Oak Ridge National Laboratory

AOGS2023 20th Annual Meeting — Asia Oceania Geosciences Society

Towards improved accuracy and efficiency of ecohydrologic processes using artificial intelligence

Elias C. Massoud[1], Forrest M. Hoffman[1], Nathan Collier[1]

1 Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Abstract ID: BG06-A006
Session ID: BG06 - Integrated Understanding of Global Carbon and Other Biogeochemical Cycles and Their Feedbacks

Contact: massoudec@ornl.gov

Abstract

Ecohydrology is the study of ecosystem and water cycle science, and understanding interactions among ecohydrologic mechanisms is challenging. Artificial Intelligence and Machine Learning (AI/ML) approaches are likely to provide new avenues for simulating mechanistic processes at different scales. In this work, we apply ML-based methods to improve simulations of vegetation processes relevant to ecohydrology, such as photosynthesis and stomatal conductance. Our results showcase improved skill and reduced computational cost of ML-based methods compared to commonly used empirical approaches. This work explores potential benefits of combining ML-based methods with existing approaches in ecohydrology.

Introduction

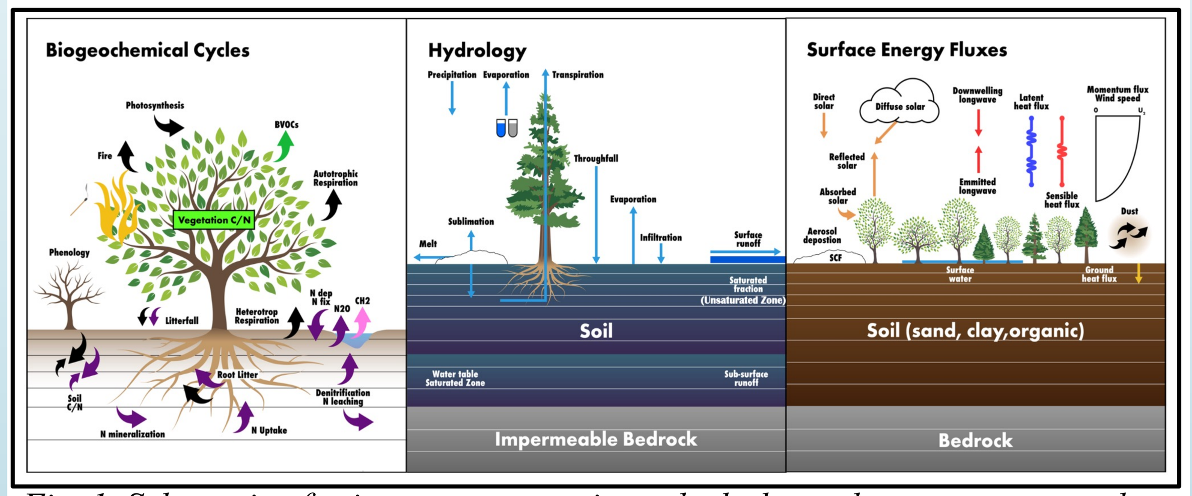

Fig. 1. Schematic of primary processes in ecohydrology, shown as an example of integrated ecohydrologic processes required for modeling across scales.

Ecohydrology bridges the gap between ecosystem ecology and water cycle science by incorporating knowledge of land surface processes, plants, atmospheric science, and hydrology.

Ecohydrologic processes operate at varying scales ranging from stomates and microorganisms to canopies, forests, watersheds, continents and the entire globe (Fig. 1 & 2).

Research is being conducted to determine the benefits of implementing and integrating ML-based methods within Earth System Models (ESM) and ecohydrologic models (Fig. 3).

ML-based simulators will not replace the representation of older empirical approaches, but instead will supplement and act as an interchangeable model choice to larger models.

ML-based methods require training from observed datasets. Datasets that capture specific ecohydrologic processes are rare, however, making it challenging to benefit from ML.

We compiled information from a collection of various leaf-level data (e.g., Lin et al. 2015; Anderegg et al. 2018; Han et al. 2022). This curated dataset is used to develop initial ML models for photosynthesis and stomatal conductance (Fig. 4). Preliminary results from our investigations are shown here (Fig. 5 & 6).

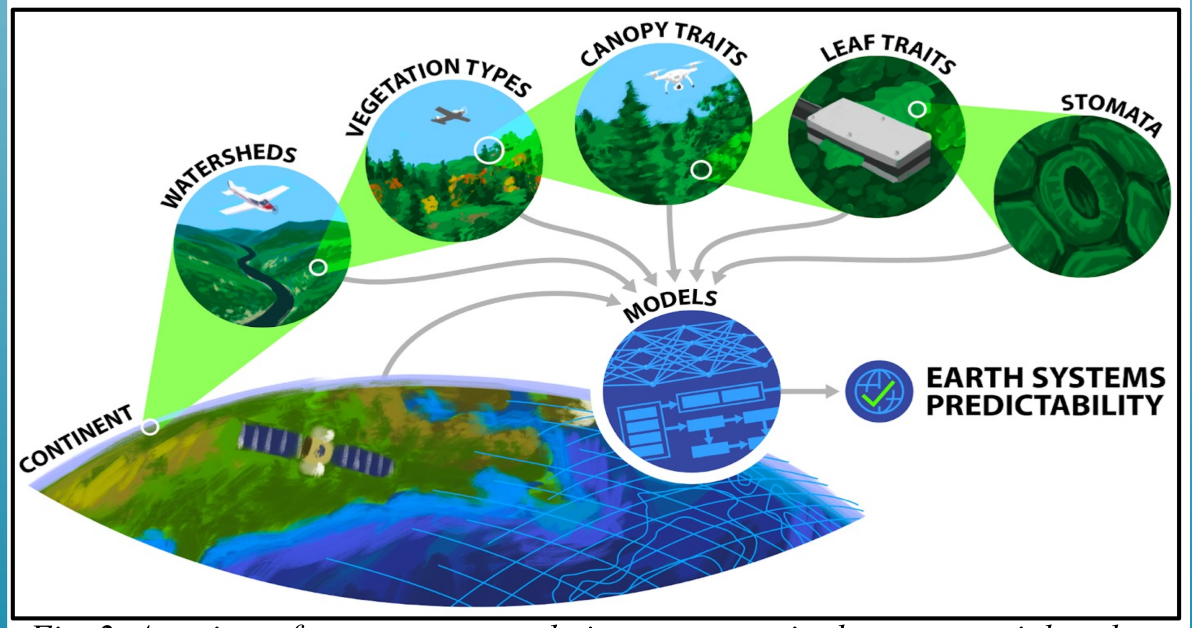

Fig. 2. A variety of measurement techniques are required across spatial scales to improve representation of ecohydrologic processes with ML approaches. In this work, we employed leaf-level data to perform the ML simulations.

Materials – Data Collection

We collected data and provided curation for leaf-level fluxes from Lin et al. 2015, Anderegg et al. 2018, and Han et al. 2022.

This data set will be used for systematic benchmarking and for ML training. The code below shows how to download the data:

```
import intake cat =
intake.open_catalog("https://raw.githubusercontent.com/nocollier/MLPhotoSynthesis/main/data/leaf-level.yaml")
df = cat['Lin2015'].read()
```

Methodology

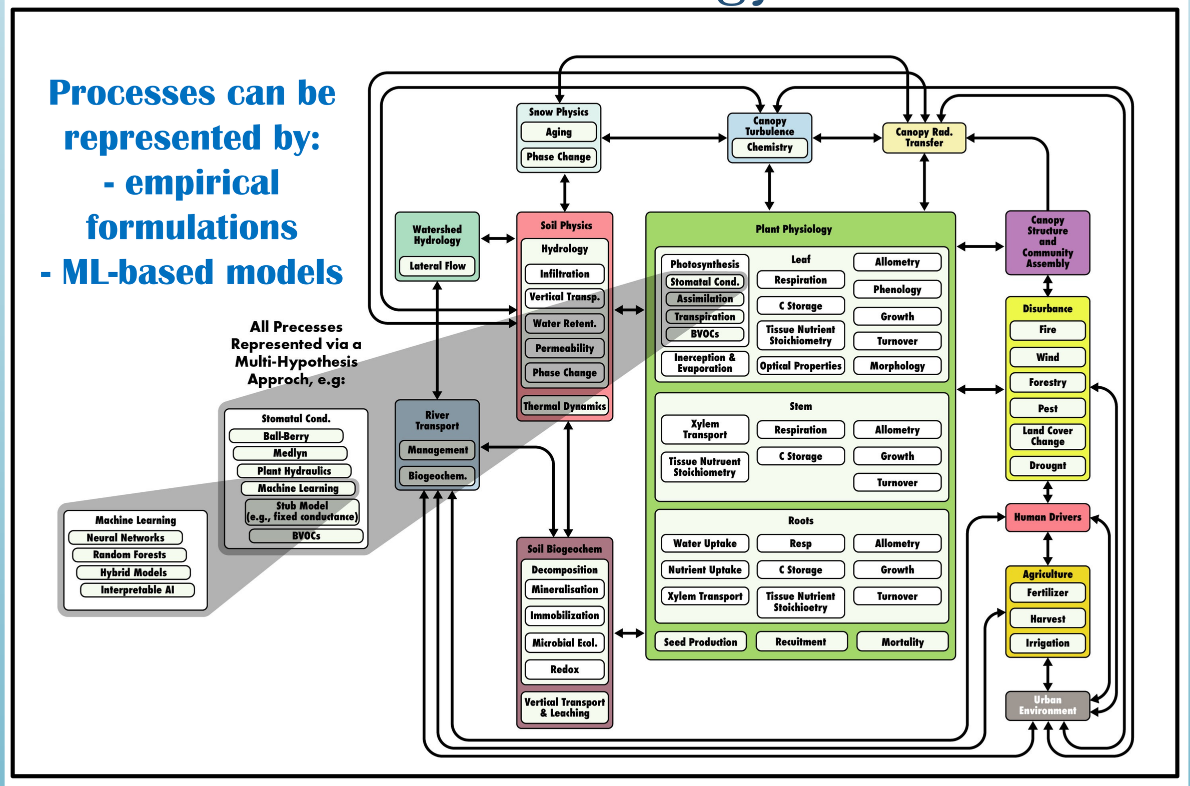

Fig. 3. A process schematic of a full-complexity land surface model. All processes are intended to allow alternative specification, including empirical or ML-derived formulations (Adapted from Fisher and Koven, 2020). For this work, we focus on building ML models to simulate photosynthesis and stomatal conductance and we compare with empirical formulations and observed data.

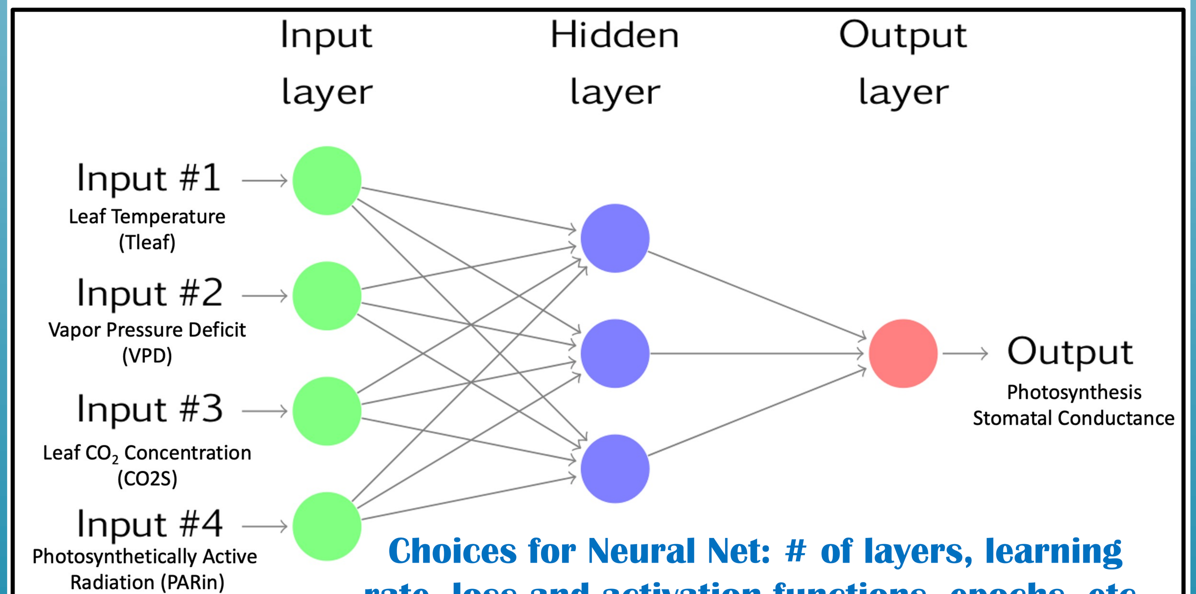

Fig. 4. Schematic of Neural Network implemented in this study, with the inputs shown on the left, and the outputs on the right. Different input layers and hidden layers were tested to optimize the prediction skill of the Neural Net.

Results: Photosynthesis

We simulated photosynthesis using empirical formulations from CLM5 (Fig. 5) via the MAAT toolbox, c.f. Walker et al., 2018. Changing the Vc,max parameter and empirical function do not allow a proper fit with the observed dataset (e.g., for high leaf temperatures). Then, we tested a Neural Network that was trained on the Lin 2015 data, and we found a much stronger fit.

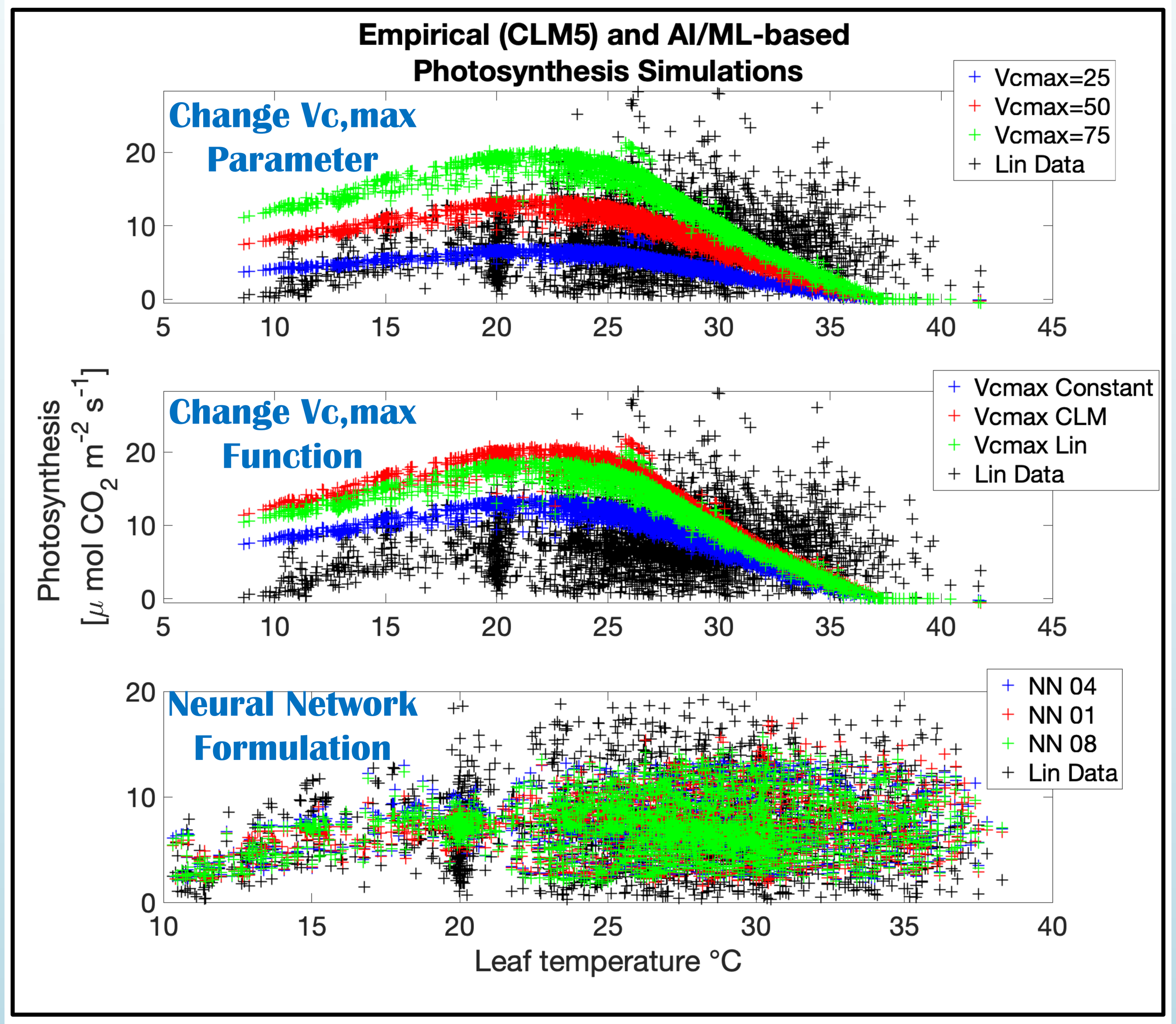

Fig. 5. Empirical (CLM5) and ML-based (Neural Network) models to simulate photosynthesis, compared to observations from Lin 2015, for Broadleaf Deciduous Temperate Trees. Changes to the Vc,max parameter (top) or the Vc,max function (center) do not allow a fit as good as the Neural Net (bottom).

Results: Stomatal Conductance

We tested commonly used empirical formulations for stomatal conductance (Ball-Berry/Medlyn) and ML-based models (Random Forests) using the same inputs as the empirical methods as well as additional inputs from dataset (Fig. 6).

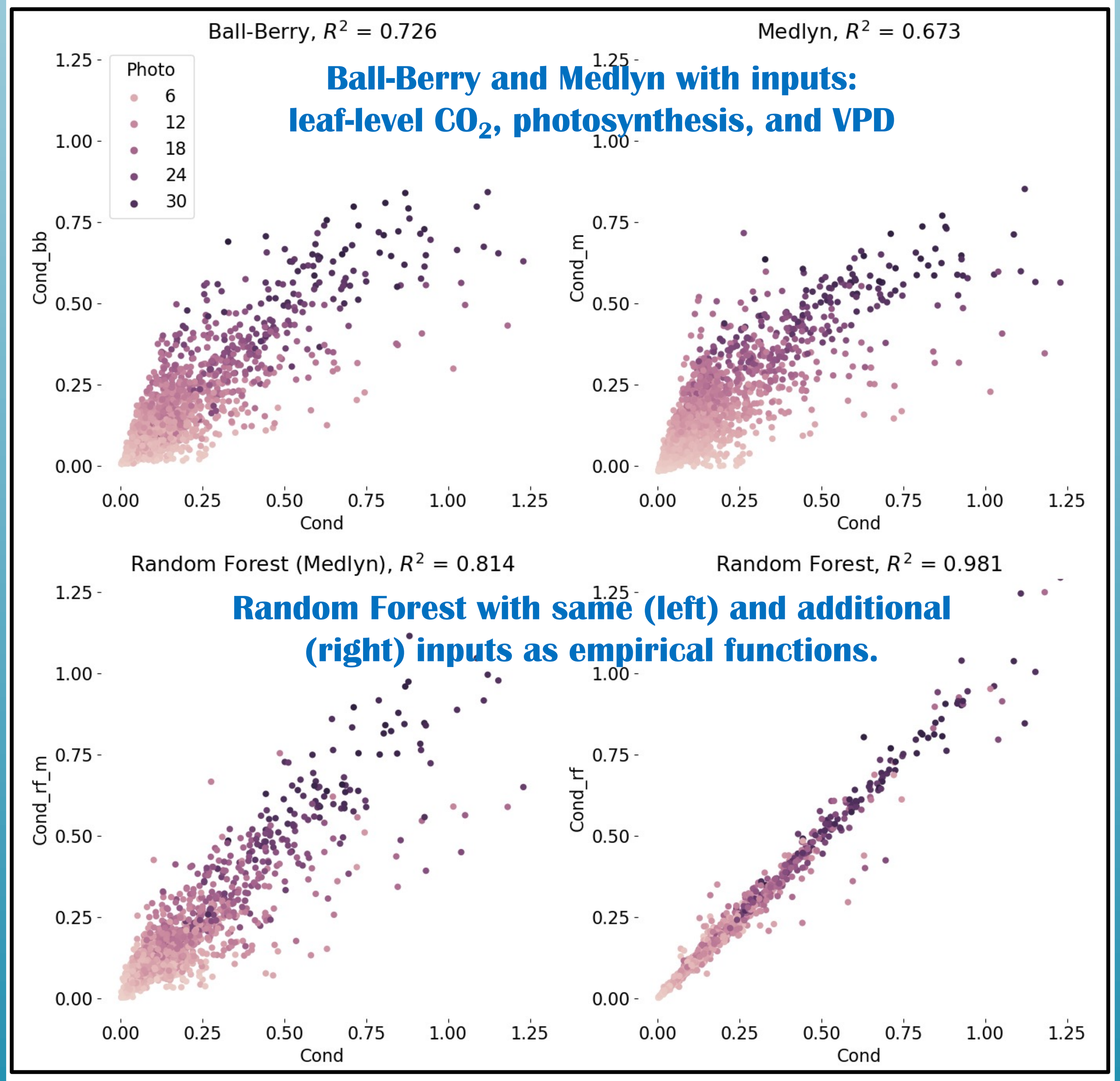

Fig. 6. Ball-Berry (top left), Medlyn (top right), and Random Forests using the same inputs (bottom left) and additional inputs (bottom right), compared to observations from Lin 2015. Inputs: leaf-level CO_2, photosynthesis, and VPD.

(Dis)Continuous Response Surfaces

Most process-based or empirical formulations have continuous response surfaces and therefore are differentiable. However, ML-based models may exhibit discontinuities in their surfaces, due to things such as gaps in the data representation.

We can see in the figure below (Fig. 7) that discontinuities exist in certain regions of the Random Forest heat map of stomatal conductance. For Medlyn, the surface is smoother. Simulations here use VPD and photosynthesis as inputs to the models.

The discontinuities shown for the Random Forest model are likely to have numerical consequences when attempting to couple and integrate ML-based formulations into larger ESMs.

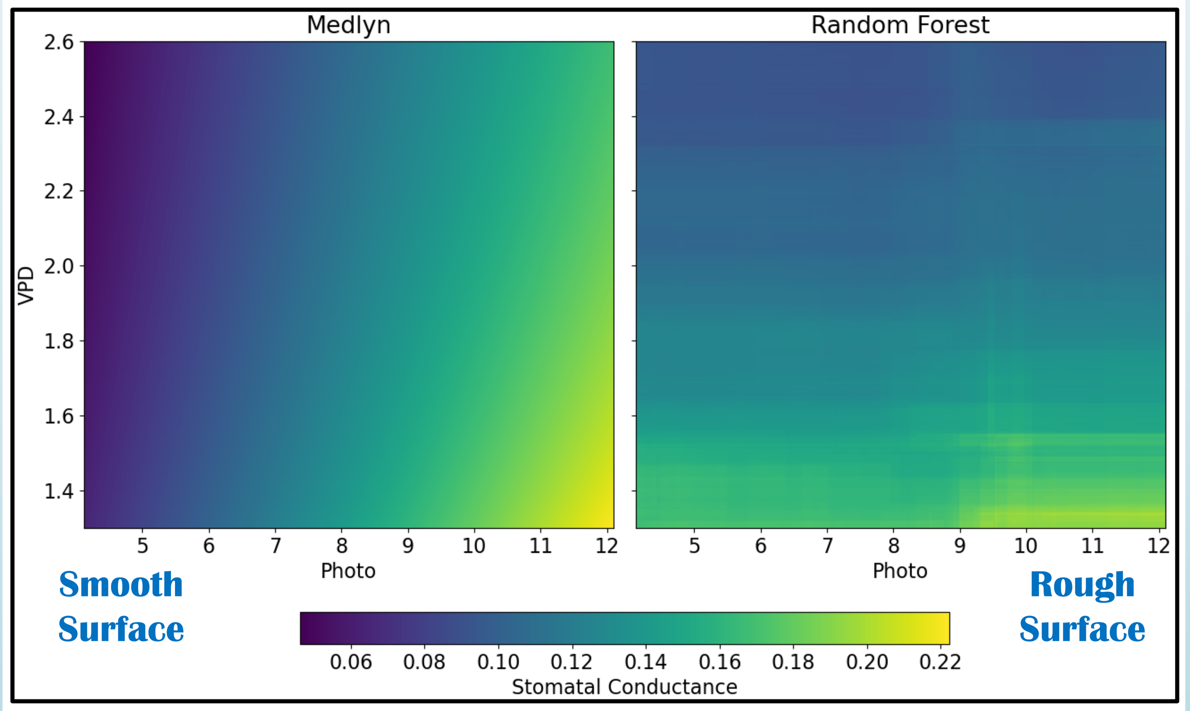

Fig. 7. Multi-dimensional response surface of the Random Forest heat map of stomatal conductance, using VPD and photosynthesis as inputs. The Medlyn surface (left) is smoother than the one generated by the Random Forest (right).

Conclusion

The evolution of ML-informed ESMs and ecohydrologic models will require a community effort, involving multiple disciplines, advanced training, and new ways of designing, implementing, parameterizing and communicating model outputs.

In this work, we developed initial ML-based prototypes for simulating photosynthesis and stomatal conductance.

Preliminary results from our investigations showed improved performance for photosynthesis using Neural Networks compared to an empirical (CLM5) formulation. We found the same benefits for stomatal conductance using Random Forests compared to Medlyn & Ball-Berry empirical methods.

Future work utilizing ML-based methods in ecohydrology will require additional and richer data to properly train the models.

The code to implement these approaches is hosted on GitHub (https://github.com/rubisco-sfa/MLEcohydrology), where future investigations and developments will also be hosted.

Citations

- Anderegg, W. R. L., A. Wolf, A. Arango-Velez, B. Choat, D. J. Chmura, S. Jansen, T. Kolb, S. Li, F. C. Meinzer, P. Pita, V. Resco de Dios, J. S. Sperry, B. T. Wolfe, and S. Pacala (2018), Woody plants optimise stomatal behaviour relative to hydraulic risk, *Ecol. Lett.*, 21:968–977, doi:10.1111/ele.12962.
- Fisher, R.A. and Koven, C.D., 2020. Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. Journal of Advances in Modeling Earth Systems, 12(4), p.e2018MS001453. https://doi.org/10.1029/2018MS001453.
- Han, J., C. Y.-Y. Chang, L. Gu, Y. Zhang, E. W. Meeker, T. S. Magney, A. P. Walker, J. Wen, O. Kira, S. McNaull, and Y. Sun (2022), The physiological basis for estimating photosynthesis from Chla fluorescence, *New Phytol.*, 234(4):1206–1219, doi:10.1111/nph.18045.
- Lin, Y.-S., B. E. Medlyn, R. A. Duursma, I. C. Prentice, H. Wang, et al. (2015), Optimal stomatal behaviour around the world, *Nat. Clim. Change*, 5(5):459–464, doi:10.1038/nclimate2550.
- Massoud, E., F. Hoffman, Z. Shi, N. Collier, C. Koven, J. Mao, et al. (2023), Perspectives on artificial intelligence for predictions in ecohydrology, *Artificial Intelligence for the Earth Systems*, in review.
- Walker, A. P., Ye, M., Lu, D., De Kauwe, M. G., Gu, L., Medlyn, B. E., Rogers, A., and Serbin, S. P.: The multi-assumption architecture and testbed (MAAT v1.0): R code for generating ensembles with dynamic model structure and analysis of epistemic uncertainty from multiple sources, Geosci. Model Dev., 11, 3159–3185, https://doi.org/10.5194/gmd-11-3159-2018, 2018.