GC21M-1096:
Is Parameter Inference a Disappearing Practice? Comparing Photosynthesis Simulations Using Perturbed Parameter Ensembles and Machine Learning

Elias C. Massoud[1], Forrest M. Hoffman[1], Nathan Collier[1], Anthony Walker[2], Bharat Sharma[2]
1 Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
2 Environmental Sciences Division, , Oak Ridge National Laboratory, Oak Ridge, TN, USA

American Geophysical Union
AGU Annual Meeting 2023
December 11-15, San Francisco, CA, USA

Introduction

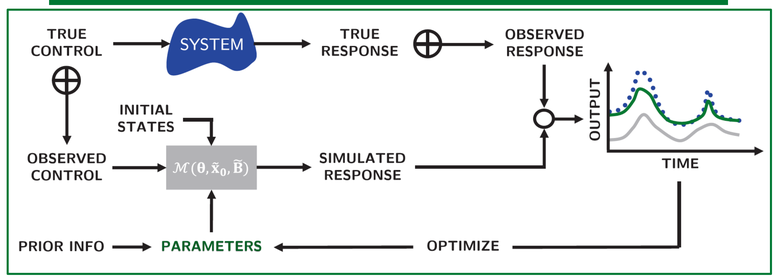

Fig. 1. Schematic overview of the parameter estimation problem (model calibration). The model parameters are adjusted iteratively so that the simulated response of the model (solid lines) approximates as closely as possible the observed response (blue dots) (Vrugt and Massoud 2018).

Parameter estimation (or model calibration) involves searching for parameter values that allow the model to mimic the observed system response. The calibrated model (Fig. 1) can then be used for simulation and prediction of different system behaviors. Simulating the original model can be very expensive, so the goal with parameter estimation is to find the best parameter values as rapidly as possible using the smallest number of model simulations. ML-models require training, which can also be costly, but then require much less compute time to perform simulations that mimic the original model or to mimic the data that it is trained on. A trained ML-model can then be used in place of the original model to estimate parameter values. Or, the ML model can be used to simulate the system without reliance on the original model.

Data Curation & ML Training

ML-based methods require training from observed datasets (Massoud et al., 2023). 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 shown in Fig. 2 and 3.

Command to load data:

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

Data-driven ML-based models can be built to simulate leaf-level processes. The code to implement these approaches is on GitHub. Future developments will be hosted here: (https://github.com/rubisco-sfa/MLEcohydrology),

References

- Vrugt, J.A., and Massoud E.C. "Uncertainty quantification of complex system models: bayesian analysis." Handbook of Hydrometeorological Ensemble Forecasting (2019): 563-636. DOI: 10.1007/978-3-642-39925-1_27.
- Massoud, E., et al. (2023), Perspectives on artificial intelligence for predictions in ecohydrology, Artificial Intelligence for the Earth Systems. *Artif. Intell. Earth Syst.*, 2, e230005, doi.org/10.1175/AIES-D-23-0005.1.
- Lin, Y.-S., et al. (2015), Optimal stomatal behaviour around the world, Nat. Clim. Change, 5(5):459–464, doi:10.1038/nclimate2550.
- Anderegg, W. R. et al. (2018), Woody plants optimise stomatal behaviour relative to hydraulic risk, Ecol. Lett., 21:968–977, doi:10.1111/ele.12962.
- Han, J. et al.,(2022), The physiological basis for estimating photosynthesis from Chla fluorescence, New Phytol., 234(4):1206–1219, doi:10.1111/nph.18045.
- Walker, A. P., et al.: 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.

Stomatal Conductance

We tested empirical models and ML models for stomatal conductance (Fig. 2).

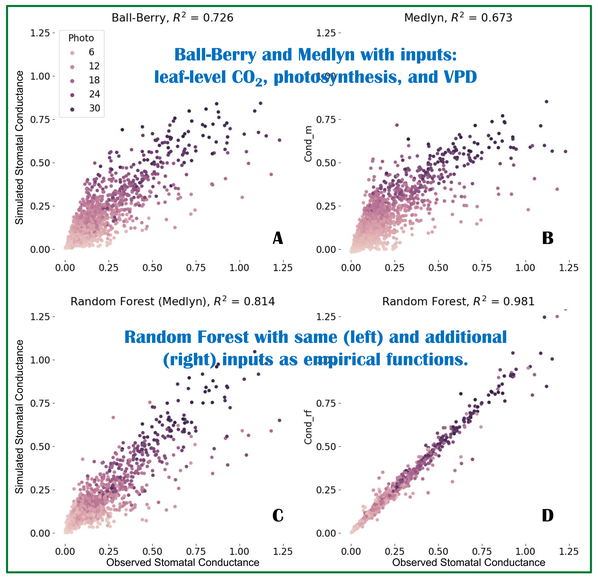

Fig. 2. Stomatal Conductance [mol m^{-2} s^{-1}] compared to data from Lin 2015 using (A) Ball-Berry, (B) Medlyn, (C) Random Forests with same inputs and (D) RF with additional inputs. Inputs to A-B-C: leaf-level CO_2, photosynthesis, and VPD. Additional inputs to D: transpiration, soil moisture index, PARin.

Photosynthesis

We tested photosynthesis simulations by changing Vc,max parameter/function in MAAT-CLM5 (Walker et al., 2018) and by using Neural Networks (Fig. 3).

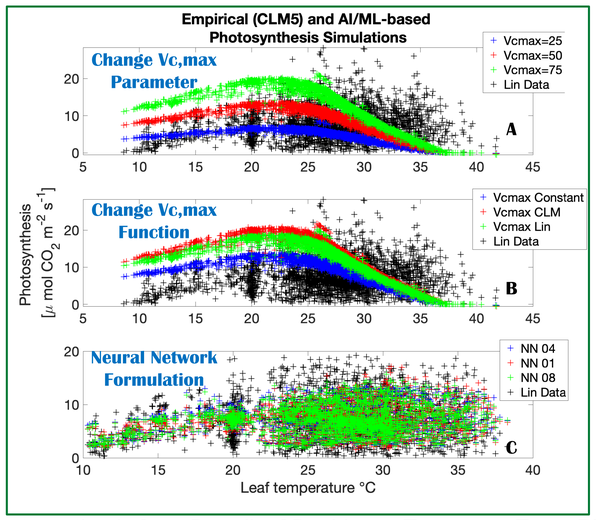

Fig. 3. Photosynthesis for a Broadleaf Deciduous Temperate Trees PFT using MAAT-CLM5 empirical formulation with (A) different Vc,max parameters and (B) functions and (C) using Neural Networks compared to data from Lin 2015.

Impact of ML on Solver Performance

Solving for photosynthesis requires an iterative solution to a non-linear problem. This is what is most costly in the simulations. Can we do better at "guessing" initial guess (x_0) using ML-based methods? (Fig. 4).

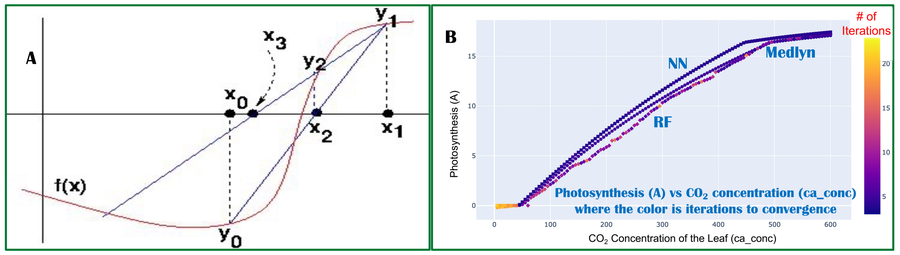

Fig. 4. (A) Schematic of iterative root-finding. (B) At low CO_2 concentrations (low photosynthesis) equations are harder to solve requiring more iterations. Random Forests (Diamonds) require more iterations to converge, Neural Networks (Squares) require less, Medlyn (Circles) is between RF and NN.

Improved Initial Guess (x0) to the Solver

We simulated E3SM-ELM-CNP at the Duke site for 158 years (I20TR). The photosynthesis solver is called millions of times (Fig. 5). We compared the total number of iterations in the solver using a simulation with original initial guess of x_0=0.7*Ca (baseline) and a run that uses a Neural Network-based initial guess (NN). Therefore, the physics of the original model is maintained.

Inputs:

```
"x0", # initial guess of the solution
"lmr_z", # leaf maintenance respiration rate (umol CO2/m**2/s)
"par_z", # par absorbed per unit lai for canopy layer (w/m**2)
"rh_can", # canopy air relative humidity
"gb_mol", # leaf boundary layer conductance (umol H2O/m**2/s)
"je", # electron transport rate (umol electrons/m**2/s)
"cair", # Atmospheric CO2 partial pressure (Pa)
```

Outputs:

```
"xf", # final value of the solution
```

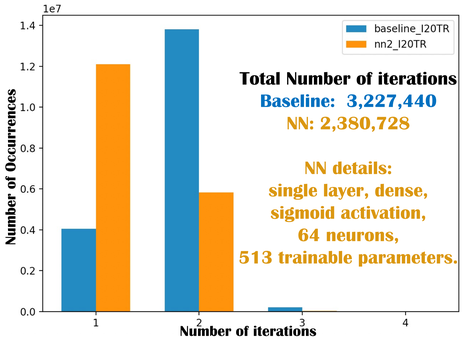

Fig. 5. Bar plots of the total number of iterations for all photosynthesis solves in the model simulation. Baseline represents the original model simulations and NN represents the simulations with the NN-based initial guess to the photosynthesis solver. Overall, NN requires less iterations.

Conclusions

Is parameter inference a disappearing practice? No. However, it is a changing practice. The use of ML-based models with ESMs will help advance parameter inference. We no longer must rely on costly ESM simulations to estimate parameter values. We can now infer parameters through data-driven machine learning approaches, by building surrogate models of original ESMs, or by trying novel methods like improving the initial guess to the photosynthesis solvers using NNs.