitation in model development and refinement; and a confirmation of model performance. We did not intend to characterize the physiology's control on ET and R_{eco} as the only data we are currently using are meteorological reanalysis data and satellite data. We agree with the reviewer more thorough and nuanced research can advance our understanding of ecosystem dynamics, and we have added additional references that help us better understand the physiology's control on ET and R_{eco} dynamics (L445-L451).

L600-L601: See comment on L525-L526.

Response: Please see our responses to Line525-Line526 and Line604-Line605 comments.

L604-L605: This implies that growing season length determines snow water storage when in fact, it's closer to the opposite i.e., air temperature and/or snow accumulation determine the onset of the growing season. See Lian et al. 2020 and Zhang et al. 2020 for examples of more recent work on this topic. Combining the Sloat et al. 2015, Wainwright et al., 2020 and Hu et al., 2010 references here also raises an important distinction. Whereas the Sloat and Wainwright references invoke fore-summer i.e., pre-monsoon drought, the Hu reference pertains to late summer drought i.e., after snowmelt water inputs have subsided. This distinction reflects the typical relative importance of snowmelt vs. monsoon precipitation at a given site e.g., snow-dominated sites may be susceptible to moisture limitation after the snowmelt pulse (late summer; Hu et al. 2010), whereas monsoon-dominated sites may be susceptible to moisture limitation before the onset of monsoon rains (early/fore-summer; Sloat et al. 2015; Wainwright et al. 2020). Please establish the typical relative importance of snow versus monsoon precipitation at the East River site and how your results may be expected to change at sites where moisture availability is typically more or less affected by snowmelt versus monsoon precipitation.

Response: We agree that L604-L605 was misleading. We have clarified the sentence. In addition, we clarified the typical relative importance of snow versus monsoon precipitation on ET at East River site in the revised manuscript (L380-L396).
With regard to the studies the reviewer is referring, we note that Sloat et al. (2015) used peak net ecosystem productivity and Wainwright et al. (2020) used peak June NDVI as measures for fore-summer periods at the East River sites whereas Hu et al. (2010) used annual carbon uptake and growing season length at Niwot Ridge. Though they have chosen different metrics in their studies, we do not think there's a distinct difference at Niwot Ridge (US-NR1) or East River that one site is more snow-dominated versus monsoon-dominated, or that one site constrained by fore-summer drought or post-monsoon drought. Here we used SNOTEL Butte (ER-BT) as a representative site for the East River Watershed due to data availability.

In fact, US-NR1 and the East River watershed share lots of similarities (e.g., in the same ecoregion). Precipitation, temperature and elevation are similar for US-NR1 and ER-BT (Table 1). Palmer drought index (PDI) and Palmer drought sensitivity index (PDSI) were used to quantify drought conditions, as documented in Sloat et al. (2015) and Wainwright et al. (2020). We did not find any quantitative measures for droughts in Hu et al. 2010. None of these three studies derived any indices to explicitly quantify post-monsoon drought conditions, so we used August PDSI to compare them. Figure 1 presents the PDSI time series obtained from Abatzoglou et al. (2018) for US-NR1 and ER-BT. Based on the U.S. drought monitor classification, a value of -1 is the threshold for droughts. And the more negative PDSI values are, the more severe the droughts are. If PDSI values are greater than -1, the ecosystems may not experience drought condition.

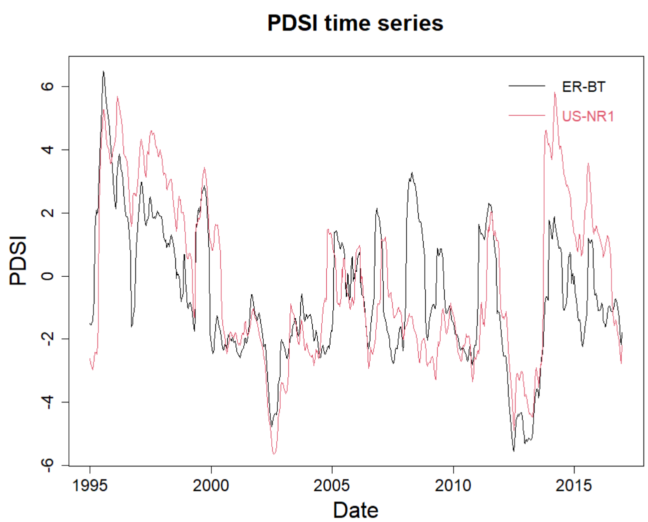


Figure 1. Time series of PDSI at ER-BT and US-NR1. Values smaller than -1 indicate drought condition.

We applied a simple linear regression of these PDSI values between US-NR1 (Hu et al. 2010) and ER-BT (Wainwright et al. 2020). We found a correlation coefficient of 0.88 ($p < 2.2e-16$), 0.82 ($p < 2.2e-16$) and 0.91 ($p < 2.2e-16$) for annual, June and August mean PDSI values between the sites, respectively. PDSI values in 2008 and 2014 differ significantly between the two sites, however that was mainly caused by unusual precipitation events and outside period with drought conditions as PDSI is greater than -1. Based on this result, we believe it is reasonable to conclude that the drought conditions for US-NR1 and East River Watershed are similar.

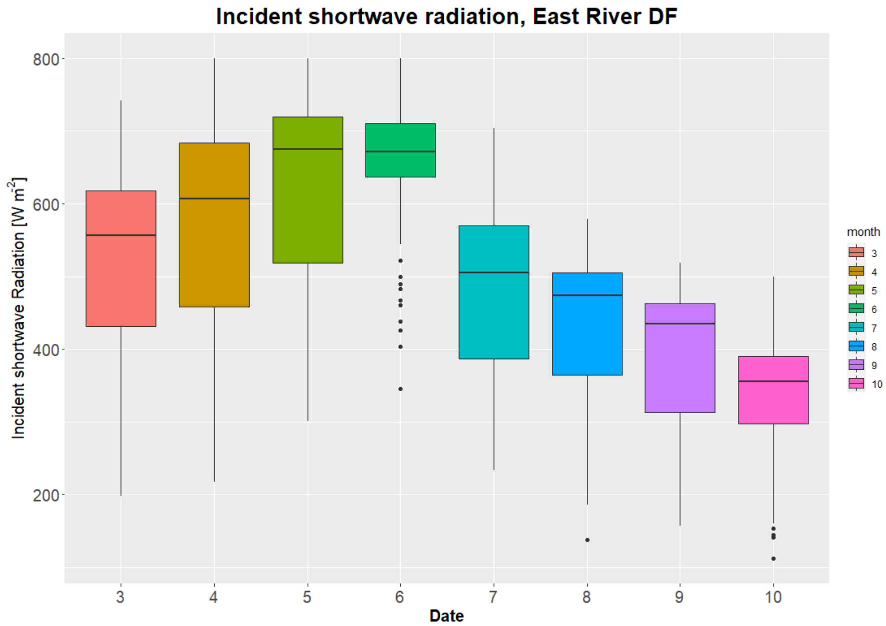


Figure 2. Net radiation distribution from 2011 to 2016 grouped by month at the East River Watershed.

We also discovered a high correlation between June PDSI and August PDSI. The correlation coefficients are 0.98 ($p < 2.2e-16$) and 0.90 ($p < 2.2e-16$) for US-NR1 and ER-BT, respectively, which indicates the coherency of fore-summer drought and post-monsoon drought if any. We want to note that PDSI has its own limitations, and we were not able to explore other data products that may be more sensitive to monsoon precipitation. Still, this result indicates occurrence of post-monsoon droughts are highly correlated with the occurrence of fore-summer droughts. Individual monsoon events may change the soil moisture condition in short terms, however may not entirely alter the drought conditions. We also want to point out to a recent work by Carroll et al. (2020), where they discovered July-September monsoon in the central Rocky Mountains may support ET in lower subalpine forests, but the monsoon precipitation also contributes to streamflow deficiencies caused by reductions in snow accumulation. They suggested that the timing and location of water input with respect to energy and water availability remain key issues. If monsoon arrives when potential ET (PET) is high and soil moisture is waning during fore-summer droughts, this water serves to moisten dry soils and is consumed very quickly by vegetation leading to increases in ET (moisture limiting condition). But if the timing of monsoon arrives late when PET is small, monsoon precipitation may contribute to streamflow rather than ET as the ecosystem is under energy limiting condition. In our study, we observed a significant decline in radiation after peak growing season regardless of the amount of monsoon precipitation. Net radiation declines by ~ 30% in August and ~ 40% in September compared to June. During the late-summer and autumn months, we think the East River Watershed is more likely to be constrained by energy rather than moisture limitation during late-summer and autumn months. We provide revised text at L378-L396; L456-L472.

L612: Hard to follow, I think "whereas" may be the wrong word here.

Response: We meant to say that earlier arrival (early-July in 2012) of monsoon precipitation help buffer the fore-summer drought condition. Correspondingly, 2012 July ET is not substantially different compared to other years.

L629: "Microclimate" is misspelled.

Response: We removed this sentence in the revised manuscript to shorten the manuscript. We have made sure spelling is accurate throughout the manuscript.

Our response to RC#3
Authors' response to RC3 review
We appreciate the anonymous reviewer for reviewing our manuscript and provide constructive for us to better improve the manuscript.
Major remarks:
1. I read parts of the manuscript several times to understand how the FLUXNET and CLM data was used (~~combined~~, separately) and how the framework 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.

Response: In the revised manuscript, we have added a section to demonstrate the four use cases and indicate the relevant data used for training and validation (L271-L284).

2. For me, the term "hybrid" is a bit confusing here. I assume that you refer to Reichstein (2019), where "(5) Surrogate modeling 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 easier to understand if you would make this clearer.*

Response: We use ‘hybrid’ in HPM to indicate the use of machine learning with mechanistic-based models/output and FLUXNET measurements integrating with other datasets, such as remote sensing. We show how HPM approach was used to 1) couple flux measurement for gap filling and time series prediction (Use case 1); 2) integrate flux measurement for spatial reconstruction and configuration in different ecoregions (Use case 2); 3) implement with physical process models (Use case 3) and 4) provide flux estimation to gain better understanding of ecosystem dynamics (Use case 4).. We have better clarified these points in the revised manuscript (L12-L14; L115-L118).

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).

Response: We have made the necessary changes.

From the HESS guidelines: “Common Latin phrases are not italicized (for example, et a;., cf., e.g., a priori, in situ, […])” (e.g., line 49, in situ).

Response: We have made the necessary changes.

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 …”.”

Response: We have made the necessary changes.

From the HESS guidelines: “Units must be written exponentially (e.g. $W\ \mathrm{m}^{-2}$).” e.g., line 380 or in axes labels, you use mm/d instead of m d^{-1}.

Response: We have made the necessary changes.

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.

Response: We have made the necessary modifications.

Time-series figures: please add a legend for all plots (pink points, red, green, blue, black lines).

Response: We have made the necessary modifications.

Symbol notation: I noticed you use “ET” for evapotranspiration and “R_{ECO}” for ecosystem respiration. I find this is inconsistent, as you either you use these as abbreviations, which are not italic (“ET” & “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 subscripts are only italic if they refer to a variable (such as in x_i, where i is an index), but not if the subscript is a name.

Response: Thank you for the comment. We have made the necessary changes.

In general, many small "not so nice" things like units written inconsistently.

Response: We appreciate the comments and have made necessary changes.

I suggest to not put "learn" in quotes (as in the model "learns") as the term is very commonly used in this context.

Response: We have made the necessary modifications.

Nice that you split the data in training, validation, and test (prediction) set! This is often not done.

Response: Thank you

The abstract is too detailed in my opinion, consider to shorten.

Response: We have made the necessary modifications (32 lines to 22 lines).

I suggest to state clearly how the approach is hybrid and why you use the approach.

Response: We have increased the clarity (L115-L118).

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

Response: We have made the necessary modifications (24 lines to 16 lines).

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

Response: We have made the necessary changes (L154).

Fig.1 Consider highlighting the SNOTEL sites visually.

Response: We used different shapes and colors to distinguish different sites (L157).

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

Response: We have made the necessary changes (L209-L221).

L260: Does "deeply connected neural networks" refer to a fully connected neural network?

Response: Yes.

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

Response: We trained the model on individual sites.
L282-L320: Consider replacing the extensive description of LSTMs with a conceptual high-level description.

Response: We have made the necessary changes (L209-L221).

L326: Would be nice to see if a smaller model does the job (but not essential here).

Response: The current configuration of neural networks does not require any super-computing power and we were satisfied with the prediction accuracy.

L331: Olah. (2015) -> Olah (2015)

Response: We have made the necessary changes (L221).

L340-L352: 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 available features?

Response: At locations dominated by snow, timing of snowmelt and bareground date is important for ET and R_{eco} dynamics. As there are only 4 or 5 features currently used, manual separation of precipitation into rain and snow may help the model establish linkages between precipitation and energy perspectives to better learn ecological memories and thus improve model performance. At locations where snow is rarely present, precipitation was directly used. We clarified this in the paper.

L355: 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".

Response: Yes. We have made the necessary changes (L264).

I suggest 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.

Response: We have made the necessary changes. A new section has been added (L271-L284).

The interpretation would be much easier if you would show the mean seasonal cycle and the interannual variability!

Response: Thank you for your comment. We intended to use the monthly mean comparisons to show seasonal cycles and interannual variability. In discussion sections, we provided more details about ET and R_{eco} at specific years.

L399-L407: This is already discussion of the results.

Response: We have made the necessary changes.

L399-L404: I would expect that the LSTM learns SM dynamics i.e., it represents it (implicitly) in its hidden state. SM would not necessarily be needed as the LSTM earns 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 advantages of using models like an LSTM, it can learn ecological memory and thus, variables such as soil moisture may not be needed!

Response: We agree with the reviewer that LSTM does has the advantage it could in theory learn the ecological memory. Still, we have to recognize that results of this study show that the use of LSTM cannot replace entirely the information present in soil moisture. Results show that ET and R_{eco} estimations at sites limited by energy condition have very high estimation accuracy, which suggests LSTM was able to capture the ecological memories. However, at sites that experiences drought conditions, some of ET and R_{eco} anomalous values are not frequent enough for LSTM to learn. These are time period where soil moisture data can be useful for this case to better inform LSTM and further increase prediction accuracy.

L405-L407: I agree that LSTMs tend to have issues with extreme values. In my opinion, 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.

Response. We agree with the reviewer. We have elaborated on this issue (L492-L497).

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

Response. We have made the necessary changes.

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

Response. We have elaborated on this (L306-L308).

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

Response. In the revision, we have adopted a consistent color scheme and palette.

L450: I don't know what an "1-D" model is, consider explaining.

Response. We were referring to the 1-dimensional CLM model developed in Tran et al. It solves physical equations in the vertical direction (L324-L326).

L475: The mechanistic HPM model?

Response. We have made the necessary changes (L341).

L479: 30m -> 30 m.

Response. We specified the resolution of remote sensing data in L82 and removed this sentence to shorten the manuscript.

L479+ Much of it is discussion.

Response. We have made the necessary changes.

L517: 17% -> 17 %

Response. We have made the necessary changes (L372).

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

Response. We think panels (a) and (b) are needed as they show the temporal trends and explain the seasonality of ET and Reco estimation at the East River Watershed for deciduous forest. Panels (a) and (b) placed the background for the following panels. Thus we decided to keep these two panels in Fig. 11

L559: You referred to "physically-model-based HPM" as "mechanistic HPM" (line 264), you may use the latter one here.

Response. We removed this sentence during revision to shorten the paragraph. But yes, it should be 'mechanistic HPM'.

L625: Again, I think you need to discuss the "memory aspect". If you have meteorological data and site-level variables (e.g. vegetation type, soil properties), and enough training data, an LSTM would learn SM implicitly. 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.*

Response. We agree with the reviewer that LSTM has been successful capturing the ecological memory effects in our study as well, and we have acknowledged this perspective in the revised manuscript (L426-L429). However, our results at certain sites suggest that drought occurrence and moisture limiting conditions may not be well captured by LSTM. We agree with the reviewer that soil moisture should be derived from the input data, but challenges still remain. There are uncertainties in the meteorological inputs (L405-L411), which increases the difficulties for LSTMs to learn soil moisture implicitly. LSTM may not be sufficiently trained upon drought conditions and longer time series may improve model performance. Soil moisture data can potentially fill the gap between atmospheric forcings and site-specific information. Thus at the current stage, we recommend to include soil moisture data when available to bypass certain limitations in data inputs and insufficient training. We have increased the clarity in the revised manuscript (L475-L481).

L651-L660: As an outlook: the model could be trained on FLUXNET and process-based simulations jointly.

Response. We have elaborated on this point (L495-L497).

L669: I cite reviewer #1: "Replace CO with Colorado, USA for the global audience."

Response. We have made the necessary changes (L506).

A Deep-Learning Hybrid-Predictive-Modeling (HPM) Approach for Estimating Evapotranspiration and Ecosystem Respiration

Jiancong Chen[1], Baptiste Dafflon[2], Anh Phuong Tran[2, 3], Nicola Falco[2], and Susan S. Hubbard[2]

[1]Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA, [2]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [3]Depatment of Water Resources Engineering and Technology, Water Resources Institute, 8, Phao Dai Lang, Dong Da, Hanoi, Vietnam

Abstract: ~~Gradual changes in meteorological forcings (such as temperature and precipitation) are~~Climate change is reshaping vulnerable ecosystems, leading to uncertain effects on ecosystem dynamics, including ~~water and carbon fluxes. Estimating~~ evapotranspiration (ET) and ecosystem respiration (~~R_{ECO}) is essential for analyzing the effect of climate change on ecosystem behavior. To obtain a better understanding of these processes, we need to improve our~~ R_{eco}). However, accurate estimation of ~~water and carbon fluxes over space and time, which is difficult within ecosystems that often have only sparse~~ET and R_{eco} still remains challenging at sparsely monitored watersheds where data~~.~~ and field instrumentation are limited. In this study, we developed a hybrid predictive modeling approach (HPM) that integrates eddy covariance measurements, physically-based model simulation results, meteorological forcings, and remote sensing datasets to estimate ~~evapotranspiration (ET) and ecosystem respiration (R_{ECO})~~ET and R_{eco} in high space-time resolution. HPM relies on a deep learning algorithm~~-~~, long ~~short-term~~shortterm memory (LSTM~~) as well as direct measurements or outputs from physically-based models.~~), and requires only air temperature, precipitation, radiation, normalized differences vegetation index (NDVI) and soil temperature (when available) as input variables. We tested and validated HPM estimation results ~~at sites within various sites. We particularly focus on testing HPM in mountainous regions, given their importance for water resources, their vulnerability to climate change, and the recognized difficulties in estimating ET and R_{ECO} in such regions. We benchmarked daily scale estimates of ET and R_{ECO} obtained from the HPM method against measurements made at FLUXNET stations and outputs from the Community Land Model (CLM) at~~ in different climate regions and developed four use cases to demonstrate the applicability and variability of HPM at various FLUXNET sites and Rocky Mountain SNOTEL ~~stations. At the mountainous~~ sites in Western North America. To test the limitations and performance of HPMs in mountainous watersheds, an expanded use case focused on the East River Watershed ~~site in the Upper~~, Colorado ~~River Basin, we explored how ET and R_{ECO} dynamics estimated from the new HPM approach vary with different vegetation and meteorological forcings.~~, USA. The results ~~of this study~~ indicate ~~that~~ HPM is capable of identifying complicated interactions among meteorological forcings, ET, and ~~R_{ECO}~~R_{eco} variables, as well as providing reliable estimation of ET and ~~R_{ECO}~~ R_{eco} across relevant spatiotemporal scales, even in challenging mountainous systems. ~~With HPM estimation of ET and R_{ECO} at the East River Watershed, we identified that HPM ET models are sensitive to temperature and radiation inputs whereas NDVI, temperature and radiation all have crucial influences over R_{ECO} dynamics. In general, our study demonstrated that the HPM approach can circumvent the typical lack of spatiotemporally dense data needed~~The study documents that HPM increases our capability to estimate ET and ~~R_{ECO} over space and time, as well as the parametric and structural uncertainty inherent in mechanistic models. While the current limitations of the HPM approach are driven by the temporal and spatial~~

~~resolution of available datasets (such as meteorological forcing and NDVI data), ongoing advances are expected to further improve accuracy and resolution of ET and R_{ECO} estimation using HPM~~R_{eco} and enhances process understanding at sparsely monitored watersheds.

1. Introduction:

~~Evapotranspiration~~Climate change has a profound influence on global and regional energy, water and carbon cycling, including evapotranspiration (ET) and ecosystem respiration ~~(R_{ECO}) are key components of ecosystem water and carbon cycles.~~(R_{eco}). ET is an important link between the water and energy cycles: dynamic changes in ET can affect precipitation, soil moisture, and surface temperature, leading to uncertain feedbacks in the environment (Jung et al., 2010; Seneviratne et al., 2006; Teuling et al., 2013). Thus, quantifying ET is particularly essential for improving our understanding of water and energy interactions ~~and~~as well as watershed ~~response~~responses to abrupt disturbances and gradual climate changes~~ in climate~~, which is critical for water resources management, agriculture, and other societal benefits (Anderson et al., 2012; Jung et al., 2010; Rungee et al., 2019; Viviroli et al., 2007; Viviroli and Weingartner, 2008). ~~R_{ECO} ,~~R_{eco}, which represents ~~–~~the ~~sum of autotrophic~~total respiration~~ and respiration by heterotrophic microorganisms~~ in a specific ecosystem, plays a vital role in the response of terrestrial ecosystem to global change (Jung et al., 2017; Reichstein et al., 2005; Xu et al., 2004). ~~As long term exchanges in R_{ECO} have pivotal influences over the climate system~~While increases in R_{eco} may contribute to accelerating global warming through positive feedbacks to the atmosphere (Cox et al., 2000; Gao et al., 2017; IPCC, 2019; Suleau et al., 2011), ~~approaches are needed to estimate~~ estimating and ~~monitor R_{ECO}~~ monitoring R_{eco} over relevant spatiotemporal scales~~.~~ is challenging. As described below, there are many different strategies for measuring and estimating ET and ~~R_{ECO}~~,R_{eco}, each of which has advantages and limitations. ~~The motivation for this~~This study is motivated by the recognition that current methods cannot provide ET and ~~R_{ECO}~~R_{eco} at space and time scales (e.g., daily) needed to improve prediction of changing terrestrial system behavior, particularly in challenging mountainous watersheds.

Several ground-based approaches have been used to provide *in situ* estimates or measurements of ET and ~~R_{ECO}~~ R_{eco}. Ground~~-~~-based flux chambers ~~capture and~~ measure trace gases emitted from the land surface, which can be used to estimate ET and ~~R_{ECO}~~R_{eco} (Livingston and Hutchinson, 1995; Pumpanen et al., 2004). ~~However, the microclimate of the environment is affected by the chamber, and the laborious acquisition process and small chamber size typically lead to information with coarse spatiotemporal resolution (Baldocchi, 2014).~~ The eddy covariance method uses a tower with installed instruments to autonomously measure fluxes of trace gases between ecosystem and atmosphere (Baldocchi, 2014; Wilson et al., 2001). ~~The covariance between the vertical velocity and mixing ratios of the target scalar is computed to obtain the fluxes of carbon, water vapor, and other trace gases emitted from the land surface.~~ ET is then calculated from the latent heat flux, and ~~R_{ECO}~~R_{eco} is calculated from the net carbon fluxes using night-time or daytime partitioning approaches (van Gorsel et al., 2009; Lasslop et al., 2010; Reichstein et al., 2005). The spatial footprint of obtained eddy covariance fluxes is on the order of hundreds of meters, and the temporal resolution of the measurements ~~range~~ranges from hours to decades (Wilson et al., 2001). ~~Such~~Tower-based in~~-~~ situ measurements of fluxes have been integrated into the global AmeriFlux (http://ameriflux.lbl.gov/) and FLUXNET (https://FLUXNET.fluxdata.org/) networks~~, where such data have greatly benefited process investigations and model development undertaken by a wide scientific community. However,~~

Formatted: Subscript

Formatted: Font: Not Italic

Formatted: Font: Not Italic

~~given the cost, effort, and power required to install and maintain a flux tower, eddy covariance towers are typically sparse relative to the scale of study sites used to address ecosystem questions. Additionally, the location of a flux tower within a watershed greatly influences measurement representativeness. For example, for logistical reasons, eddy~~. Eddy covariance towers are usually installed at valley bottoms of mountainous watersheds (Strachan et al., 2016). ~~However, microclimate caused by complex mountainous terrains (e.g., slope, aspect and elevation) can have different radiation inputs and moisture dynamics compared to flat areas where flux towers are mostly installed. Flux measurements from eddy covariance towers provide a representation of major driver and controls on ET and R_{ECO} in an ecoregion while meteorological forcing variability needs to be accounted to possibly represent various aspects introduced by complex terrain. Thus, though measurements from a single flux tower may not capture heterogeneity in ET and R_{ECO} due to complex terrain, they~~Data from flux towers should also be used carefully as flux footprints may vary significantly across sites and through time depending on site-specific information, turbulent states of the atmosphere and underlying surface characteristics (Chu et al., 2021). Given the cost and efforts required to install and maintain a flux tower, eddy covariance towers are typically sparse and may not capture complex fluxes at sites with complex terrains, such as montane environments. Though measurements from a single flux tower may not capture heterogeneity in ET and R_{eco} due to complex terrains, they can support the development of statistical or physical-based models integrated with other types of data to provide ET and ~~R_{ECO}~~R_{eco} estimation ~~in high resolution over space and time~~as we describe herein.

Physically-based numerical models, which ~~numerically~~ represent land-surface energy and water balance, have also been used to estimate ET and ~~R_{ECO}~~R_{eco} (Tran et al., 2019; Williams et al., 2009)~~. These physically-based models solve physical equations to simulate the exchanges of energy, heat, water and carbon across atmosphere-canopy-soil compartments. Examples include~~, such as the Community Land Model (CLM, Oleson et al., 2013). Performance of these models ~~depend~~depends on the accuracy of inputs and parameters, such as soil type and leaf area index, which can be difficult to obtain at a sufficiently high spatiotemporal resolution. The lack of measurements to infer parameters needed for models often leads to large discrepancies between model-based and flux-tower-based ET and ~~R_{ECO}~~R_{eco} estimates. Conceptual model uncertainty inherent in mechanistic models can also lead to ET and R_{ECO} estimation uncertainty and errors. For example, Keenan et al. (2019) suggested that current terrestrial carbon cycle models neglect inhibition of leaf respiration that occurs during daytime, which can result in a bias of up to 25 %. Chang et al. (2018) ~~used virtual experiments with 3-D terrestrial integrated modeling system to investigate why a lower ratio of transpiration to ET is always produced by large scale land surface models. Their study suggested heterogeneous fluxes caused by uneven hydraulic distribution due to complex terrain are not always considered in process-based models. These conceptual uncertainties, in addition to data sparseness and data uncertainty, further limit the applicability of physically-based models to estimate ET and R_{ECO} at high spatiotemporal scales. Semi-analytical formulations based on combinations of meteorological and empirical parameters provide a reference condition for the water and energy balance. Examples used to estimate potential ET include~~suggested that process-based models may not represent transpiration accurately due to challenges in simulating the uneven hydraulic distribution caused by complex terrain. Semi-analytical formulations are also commonly used to estimate ET, including the Budyko framework and its extensions (Budyko, 1961; Greve et al.,

Formatted: Subscript

2015; Zhang et al., 2008); the Penman-Monteith's equation (Allen et al., 1998), and the Priestley-Taylor equation ~~(Priestley and Taylor, 1972)~~(Priestley and Taylor, 1972). ~~Actual ET can then be approximated by multiplying a coefficient associated with water deficit (De Bruin, 1983; Williams & Albertson, 2004). However, even with these empirical formulations many attributes are still difficult to obtain globally at high temporal scales, such as water-vapor deficit, leaf area index, and aerodynamic conductance of different plants.~~However, these conceptual uncertainties, in addition to data sparseness and data uncertainty, still limit the applicability of these approaches.

Remote sensing products, such as Landsat imagery (Irons et al., 2012), Sentinel-2 (Main-Knorn et al., 2017) and the moderate-resolution imaging spectroradiometer (MODIS, NASA. 2008), have also been integrated to estimate ET and ~~R_{ECO} with empirical, statistical, or semi-physical relations~~R_{eco} (Abatzoglou et al., 2014; Daggers et al., 2018; Mohanty et al., 2017; Paca et al., 2019). ~~Due to the high spatial coverage of remote sensing products, global-scale estimates of ET and R_{ECO} have become feasible. For example,~~ Ryu et al. (2011) proposed the 'Breathing Earth System ~~Simulator~~Simulator' approach, which integrates mechanistic models and MODIS data to quantify ET and GPP with a spatial resolution of 1-5 km and a temporal resolution of 8 days. Ai et al. (2018) extracted ~~enhanced vegetation index, fraction of absorbed photosynthetically active radiation, and leaf area index~~indices from the MODIS dataset—and used the rate-temperature curve and strong correlations between terrestrial carbon exchange and air temperature to estimate ~~R_{ECO}~~R_{eco} at 1 km spatial resolution and 8-day temporal resolution. Ma et al. (2018) developed a data fusion scheme that fused Landsat-like-scale datasets and MODIS data to estimate ET and irrigation water efficiency at a spatial scale of ~100 meters. However, even though remote sensing data cover large areas of the earth surface, they typically do not provide information over both high spatial and temporal resolution, and ~~are also~~data quality is subject to ~~cloudy~~cloud conditions. For example, Landsat has average return periods of 16 days with a spatial resolution of 30 m (visible and near-infrared), whereas MODIS has 1-2 days temporal resolution with a 250 m or 1 km spatial resolution depending on the sensors. These resolutions are typically too coarse to enable exploration of how aspects such as plant phenology, snowmelt, and rainfall ~~impact integrated ecosystem~~influence water and energy dynamics~~.~~ of an ecosystem.

Combining machine-learning models with remote sensing products and meteorological inputs offers another option for large-scale estimation of ET and ~~R_{ECO}.~~ R_{eco}. Remotely sensed data can be good proxies for plant productivity and can be easily implemented into machine-learning models for ET and R_{ECO} estimation, such as for an enhanced vegetation index, land surface water index and normalized differences vegetation index (NDVI) (Gao et al., 2015; Jägermeyr et al., 2014; Migliavacca et al., 2015). Li and Xiao (2019) developed a data-driven model to estimate gross primary production at a spatial and temporal resolution of 0.05° and 8 days~~, respectively, using MODIS and meterological reanalysis data.~~. Berryman et al. (2018) demonstrated the value of a Random Forest model to predict growing season soil respiration from subalpine forests in the Southern Rocky Mountains ecoregion. Jung et al. (2009) developed a model tree ensemble approach to upscale FLUXNET data, where they successfully estimated ET and GPP. Other methods have used support vector machines, artificial neural networks, random forest, and piecewise regression (Bodesheim et al., 2018; Metzger et al., 2013; Xiao et al., 2014; Xu et al., 2018). These models were trained with ground-measured flux observations and other variables, and then applied to estimate ET over continental or global scales with remote sensing and meteorological inputs. Some of the most important inputs include the enhanced vegetation index, aridity index, air temperature, and precipitation. ~~However, the~~The spatiotemporal resolution of these approaches is constrained by the resolution of remote sensing products and meteorological inputs. Additionally, parameters such as leaf area index, cloudiness, and the vegetation types required by those models may not be available at the required resolution, accuracy or location. For example, in systems that have significant elevation gradients, errors may ~~result~~occur when valley-based FLUXNET data are used for training and then applied to hillslope or ridge ET and ~~R_{ECO}~~R_{eco} estimation.

Development of hybrid models that link direct measurements and/or ~~interpretable~~ mechanistic models with data-driven methods can benefit ET and ~~R_{ECO}~~R_{eco} estimation (Reichstein et al., 2019). While remote sensing data that cover large regions provide promise for informing models, quantitative interpretation of these data needed for input into mechanistic models is still challenging (Reichstein et al., 2019). Physically~~-~~ based models can provide estimates of ET and ~~R_{ECO},~~R_{eco}, but the estimate error can be high, owing to parametric, structural, and conceptual uncertainties as described above. Hybrid data-driven frameworks are ~~potentially~~ advantageous because they enable the integration of remote sensing datasets, meteorological forcings, and mechanistic model outputs of ET and ~~R_{ECO}~~R_{eco} into one model. Machine-learning approaches ~~are~~can then be applied to extract the spatiotemporal patterns for ET and ~~R_{ECO}~~R_{eco} prediction. ~~Hybrid models can utilize the high spatial coverage~~The integration of ~~remote sensing~~ multi-model and multi-data ~~(e.g., 30 m of Landsat) and high temporal resolution of direct measurement from flux towers or simulation results from mechanistic models (e.g., daily or hourly scales), thus providing alternative approaches for next-stage, more accurate estimation of~~can increase our modeling capability to estimate ET and ~~R_{ECO} at greater spatial~~R_{eco} and ~~finer temporal scales, and enhancing~~enhance our process understanding of ecosystem water and carbon cycling under climate change.

In this study, we developed a hybrid predictive modeling approach (HPM) to ~~better~~ estimate daily ET and R_{ECO} with easily acquired meteorological data (i.e., air temperature, precipitation and radiation) and remote sensing products (i.e., NDVI). HPM is hybrid as it can ~~use deep learning models to~~flexibly integrate direct measurements from flux towers and/or physically-based model results (e.g., CLM) ~~with~~and utilize deep learning long-short term memory recurrent neural network (LSTM) to establish statistical relationships among fluxes, meteorological and remote sensing inputs~~ to capture~~. Once developed, the ~~complex physical interactions within the watershed ecosystem. After development, we validated~~corresponding HPM ~~performance with the FLUXNET dataset and benchmarked the CLM model at select sites. We then~~can be used ~~the HPM for~~as a modeling tool to estimate ET and ~~R_{ECO} estimation at the mountainous East River Watershed in Colorado, USA~~R_{eco} over space and ~~investigated how ET~~time. We developed four use cases to demonstrate the applicability of HPM based on site-specific data and ~~R_{ECO} dynamics varies within the East River Watershed.~~

model availability. The remainder of this paper is organized as follows. Section 2 mainly describes the sites considered in this study and how data were acquired and processed. Section 3 presents the methodology of the HPM approach, followed by the results of various use cases presented in Section 4. Discussion and conclusion are provided in Sections 5 and 6, respectively.

2. Site Information, Data Acquisition and Processing

Formatted: Subscript

The HPM method was tested using data from a range of different ecosystem types to explore its performance under different conditions. ~~However, we~~We place a particular focus on mountainous sites, given their regional and global importance yet challenges associated with ET and ~~R_{ECO}~~R_{eco} in these regions, as described above.

2.1 FLUXNET Stations and Ecoregions

Nine FLUXNET stations, which cover a wide range of climate and elevations, were selected for this study (Table 1 and Figure 1~~), which cover a wide range of climate and elevations.~~). These stations have elevations from 129 m (US-Var) to 3050 m (US-NR1), mean annual air temperature from 0.34℃ (CA-Oas) to 17.92℃ (US-SRM), and mean annual precipitation from 320 mm (US-Whs) to 800 mm (US-NR1). These FLUXNET stations also cover a wide range of vegetation types (i.e., evergreen forest, deciduous forest, and shrublands). As indicated by Hargrove et al. (2003), FLUXNET stations ~~provide a good representation of~~were maintained to capture watershed dynamics at different ecoregions, which are areas that display recurring patterns of similar combinations of soil, vegetation and landform characteristics (Omernik, 2004). Omernik & Griffith. (2014) delineated the boundaries of ecoregions through pattern analysis that consider the spatial correlation of both physical and biological factors (i.e., soils, physiography, vegetation, land use, geology and hydrology) in a hierarchical level. FLUXNET stations considered in this study mainly locate in ~~4~~four unique ecoregions (Table 1). As is described below, we developed local-scale (i.e., point scale) HPM that are representative for different ecoregions using data provided at these FLUXNET stations to estimate ET and R_{ECO}, and validated the HPM estimates with measurements from stations within the same ecoregion.

2.2 SNOTEL Stations

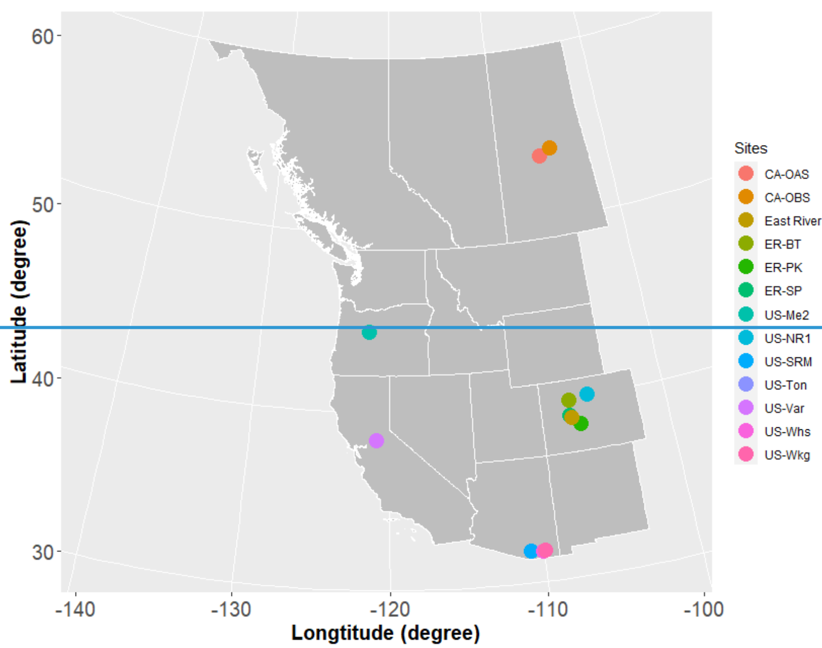
For reasons described below, we performed a deeper exploration of HPM performance within one of the mountainous watershed sites (the East River Watershed of the Upper Colorado River Basin, USA), which is located in the "western cordillera" ecoregion. At this site, we utilized meteorological forcings data from three snow telemetry (SNOTEL) stations. These sites include the Butte (ER-BT, id: 380), Porphyry Creek (ER-PK, id: 701) and Schofield Pass (ER-SP, id: 737) sites. ~~A~~A one-dimensional (vertical) CLM model was developed at these SNOTEL stations that provides physically-model-based ET estimation (Tran et al., 2019). Table 1 summarizes the SNOTEL stations used in this study and the corresponding climate characteristics. Figure 1 shows the geographical locations of FLUXNET and SNOTEL stations selected in this study.

Table 1. Summary of FLUXNET stations and SNOTEL stations information. * denotes SNOTEL stations and all others are FLUXNET stations. Dfc, Bsk, Csa represent subarctic or boreal climates, semi-arid climate, Mediterranean hot summer climates, respectively. ENF, DBF, WSA, GRA, and OSH represent evergreen needleleaf forest, deciduous broadleaf forests, woody savannas, grasslands, open shrubland, respectively. FLUXNET data were obtained from the FLUXNET2015 database.

Site ID	~~Site Name~~	Latitude, Longitude	Elevation (m)	Mean Annual air temperature (℃)	Mean Annual Precipitation (m)	Climate Koeppen	Vegetation IGBP	~~Ecoregions~~Ecoregion (Level II)	Period of Record
US-NR1	~~Niwot Ridge~~	(40.0329, -105.5464)	3050	1.5	800	Dfc	ENF	Western Cordillera	2000-2014
CA-	~~Saskatchewan-Aspen~~	(53.6289, -	530	0.34	428.53	Dfc	DBF	Boreal Plain	1997-

Formatted: Line spacing: 1.5 lines
Deleted Cells
Formatted: Line spacing: 1.5 lines
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Line spacing: 1.5 lines
Formatted: Line spacing: 1.5 lines
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Line spacing: 1.5 lines
Formatted: Line spacing: 1.5 lines
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold
Formatted: Font: 9 pt, Bold

Oas		106.1978)							2010
CA-Obs	~~Saskatchewan-Black Spruce~~	(53.9872, -105.1178)	628.94	0.79	405.6	Dfc	ENF	Boreal Plain	1999-2010
US-SRM	~~Santa Rita Mesquite~~	(31.8214, -110.8661)	1120	17.92	380	Bsk	WSA	Western Sierra Madre Piedmont	2005-2015
US-Ton	~~Tonzi Ranch~~	(38.4316, -120.9660)	177	15.8	559	Csa	WSA	Mediterranean California	2002-2015
US-Var	~~Vaira Ranch-Ione~~	(38.4133, -120.9507)	129	15.8	559	Csa	GRA	Mediterranean California	2002-2015
US-Whs	~~Walnut Gulch Lucky Hills Shrub~~	(31.7438, -110.0522)	1370	17.6	320	Bsk	OSH	Western Sierra Madre Piedmont	2008-2015
US-Wkg	~~Walnut Gulch Kendall Grasslands~~	(31.7365, -109.9419)	1531	15.64	407	Bsk	GRA	Western Sierra Madre Piedmont	2005-2015
US-Me2	~~Metolius mature ponderosa pine~~	(44.4523, -121.5574)	1253	6.28	523	Csb	ENF	Western Cordillera	2012-2015
ER-BT*	~~East River-Butte~~	(38.894, -106.945)	3096	2.38	821	Dfc	N/A	Western Cordillera	1995-2017
ER-SP*	~~East River-Schofield Pass~~	(39.02, -107.05)	3261	2.46	1064	Dfc	N/A	Western Cordillera	1995-2017
ER-PK*	~~East River-Porphyry Creek~~	(38.49, -106.34)	3280	1.97	574	Dfc	N/A	Western Cordillera	1995-2017



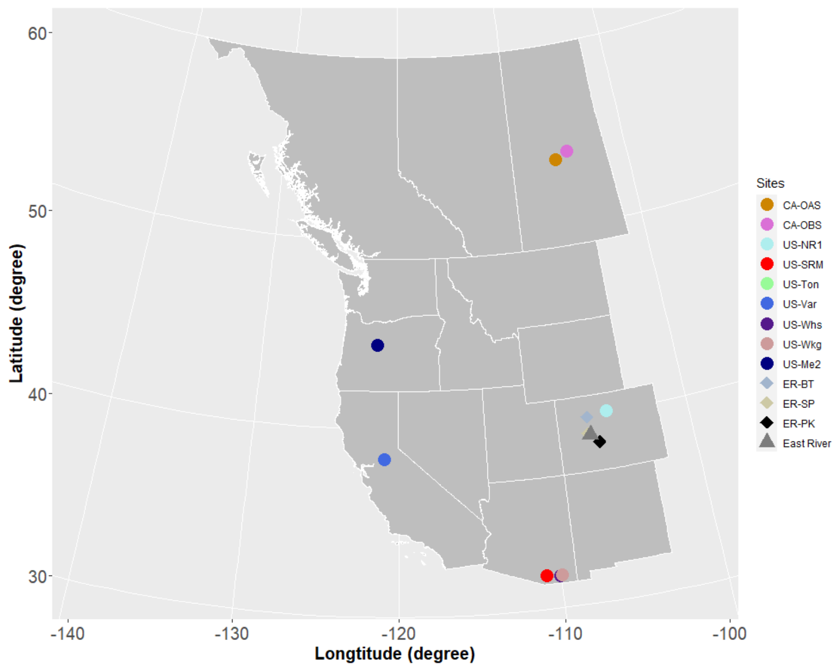


Figure 1. Location of sites considered in this study. Note: US-Ton and US-Var; US-Whs and US-Wkg are ~~at the same locations.~~ closed to each other. East River Watershed is located next to ER-BT. The white lines delineate Western US states and Canadian provinces. Circles represent FLUXNET sites, diamonds represent SNOTEL sites and triangle represents the East River Watershed.

2.3 East River Watershed Characteristics and Previous Analyses

Data from the East River Watershed were used to explore how ET and ~~R_{ECO}~~R_{eco} dynamics estimated from the developed HPM vary with different vegetation and meteorological forcings. The East River Watershed is located northeast of the town of Crested Butte, Colorado. This watershed has an average elevation of 3266 m, with significant gradients in topography, hydrology, geomorphology, vegetation, and weather. ~~The watershed has a mean annual temperature around 0°C , with an average of 1200 mm yr^{-1}~~The mean annual air temperature in the East River is ~2.4°C, with average daily air temperatures of −7.6°C and 13.4°C in December and July respectively (Kakalia et al., 2020) and an average of 1200 mm yr^{-1} total precipitation (Hubbard et al., 2018). Consisting of montane, subalpine, and alpine life zones, each with distinctive vegetation biodiversity, the East River Watershed is a testbed for the US Department of Energy Watershed Function Scientific Focus Area Project, led by the Lawrence Berkeley National Laboratory (Hubbard et al., 2018). The project has acquired a range of datasets, including hydrological, biogeochemical, remote sensing, and geophysical datasets.

Recently completed studies at the East River Watershed were used in this study to inform HPM and to assess the results. For example, physically-model-based estimations of ET at this site (Tran et al., 2019) were used

Formatted: Subscript

herein for HPM development and validation. Falco et al. (2019) used machine-learning-based remote sensing methods to characterize the spatial distribution of vegetation types, slopes, and aspects within a hillslope at the East River Watershed, which were used with obtained HPM estimates to explore how vegetation heterogeneity influences ET and R_{ECO} dynamics. To perform this assessment, we computed the spatial distribution of vegetation types at watershed scale based on Falco et al. (2019). We evaluated manually and selected 16 locations within the East River Watershed having different vegetation types and slope aspects. These 16 locations were chosen to be at the center of vegetation patched and covered by one vegetation type. A summary of the locations is presented in Table 2; the spatial distribution of the locations is shown in Figure 2.

Table 2: Location and vegetation types of East River Watershed sampling points (Figure 2)

Easting (m)	Northing (m)	Vegetation Type	Aspect	Elevation (m)
327085	4309878	Deciduous Forest	South	2983
326288	4312504	Deciduous Forest	South	3177
330012	4313132	Deciduous Forest	North	3108
326854	4313192	Deciduous Forest	South	3098
328246	4312832	Meadow	South	3095
327010	4315059	Meadow	South	2790
328738	4306139	Meadow	North	2890
334270	4309465	Meadow	North	2929
333406.5	4308340	Riparian Shrubland	South	2760
327846	4312497	Riparian Shrubland	South	2723
334641	4305632	Riparian Shrubland	North	2740
330760	4310097	Riparian Shrubland	South	2855
329573	4314569	Evergreen Forest	South	3026
333106	4307313	Evergreen Forest	North	3102
325056	4310456	Evergreen Forest	South	2961
335141	4309614	Evergreen Forest	North	3131

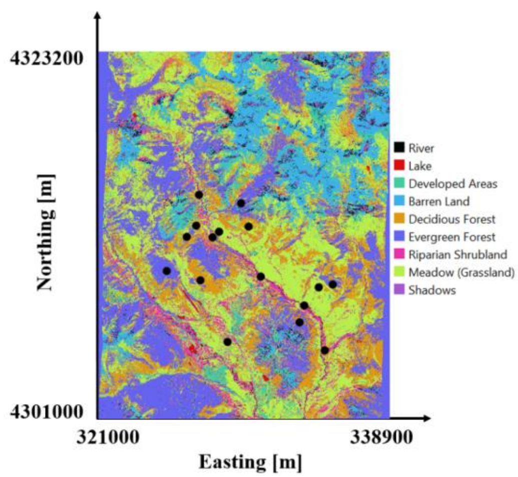


Figure 2: Vegetation classification of the East River, CO Watershed from Falco et al. (2019). East River sites selected in this study are denoted by black circles.

2.4 Data Collection and Processing

To enhance transferability of the developed HPM strategy to less intensively characterized watersheds, we selected only "easy to measure" or "widely available" attributes, such as precipitation, air temperature, radiation and NDVI, as inputs to the HTM model. Soil temperature was used when available. The data sources used for these inputs include FLUXNET data (https://fluxnet.fluxdata.org/), SNOTEL data (https://www.wcc.nrcs.usda.gov/snow/) and developed CLM model (Tran et al., 2019) at SNOTEL stations, DAYMET meteorological inputs (Thornton et al., 2017) and remote sensing data from Landsat imageries (Irons et al., 2012).

~~A variety of measured data and model outputs were used to train and validate HPM. We obtained daily meteorological data, including air temperature, precipitation, radiation, ET, and R_{ECO} data, from the FLUXNET database at the selected FLUXNET sites. The pipeline of data processing for FLUXNET dataset is provided at https://FLUXNET.fluxdata.org/.~~ We identified some data gaps and erroneous data (especially during winter seasons) for the ET estimates at US-NR1, which were cleaned following the procedures presented in Rungee et al. (2019). ~~The meteorological data were used as inputs for HPM development, and ET and R_{ECO} data from these sites were used for HPM validation.~~ At the three selected SNOTEL stations, ~~we obtained~~ air ~~temperature, precipitation, and snow-water-equivalent data from the SNOTEL database. Air~~ temperature data at these three SNOTEL stations were processed following Oyler et al. (2015)~~, given potential systematic artifacts. Snow-water-equivalent data are not easily acquired, and thus were not considered as inputs for HPM. However, a categorical variable was constructed to assimilate information regarding snow (Section 3.2.1). CLM models were generated following Tran et al. (2019) for the SNOTEL stations and US-NR1 to assess the spatiotemporal variability of ET at the East River Watershed and for training and validating HPM (Section 4.3). The DAYMET dataset (Thornton et al., 2017) provided gridded daily weather-forcings-attribute estimates at a 1 km spatial resolution. We obtained the incident radiation data from DAYMET at the SNOTEL stations as inputs for HPM. For the East River Watershed sites, meteorological forcings data, including air temperature, precipitation and radiation, were also obtained from DAYMET. The low spatial resolution of DAYMET data introduces uncertainty in HPM estimation of ET and R_{ECO}, which will be discussed in the following sections. We calculated the NDVI time series from the red band (RED) and near-infrared band (NIR) from Landsat 5, 7, and 8 images at all selected FLUXNET sites, SNOTEL stations, and East River Watershed sites at a spatial scale of 30 m.~~

~~Since cloud conditions can severely decrease data quality, we~~ and radiation data was obtained from DAYMET. CLM models were generated following Tran et al. (2019) for the SNOTEL stations and US-NR1. At the East River Watershed sites, data were obtained from DAYMET. NDVI time series were calculated from the red band and near-infrared band from Landsat 5, 7 and 8 images at all sites. We used the cloud-scoring algorithm provided in the Google Earth Engine to mask clouds in all retrieved data, only selecting the ones that had a simple cloud score below 20 to ensure data quality. Given the different calibration sensors used in Landsat 5, 7, and 8, we also followed the processes described in Homer et al. (2015) and Vogelmann et al. (2001) to keep NDVI computations consistent over time. Landsat satellites have a return period of 16 days, and thus we performed a reconstruction of NDVI time series to obtain daily scale time data (Section 3.2.2).

Formatted: Indent: First line: 0.5"

3. Hybrid Predictive Modeling Framework

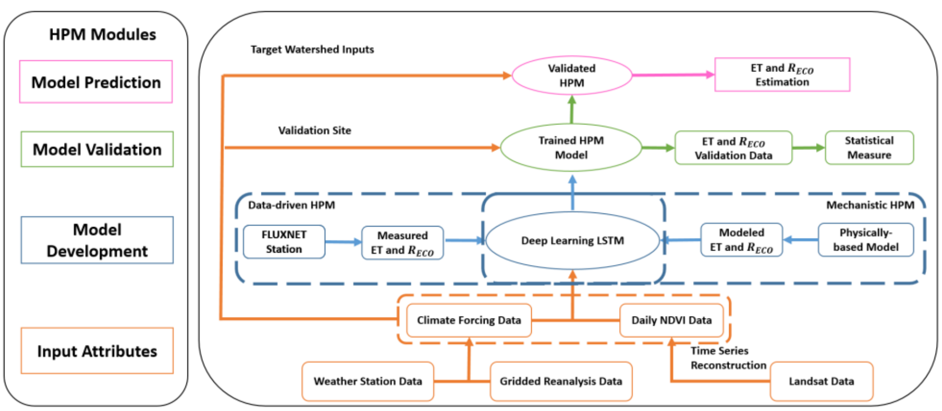
In this section, we illustrate the steps for building an HPM model for ET and ~~R_{ECO}~~ R_{eco} estimation over time and space. Figure 3 presents the general framework of HPM, which includes modules for data preprocessing, model development, model validation, and predictive modeling.

3.1 Model Framework

HPM establishes relationships among meteorological forcings attributes, NDVI, ET~~,~~ and ~~R_{ECO}~~ R_{eco} (Figure 3)~~. Both input data (e.g., meteorological forcings) and output data (ET and R_{ECO}) used for training and validation are preprocessed for gap filling, smoothing, and data updating. HPM "learns" the complex space-time relationship among meteorological forcings, NDVI, ET, and R_{ECO})~~ using ~~a~~ deep-learning-based module (~~deeply~~fully connected deep neural networks and ~~a~~ long short-term memory recurrent neural ~~network). HPM then can be used for ET and R_{ECO} estimation at sparsely monitored watersheds. Individual HPM models can be trained in two different ways using ET and R_{ECO} information: with data obtained from flux towers ("data-driven HPM") or with outputs from 1-D physically-based models ("mechanistic HPM"). In both cases, the models obtained with local data are then used to estimate ET and R_{ECO} at other sites in the same ecoregion (see Section 2.1). For ecoregions not represented by FLUXNET sites, it is necessary to develop mechanistic HPM that enables ET and R_{ECO} estimation over space and time.~~

~~HPM has several additional modules, including model development, model validation, and model prediction modules. In the HPM model development module, deep-learning algorithms are trained with input features and response data until a pre-defined "stopping criteria" (e.g., root mean squared error, RMSE) is met, indicating subsequent training would lead to minimal improvement. In the validation module, estimation outputs from the "trained HPM models" are compared with other ET and R_{ECO} data obtained from other independent sites or mechanistic models within the same ecoregion. Statistical measures, including adjusted R^2 and mean absolute error (MAE), are computed to evaluate the performance of HPM models. In the predictive model module, meteorological forcings data and remote sensing data are processed at target sites of interest, and the validated HPM model is used to estimate ET and R_{ECO} at these sites. ET and R_{ECO} outputs estimated from HPM at sparsely monitored watersheds then provide alternative datasets for process understanding within the target watersheds.~~

Formatted: Indent: First line: 0.5"



~~**Figure 3: Hybrid Predictive Model (HPM) Framework. The HPM model mainly consists of four modules: Input Attributes, Model Development, Model Validation and Model Prediction, represented by rectangles with colors. Arrows represent the linkages among different modules. Choices of data-driven HPM or mechanistic HPM depend on the ecoregion of target watershed and data availability.**~~

networks). Long short-term memory (LSTM, Hochreiter & Schmidhuber, 1997) is ~~capable of identifying long-term dependencies between climate and environmental data.~~ a type of recurrent neural network (RNN) capable of learning temporal dependence without suffering from optimization difficulties (e.g., vanishing errors). An LSTM layer consists of memory blocks and unique cell states that are controlled by three multiplicative units, including the input, output and forget gates. These gates regulate the flow of information and decide which data in a sequence is important to keep or throw away. Through the LSTM structure, even information from the earlier time steps can make its way to later time steps, reducing the effects of short-term memory and thus capturing long-term dependence. LSTM has been previously used to capture such dependencies between climate and environmental data. For example, Kratzert et al. (2018) successfully used LSTM to learn the long-term dependencies in hydrological data (e.g., storage effects within catchments, time lags between precipitation inputs and runoff generation) for rainfall-runoff modeling. LSTM has also been used for gap filling in hydrological monitoring networks in the spatiotemporal domain (Ren et al., 2019). ~~In this study, the outputs (ET or R_{ECO}) denoted as y are predicted from the input $x = [x_1, x_2, \ldots, x_T]$, consisting of the last T consecutive time steps of attributes, such as meteorological forcings attributes (e.g., air temperature and precipitation) and remote sensing attributes (i.e., NDVI). In a recurrent neural network (RNN), h_t represents the internal state at every time step t that takes in current input value x_t and previous internal state h_{t-1}, and is recomputed along the time axis using the following equation:~~ More information about the LSTM-RNN method is provided by Olah (2015).

$$\cancel{h_t = g(Wx_t + Uh_{t-1} + b),} \qquad \cancel{(1)}$$

~~where g represents the hyperbolic tangent activation function, W and U are trainable weight metrices of the hidden state h, and b is a bias vector. W, U and b are all trainable through optimization. LSTM introduces the cell state c_t, which makes LSTM powerful in identifying long-term dependencies in a statistical manner. The cell state c_t has three gates structures, including "forget gates" (which determine what information from previous cell states will be forgotten), "~~HPM modules include input ~~gates" (which determine what information will be conveyed from the forget gate) and "output gates" (which return information from cell state c_t to a new state h_t). With these gate structures, the cell state c_t controls what information will be forgotten, conveyed, and updated over time. The forget gate is formulated as follows:~~

$$\cancel{f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right),} \qquad \cancel{(2)}$$

~~where f_t results in a value between 0 and 1 indicating the degree of information to be forgotten; σ is the logistic sigmoid function, $and\ W_f, U_f$ and b_f are trainable parameters. Next, the input gate decides which values will be updated in the current cell state, and creates a vector of candidate values $\widetilde{c_t}$ in the range of (-1, 1) through a $tanh$ layer, which will be used to update the current state. With the candidate values calculated from the current state, and the information conveyed from the forget gate, we can calculate the current cell state as follows:~~

$$\cancel{i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i),} \qquad \cancel{(3)}$$

$$\cancel{\widetilde{c_t} = tan\,h(W_c x_t + U_c h_{t-1} + b_c),} \qquad \cancel{(4)}$$

$$\cancel{c_t = f_t * c_{t-1} + i_t * \widetilde{c_t},} \qquad \cancel{(5)}$$

~~where i_t is the input gate that defines which information of $\widetilde{c_t}$ will be used to update the current cell state and is in the range of (0, 1); c_t represents the current cell state; and $W_c, U_c, b_c, W_i, U_i,\ and\ b_i$ are trainable parameters.~~

Formatted: Indent: First line: 0"

~~Finally, the output gate o_t controls the information of cell state c_t to a new hidden state h_t, which is computed using the following equation:~~

$$\cancel{o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o),} \qquad \cancel{(6)}$$

$$\cancel{h_t = \tanh(c_t) * o_t,} \qquad \cancel{(7)}$$

~~With the new hidden state calculated, ET and~~ $\cancel{R_{ECO}}$ ~~can be calculated using a one unit dense layer:~~

$$\cancel{y_t = W_d h_t + b_d,} \qquad \cancel{(8)}$$

~~where~~ $\cancel{W_d}$ ~~and~~ $\cancel{b_d}$ ~~are additional trainable parameters. In summary, the LSTM unit calculates the internal state using current meteorological forcings and remote sensing data at every time step. The forget gate, input gate, and output gate decide what information from previous time steps will be kept, updated, and conveyed to the new hidden state. Finally, with a single dense layer, the algorithm will output ET and~~ $\cancel{R_{ECO}}$ ~~estimation from the trained~~ attributes, model~~.~~

~~A 70%-30% split between training and~~ development, validation ~~time series~~ , and prediction. Based on data availability, input features are obtained from flux towers, gridded meteorological data ~~was applied here, where the first 70% of the data were used~~, and remote sensing data; all data are preprocessed for gap filling, smoothing, and updating. In the HPM model development ~~as a learning process, and 30% of the data were used as validation sets at~~module, individual ~~sites. At the East River Watershed, HPM results were also validated~~ HPM models can be trained in two different ways based on data availability: with data obtained from flux towers ("data-driven HPM") or with ~~benchmark CLM~~ outputs from physically-based models ("mechanistic HPM"). Seventy percent of these data are used for training LSTM to learn the interactions among input features, ET, and R_{eco}, until a pre-defined "stopping criteria" (e.g., root mean squared error, RMSE) is met, indicating subsequent training would lead to minimal improvement. ~~Tran et al. (2019) and FLUXNET measurements. We used the mean absolute error (MAE), and adjusted~~ $\cancel{R^2}$ ~~as the statistical measure to determine model performance.~~ In most models, the configuration of the neural networks includes a first LSTM layer with 50 units, a second LSTM layer with 25 units, and a dense layer with 8 units having L2 regularizers, and a final output dense layer. Dropout layers are also embedded in the model to prevent overfitting. There are 11600 and 7600 parameters for the first and second LSTM layers; 208 and 9 for the first and second dense layers and no parameters for the dropout layers. Other configurations of networks may provide better estimation results; however, they are not assessed in this study as the proposed configuration already provide reasonable results. ~~More information about the LSTM-RNN method is provided by Olah. (2015).~~

In the validation module, we implemented a validation procedure that uses the remaining 30 % of the data to assess model performance. Estimation outputs from the trained HPM models are also compared with other ET and R_{eco} data obtained from other independent sites or mechanistic models within the same ecoregion. Statistical measures such as adjusted R^2 and mean absolute error (MAE) are computed to evaluate the performance of HPM models. In the predictive model module, meteorological forcings data and remote sensing data are processed at target sites of interest, and the validated HPM model is used to estimate ET and R_{eco} at these sites. ET and R_{eco} outputs estimated from HPM at sparsely monitored watersheds then provide alternative datasets for process understanding within the target watersheds.

Formatted: Indent: First line: 0.5"

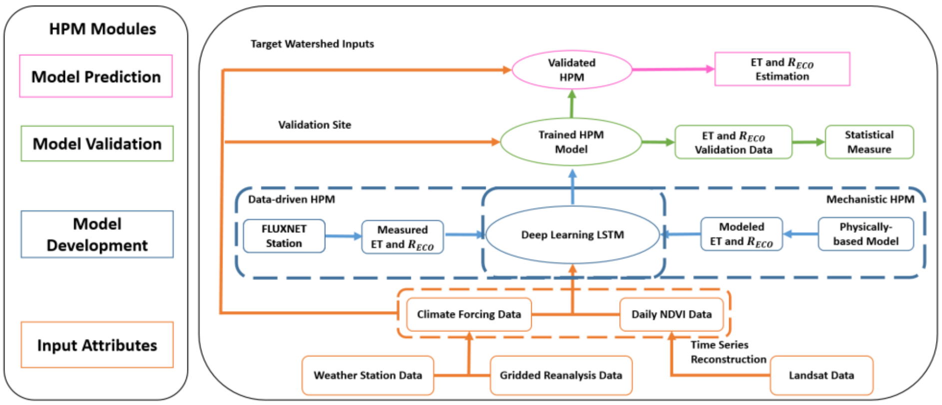


Figure 3: Hybrid Predictive Model (HPM) Framework. The HPM model mainly consists of four modules: Input Attributes, Model Development, Model Validation and Model Prediction, represented by rectangles with colors. Arrows represent the linkages among different modules. Choices of data-driven HPM or mechanistic HPM depend on the ecoregion of target watershed and data availability.

3.2 Feature Selection

~~Key properties influencing ET and R_{ECO} dynamics are linked to snow processes, plant dynamics, moisture stresses, radiation inputs and other relevant processes. However, at~~ At sparsely monitored watersheds, only weather reanalysis data and remote sensing data are commonly available. Thus, we mainly considered air temperature, radiation, precipitation, vegetation indices (e.g., NDVI) and variables inferred from these data as inputs for HPM. Soil temperature when available is used at FLUXNET sites. Other key attributes that depend on depth and site-specific characteristics such as soil moisture and snow depth are not used in current HPM models due to data availability.

3.2.1 Snow information

In snow-influenced mountainous watersheds, ~~snow dynamics significantly influence water and carbon fluxes. Because of the difficulties in measuring snow time series over space, we did not directly use attributes such as snow water equivalent as input to HPM. Instead,~~ we separated precipitation data into snow precipitation (when air temperature < 0) and rainfall precipitation (when air temperature > 0~~). This~~), which is in line with what has been used in hydrological models such as CLM (Oleson et al., 2013). ~~Note that for certain sites in this study, snow is not present (e.g., US-Ton). In order to capture the dynamics of snow processes, such as accumulation and melting, we constructed a categorical variable (sn), as follows:~~Knowles et al. (2016) discovered a significant correlation between day of peak snow accumulation, snowmelt and air temperature. To capture snow related dynamics (e.g., snowmelt), we constructed a categorical variable (sn) based on air and soil temperature thresholds. Note: this may not be needed if snow data becomes available and at sites where snow is rarely present.

$$sn = \begin{cases} 0, during\ snow\ accumlation;\ \ SWE > 0\ and\ SWE < peak\ SWE \\ 1, during\ snow\ melting; SWE > 0\ and\ \ SWE \leq peak\ SWE \\ 2, no\ snow; SWE = \ 0 \end{cases}, \quad (9)$$

~~Since data on peak SWE are rarely available because of the difficulties in measuring snow, we also define a proxy categorical variable, ***sn***. When no SWE measurements were available, we estimated ***sn*** using air and soil temperature data following Knowles et al. (2016), who found significant correlations between the day of peak snow accumulation and first day of air temperature above 0 degrees Celsius, as follows:~~

$$\boldsymbol{sn} = \begin{cases} 0, during\ snow\ accumlation;\ \ Air\ Temperature < 0 \\ 1, during\ snow\ melting; Air\ Temperature > 0\ while\ Soil\ Temperature \leq 0, \\ 2, no\ snow; Air\ Temperature\ and\ Soil\ Temperature > 0 \end{cases}$$ ~~(10~~ (1)

3.2.2 Vegetation information

~~To mitigate the long return periods of satellites and the presence of clouds, we~~ We reconstructed daily NDVI values based on meteorological ~~forcings~~forcing data (e.g., air temperature, precipitation, radiation) using ~~deep learning recurrent neural networks, leading~~LSTM to ~~estimates of NDVI at daily~~ increase the temporal ~~resolution. For example,~~coverage of NDVI. Figure 4 represents Landsat-derived NDVI and reconstructed NDVI values for two sites at the East River, CO watershed: Butte (ER-BT), and Schofield Pass (ER-SP). ~~Figure 4 reveals that based on meteorological forcings data only, the reconstructions achieved an adjusted R^2 of 0.65.~~ Though not ideal, as satellites continue to advance and more training data becomes available, the accuracy of NDVI temporal reconstruction is expected to increase.

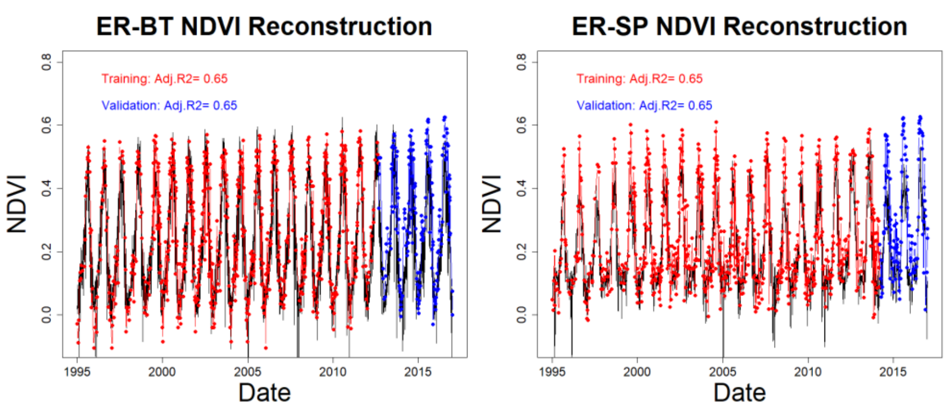


Figure 4: Temporal reconstruction of NDVI at ER-BT (left) and ER-SP (right). Black lines represent reconstructed daily NDVI. Red points are used for training and blue points are used for validation

3.3 Use Cases

We developed four different use cases to demonstrate the applicability of HPMs based on site-specific data and model availability. Use case 1 focuses on ET and R_{eco} in the time domain, where a HPM is trained on direct measurements from flux tower. A 70%-20%-10% training-validation-prediction split of the data was used. These HPMs are useful for time series gap filling and future prediction. Use case 2 and use case 3 have emphasis on providing ET and R_{eco} over space, where use case 2 uses data-driven HPM and use case 3 utilizes mechanistic HPM. Data-driven HPM is trained with data from flux tower and mechanistic HPM is trained upon outputs from a mechanistic model (e.g., CLM). These HPMs are usually trained at well monitored watersheds where either flux data is available or data support the development of a mechanistic model. After training, these HPMs integrate meteorological and remote sensing inputs to provide ET and R_{eco} at target sparsely monitored watersheds within the same ecoregion. For both use case 2 and 3, we validated the HPM estimations against data from other sites within the same ecoregion. Use case 4 focuses on the East River Watershed, where we demonstrate how HPM can increase our understanding of ecosystem fluxes and explore the limitations of HPM in mountainous watersheds. Use case 4 estimations were validated against data extracted from other studies.

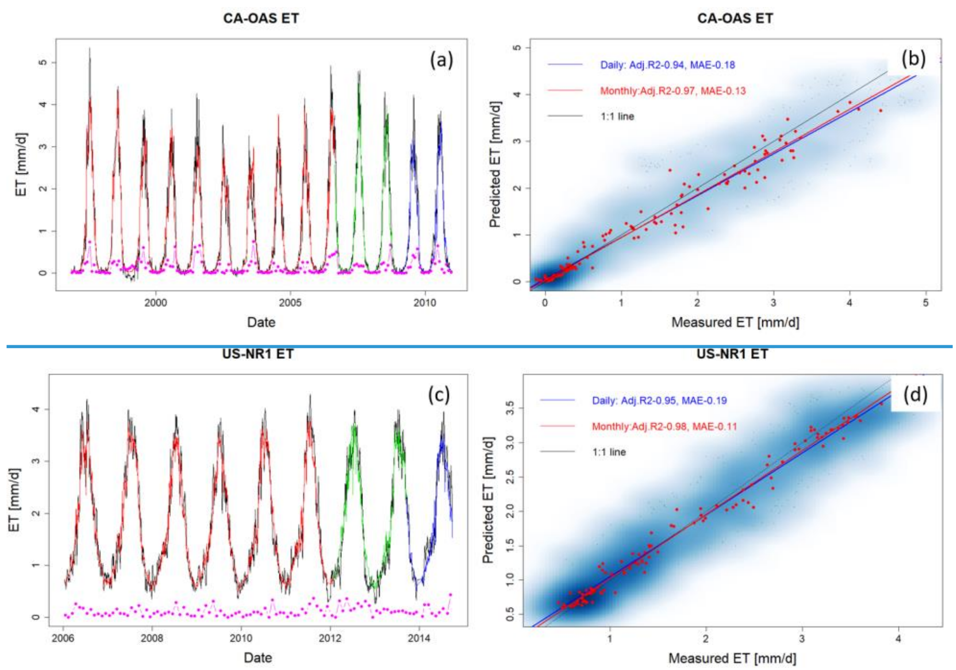
4. Results

~~We tested HPM's capabilities using four different use cases to explore different conditions. First, we tested the capability of HPM to estimate long term temporal dependency among meteorological forcings, ET, and R_{ECO} (Use Case 1; presented in Section 4.1). Second, we validated HPM's capability to estimate the spatial distribution of ET and R_{ECO} over space in selected watersheds, where we developed HPM using existing FLUXNET data (Use Case 2; data-driven HPM, Section 4.2) or outputs from a mechanistic model (Use Case 3; physical-model-based HPM, Section 4.3). In Use Case 4, HPM was used to estimate ET and R_{ECO} at selected sites within the East River Watershed and to distinguish how ET and R_{ECO} dynamics varies in the East River Watershed (Section 4.4). Temporal resolution of HPM models for all Use Cases are at daily scale and the spatial resolution depends on the use of meteorological forcing data. These four use cases illustrate and demonstrate how HPM can be developed and applied at target watersheds where data are sparse.~~

4.1 Use Case 1: ET and R_{ECO} Time Series Estimation with HPM Developed at FLUXNET Sites

Local HPMs were developed to estimate ET and ~~R_{ECO}~~ R_{eco} using flux tower data obtained from FLUXNET sites listed in Table 1. At all FLUXNET sites, air temperature, precipitation, net radiation, NDVI and soil temperature were used. For US-NR1, CA-Oas and CA-Obs, *sn* is also included. The results, which are shown in ~~Figure~~Fig. 5, A1-A4 and Table 3, reveal that the HPM approach was effective for estimating ET~~. Adjusted R^2 between the HPM estimates~~ and ~~flux tower measurements are above 0.85 for all sites, and mean absolute errors are small at a level of ~0.2 mm/d. Figure 5 displays the daily scale estimation of ET from HPM US-NR1 and CA-OAS (other sites provided in supplementary material), and presents monthly mean ET values of measurements, HPM estimations, and differences.~~ R_{eco}. The long-term trends in ET and R_{eco} are well captured by HPM. ~~At larger temporal scales (monthly or yearly), HPM provides reasonable estimation of ET at these sites.~~ However, short-term fluctuations in ET and R_{eco} during the summer periods are also not well captured by ~~ET, specifically at California sites during the periods when plant transpiration~~HPM. For example, at US-Ton and US-Var, we observed an increasing discrepancy in summer month ET and R_{eco}. This is mainly caused by insufficient training for summer extremes. At US-Me2, we observed significant increasing errors in the validation set, especially for R_{eco} that are caused by significant differences in raw data between 2002-2010 (data used for training) and ~~soil evaporation are constrained by soil moisture (Figure A2 panel a~~post-2011 (data used for validation).

Formatted: Font: Italic


CA-OAS ET
(a)
ET [mm/d]
Date
CA-OAS ET
(b)
Daily: Adj.R2-0.94, MAE-0.18
Monthly:Adj.R2-0.97, MAE-0.13
1:1 line
Predicted ET [mm/d]
Measured ET [mm/d]
US-NR1 ET
(c)
ET [mm/d]
Date
US-NR1 ET
(d)
Daily: Adj.R2-0.95, MAE-0.19
Monthly:Adj.R2-0.98, MAE-0.11
1:1 line
Predicted ET [mm/d]
Measured ET [mm/d]

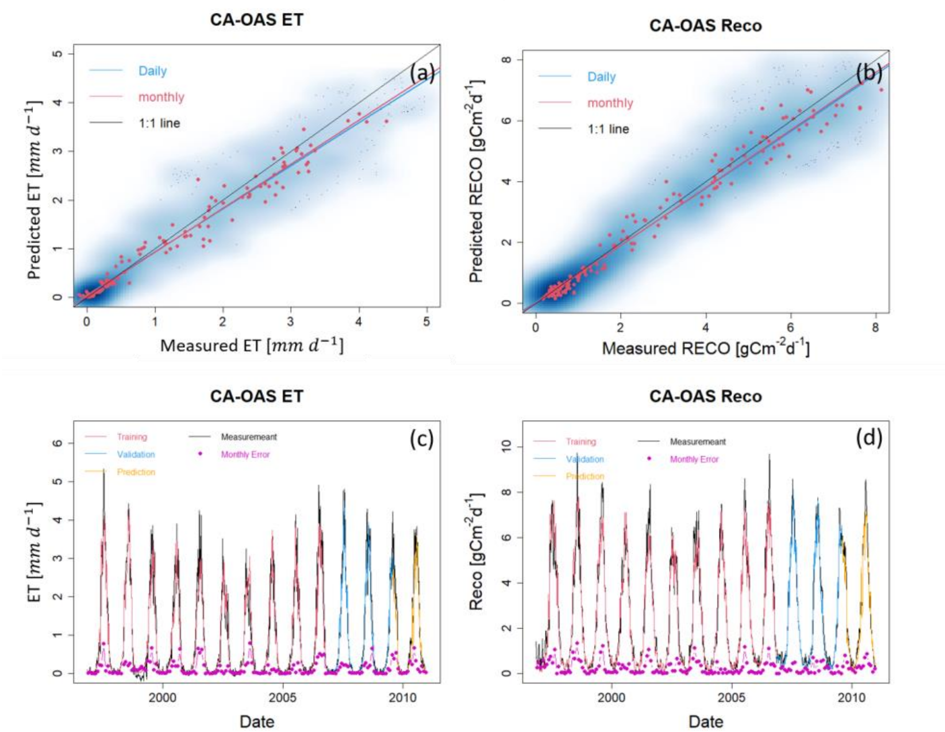
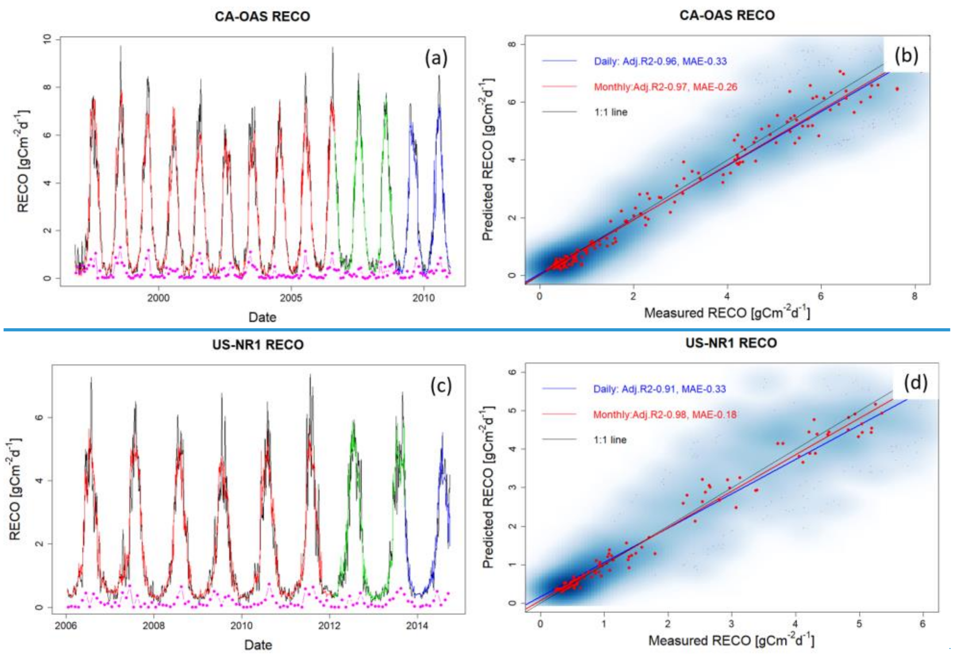


Figure 5: ET and R_{eco} estimation with data from FLUXNET sites at CA-OAS~~ and US-NR1~~. Panels (a) and (~~c) illustrate the daily estimation of ET with red, green, and blue lines representing data used for training, validation, and prediction, respectively, and the black line showing the eddy covariance measurements. Pink points describe monthly mean difference between HPM estimation and measured data. Panels (b) and (d~~b) show the scatter plots of daily (blue) and monthly (red) ET~~,~~ and R_{eco} between HPM estimation and FLUXNET data. Darker blue clouds represent greater density of data points. Panels (c) and (d) present the daily HPM estimation of ET and R_{eco} separated by training, validation and prediction sets. Pink points depict monthly error between HPM estimation and FLUXNET data. Results for other sites are included in supplementary materials below (~~Figures~~Fig. A1~~ and~~, A2~~)~~.

~~Similarly, Table 3 and Figure 6 reveal that HPM was also effective in estimating R_{ECO}, leading to small MAE and adjusted R^2 of 0.8 between estimated and measured R_{ECO} except for US-Ton and US-Var. Figure 6 presents HPM-estimated R_{ECO} at US-NR1 and CA-OAS, with other sites presented in Figures~~, **A3 and A4**~~. Long-term dynamics of R_{ECO} are also successfully captured by HPM; however, HPM does not accurately capture R_{ECO} during peak growing seasons. For example, we observed an over estimation of R_{ECO} during 2012 summer at US-Whs, whereas at US-NR1 HPM estimation during peak growing season are smaller than measured values. While soil moisture is important for R_{ECO} during peak growing season (Ng et al., 2014; Wang et al., 2014), the developed HPM currently does not include soil moisture as a key attribute. HPM R_{ECO} estimation at US-Ton and US-Var show higher uncertainties (i.e., $MAE > 0.4$ and Adj. $R^2 < 0.8$). At these sites limited by water conditions (e.g., US-Ton)~~

Formatted: Font: 9 pt, Bold

~~and sites with seasonally dry periods (e.g., US-Whs), it is necessary to include variables that could provide information regarding moisture stresses in the subsurface. Soil moisture that directly quantify water stress can be helpful to increase R_{ECO} prediction accuracy (Noormets et al., 2008). Underestimation of peak growing season R_{ECO} can also come from biases within LSTM training, which is strong in capturing long-term temporal trends but less effective in obtaining peak values, and thus lead to increasing prediction errors during growing season compared to other periods of time.~~



~~**Figure 6: R_{ECO} estimation with data from FLUXNET sites at CA-OAS and US-NR1. Panels (a) and (c) present daily estimation of R_{ECO} with red, green, and blue lines representing data used for training, validation, and prediction, and the black line shows the eddy covariance measurements. Pink points describe monthly mean difference between HPM estimation and measured data. Panels (b) and (d) show the scatter plots of daily (blue) and monthly (red) R_{ECO}. Darker blue clouds represent greater density of data points. Results for other sites are included in supplementary materials below (Figures A3 and A4).**~~**).**

Table 3: Statistical measures of HPM estimation of ET and ~~R_{ECO}~~R_{eco}

Site ID	Train MAE -ET [mm ~~/d~~ d^{-1}]	Test MAE - ET [mm ~~/d~~ d^{-1}]	Train Adj. R^2 - ET	Test Adj. R^2 - ET	Train MAE $-$~~R_{ECO}~~R_{eco} [$gCm^{-2}d^{-1}$]	Test MAE $-$~~R_{ECO}~~R_{eco} [$gCm^{-2}d^{-1}$]	Train Adj. R^2 $-$~~R_{ECO}~~R_{eco}	Test Adj. R^2 $-$~~R_{ECO}~~R_{eco}
US-NR1	0.19	0.11	0.95	0.98	0.33	0.18	0.91	0.98
CA-Oas	0.18	0.13	0.94	0.97	0.33	0.26	0.96	0.97
CA-Obs	0.12	0.09	0.95	0.96	0.29	0.25	0.96	0.97
US-SRM	0.22	0.17	0.92	0.94	0.24	0.19	0.80	0.87
US-Ton	0.22	0.17	0.92	0.94	0.43	0.36	0.76	0.82
US-Var	0.15	0.12	0.92	0.95	0.49	0.38	0.81	0.88

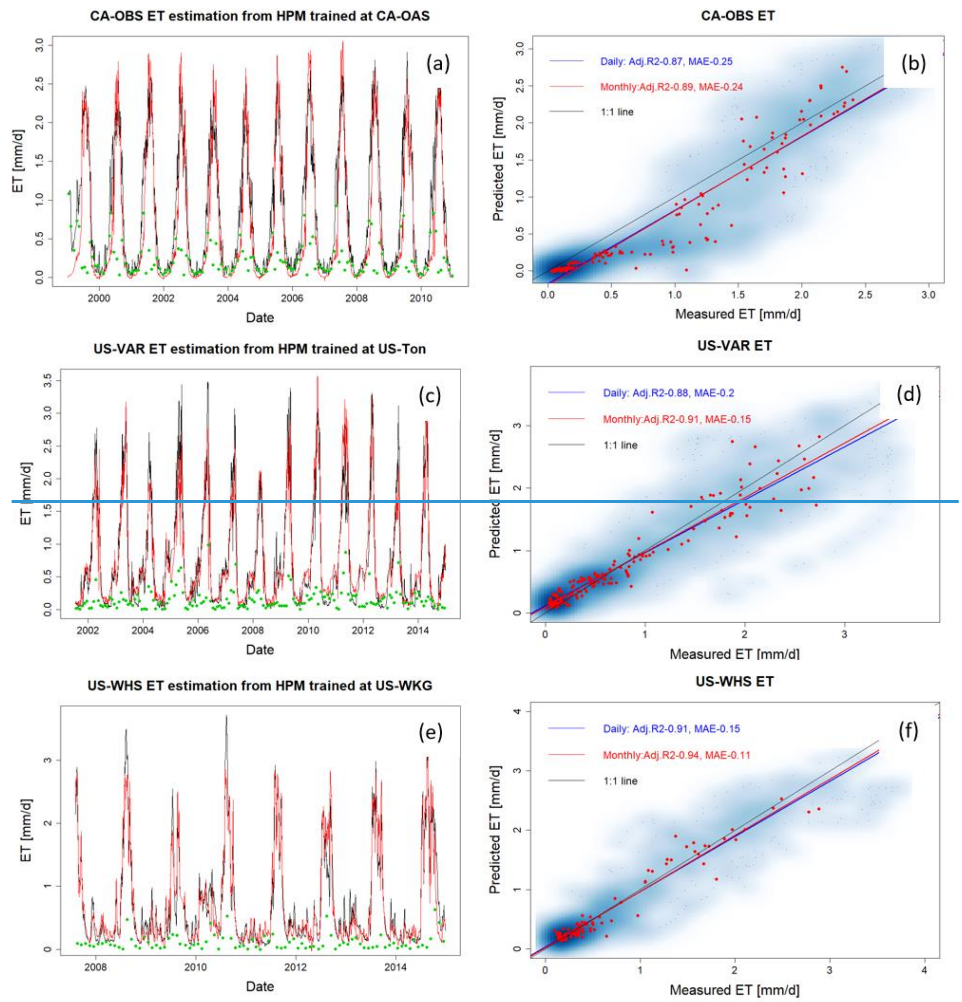
Formatted: Subscript

US-Whs	0.13	0.09	0.93	0.96	0.12	0.09	0.84	0.89
US-Wkg	0.19	0.15	0.87	0.91	0.18	0.15	0.85	0.91
US-Me2	0.36	0.43	0.81	0.75	0.75	0.83	0.88	0.85

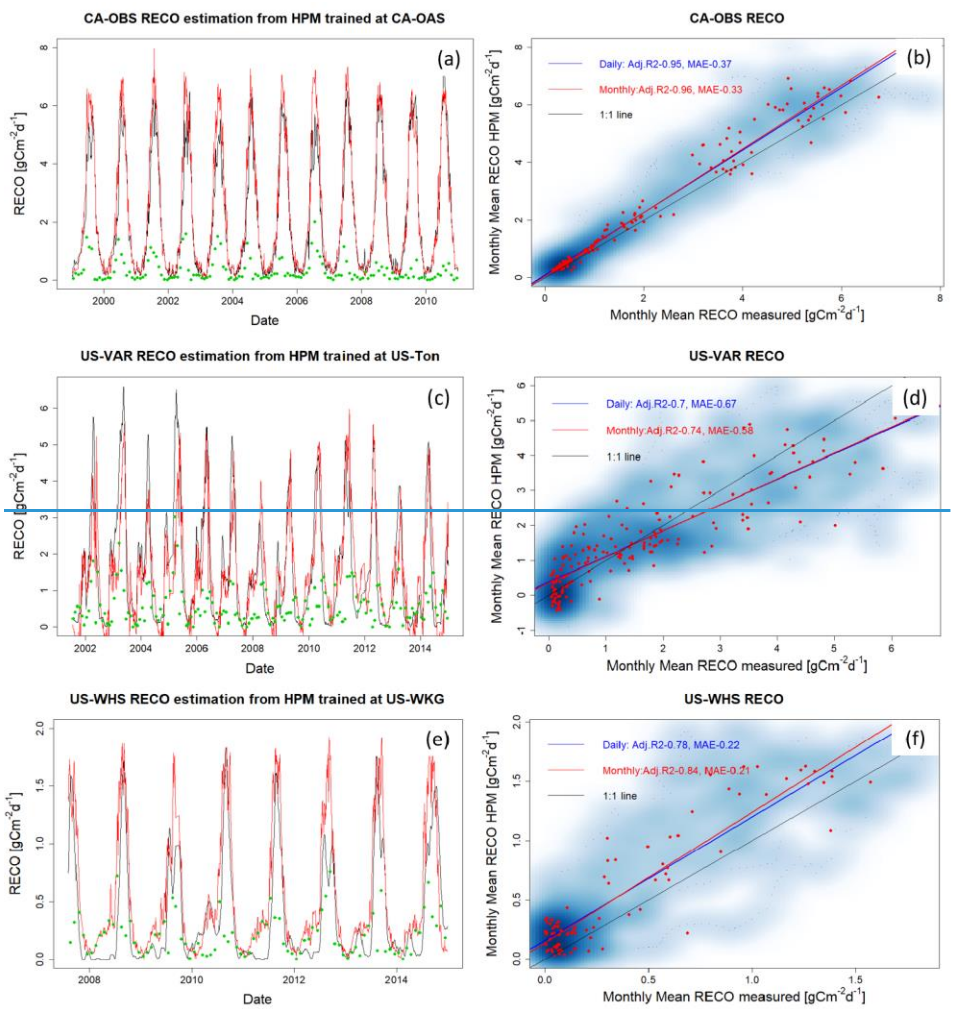
4.2 Use Case 2: Ecoregion-Based, Data-Driven HPM Model for ET and R_{ECO} Estimation

~~While the effort and cost involved in establishing flux towers naturally limit the spatial coverage of obtained measurements, point scale measurements from one FLUXNET station provides representative information about ecosystem dynamics at other locations within the same ecoregion.~~ In this section, we explored the use of a data-driven HPM trained with one FLUXNET station to estimate ET and ~~R_{ECO}~~R_{eco} at other locations within the same ecoregion. ~~To test this approach, we first trained HPM at a selected FLUXNET stations and validated these HPM models at other FLUXNET stations (ET and R_{ECO} data at testing sites were only used for comparison with HPM prediction) within the same ecoregion.~~ Specifically, we developed HPM models at US-Ton, CA-Oas and US-Wkg, and provided ET and ~~R_{ECO}~~R_{eco} estimations at US-Var, CA-Obs and US-Whs at three ecoregions, respectively.

Table 4 summarizes how we developed the data-driven HPM models for spatially distributed estimation of ET and ~~R_{ECO}~~R_{eco} as well as the corresponding statistical summaries. Figures ~~7~~6 and ~~8~~A5 present the time series of HPM-estimated ET and R_{ECO} compared to measurements from flux towers. HPM estimation at US-Obs, US-Whs and US-Var achieved an adjusted R^2 of 0.87, 0.88 and 0.91 for ET and 0.95, 0.70 and 0.78 for R_{ECO}, respectively. These results show that HPM captures the seasonal and long-term dynamics of ET and R_{ECO}. However, at sites that experience seasonally dry periods (e.g., US-Whs), prediction accuracy decreases during the peak growing season. ~~For example, we observed large errors in HPM-based estimations compared to measurements during peak growing seasons (e.g., a 0.5 mm discrepancy in June mean ET). We interpret this discrepancy as the result that current HPM models did not capture water stress conditions, and it is necessary to include other key attributes (e.g., soil moisture) to improve prediction accuracy, especially at these sites with seasonally dry periods.~~ Although the prediction accuracy is not as high as Use Case 1 (Section 4.1), this use case demonstrates that HPM can learn the complicated relationships between responses and features successfully, and that a local data-driven HPM can be used to fuse with data from other subsites for long-term estimation of ET and R_{ECO} within the same ecoregions.



~~**Figure 7. ET estimation at CA-Oas (a), US-Var (c), and US-Whs (e) with HPM trained at US-Ton, US-Wkg, and CA-Oas, respectively. Red and black lines represent HPM estimation and real measurements, with green points denoting the monthly mean difference between HPM estimationss and measurements. Panels (b), (d), and (f) show the scatter plots of daily (blue) and monthly (red) ET at these three sites. Darker blue clouds represent greater density of data points.**~~


CA-OBS RECO estimation from HPM trained at CA-OAS
(a)
RECO [gCm-2d-1]
Date
CA-OBS RECO
(b)
Daily: Adj.R2-0.95, MAE-0.37
Monthly:Adj.R2-0.96, MAE-0.33
1:1 line
Monthly Mean RECO HPM [gCm-2d-1]
Monthly Mean RECO measured [gCm-2d-1]
US-VAR RECO estimation from HPM trained at US-Ton
(c)
RECO [gCm-2d-1]
Date
US-VAR RECO
(d)
Daily: Adj.R2-0.7, MAE-0.67
Monthly:Adj.R2-0.74, MAE-0.58
1:1 line
Monthly Mean RECO HPM [gCm-2d-1]
Monthly Mean RECO measured [gCm-2d-1]
US-WHS RECO estimation from HPM trained at US-WKG
(e)
RECO [gCm-2d-1]
Date
US-WHS RECO
(f)
Daily: Adj.R2-0.78, MAE-0.22
Monthly:Adj.R2-0.84, MAE-0.21
1:1 line
Monthly Mean RECO HPM [gCm-2d-1]
Monthly Mean RECO measured [gCm-2d-1]

~~**Figure 8. R_{ECO} estimation at CA-Oas (a), US-Var (c), and US-Whs (e) with HPM trained at US-Ton, US-Wkg, and CA-Oas,**~~

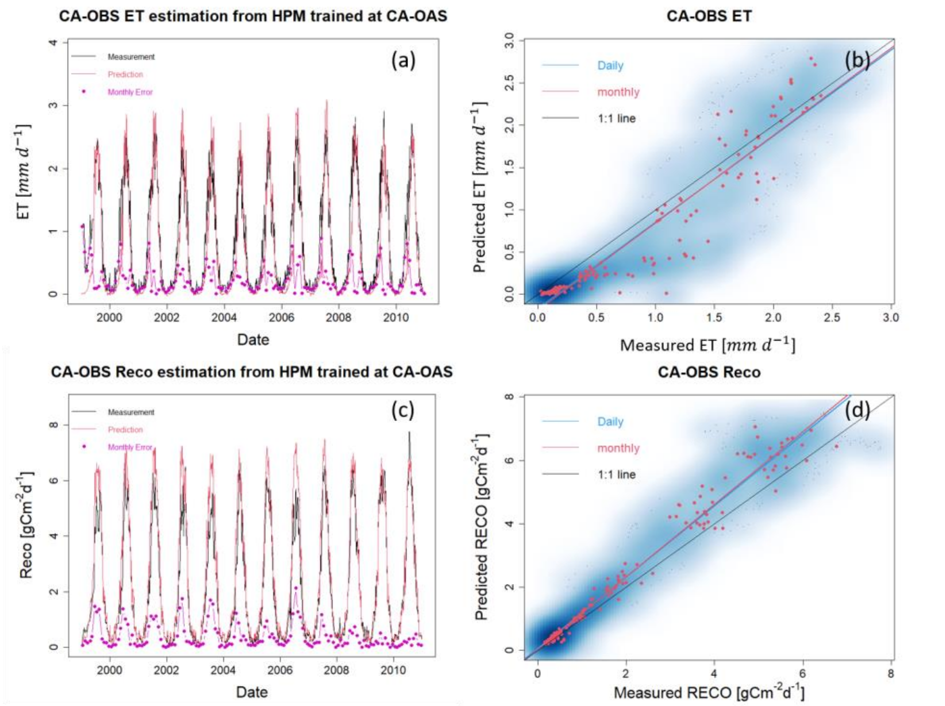


Figure 6. ET and R_{eco} estimation at CA-Obs using HPM trained at CA-Oas. Other sites are presented in Fig. A5.

~~, respectively. **Red and black lines represent HPM estimations and real measurements; green points denote the monthly mean difference between HPM estimation and measurements. Panels (b), (d), and (f) show the scatter plots of daily (blue) and monthly (red) R_{ECO} at these three sites. Darker blue clouds represent greater density of data points.**~~

Formatted: Font: 10 pt, Not Bold

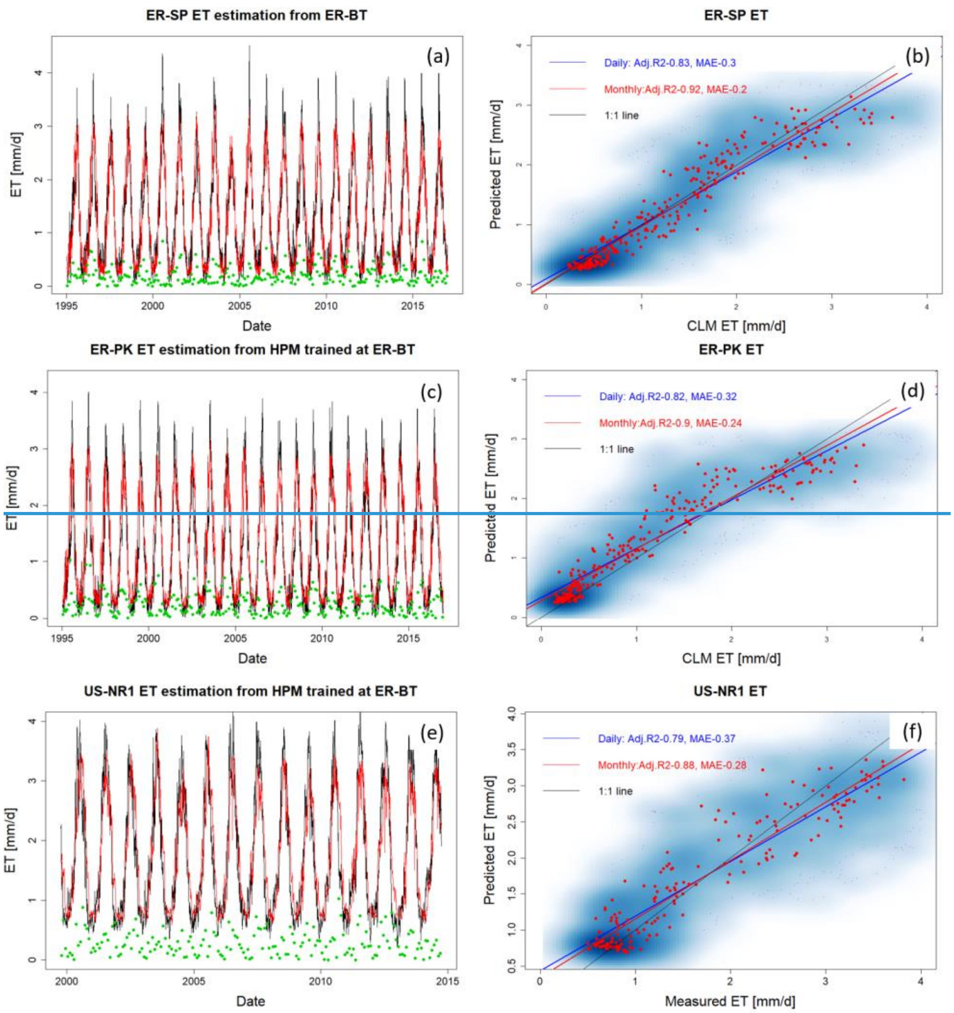
4.3 Use Case 3: Ecoregion-Based, Mechanistic HPM Estimation of ET

Mechanistic HPM, which is trained ~~with ET estimates from 1-D physically-based-model simulations, provides an avenue for estimating ET in ecoregions where direct measurements from eddy covariance tower are not available. In order to test the effectiveness of the mechanistic HPM, we focused on the three SNOTEL stations and US-NR1, which locates in the "Western Cordillera" ecoregion. Mechanistic HPM is coupled with CLM simulations at these sites (Tran et al., 2019). To ensure the CLM physically-based-model simulations can provide alternative datasets to develop mechanistic HPMs, we compared CLM estimation and direct measurements of ET at US-NR1 (Figure S2). The consistent results between measured ET and CLM-estimated ET (adjusted $R^2 = 0.88$; $k = 0.95$) indicate independent CLM simulations can be effectively used to develop the mechanistic HPM.~~

~~We applied mechanistic HPM trained with 1-D~~upon physically-based model simulations, provides an avenue for estimating fluxes in ecoregions where flux towers are not available. Consistent results between measured ET and CLM-estimated ET at US-NR1 (adjusted $R^2 = 0.88$; $k = 0.95$, Fig. S1) indicate independent CLM simulations can be effectively used to develop the mechanistic HPM. We applied mechanistic HPM trained with 1-Dimensional (vertical) CLM developed at ER-BT (Tran et al., 2019) to estimate ET at sites classified as part of the ~~same~~western Cordillera ecoregion (i.e., ER-SP, ER-PK and US-NR1). We then compared ET estimation from HPM to independent CLM-based ET estimations at ER-SP and ER-PK~~ and to direct measurements at US-NR1.~~. Figure 97 shows a high consistency between HPM estimation and the validation data. For all scenarios, an adjusted R^2 of 0.8 or greater is observed (Table 4), which strongly indicates that mechanistic HPM can provide accurate ET estimation at sites of similar ecoregions. These results suggest the broad applicability of mechanistic HPM to estimate ET based on ecoregion characteristics. This approach is expected to be particularly useful for regions where flux towers are difficult to install or where measured fluxes are not representative of the landscape, such as in mountainous watersheds.

Table 4. Statistical summary of HPM estimation over space with FLUXNET sites and SNOTEL stations with CLM

Target Site	Training Site	Level II Ecoregion	ET MSE (monthly)[mm~~/d]~~ d^{-1}]	ET Adj. R^2	~~R_{ECO}~~R_{eco} MSE(monthly)[$gCm^{-2}d^{-1}$]	~~R_{ECO}~~R_{eco} Adj. R^2
CA-Obs	CA-Oas	Boreal Plain	0.39	0.88	0.36	0.97
US-Var	US-Ton	Mediterrean California	0.34	0.70	0.67	0.70
US-Whs	US-Wkg	Western Serra Madre Pidemont	0.13	0.94	0.17	0.85
ER-SP	ER-BT	Western Cordillera	0.20	0.92	-	-
ER-PK	ER-BT	Western Cordillera	0.24	0.90	-	-
~~US-NR1~~	~~ER-BT~~	~~Western Cordillera~~	~~0.23~~	~~0.90~~		


ER-SP ET estimation from ER-BT
(a)
ET [mm/d]
Date
ER-SP ET
(b)
Daily: Adj.R2-0.83, MAE-0.3
Monthly:Adj.R2-0.92, MAE-0.2
1:1 line
Predicted ET [mm/d]
CLM ET [mm/d]
ER-PK ET estimation from HPM trained at ER-BT
(c)
ET [mm/d]
Date
ER-PK ET
(d)
Daily: Adj.R2-0.82, MAE-0.32
Monthly:Adj.R2-0.9, MAE-0.24
1:1 line
Predicted ET [mm/d]
CLM ET [mm/d]
US-NR1 ET estimation from HPM trained at ER-BT
(e)
ET [mm/d]
Date
US-NR1 ET
(f)
Daily: Adj.R2-0.79, MAE-0.37
Monthly:Adj.R2-0.88, MAE-0.28
1:1 line
Predicted ET [mm/d]
Measured ET [mm/d]

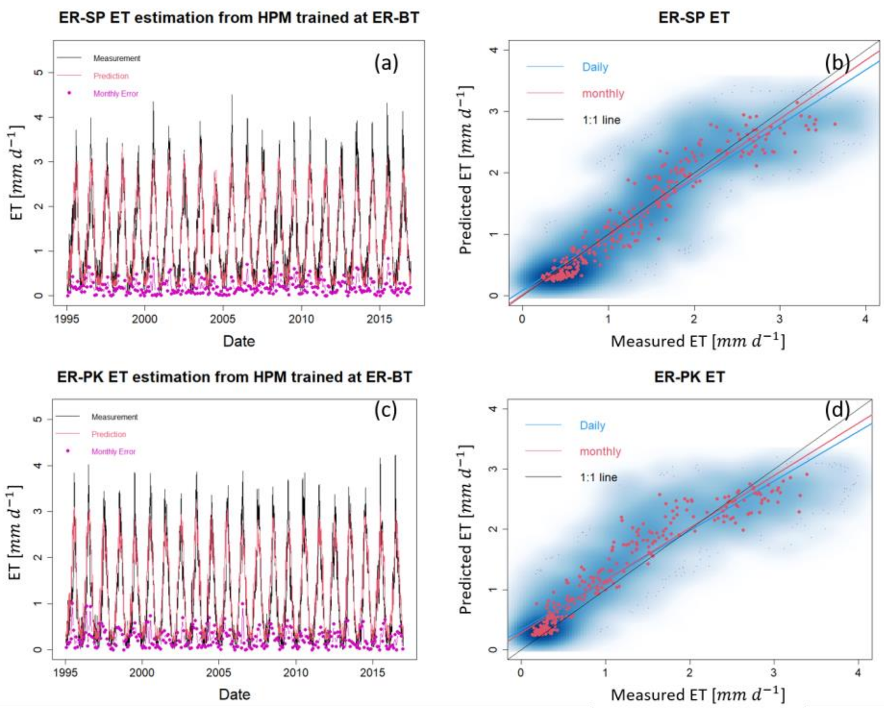


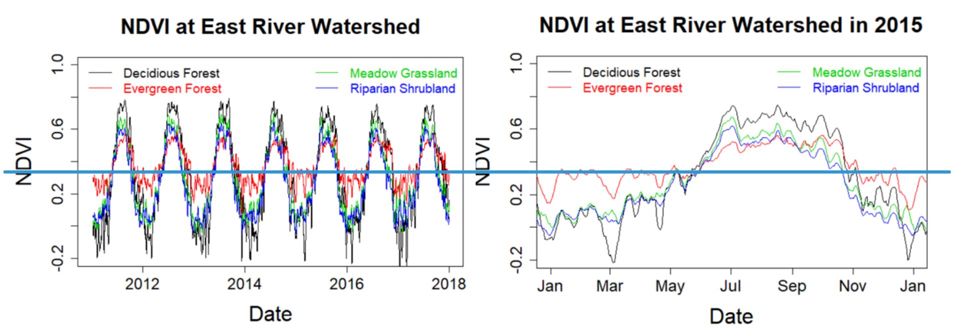
Figure ~~9~~7. HPMs trained with CLM simulation at ER-BT are used to estimate ET at ER-SP~~, ER-PK,~~ and ~~US-NR1~~ER-PK. Panels (a~~), (c),~~) and (~~e~~c) display the time series of HPM estimation of ET (red lines), ~~as well as~~and independent CLM estimation at ER-SP~~,~~ and ER-PK~~, and eddy covariance measurements at US-NR1 (black lines).~~. Panels (b~~),~~) and (d~~), and (f~~) show the scatter plots of daily (blue) and monthly (red) ET at these ~~three~~ sites. Darker blue clouds represent greater density of data points.

4.4 ~~Exploration of How ET and R_{ECO} Varies with Meteorological forcings~~Use Case 4: HPM approach improved our prediction capability and ~~Vegetation Heterogeneity~~process understanding at the East River Watershed

~~ET and R_{ECO} estimated from the HPM model at the mountainous East River Watershed in CO enabled us to analyze how vegetation heterogeneity and meteorological forcings heterogeneity influence estimated ET and R_{ECO} dynamics, and to identify limitations in the developed approach for estimating ET and R_{ECO} across mountainous and heterogeneous watersheds.~~

~~NDVI time-series data provide high-resolution (30m scale) information about vegetation variability across the East River Watershed. The spatial distribution of vegetation cover presented in Figure 2 (from Falco et al. 2019) enables us to distinguish different patches of deciduous forests, evergreen forests, meadow grassland and riparian shrublands and retrieve corresponding NDVI time series. NDVI time series is related with snowmelt processes, whereas earlier snowmelt triggers earlier vegetation growth and result in earlier rise NDVI values (Pedersen et al., 2018). Figure 10 shows Landsat-derived and reconstructed NDVI values for the four different vegetation types~~

~~within the East River Watershed. March, April and May mean NDVI values in 2012 for site DF1 are 0.07, 0.22 and 0.37 respectively compared to 0.06, 0.15 and 0.33 in 2015. The early rise of NDVI values observed in April 2012 is consistent with the fact that snowmelt occurred much earlier in 2012 than in 2015, as recorded by the SNOTEL Butte station. Earlier increase of NDVI in earlier snowmelt year (2012) was also observed for other vegetation types. In addition, evergreen forests have an extended growing season compared to the other vegetation types. For example, March mean NDVI for EF1, RS1 and MS1 in 2012 are 0.30, 0.13, 0.11 compared to 0.28, 0.11, 0.08 in 2015, respectively whereas May mean NDVI for EF1, RS1 and MS1 in 2012 are 0.38, 0.33, 0.35 compared to 0.34, 0.29 and 0.31 in 2015, respectively. Though earlier snowmelt triggers earlier increase in vegetation growth, significant faster greenness was observed for deciduous forests, meadow grasslands and shrublands compared to evergreen forests, where NDVI increased by 0.08, 0.20, 0.24 and 0.30 for evergreen forests, shrublands, grasslands and deciduous forests in 2012, respectively. In addition, peak NDVI is generally smaller in evergreen forests compared to deciduous forests, meadow grasslands and riparian shrublands. NDVI ranges from 0.2 to 0.6 for evergreen forests, whereas larger fluctuations in NDVI are observed for deciduous forests, shrublands and grasslands. The NDVI values during the winter are likely sensing both snow and forest density, due to pixel spatial averaging from Landsat images. Similar to Qiao et al. (2016), we also found that the NDVI of deciduous forests exhibits a significant increase during the growing season, followed by a sharp decline (likely caused by defoliation), and that evergreen forests had a more stable NDVI. Similar sharp decreases in the NDVI of riparian shrublands and meadow grasslands are observed.~~



~~**Figure 10**~~With the proposed HPM approach (e.g., mechanistic HPM), we were able to estimate ET and R_{eco} at selected locations at the East River Watershed, CO, USA with only meteorological forcings and remote sensing data. Our estimations are comparable to other independent studies, such as Mu et al. (2013) (Fig. S2) and Berryman et al. (2018). HPM estimations enhanced our understanding of watershed processes and enabled us to explore the limitations in the developed HPM approach especially at mountainous watersheds.

Physiology differences among vegetation types and dynamic changes in meteorological conditions were well captured by input features and HPM at the East River Watershed. Not surprisingly, the reconstructed NDVI indicated that deciduous forests have the highest peak NDVI followed by grasslands, shrublands and evergreen forests whereas annual variation of NDVI in evergreen forests is smaller than the other vegetation types (Fig. 8).

Year 2012 is regarded as a fore-summer drought year with earlier than normal snowmelt, and year 2015 is regarded as a normal water year. The Palmer drought severity index (PDSI) is -5.2 and -1.5 for June and -4.6 and 1.1 for August in 2012 and 2015, respectively. Dynamic changes in meteorological conditions between 2012 and 2015 were also reflected in the reconstructed NDVI time series. We observed an earlier rise of NDVI in 2012: March, April and May mean NDVI values for deciduous forest sites are 0.07, 0.2 and 0.37 compared to 0.06, 0.15 and 0.33 in 2015. Similar trends were observed for other vegetation types during spring months as well. NDVI values remain high during the peak growing season (deciduous forest > grassland > shrubland > evergreen forest) for both 2012 and 2015. However, we observed NDVI declines for grasslands and shrublands since August in 2012 but not until September in 2015. During autumn periods, NDVI declines significantly following the sharp decline in radiation.

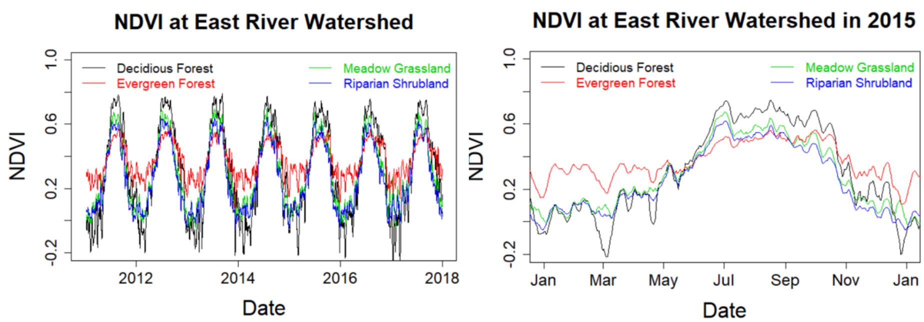
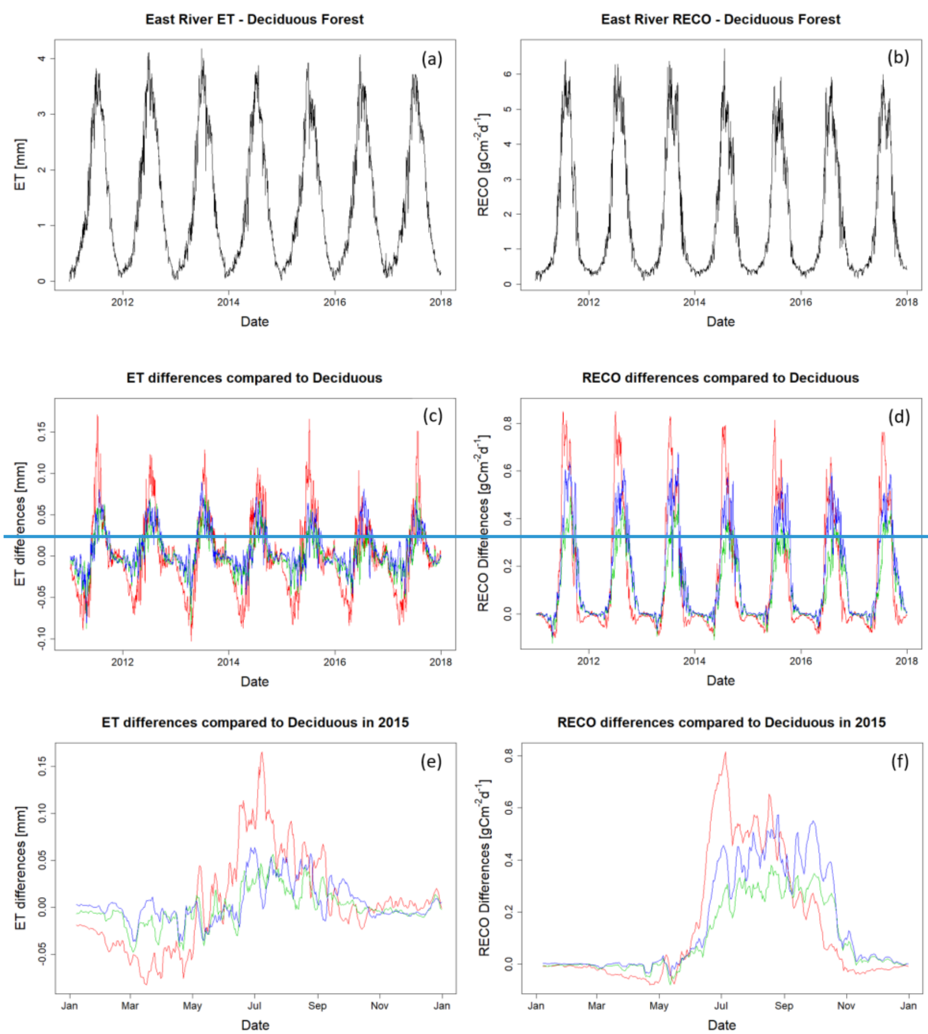


Figure 8: Reconstructed NDVI time series at selected locations in the East River Watershed for 2011 to 2018 (panel a) and for 2015 (panel b, normal water year). Black, red, green, and blue lines represent the time series of NDVI for deciduous forests, meadow grasslands, evergreen forests and riparian shrubland, respectively.

HPM-estimated ET and ~~R_{ECO}~~R_{eco} also show different dynamics with different vegetation types ~~as a result of differences in snowmelt timing,~~and meteorological ~~forcing and vegetation heterogeneity.~~conditions. Figure ~~11a~~9a and ~~11b~~9b present the time series of estimated ET and ~~R_{ECO}~~R_{eco} associated with deciduous forests, respectively. Figure ~~11c~~9c and ~~d~~9d present the ET and ~~R_{ECO}~~R_{eco} differences between deciduous forests sites and evergreen forests, shrublands and grasslands. Before peak growing season, evergreen forests have the greatest ET and ~~R_{ECO}~~R_{eco} compared to the other vegetation types. ET of evergreen forests is about 10% greater than deciduous forests, whereas ET of deciduous forests during peak growing season is greater than evergreen forests, shrublands and meadows. After growing season, the NDVI of deciduous forests is less than 0.2 (loss of leaves) compared to the NDVI of evergreen forests. Before peak growing season, ~~R_{ECO}~~R_{eco} of evergreen forests is slightly greater than deciduous forests, meadow grasslands and shrublands. During peak growing season, we observed largest ~~R_{ECO}~~R_{eco} for deciduous forests sites (~ 6 $gCm^{-2}d^{-1}$) followed by meadows, shrublands and evergreen forests. ~~R_{ECO}~~R_{eco} of deciduous forests is around 17 % greater than ~~R_{ECO}~~ R_{eco} of evergreen forests. However, we did not observe significant differences in annual ET among these four vegetation types (e.g., ~~DF1~~DF: 535 to 573 mm, ~~MS1~~MS: 534 to 570 mm, ~~RS1~~RS: 532 to 567 mm and ~~EF1~~EF: 532 to 569 mm across 7 years in this study). Total annual ~~R_{ECO}~~R_{eco} of deciduous forests is greater than the other vegetation types (DF1: 642 to 698 gCm^{-2}, MS1: 588 to 636 gCm^{-2}, RS1: 589 to 636 gCm^{-2} and EF1: 592 to 639gCm^{-2}). These results indicate HPM R_{eco} models are sensitive to vegetation types and HPM ET models are mostly constrained by meteorological conditions.

Considering the inter-annual variability in meteorological forcings, we further selected year 2014 (large snow precipitation ~ 587 mm but small rain precipitation ~ 275 mm) in addition to 2012 (drought year) and 2015 (small snow precipitation ~ 383 mm and large rain precipitation ~ 477 mm) to test HPM performance. As HPM does not have the capability to identify snow and monsoon precipitation's contribution to fluxes, we separated annual ET and R_{eco} into pre-June (January-June) and post-July (July-December) to quantify the contribution from snow and monsoon. Earlier snowmelt that occurred in 2012 boosted spring ET and R_{eco} and we observed larger March-mean ET and R_{eco} compared to 2014 and 2015 that are characterized by later snowmelt. Occurrences of fore-summer drought in 2012 led to moisture limiting conditions, resulting in large fluctuations of ET and R_{eco} during May and June. ET fluctuated from 2.9 to 1.9 $mm\ d^{-1}$ during late May, and 3.53 to 2.6 $mm\ d^{-1}$ during early June. However, early occurrence of monsoon in 2012 led to a peak ET in early July. Due to late snowmelt, ET did not significantly fluctuate in 2014 and 2015. However, peak ET shifted towards late July in 2014. Regarding R_{eco} dynamics, fore-summer drought conditions led to variations in R_{eco} from ~ 4 to 6 $gCm^{-2}\ d^{-1}$ in 2012. In 2014, we observed more steady increase of R_{eco} during the early and peak growing seasons. For late-summer and autumn months (August – October), ET decreased steadily in all three years regardless of monsoon precipitation inputs, following the significant decline in radiation. Pre-June ET and R_{eco} (255mm and 217 $gCm^{-2}\ d^{-1}$) were both greater in 2012 compared to 2014 (223 mm and 178 $gCm^{-2}\ d^{-1}$) and 2015 (230 mm and 197 $gCm^{-2}\ d^{-1}$) in deciduous forests. While there were no significant differences in post-July ET among the three years (318, 316 and 306 mm), 2012 was the highest. Within deciduous forests and annually over 2012, 2014 and 2015, ET was 573 mm, 539 mm and 536 mm and R_{eco} was 698 gCm^{-2}, 642 gCm^{-2} and 652 gCm^{-2}, respectively. ~~Considering the inter-annual variability in snow dynamics, we observed annual ET at 569~~ ~~mm and 532 mm and annual R_{ECO} at 639 gCm^{-2} and 602 gCm^{-2} at EF1 for 2012 and 2015, respectively. We observed an earlier increase in ET and R_{ECO} in 2012 with March-mean ET and R_{ECO} at 0.69 mm/day and 0.51 $gCm^{-2}d^{-1}$ compared to 0.60 mm/day and 0.47 $gCm^{-2}d^{-1}$ in 2015. During peak growing season, we observed July-mean ET at 3.43 and 3.33 mm/day and R_{ECO} at 4.73 and 4.47 0.47 $gCm^{-2}d^{-1}$ for 2012 and 2015, respectively. Though earlier snowmelt usually triggers summer drought conditions, we observed a significantly greater amount of monsoon precipitation in 2012 (3.06mmd^{-1}) compared to 2015 (1.87mmd^{-1}). Water stress situation caused by earlier snowmelt was largely compensated by earlier monsoon in 2012, and thus we observed higher March, July and annual ET and R_{ECO} compared to 2015. Similar trends have also been observed for deciduous forests, shrublands and meadows in 2012 and 2015.~~

Formatted: Font: 10 pt, Not Bold



. Similar trends were observed for other vegetation types.

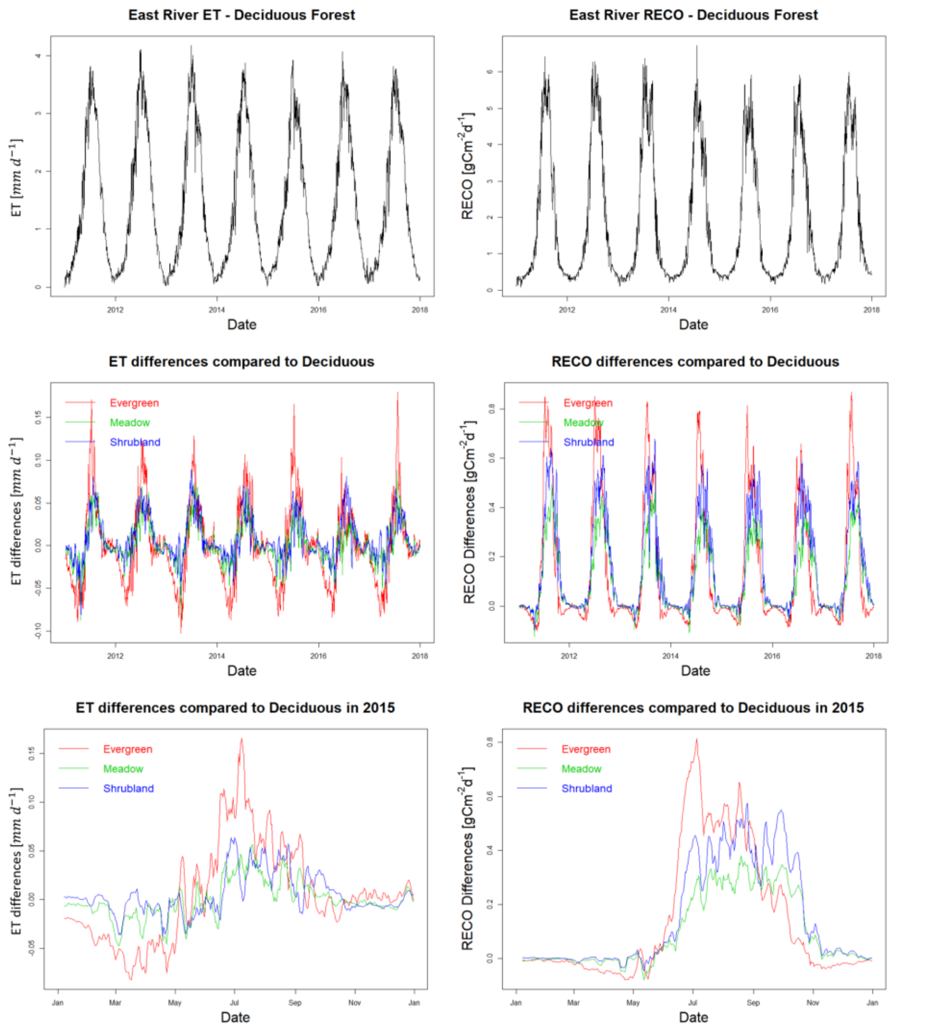
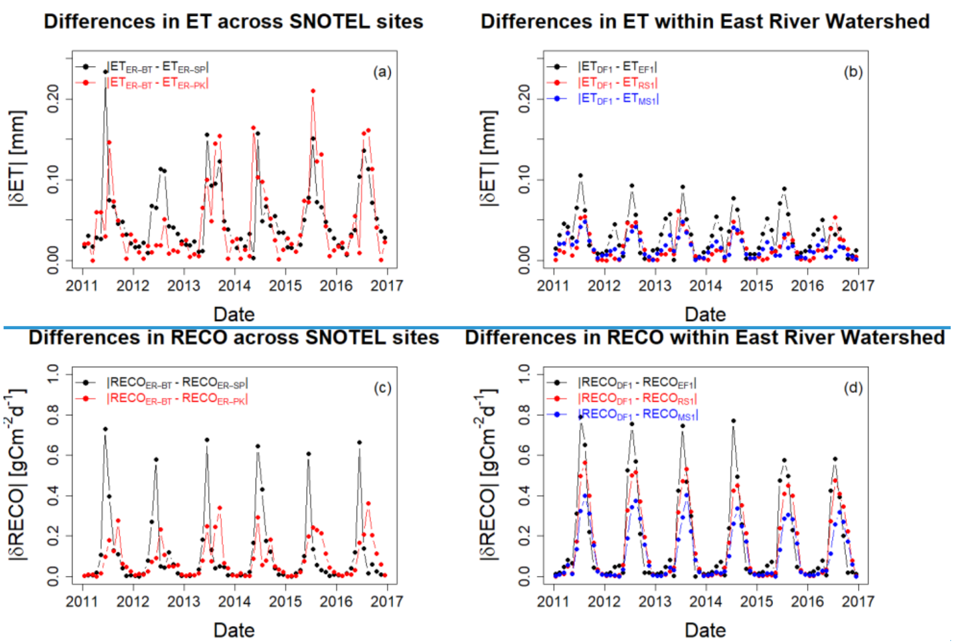


Figure ~~11~~9: ET (a) and ~~R_{ECO}~~R_{eco} (b) estimation for the deciduous forest site DF1 at the East River Watershed. Panels (c) and (d) show the differences in ET and ~~R_{ECO}~~R_{eco} among various vegetation types and ~~DF1.~~deciduous forest. Red, green, and blue lines represent the differences in evergreen forest, meadow, and riparian shrubland compared to ~~DF1.~~deciduous forest. Panels (e) and (f) zoom into 2015 to better display seasonal variations.

~~ET and R_{ECO} estimation at the East River Watershed from the HPM model further enabled~~Though HPM estimations allowed us to ~~assess the role of input attributes and~~explore differences in ET and R_{eco} across vegetation types and meteorological forcings heterogeneity, it is necessary to investigate the limitations of ~~the~~ HPM approach. Figure ~~12~~10 shows the absolute value of monthly mean difference in ET (Fig. ~~12a~~10a and Fig. ~~12b~~10b) and

~~R_{ECO}~~R_{eco} (Fig. ~~12c~~10c and Fig. ~~12d~~10d) across SNOTEL stations (ER-BT, ER-SP and ER-PK) and within selected East River locations. ~~Landsat data enabled us to capture NDVI differences at these sites (Figure 10), but we have identified the insufficient resolution of input meteorological forcing data at the East River sites. We observed a~~ We observed greater differences in air temperature and radiation at the SNOTEL sites ~~whereas there's~~and very small differences at the East River sites (Figure ~~S3). Summer~~S4). June air temperature differences among SNOTEL sites ~~can be~~were occasionally over 3 ℃ ~~but there's a barely 0.2 ℃ differences in~~, while the DAYMET data ~~used for~~from the East River ~~sites~~rarely revealed 0.2 ℃ differences. In addition, a ~80 W ~~$/m^2$~~ m^{-2} of radiation differences ~~is~~was observed with SNOTEL data whereas radiation differences stays around 30 W ~~$/m^2$~~ m^{-2} for East River sites. Correspondingly, we observed 2.5 times greater differences in ET across SNOTEL stations compared to the sites within the East River watershed. We observed similar level of differences (around 0.8 gCm^{-2}) in ~~R_{ECO}~~R_{eco} within East River Watershed and across SNOTEL stations. Landsat data enabled us to capture NDVI differences at these sites, but we have identified the insufficient resolution of input meteorological forcing data at the East River sites. These results indicate uncertainties in meteorological forcing attributes (e.g., radiation and air temperature) can have a huge influence over HPM ET estimation and HPM R_{eco} model is more sensitive to ~~temperature and radiation inputs whereas NDVI, temperature and radiation are all influential for HPM R_{ECO} models. Differences in ET and R_{ECO} among SNOTEL sites and East River sites are resulted from the differences in input meteorological forcing data~~NDVI datasets. If high resolution meteorological data becomes available for the East River watershed, we believe the HPM approach can better capture heterogeneities in ET and ~~R_{ECO}~~R_{eco} at the East River watershed and better distinguish the roles of meteorological forcing and vegetation heterogeneity on ET and ~~R_{ECO}~~R_{eco} distribution.



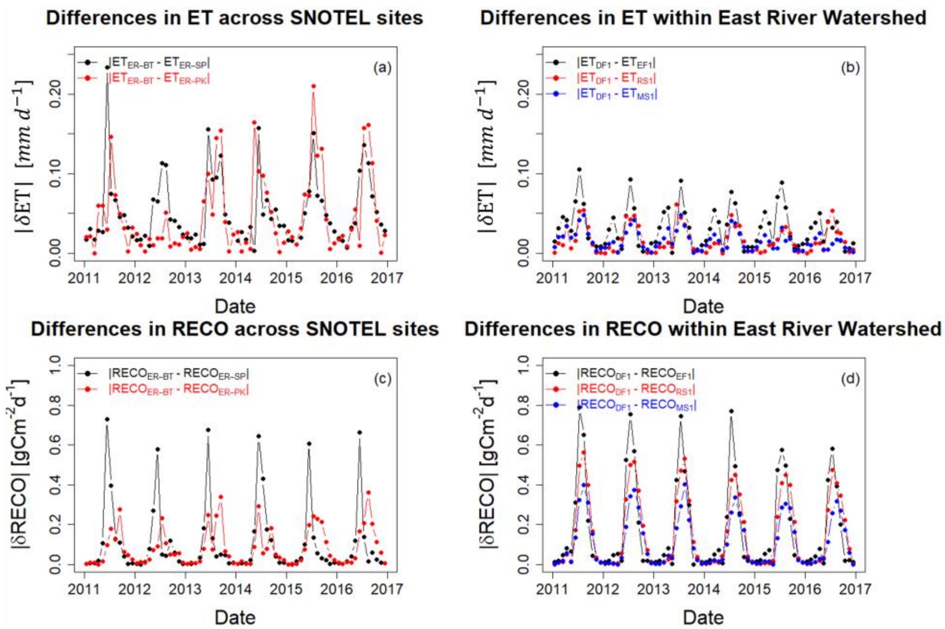


Figure ~~12~~10. Absolute differences in monthly mean ET and ~~R_{ECO}~~R_{eco} across SNOTEL stations and within East River Watershed. Panels (a) and (c) describe the absolute differences in monthly mean ET and ~~R_{ECO}~~R_{eco} between ER-BT, ER-SP, and ER-PK. Panels (b) and (d) describe the absolute differences in monthly mean ET and ~~R_{ECO}~~R_{eco} within East River Watershed between deciduous ~~forest (DF1),~~forests, evergreen ~~forest (EF1),~~forests, meadow ~~(MS1),~~grasslands, and riparian ~~shrubland (RS1).~~shrublands.

5. Discussion

Our study demonstrates that HPM provides reliable estimations of ET and ~~R_{ECO}~~R_{eco} under various climate and vegetation conditions~~, including data-based HPMs that are trained with FLUXNET data as well as physical-model-based HPMs that are coupled with simulations results~~. The unique gated structures and cell states of LSTM allow HPM to track information from ~~mechanistic models (i.e., CLM in this study). With 70~~earlier times and decide which information to pass along and which information to forget. This effective configuration allows LSTM to effectively capture the long-term dependencies and ecological memory effects among meteorological forcings, NDVI, ET and R_{eco}. With 70 % of the data used for training (model development), ET and ~~R_{ECO}~~R_{eco} estimation from HPM achieves an average adjusted R^2 of 0.9 compared to ~~eddy covariance~~flux tower measurements. ~~With this high estimation accuracy, we demonstrated that this approach could be used~~To demonstrate HPM's applicability for ~~predicting ET and R_{ECO} over time. HPM is capable of "learning" the complex interactions among meteorological forcings, vegetation dynamics, and water and carbon fluxes. The underlying relationships acquired by HPM can serve as a local ecohydrological model for long-term monitoring of ET and R_{ECO} with the aid of remote sensing data, and can fill in gap data during occasional equipment failure. HPM was also successful at estimating the spatial~~

~~distribution of~~providing ET and ~~R_{ECO} through exploiting an ecoregion concept. Using the representative FLUXNET sites in different ecoregions, HPM provided estimates of~~R_{eco} estimation at sparsely monitored watersheds, we presented four use cases, including prediction ET and ~~R_{ECO} at locations using learned relationships~~R_{eco} in the time domain, data-driven HPMs and mechanistic HPMs. Results from ~~other sites having the same ecoregion classification. For conditions where no FLUXNET sites are within the same ecoregion, our study showed that physically-based models that utilize weather forcings data can provide alternatives for developing mechanistic HPM to~~ the four use cases suggest HPM is a powerful approach to estimate ET and ~~R_{ECO}.~~

~~With the proposed HPM approach, we investigated the variability in ET and R_{ECO} estimations across different proportions of the East River Watersheds. While we currently do not have continuous measurements of ET~~ R_{eco} at target watersheds requiring only 5 commonly available input data and ~~R_{ECO} at the East River Watershed for validation, our results are comparable to other studies that focus on sites within the same ecoregion. HPM-based ET estimation at East River Watershed is comparable to Mu et al. (2013), where ET is computed based upon the logic of the Penman-Monteith equation and MODIS remote sensing data (Figure S1), and the HPM-based R_{ECO} estimation is comparable to what Berryman et al. (2018) discovered, with growing season R_{ECO} ranging between 555 to 607 gCm^{-2} and mean growing season R_{ECO} ranging between 3.01 to 3.30 gCm^{-2}. Annual ET between deciduous forests and evergreen forests are not statistically different, which is similar to Mu et al. (2013). Annual R_{ECO} differences between evergreen forests and deciduous forests are around 50 gCm^{-2}, which is comparable to Berryman et al. 2018.~~can advance our understanding of watershed processes.

~~We confirmed the important role of vegetation heterogeneity in modeling ET and R_{ECO} dynamics, which further enabled us to better understand ecosystem dynamics at the East River Watershed. As indicated~~ HPM was capable incorporating information from NDVI time series ~~(Fig 10),~~ to delineate the physiological differences among deciduous forests, evergreen forests, shrublands and grasslands. In our study, NDVI data indicated evergreen forests have a longer growing season compared to other vegetation types~~; however,~~ and deciduous forests have higher peak NDVI values. Correspondingly, we also observed an earlier increase in ET and ~~R_{ECO}~~R_{eco} for evergreen forests (before May), but larger ET and ~~R_{ECO}~~R_{eco} for deciduous forests during peak growing season (around June and July). ~~Similar dynamics were also observed at regions that have different climate conditions. Through assessing the differential mechanisms of deciduous forests and evergreen forests at various sites under Mediterranean climates,~~ Baldocchi et al. (2010) found that deciduous forests had a shorter growing season, but showed a greater capacity for assimilating carbon during the growing season. Evergreen forests, on the other hand, had an extended growing season but with a smaller capacity for gaining carbon. ~~These results were identified through analyzing the relationships among leaf ages, leaf nitrogen level, leaf area, and water use efficiencies of these tree species at the selected Mediterranean sites.~~ They found older leaves tend to have smaller leaf nitrogen and stomata conductance that lead to smaller ET and ~~R_{ECO}~~R_{eco} during peak growing seasons. ~~Though our approach were not able to quantify the physiology differences among vegetation types, HPM estimation indicated evergreen forests that maintain leaves throughout the year have smaller ET and R_{ECO} during peak growing season compared to other vegetation types.~~

~~Dynamic changes in the inter-annual variability of meteorological conditions result in varying growing season length and spatiotemporal variability in ET and R_{ECO}. Earlier snowmelt triggers earlier growth of vegetation,~~

~~causing earlier rise in ET and R_{ECO}. However, earlier growth in vegetation and increasing demand for water results in drought conditions (Sloat et al., 2015; Wainwright et al., 2020) that decrease ET and R_{ECO}. Timing and amount of monsoon precipitation are also important monsoons can relieve water stress and lead to increases in ET and R_{ECO}. Combination of these events jointly determine the magnitude of annual ET and R_{ECO}.~~ Hu et al. (2010) ~~analyzed flux data at US-NR1 to determine the relationships between growing season lengths and carbon sequestration, and found that extended growing season length resulted in less annual CO_2 uptake. They found that the duration of growing seasons substantially decreases snow water storage, which significantly decreases forest carbon uptake. Wieder et al. (2017) used point-scale CLM to better understand how complex terrain controls landscape-level variation of water, carbon and energy fluxes in the Niwot Ridge mountain ecosystems. With synthetic scenarios (e.g., different snow accumulation dynamics, fluctuations in air temperature), their simulation indicated earlier snowmelt and warmer summertime temperatures might drive divergent plant responses across the landscape. In our study, the combination of early snowmelt and early vegetation growth resulted in higher March ET and R_{ECO} in 2012 compared to 2015. The earlier start of growing season led to occurrences of fore-summer drought that decreases ET and R_{ECO}. However, the substantial earlier monsoon precipitation in 2012 relieved subsurface water stress whereas we observed higher July ET and R_{ECO} compared to other years. In addition, we observed smaller annual ET and R_{ECO} for evergreen forests that have longer growing season compared to other vegetation types. These results suggested HPM is capable of translating these variabilities in meteorological forcing and vegetation variables to ET and R_{ECO} dynamics.~~ found that extended growing season length resulted in less annual CO_2 uptake at Niwot Ridge, USA. They found increasing growing season length is usually correlated with decreasing snow water storage and decreasing forest carbon uptake. Xu et al. (2020) suggested canopy photosynthetic capacity is the driving force that lead to different resources use efficiencies (RUEs) between deciduous forests and evergreen forests. Novick et al. (2015) focused on the net ecosystem exchange of CO_2 and also suggested seasonality is less important for evergreen forests, where significant amounts of carbon were assimilated outside of active season. These findings are similar to what we found in HPM estimations, where we observed a greater ET and R_{eco} contribution during early and later seasons for evergreen forests compared to deciduous forests that have significantly greater peak ET and R_{eco} during peak growing season. As HPM only requires 5 input features and NDVI is the only variable related with vegetation types, we were not able to perform detailed analysis delinearing the physiological control on ET and R_{eco} dynamics. But we believe HPM models are still useful as they can be provide initial ET and R_{eco} estimation that help with site selection and field campaign designs.

~~Through comparing the HPM estimation results at different ecoregions, we also identified and assessed the limitations of current selection of input parameters. In the current study, we only used meteorological forcing and remote sensing based variables as inputs for HPM models, because these data are generally acquirable from weather reanalysis datasets and remote sensing products. HPM models with these variables provided reasonable estimates of ET and R_{ECO} for ecoregions limited by energy conditions, however we observed a decreasing prediction accuracy for ecoregions that experience seasonally dry periods. For example, HPM estimates at US-NR1 and CA-OAS achieved very high R^2 and small MAE; but prediction accuracy decreases especially during peak growing season at US-Ton and other water-limiting sites. These results indicate other key variables are necessary in order to capture~~

~~dynamics during the seasonally dry periods, such as soil moisture measurement. The current HPM models did not use soil moisture as an input variable due to data availability reasons, but we believe and recommend adding soil moisture as well as other key variables to HPMs to further improve model performance at these seasaonlly dry ecoregions when such data becomes available.~~

~~Parameterization and spatiotemporal resolution of meteorological forcing data still remain a challenge for improving ET and R_{ECO} estimation at sparsely monitred watersheds. Microcliamte and heterogeneities in meteorological forcing attributes control the mangnitude and timing of ET and R_{ECO} dynamics. Other field~~Temporal variability in meteorological conditions also leads to unique ET and R_{eco} responses at the East River Watershed, as shown by HPM estimations. Three years with a diverse combination of snow and rain precipitation were analyzed. In 2012, a year that experienced earlier snowmelt, both ET and R_{eco} increased early in the season. However, earlier growth in vegetation and increasing demand for water resulted in fore-summer drought conditions that led to decreases in ET and R_{eco} in late May and June. In 2014, HPM estimated a steady increase in ET and R_{eco} during spring months following radiation and air temperature trends, with no subsequent significant decline in ET and R_{eco}. This indicates that energy was still the key limiting factor for spring dynamics in 2014, leading to a smaller pre-June ET and R_{eco} compared to 2012. Following an earlier arrival of monsoon in 2012 compared to 2014 and 2015, we observed higher mean ET and R_{eco} in July than in June, which indicates the earlier arrival of monsoon precipitation greatly reduced the moisture limiting condition caused by fore-summer drought and led to subsequent increase in ET and R_{eco}. During late summer and autumn months, radiation declined significantly with ~ 30 % decrease in August and ~ 40 % decrease in September. Though 2012, 2014 and 2015 had diverse monsoon precipitation during these periods, HPM did not estimate significant differences in post-July ET. This result indicates the East River watershed is mainly under energy-limiting rather than moisture-limiting conditions during late-summer and autumn; and timing of monsoon arrival is more important than the absolute amount of monsoon precipitation for ET dynamics. This result is consistent with findings in Carroll et al. (2020). Their study also indicated earlier arrival of summer monsoon was effectively supporting ET and that the monsoon precipitation was quickly consumed by vegetation, whereas later arrival of summer monsoon water mainly contributed to streamflow under energy-limiting conditions.

Uncertainties of HPM models arise from several aspects. First, current choices of only five input features based on data availability may decrease estimation accuracy in certain environments, such as sites with seasonally dry periods. Though the LSTM component within HPMs can capture the memory effects and long-term dependencies of watershed dynamics, rare extreme values are difficult to be captured by LSTM due to insufficient training data for such cases. For example, we observed a decreasing prediction accuracy for ET and R_{eco} estimation at sites that experience drought conditions. Current use of meteorological forcings data and NDVI may not provide sufficient data for LSTM to identify droughts implicitly. Other key variables (e.g., soil moisture) when available can potentially be useful to help LSTM better quantify these rare events and increase model performance. Secondly, parameterization and insufficient spatiotemporal resolution of meteorological data still remain a challenge. Field observations along the Rocky Mountain ranges have shown that south-facing hillslopes have significantly earlier snowmelt compared to north-facing hillslopes (Kampf et al., 2015; Webb et al., 2018)~~, which are hypothesized to result in significant differences in ET and R_{ECO} dynamics. We compared ET and R_{ECO} differences among SNOTEL~~

~~sites and East River sites and identified ET differences among SNOTEL sites are greater than the differences among East River sites but R_{ECO} differences are similar between the two groups. Data from weather stations (SNOTEL sites) captured the spatiotemporal heterogeneity in radiation and temperature, however DAYMET data suggested very small differences in radiation and temperature (Figure S3 and S4). The insufficient spatial resolution of input meteorological forcing data limits HPM performance at the East River Watershed. Uncertainties in meteorological inputs can result in large errors (i.e., >20% MAE) and reduce accuracy by 10-30% in ET and R_{ECO} estimations as suggested by~~. However, we did not observe same level of heterogeneities in radiation and air temperature in reanalysis data compared to weather station data (Fig. S4 and S5). Mu et al. (2013) and Zhang et al. (2019)~~. Thus, there is still a significant need for high-spatial-resolution~~ suggested uncertainties in meteorological~~ forcing data products to enable better estimates of ET~~ inputs can result in large errors (i.e., > 20 % MAE) and ~~R_{ECO} and assess the governing factors that regulate their spatiotemporal variability.~~

~~In addition to the quality of meteorological data~~reduce accuracy by 10 – 30 %. Additionally, HPM is also influenced by remote sensing inputs accuracy~~. Incorrectly calculated or pixel-averaged NDVI values from Landsat images can greatly alter HPM outputs for ET and R_{ECO}. Satellite images with different~~, including but not limited to insufficient resolution, cloud ~~cover have a slight influence over the NDVI values calculated, which do not represent real-time vegetation~~ conditions~~. Algorithms used to reconstruct daily NDVI time series are also subject to uncertainties~~, spatial averaging, temporal reconstruction, any other algorithms involved. But with recent advances in remote sensing and satellite technologies (McCabe et al., 2017) and harmonized Landsat-Sentinel datasets (Claverie et al., 2018), the spatial and temporal resolution should greatly increase in the future (i.e., 3 m resolution and daily). ~~These advances will lead to more accurate classification of vegetation types and NDVI calculations, which are expected to decrease uncertainty associated with flux estimation~~Finally, errors can stem from the HPM hybrid approaches and conceptual model uncertainties. Any original errors in mechanistic models will be passed onto HPM estimations of ET and R_{eco}. We recommend to train data-driven HPM and mechanistic HPM using long time series (e.g., > 5 years) with high quality data or simulations, which enables HPMs to better memorize long-term dependencies of ecosystem dynamics. Though some of the uncertainties still remain a challenge, efforts have been made to minimize them through the technical advances described herein. Future HPM models can potentially be jointly trained on FLUXNET and process-based simulations to bypass certain limitations and provide more accurate ET and R_{eco} at sparsely monitored watersheds.

~~Another source of uncertainty in HPM arises from the choice of hybrid approaches and any parameter uncertainties in mechanistic models. Since HPM relies on accurate ET and R_{ECO} inputs from flux towers or mechanistic models, any uncertainties in measuring or modeling ET and R_{ECO} will propagate to HPM. If HPM is developed with a mechanistic model that has such missing components, these biases will be passed on to HPM estimation of ET and R_{ECO}. Parameter and conceptual model uncertainties in mechanistic models also restrict HPM's ability to "learn" the ecosystem dynamics. In order to reduce potential biasedness, we trained data-based HPM and physical-model-based HPM upon long time series (e.g., > 5 years) with quality assessed data or simulation results, which also enables HPM to better memorize long time dependencies of ecosystem dynamics.~~

~~Though the quantification of uncertainties remains challenging, efforts have been made to lower these uncertainties using the technical advances described here.~~

6. Conclusion

In this study, we developed and tested a Hybrid Predictive Modeling ~~(HPM)~~ approach for ET and ~~R_{ECO}~~R_{eco} estimation, with ~~a~~an enhanced focus on ~~mountainous watersheds~~a watershed in the Rocky Mountains. We developed individual HPM models at various FLUXNET sites and at sites where data ~~can supports~~could support the proper development of a mechanistic model (e.g., CLM). These models were validated against eddy covariance measurements and CLM outputs. We further used these models for ET and ~~R_{ECO}~~R_{eco} estimation at watersheds within the same ecoregion to test HPM's capability of providing estimation over space, where only meteorological forcings data and remote sensing data were available. Lastly, we applied the HPM to provide long-term estimation of ET and ~~R_{ECO}~~R_{eco} and test the sensitivity of HPM to various vegetation ~~types at various sites~~and meteorological conditions within the East River Watershed of CO, USA.

Formatted: Subscript

Given the promising results of HPM, ~~this work~~the approach offers an avenue for estimating ET and ~~R_{ECO}~~ R_{eco} using easy-to-acquire or commonly available datasets. This study also suggests that the spatial heterogeneity of meteorological forcings and vegetation dynamics have significant impacts on ET and ~~R_{ECO}~~R_{eco} dynamics, which may be currently underestimated due to typically coarse spatial resolution of data inputs. Parameters related to energy and soil moisture conditions can be implemented into HPM to increase HPM's accuracy, especially for sites in ecoregions limited by soil moisture conditions. Lastly, it should be pointed out that HPM is not restricted to estimation of ET and ~~R_{ECO} only. We focused here on developing HPM for ET and R_{ECO}, but~~R_{eco} only. HPM also has great potential for estimating other parameters important for water and carbon cycles given the right choice of input variables~~. Indeed, other attributes~~, such as net ecosystem exchange (Figure A6~~)~~). Thus, we believe the proposed HPM model can improve our prediction capabilities of ET and ~~sensible heat flux, might also be accurately captured~~R_{eco} at sparsely monitored watersheds and ~~represented with HPM, given the right choice of features~~advance our understanding of watershed dynamics.

Data availability. The data used in this study are from publicly available datasets. FLUXNET measurements can be accessed at https://FLUXNET.fluxdata.org. SNOTEL data are available at https://www.wcc.nrcs.usda.gov/snow/. DAYMET data can be found at (Thornton et al., 2017) or via Google Earth Engine. Landsat data are available on Google Earth Engine. All data and simulated results and model parameters associated with this article can be found at https://data.ess-dive.lbl.gov/view/doi:10.15485/1633810.

~~Acknowledgement.~~Acknowledgements. This material is based upon work supported as part of the Watershed Function Scientific Focus Area funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-AC02-05CH11231. We thank Haruko Wainwright and Bhavna Arora for providing comments on East River estimations. We also greatly appreciate all the guidance provided by Professor Yoram Rubin and Professor Dennis Baldocchi at UC Berkeley to the first author. We also acknowledge the Jane Lewis Fellowship Committee of the UC Berkeley for providing fellowship support to the first author.

References

Abatzoglou, J. T., Barbero, R., Wolf, J. W. and Holden, Z. A.: Tracking Interannual Streamflow Variability with

Drought Indices in the U.S. Pacific Northwest, J. Hydrometeorol., doi:10.1175/jhm-d-13-0167.1, 2014.

Ai, J., Jia, G., Epstein, H. E., Wang, H., Zhang, A. and Hu, Y.: MODIS-Based Estimates of Global Terrestrial Ecosystem Respiration, J. Geophys. Res. Biogeosciences, 123(2), 326–352, doi:10.1002/2017JG004107, 2018.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: Crop evapotranspiration: Guidelines for computing crop requirements., 1998.

Anderson, M. C., Allen, R. G., Morse, A. and Kustas, W. P.: Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources, Remote Sens. Environ., doi:10.1016/j.rse.2011.08.025, 2012.

Baldocchi, D.: Measuring fluxes of trace gases and energy between ecosystems and the atmosphere - the state and future of the eddy covariance method, Glob. Chang. Biol., doi:10.1111/gcb.12649, 2014.

Baldocchi, D. D., Ma, S., Rambal, S., Misson, L., Ourcival, J. M., Limousin, J. M., Pereira, J. and Papale, D.: On the differential advantages of evergreenness and deciduousness in mediterranean oak woodlands: A flux perspective, Ecol. Appl., 20(6), 1583–1597, doi:10.1890/08-2047.1, 2010.

Berryman, E. M., Vanderhoof, M. K., Bradford, J. B., Hawbaker, T. J., Henne, P. D., Burns, S. P., Frank, J. M., Birdsey, R. A. and Ryan, M. G.: Estimating Soil Respiration in a Subalpine Landscape Using Point, Terrain, Climate, and Greenness Data, J. Geophys. Res. Biogeosciences, 123(10), 3231–3249, doi:10.1029/2018JG004613, 2018.

Bodesheim, P., Jung, M., Gans, F., Mahecha, M. D. and Reichstein, M.: Upscaled diurnal cycles of land-Atmosphere fluxes: A new global half-hourly data product, Earth Syst. Sci. Data, 10(3), 1327–1365, doi:10.5194/essd-10-1327-2018, 2018.

~~De Bruin, H. A. R.: A model for the Priestley-Taylor parameter alpha., J. Clim. Appl. Meteorol., doi:10.1175/1520-0450(1983)0222.0.CO;2, 1983.~~

Budyko, M. I.: The Heat Balance of the Earth's Surface, Sov. Geogr., 2(4), 3–13, doi:10.1080/00385417.1961.10770761, 1961.

Carroll, R. W. H., Gochis, D. and Williams, K. H.: Efficiency of the Summer Monsoon in Generating Streamflow Within a Snow-Dominated Headwater Basin of the Colorado River, Geophys. Res. Lett., 47(23), doi:10.1029/2020GL090856, 2020.

Chang, L. L., Dwivedi, R., Knowles, J. F., Fang, Y. H., Niu, G. Y., Pelletier, J. D., Rasmussen, C., Durcik, M., Barron-Gafford, G. A. and Meixner, T.: Why Do Large-Scale Land Surface Models Produce a Low Ratio of Transpiration to Evapotranspiration?, J. Geophys. Res. Atmos., doi:10.1029/2018JD029159, 2018.

Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., Torn, M. S., Metzger, S., Kumar, J., Arain, M. A., Arkebauer, T. J., Baldocchi, D., Bernacchi, C., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Bracho, R., Brown, S., Brunsell, N. A., Chen, J., Chen, X., Clark, K., Desai, A. R., Duman, T., Durden, D., Fares, S., Forbrich, I., Gamon, J. A., Gough, C. M., Griffis, T., Helbig, M., Hollinger, D., Humphreys, E., Ikawa, H., Iwata, H., Ju, Y., Knowles, J. F., Knox, S. H., Kobayashi, H., Kolb, T., Law, B., Lee, X., Litvak, M., Liu, H., Munger, J. W., Noormets, A., Novick, K., Oberbauer, S. F., Oechel, W., Oikawa, P., Papuga, S. A., Pendall, E., Prajapati, P., Prueger, J., Quinton, W. L., Richardson, A. D., Russell, E. S., Scott, R. L., Starr, G., Staebler, R., Stoy, P. C., Stuart-Haëntjens, E., Sonnentag, O., Sullivan, R. C., Suyker, A., Ueyama, M., Vargas, R., Wood, J. D. and Zona, D.: Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites, Agric. For. Meteorol., 301–302(February), doi:10.1016/j.agrformet.2021.108350, 2021.

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J. C., Skakun, S. V. and Justice, C.: The Harmonized Landsat and Sentinel-2 surface reflectance data set, Remote Sens. Environ., doi:10.1016/j.rse.2018.09.002, 2018.

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. and Totterdell, I. J.: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, Nature, doi:10.1038/35041539, 2000.

Daggers, T. D., Kromkamp, J. C., Herman, P. M. J. and van der Wal, D.: A model to assess microphytobenthic primary production in tidal systems using satellite remote sensing, Remote Sens. Environ., 211(April), 129–145, doi:10.1016/j.rse.2018.03.037, 2018.

Falco, N., Wainwright, H., Dafflon, B., Léger, E., Peterson, J., Steltzer, H., Wilmer, C., Rowland, J. C., Williams, K. H. and Hubbard, S. S.: Investigating Microtopographic and Soil Controls on a Mountainous Meadow Plant Community Using High-Resolution Remote Sensing and Surface Geophysical Data, J. Geophys. Res. Biogeosciences, doi:10.1029/2018JG004394, 2019.

Gao, X., Mei, X., Gu, F., Hao, W., Li, H. and Gong, D.: Ecosystem respiration and its components in a rainfed spring maize cropland in the Loess Plateau, China, Sci. Rep., doi:10.1038/s41598-017-17866-1, 2017.

Gao, Y., Yu, G., Li, S., Yan, H., Zhu, X., Wang, Q., Shi, P., Zhao, L., Li, Y., Zhang, F., Wang, Y. and Zhang, J.: A remote sensing model to estimate ecosystem respiration in Northern China and the Tibetan Plateau, Ecol. Modell., doi:10.1016/j.ecolmodel.2015.03.001, 2015.

van Gorsel, E., Delpierre, N., Leuning, R., Black, A., Munger, J. W., Wofsy, S., Aubinet, M., Feigenwinter, C., Beringer, J., Bonal, D., Chen, B., Chen, J., Clement, R., Davis, K. J., Desai, A. R., Dragoni, D., Etzold, S., Grünwald, T., Gu, L., Heinesch, B., Hutyra, L. R., Jans, W. W. P., Kutsch, W., Law, B. E., Leclerc, M. Y., Mammarella, I., Montagnani, L., Noormets, A., Rebmann, C. and Wharton, S.: Estimating nocturnal ecosystem respiration from the vertical turbulent flux and change in storage of CO 2, Agric. For. Meteorol., 149(11), 1919–1930, doi:10.1016/j.agrformet.2009.06.020, 2009.

Greve, P., Gudmundsson, L., Orlowsky, B. and Seneviratne, S. I.: Introducing a probabilistic Budyko framework, Geophys. Res. Lett., 42(7), 2261–2269, doi:10.1002/2015GL063449, 2015.

Hargrove, W. W. and Hoffman, F. M.: Using multivariate clustering to characterize ecoregion borders, Comput. Sci. Eng., 1(4), 18–25, doi:10.1109/5992.774837, 1999.

Hargrove, W. W., Hoffman, F. M. and Law, B. E.: New analysis reveals representativeness of the amerflux network, Eos (Washington. DC)., 84(48), doi:10.1029/2003EO480001, 2003.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Comput., doi:10.1162/neco.1997.9.8.1735, 1997.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J. and Megown, K.: Completion of the 2011 national land cover database for the conterminous United States – Representing a decade of land cover change information, Photogramm. Eng. Remote Sensing, doi:10.1016/S0099-1112(15)30100-2, 2015.

Hu, J., Moore, D. J. P., Burns, S. P. and Monson, R.: Longer growing seasons lead to less carbon sequestration by a subalpine forest, Glob. Chang. Biol., 16(2), 771–783, doi:10.1111/j.1365-2486.2009.01967.x, 2010.

Hubbard, S. S., Williams, K. H., Agarwal, D., Banfield, J., Beller, H., Bouskill, N., Brodie, E., Carroll, R., Dafflon, B., Dwivedi, D., Falco, N., Faybishenko, B., Maxwell, R., Nico, P., Steefel, C., Steltzer, H., Tokunaga, T., Tran, P. A., Wainwright, H. and Varadharajan, C.: The East River, Colorado, Watershed: A Mountainous Community Testbed for Improving Predictive Understanding of Multiscale Hydrological–Biogeochemical Dynamics, Vadose Zo. J., 17(1), 0, doi:10.2136/vzj2018.03.0061, 2018.

IPCC: IPCC 2019- Special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystem, Res. Handb. Clim. Chang. Agric. Law, doi:10.4337/9781784710644, 2019.

Irons, J. R., Dwyer, J. L. and Barsi, J. A.: The next Landsat satellite: The Landsat Data Continuity Mission, Remote Sens. Environ., doi:10.1016/j.rse.2011.08.026, 2012.

Jägermeyr, J., Gerten, D., Lucht, W., Hostert, P., Migliavacca, M. and Nemani, R.: A high-resolution approach to estimating ecosystem respiration at continental scales using operational satellite data, Glob. Chang. Biol., doi:10.1111/gcb.12443, 2014.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., De Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S. and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, doi:10.1038/nature09396, 2010.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana, G., Viovy, N., Wang, Y. P., Weber, U., Zaehle, S. and Zeng, N.: Compensatory water effects link yearly global land CO 2 sink changes to temperature, Nature, 541(7638), 516–520, doi:10.1038/nature20780, 2017.

Kakalia, Z., Varadharajan, C., Alper, E., Brodie, E., Burrus, M., Carroll, R., Christianson, D., Hendrix, V., Henderson, M., Hubbard, S., Johnson, D., Versteeg, R., Williams, K. and Agarwal, D.: The East River Community Observatory Data Collection: Diverse, multiscale data from a mountainous watershed in the East River, Colorado, , 1–17 [online] Available from: https://doi.org/10.22541/au.160157556.64095872, 2020.

Kampf, S., Markus, J., Heath, J. and Moore, C.: Snowmelt runoff and soil moisture dynamics on steep subalpine hillslopes, Hydrol. Process., 29(5), 712–723, doi:10.1002/hyp.10179, 2015.

Keenan, T. F., Migliavacca, M., Papale, D., Baldocchi, D., Reichstein, M., Torn, M. and Wutzler, T.: Widespread inhibition of daytime ecosystem respiration, Nat. Ecol. Evol., 3(3), 407–415, doi:10.1038/s41559-019-0809-2, 2019.

Knowles, J. F., Blanken, P. D. and Williams, M. W.: Wet meadow ecosystems contribute the majority of overwinter soil respiration from snow-scoured alpine tundra, J. Geophys. Res. G Biogeosciences, 121(4), 1118–1130, doi:10.1002/2015JG003081, 2016.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K. and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22(11), 6005–6022, doi:10.5194/hess-22-6005-2018, 2018.

Lasslop, G., Reichstein, M., Papale, D., Richardson, A., Arneth, A., Barr, A., Stoy, P. and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: Critical issues and global evaluation, Glob. Chang. Biol., 16(1), 187–208, doi:10.1111/j.1365-2486.2009.02041.x, 2010.

Livingston, G. P. and Hutchinson, G. L.: Enclosure-based measurement of trace gas exchange: applications and sources of error., 1995.

Ma, Y., Liu, S., Song, L., Xu, Z., Liu, Y., Xu, T. and Zhu, Z.: Estimation of daily evapotranspiration and irrigation water efficiency at a Landsat-like scale for an arid irrigation area using multi-source remote sensing data, Remote Sens. Environ., 216(August), 715–734, doi:10.1016/j.rse.2018.07.019, 2018.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U. and Gascon, F.: Sen2Cor for Sentinel-2., 2017.

McCabe, M. F., Aragon, B., Houborg, R. and Mascaro, J.: CubeSats in Hydrology: Ultrahigh-Resolution Insights Into Vegetation Dynamics and Terrestrial Evaporation, Water Resour. Res., 53(12), 10017–10024, doi:10.1002/2017WR022240, 2017.

Metzger, S., Junkermann, W., Mauder, M., Butterbach-Bahl, K., Trancón Y Widemann, B., Neidl, F., Schäfer, K., Wieneke, S., Zheng, X. H., Schmid, H. P. and Foken, T.: Spatially explicit regionalization of airborne flux measurements using environmental response functions, Biogeosciences, 10(4), 2193–2217, doi:10.5194/bg-10-2193-2013, 2013.

Migliavacca, M., Reichstein, M., Richardson, A. D., Mahecha, M. D., Cremonese, E., Delpierre, N., Galvagno, M., Law, B. E., Wohlfahrt, G., Andrew Black, T., Carvalhais, N., Ceccherini, G., Chen, J., Gobron, N., Koffi, E., William Munger, J., Perez-Priego, O., Robustelli, M., Tomelleri, E. and Cescatti, A.: Influence of physiological phenology on the seasonal pattern of ecosystem respiration in deciduous forests, Glob. Chang. Biol., 21(1), 363–376, doi:10.1111/gcb.12671, 2015.

Mohanty, B. P., Cosh, M. H., Lakshmi, V. and Montzka, C.: Soil moisture remote sensing: State-of-the-science, Vadose Zo. J., doi:10.2136/vzj2016.10.0105, 2017.

Mu, Q., Zhao, M. and Running, S. W.: MODIS Global Terrestrial Evapotranspiration (ET) Product (MOD16A2/A3), Algorithm Theor. Basis Doc., 2013.

NASA: Moderate Resolution Imaging Spectroradiometer (MODIS) Overview, Nasa, 2008.

~~Ng, G., Bedford, D. and Miller, D.: A mechanistic modeling and data assimilation framework for Mojave Desert ecohydrology, Water Resour. Res., 4662–4685, doi:10.1002/2014WR015281.Received, 2014.~~

~~Noormets, A., Desai, A. R., Cook, B. D., Euskirchen, E. S., Ricciuto, D. M., Davis, K. J., Bolstad, P. V., Schmid, H. P., Vogel, C. V., Carey, E. V., Su, H. B. and Chen, J.: Moisture sensitivity of ecosystem respiration: Comparison of 14 forest ecosystems in the Upper Great Lakes Region, USA, Agric. For. Meteorol., doi:10.1016/j.agrformet.2007.08.002, 2008.~~

Novick, K. A., Oishi, A. C., Ward, E. J., Siqueira, M. B. S., Juang, J. Y. and Stoy, P. C.: On the difference in the net ecosystem exchange of CO2 between deciduous and evergreen forests in the southeastern United States, Glob. Chang. Biol., 21(2), doi:10.1111/gcb.12723, 2015.

Olah, C.: Understanding LSTM Networks, https//colah.github.io/posts/2015-08-Understanding-LSTMs/, 2015, 2015.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, J., Subin, Z. M., Swenson, S. C., Thornton, P. E., Bozbiyik, A., Fisher, R. A., Heald, C. L., Kluzek, E., Lamarque, J.-F., Lawrence, P. J., Leung, L. R., Lipscomb, W., Muszala, S., Ricciuto, D. M., Sacks, W. J., Sun, Y., Tang, J. and Yang, Z.-L.: Technical Description of version 4.5 of the Community Land Model (CLM)., 2013.

Omernik, J. M.: Perspectives on the nature and definition of ecological regions., Environ. Manage., doi:10.1007/s00267-003-5197-2, 2004.

Omernik, J. M. and Griffith, G. E.: Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework, Environ. Manage., doi:10.1007/s00267-014-0364-1, 2014.

Oyler, J. W., Dobrowski, S. Z., Ballantyne, A. P., Klene, A. E. and Running, S. W.: Artificial amplification of warming trends across the mountains of the western United States, Geophys. Res. Lett., doi:10.1002/2014GL062803, 2015.

Paca, V. H. da M., Espinoza-Dávalos, G. E., Hessels, T. M., Moreira, D. M., Comair, G. F. and Bastiaanssen, W. G. M.: The spatial variability of actual evapotranspiration across the Amazon River Basin based on remote sensing products validated with flux towers, Ecol. Process., 8(1), doi:10.1186/s13717-019-0158-8, 2019.

~~Pedersen, S. H., Liston, G. E., Tamstorf, M. P., Abermann, J., Lund, M. and Schmidt, N. M.: Quantifying snow controls on vegetation greenness, Ecosphere, doi:10.1002/ecs2.2309, 2018.~~

~~PRIESTLEY~~Priestley, C. H. B. and ~~TAYLOR~~Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Mon. Weather Rev., doi:10.1175/1520-

0493(1972)100<0081:otaosh>2.3.co;2, 1972.

Pumpanen, J., Kolari, P., Ilvesniemi, H., Minkkinen, K., Vesala, T., Niinistö, S., Lohila, A., Larmola, T., Morero, M., Pihlatie, M., Janssens, I., Yuste, J. C., Grünzweig, J. M., Reth, S., Subke, J. A., Savage, K., Kutsch, W., Østreng, G., Ziegler, W., Anthoni, P., Lindroth, A. and Hari, P.: Comparison of different chamber techniques for measuring soil CO 2 efflux, Agric. For. Meteorol., doi:10.1016/j.agrformet.2003.12.001, 2004.

~~Qiao, Z., Xu, X., Zhao, M., Wang, F. and Liu, L.: The application of a binary division procedure to the classification of forest subcategories using MODIS time-series data during 2000–2010 in China, Int. J. Remote Sens., doi:10.1080/01431161.2016.1176269, 2016.~~

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J. M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D. and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm, Glob. Chang. Biol., doi:10.1111/j.1365-2486.2005.001002.x, 2005.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566(7743), 195–204, doi:10.1038/s41586-019-0912-1, 2019.

Ren, H., Cromwell, E., Kravitz, B. and Chen, X.: Using Deep Learning to Fill Spatio-Temporal Data Gaps in Hydrological Monitoring Networks, Hydrol. Earth Syst. Sci. Discuss., (May), 1–20, doi:10.5194/hess-2019-196, 2019.

Rungee, J., Bales, R. and Goulden, M.: Evapotranspiration response to multiyear dry periods in the semiarid western United States, Hydrol. Process., doi:10.1002/hyp.13322, 2019.

Ryu, Y., Baldocchi, D. D., Kobayashi, H., Van Ingen, C., Li, J., Black, T. A., Beringer, J., Van Gorsel, E., Knohl, A., Law, B. E. and Roupsard, O.: Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales, Global Biogeochem. Cycles, 25(4), 1–24, doi:10.1029/2011GB004053, 2011.

Seneviratne, S. I., Lüthi, D., Litschi, M. and Schär, C.: Land-atmosphere coupling and climate change in Europe, Nature, doi:10.1038/nature05095, 2006.

~~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.~~

Strachan, S., Kelsey, E. P., Brown, R. F., Dascalu, S., Harris, F., Kent, G., Lyles, B., McCurdy, G., Slater, D. and Smith, K.: Filling the Data Gaps in Mountain Climate Observatories Through Advanced Technology, Refined Instrument Siting, and a Focus on Gradients, Mt. Res. Dev., 36(4), 518–527, doi:10.1659/mrd-journal-d-16-00028.1, 2016.

Suleau, M., Moureaux, C., Dufranne, D., Buysse, P., Bodson, B., Destain, J. P., Heinesch, B., Debacq, A. and Aubinet, M.: Respiration of three Belgian crops: Partitioning of total ecosystem respiration in its heterotrophic, above- and below-ground autotrophic components, Agric. For. Meteorol., doi:10.1016/j.agrformet.2011.01.012, 2011.

Teuling, A. J., Van Loon, A. F., Seneviratne, S. I., Lehner, I., Aubinet, M., Heinesch, B., Bernhofer, C., Grünwald, T., Prasse, H. and Spank, U.: Evapotranspiration amplifies European summer drought, Geophys. Res. Lett., doi:10.1002/grl.50495, 2013.

Thornton, P. E., Thornton, M. M., Mayer, B. W., Wei, Y., Devarakonda, R., Vose, R. S. and Cook, R. B.: Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3, ORNL DAAC, Oak Ridge, Tennessee, USA, 2017.

Tran, A. P., Rungee, J., Faybishenko, B., Dafflon, B. and Hubbard, S. S.: Assessment of spatiotemporal variability of evapotranspiration and its governing factors in a mountainous watershed, Water (Switzerland), 11(2), doi:10.3390/w11020243, 2019.

U.S. Environmental Protection Agency: Level III Ecoregions of the Continental United States, Environ. Prot., 2003.

Visser, A., Thaw, M., Deinhart, A., Bibby, R., Safeeq, M., Conklin, M., Esser, B. and Van der Velde, Y.: Cosmogenic Isotopes Unravel the Hydrochronology and Water Storage Dynamics of the Southern Sierra Critical Zone, Water Resour. Res., doi:10.1029/2018WR023665, 2019.

Viviroli, D. and Weingartner, R.: "Water towers"—A global view of the hydrological importance of mountains, in Advances in Global Change Research., 2008.

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M. and Weingartner, R.: Mountains of the world, water towers for humanity: Typology, mapping, and global significance, Water Resour. Res., 43(7), 1–13, doi:10.1029/2006WR005653, 2007.

Vogelmann, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K. and Van Driel, N.: Completion of the 1990s National Land Cover Data set for the conterminous United States from Landsat thematic mapper data and ancillary data sources, Photogramm. Eng. Remote Sensing, 2001.

~~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.~~

~~Wang, B., Zha, T. S., Jia, X., Wu, B., Zhang, Y. Q. and Qin, S. G.: Soil moisture modifies the response of soil respiration to temperature in a desert shrub ecosystem, Biogeosciences, 11(2), 259–268, doi:10.5194/bg-11-259-2014, 2014.~~

Webb, R. W., Fassnacht, S. R. and Gooseff, M. N.: Hydrologic flow path development varies by aspect during spring snowmelt in complex subalpine terrain, Cryosphere, 12(1), 287–300, doi:10.5194/tc-12-287-2018, 2018.

Wieder, W. R., Knowles, J. F., Blanken, P. D., Swenson, S. C. and Suding, K. N.: Ecosystem function in complex mountain terrain: Combining models and long-term observations to advance process-based understanding, J. Geophys. Res. Biogeosciences, doi:10.1002/2016JG003704, 2017.

Williams, ~~C. A. and Albertson, J. D.: Soil moisture controls on canopy-scale water and carbon fluxes in an African savanna, Water Resour. Res., doi:10.1029/2004WR003208, 2004.~~

~~Williams,~~ M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. and Wang, Y.-P.: Improving land surface models with FLUXNET data, Biogeosciences Discuss., doi:10.5194/bgd-6-2785-2009, 2009.

Wilson, K. B., Hanson, P. J., Mulholland, P. J., Baldocchi, D. D. and Wullschleger, S. D.: A comparison of methods for determining forest evapotranspiration and its components: Sap-flow, soil water budget, eddy covariance and catchment water balance, Agric. For. Meteorol., 106(2), 153–168, doi:10.1016/S0168-1923(00)00199-4, 2001.

Xiao, J., Ollinger, S. V., Frolking, S., Hurtt, G. C., Hollinger, D. Y., Davis, K. J., Pan, Y., Zhang, X., Deng, F., Chen, J., Baldocchi, D. D., Law, B. E., Arain, M. A., Desai, A. R., Richardson, A. D., Sun, G., Amiro, B., Margolis, H., Gu, L., Scott, R. L., Blanken, P. D. and Suyker, A. E.: Data-driven diagnostics of terrestrial carbon dynamics over North America, Agric. For. Meteorol., doi:10.1016/j.agrformet.2014.06.013, 2014.

Xu, H., Xiao, J., Zhang, Z., Ollinger, S. V., Hollinger, D. Y., Pan, Y. and Wan, J.: Canopy photosynthetic capacity drives contrasting age dynamics of resource use efficiencies between mature temperate evergreen and deciduous forests, Glob. Chang. Biol., 26(11), doi:10.1111/gcb.15312, 2020.

Xu, L., Baldocchi, D. D. and Tang, J.: How soil moisture, rain pulses, and growth alter the response of ecosystem respiration to temperature, Global Biogeochem. Cycles, 18(4), 1–10, doi:10.1029/2004GB002281, 2004.

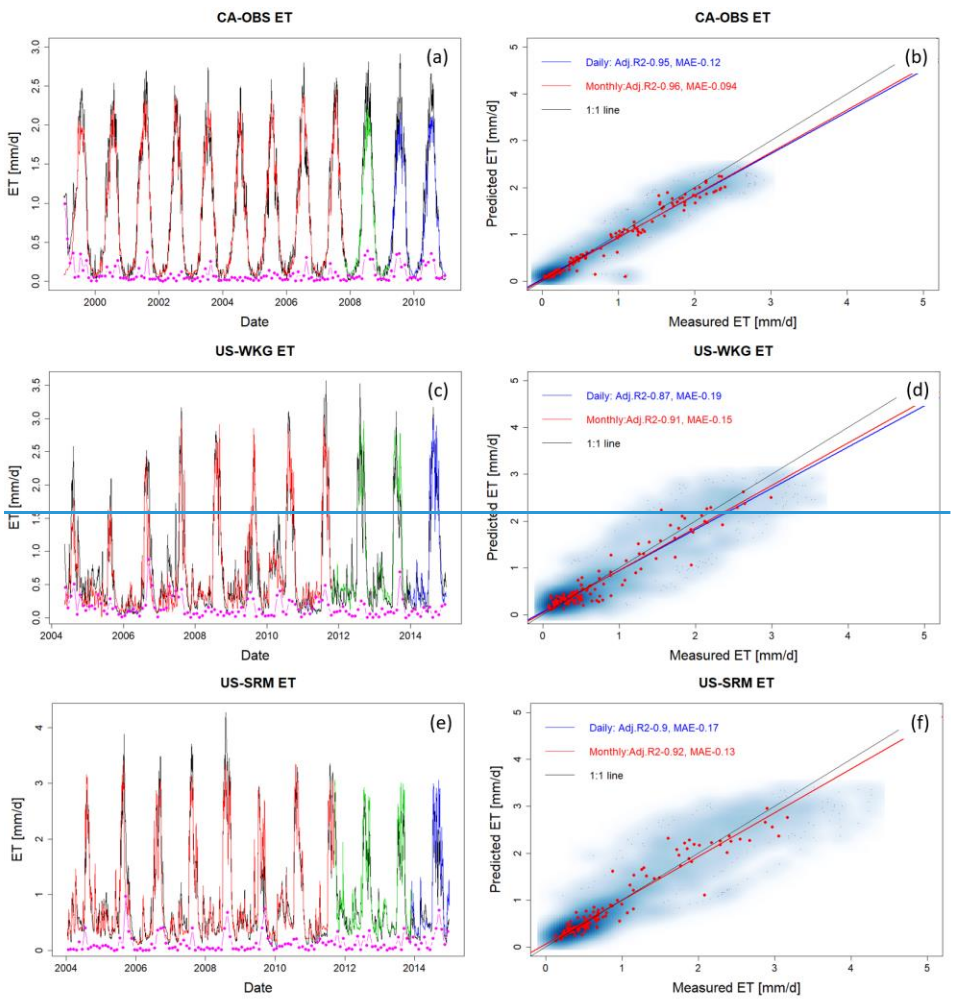
Xu, T., Guo, Z., Liu, S., He, X., Meng, Y., Xu, Z., Xia, Y., Xiao, J., Zhang, Y., Ma, Y. and Song, L.: Evaluating Different Machine Learning Methods for Upscaling Evapotranspiration from Flux Towers to the Regional Scale, J. Geophys. Res. Atmos., 123(16), 8674–8690, doi:10.1029/2018JD028447, 2018.

Zhang, L., Potter, N., Hickel, K., Zhang, Y. and Shao, Q.: Water balance modeling over variable time scales based on the Budyko framework - Model development and testing, J. Hydrol., 360(1–4), 117–131, doi:10.1016/j.jhydrol.2008.07.021, 2008.

Zhang, Y., Kong, D., Gan, R., Chiew, F. H. S., McVicar, T. R., Zhang, Q. and Yang, Y.: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017, Remote Sens. Environ., 222(May 2018), 165–182, doi:10.1016/j.rse.2018.12.031, 2019.

Appendix

1. ET and R_{ECO} Estimation over Time at other Fluxnet sites


CA-OBS ET
(a)
ET [mm/d]
Date
CA-OBS ET
(b)
Daily: Adj.R2-0.95, MAE-0.12
Monthly:Adj.R2-0.96, MAE-0.094
1:1 line
Predicted ET [mm/d]
Measured ET [mm/d]
US-WKG ET
(c)
ET [mm/d]
Date
US-WKG ET
(d)
Daily: Adj.R2-0.87, MAE-0.19
Monthly:Adj.R2-0.91, MAE-0.15
1:1 line
Predicted ET [mm/d]
Measured ET [mm/d]
US-SRM ET
(e)
ET [mm/d]
Date
US-SRM ET
(f)
Daily: Adj.R2-0.9, MAE-0.17
Monthly:Adj.R2-0.92, MAE-0.13
1:1 line
Predicted ET [mm/d]
Measured ET [mm/d]

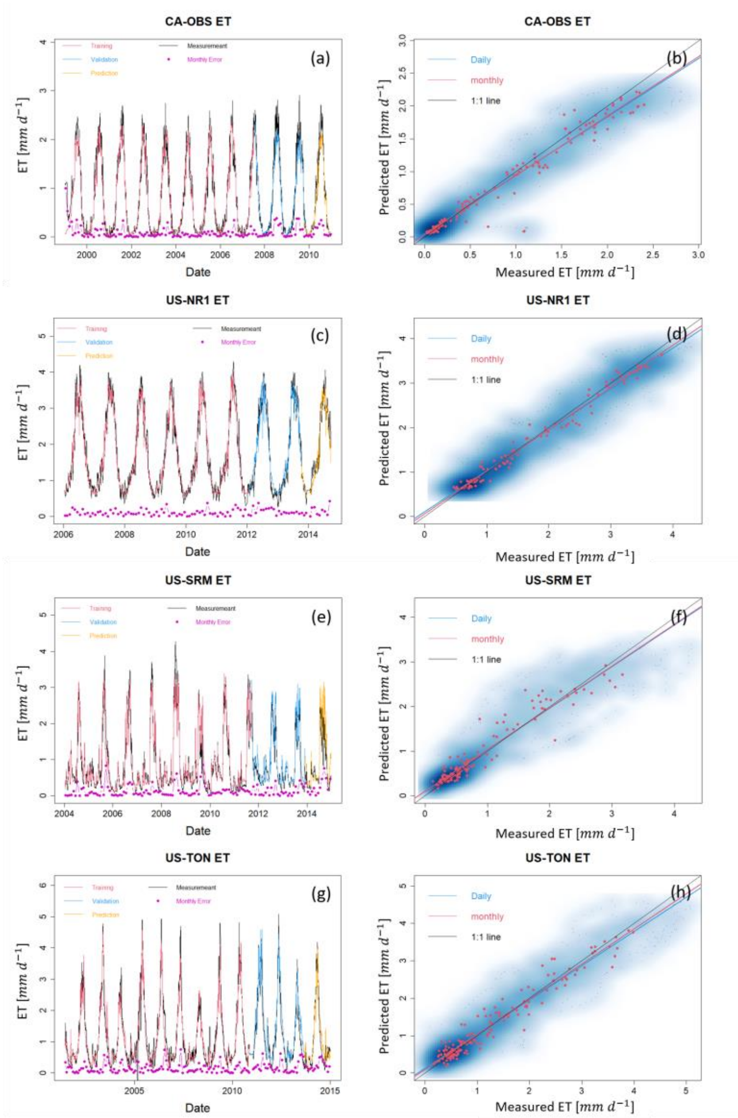
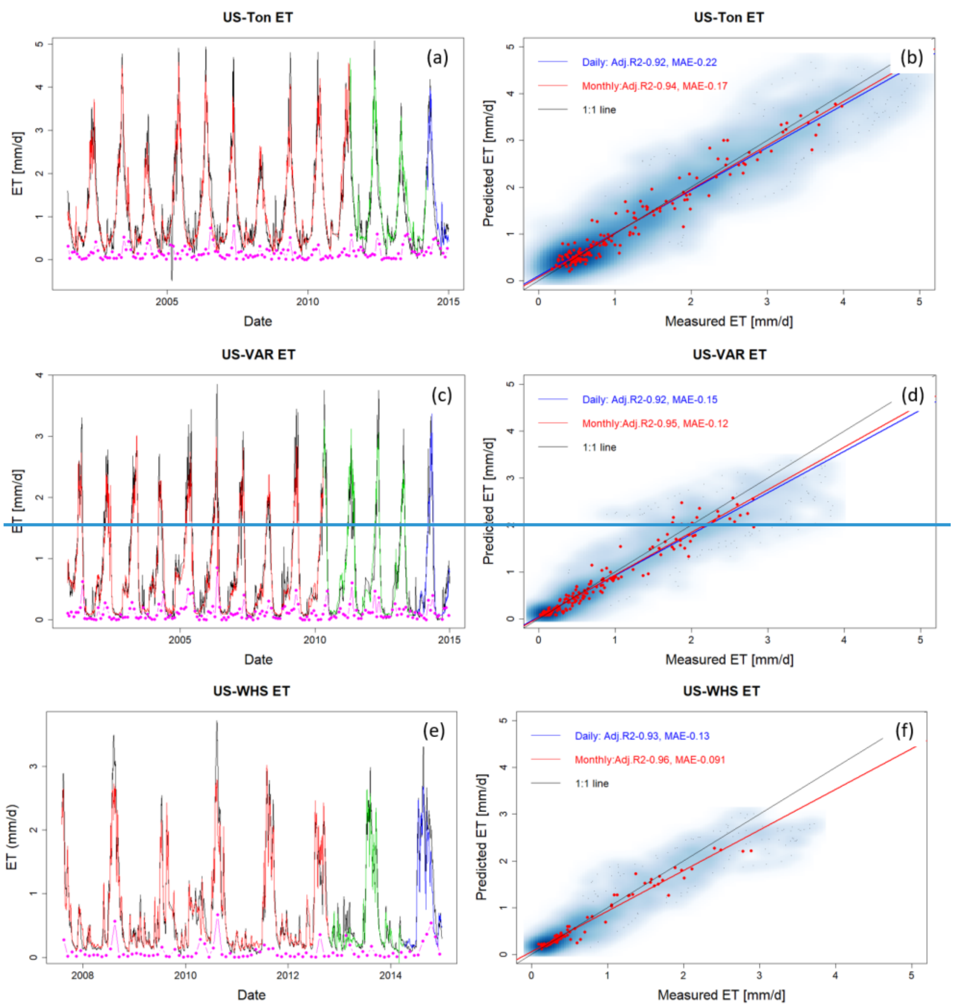


Figure A1: ET estimation with data from selected FLUXNET sites at CA-OBS, US-~~Wkg~~NR1, US-SRM, and US-~~SRM~~Ton. Panels (a), (c), (e) and (~~e~~g) present daily estimations of ET ~~with red, green, and blue lines representing data used~~separated for training, validation, and prediction~~, respectively, and the black line representing the eddy covariance measurement.~~. Pink points ~~describe~~depict monthly ~~mean difference~~error between HPM estimation and ~~measured~~FLUXNET data. Panels

(b), (d), (f) and (~~f~~h) show the scatter plots of daily (blue) and monthly (red) ET. Darker blue clouds represent greater density of data points.



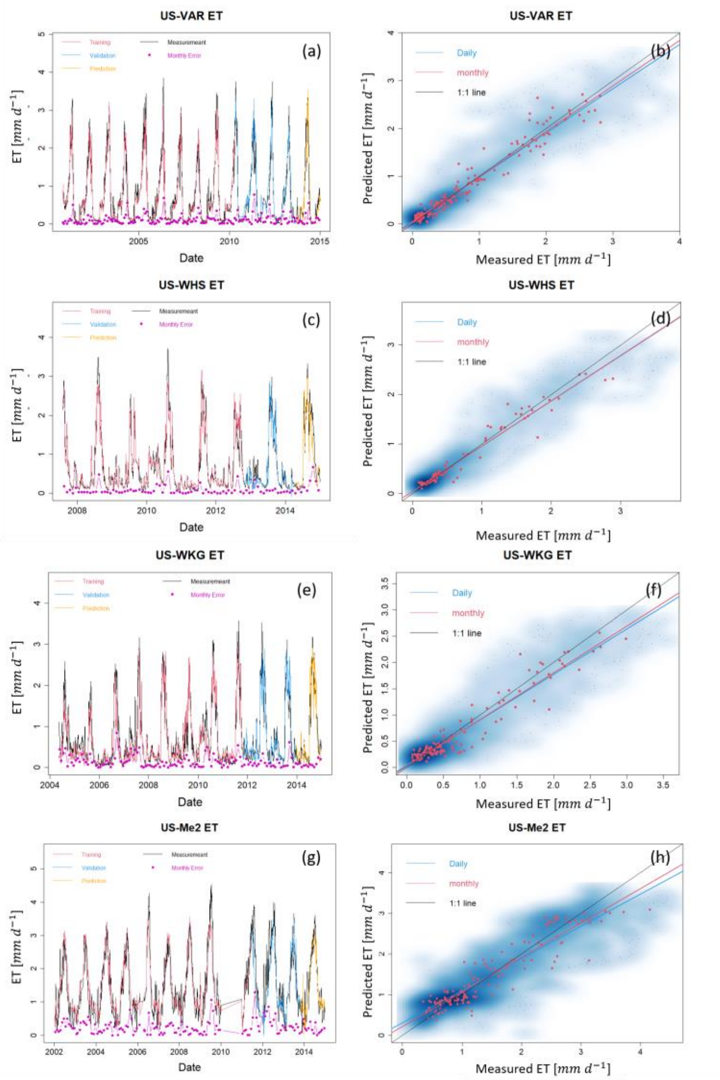
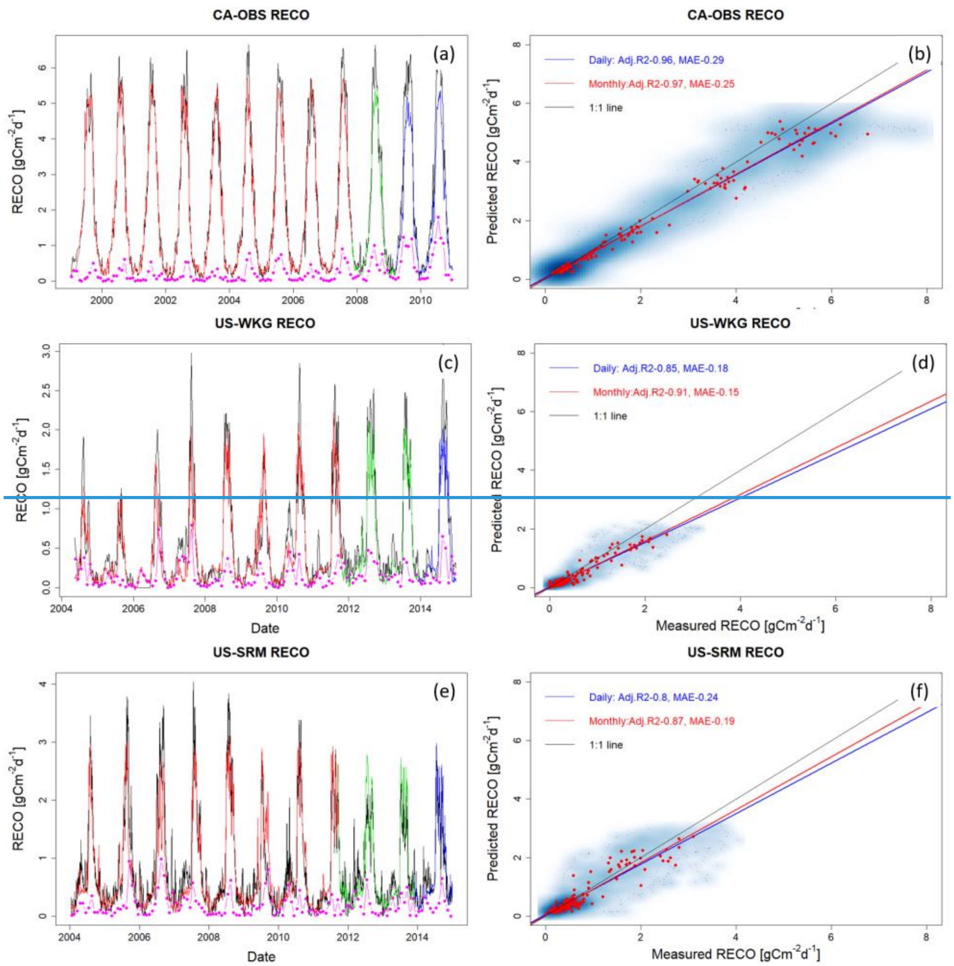


Figure A2: ET estimation with data from selected FLUXNET sites at US-~~Ton, US-~~Var, ~~and~~ US-Whs~~.~~, US-Wkg and US-Me2. Panels (a), (c), ~~and~~ (e) and (g) present daily estimations of ET ~~with red, green, and blue lines representing data used~~separated for training, validation, and prediction~~, respectively, and the black line representing the eddy covariance measurement.~~. Pink points ~~describe~~depict monthly ~~mean difference~~error between HPM estimation and ~~measured~~FLUXNET data. Panels (b), (d), (f) and (~~f~~h) show the scatter plots of daily (blue) and monthly (red) ET. Darker blue clouds represent greater density of data points.


CA-OBS RECO
CA-OBS RECO
(a)
(b)
Daily: Adj.R2-0.96, MAE-0.29
Monthly:Adj.R2-0.97, MAE-0.25
1:1 line
US-WKG RECO
US-WKG RECO
(c)
(d)
Daily: Adj.R2-0.85, MAE-0.18
Monthly:Adj.R2-0.91, MAE-0.15
1:1 line
US-SRM RECO
US-SRM RECO
(e)
(f)
Daily: Adj.R2-0.8, MAE-0.24
Monthly:Adj.R2-0.87, MAE-0.19
1:1 line
RECO [gCm-2d-1]
Predicted RECO [gCm-2d-1]
Measured RECO [gCm-2d-1]
Date

1200

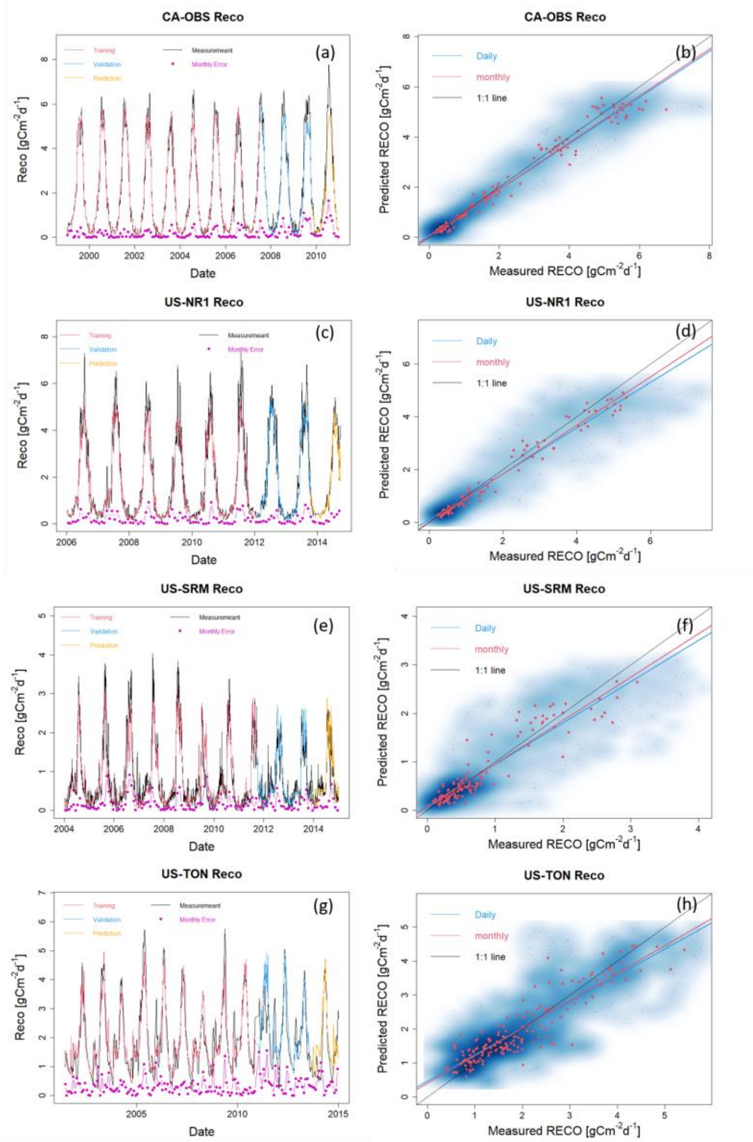
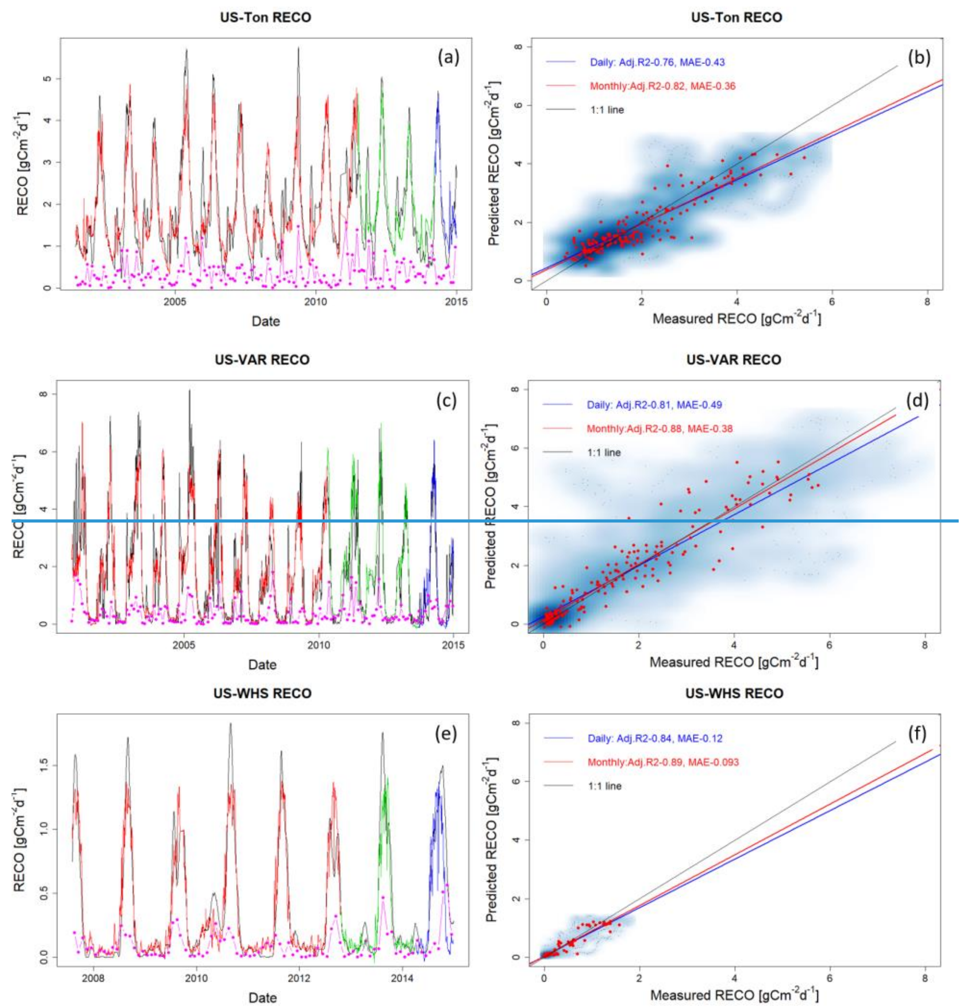


Figure A3: ~~R_{ECO}~~R_{eco} estimation with data from selected FLUXNET sites at CA-OBS, US-~~Wkg~~NR1, US-SRM, and US-~~SRM~~Ton. Panels (a), (c), ~~and~~ (e) and (g) present daily estimations of ~~R_{ECO} with red, green, and blue lines representing data used~~R_{eco} separated for training, validation, and prediction~~, respectively, and the black line is eddy covariance measurement.~~. Pink points ~~describe the~~depict monthly ~~mean difference~~error between HPM estimation and

~~measured~~FLUXNET data. Panels (b), (d), ~~and~~ (f) and (h) show the scatter plots of daily (blue) and monthly (red) ~~R_{ECO}~~ R_{eco}. Darker blue clouds represent greater density of data points.



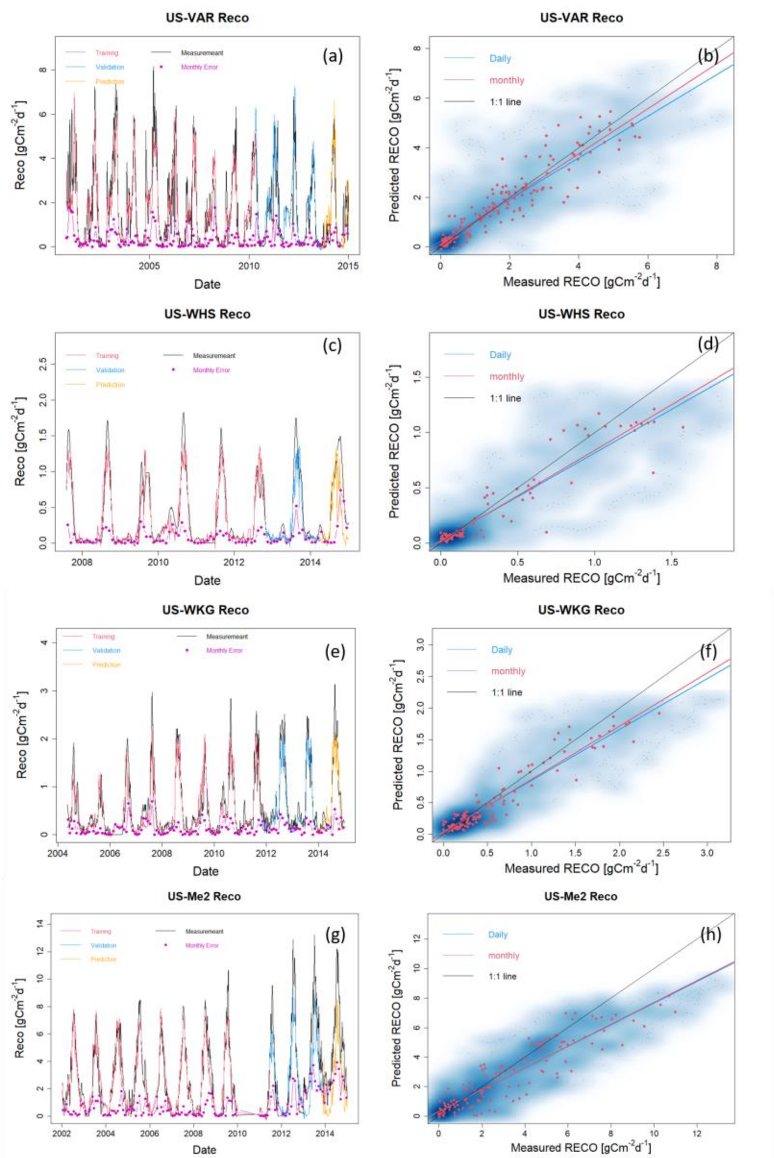
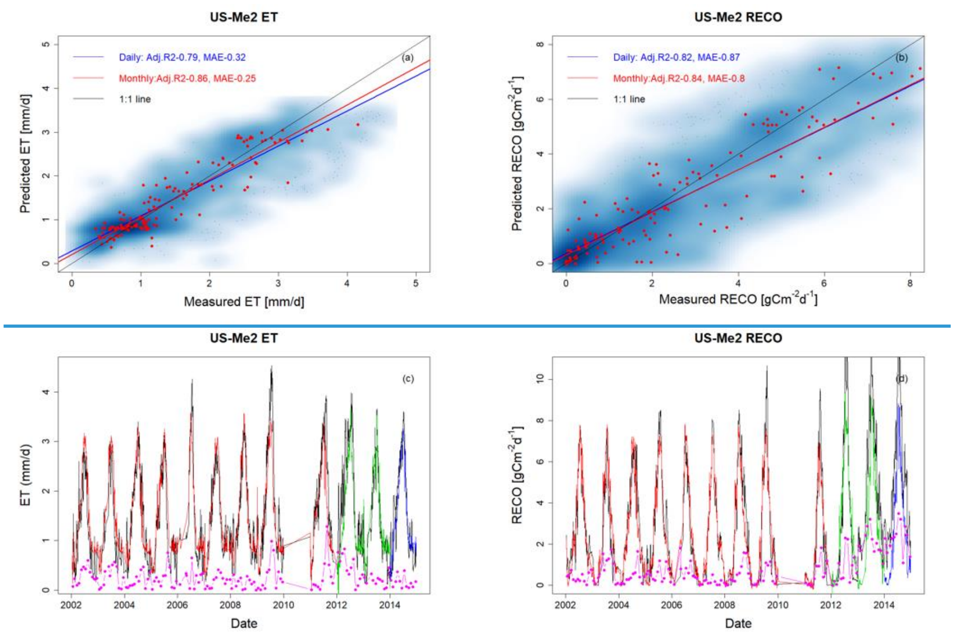


Figure A4: ~~R_{ECO}~~R_{eco} estimation with data from selected FLUXNET sites at US-~~Ton, US-~~Var, ~~and~~ US-Whs~~.~~, US-Wkg and US-Me2. Panels (a), (c), ~~and~~ (e) and (g) present daily estimations of ~~R_{ECO} with red, green, and blue lines representing data used~~R_{eco} separated for training, validation, and prediction~~, respectively, and the black line representing the eddy covariance measurement.~~. Pink points ~~describe~~depict monthly ~~mean difference~~error between HPM estimation and

~~measured~~FLUXNET data. Panels (b), (d), ~~and~~ (f) and (h) show the scatter plots of daily (blue) and monthly (red) ~~R_{ECO}.~~ R_{eco}. Darker blue clouds represent greater density of data points.



~~Figure A5: ET and R_{ECO} estimation at US-Me2. Panels (a) and (b) show the scatter plots of daily (blue) and monthly (red) ET and R_{ECO}. Darker blue clouds represent greater density of data points. Panels (c), and (d) present daily estimations of R_{ECO} with red, green, and blue lines representing data used for training, validation, and prediction, respectively, and the black line representing the eddy covariance measurement. Pink points describe monthly mean difference between HPM estimation and measured data.~~

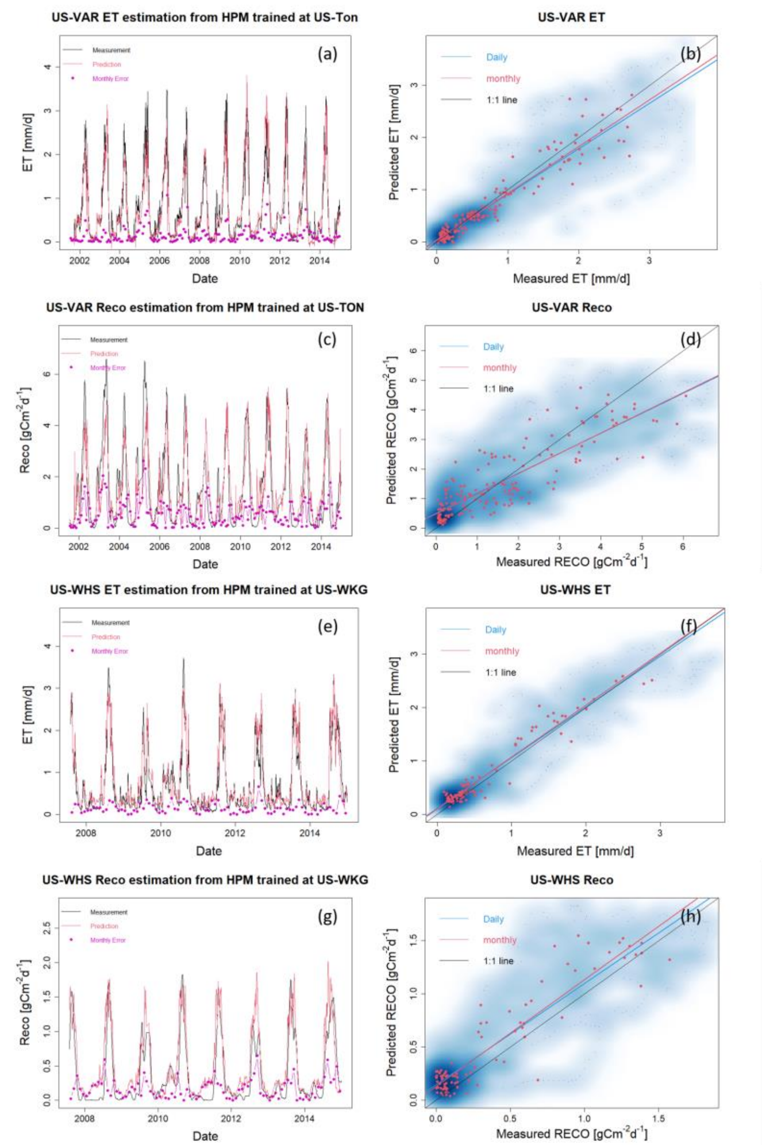
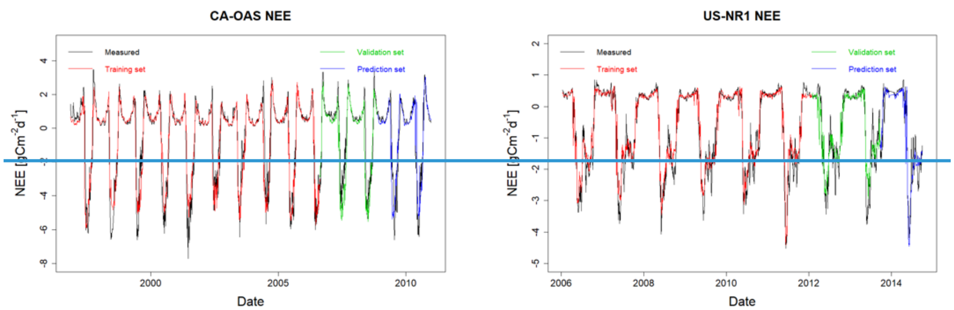


Figure A5: Use case 2. ET and R_{eco} estimation at US-Var and US-Whs from HPM trained at US-Ton and US-Wky, respectively.

2. Tested NEE Estimation over Time at CA-OAS and US-NR1



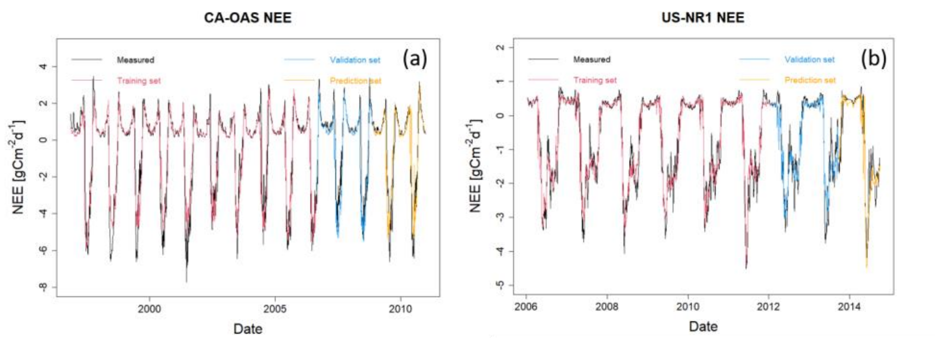


Figure A6. HPM estimate of NEE at CA-OAS and US-NR1. R^2 between estimation and measurements are 0.87, 0.83 and 0.81 at CA-OAS; 0.94, 0.88 and 0.90 at US-NR1 for the training set, validation set and prediction set, respectively. Model inputs include air temperature, soil temperature, sn, precipitation and radiation.

Formatted: Line spacing: single