Process-based Neural Network to Forecast Vegetation Dynamics

Chonggang Xu, Youzuo Lin, Nishant Panda, Monty Vesselinov, Humberto Godinez Vazquez
Los Alamos National Laboratory

Focal Area 2

This white paper calls for an AI-based emulator of DOE's dynamic vegetation model that is computationally efficient and accommodates/parameterizes the governing processes built in the model. This AI-based emulator is needed for model calibration, sensitivity analysis, prediction, uncertainty quantification, and decision making. This AI-based emulator will incorporate physics and biological information in the form of conservation laws and constitutive relationships.

Science Challenge

DOE's demographic vegetation model, the Functionally Assembled Terrestrial Simulator (FATES), allows comparison with many more observable vegetation processes than the first generation 'big leaf' vegetation models, but also faces two key challenges. First, FATES contains more degrees of freedom leading to greater complexity and more uncertainties in the parameter estimations. Second, because more processes (e.g., vegetation recruitment, growth and mortality) are introduced, FATES is computationally much more expensive. These two challenges make it difficult to calibrate the model against observations using traditional approaches for parameter estimation (e.g., maximum likelihood or Markov Chain Monte-Carlo). One solution is to build emulators to efficiently mimic these demographic vegetation models; however, due to the larger number of parameters and model complexity, it is only feasible to build emulators with a constrained and limited number of key parameters based on traditional kernel methods [1].

Rationale

Earth System models (ESMs) are evolved from the general circulation models (GCMs), which mainly focus on the physical process of ocean and atmospheric circulations, with a new focus on the climate feedbacks from biological systems[2]. An important biological component is vegetation, which plays a critical role in global and regional water cycles[3]. Specifically, the reduction in plant productivity and increasing vegetation mortality resulting from extreme climate conditions could substantially affect the regional and global water cycles [3,4].

The first-generation vegetation models represent plant communities by area-averaged leaf layers of different plant functional types (PFTs) within each land grid cell. However, these big-leaf vegetation models do not represent the coexistence of different sizes of plants or PFTs, meaning that these models cannot express the competition for light, water, and nutrients among plants. To overcome these limitations, scientists have incorporated ecosystem demographic models into ESMs[5,6]. These new models include an ecosystem's demographics by explicitly simulating plant size, diameter growth, mortality, and recruitment based on competition for light, nutrients, and water[5]. Because these demographic vegetation models explicitly represent the coexistence and competition among different size groups of PFTs, they are expected to better represent changes in regional and global water cycles associated with disturbances.

Specifically, DOE's FATES is considered a next-generation vegetation model for E3SM, with a size-structured group of plants (cohorts) and successional trajectory-based patches using the ecosystem demography approach[7]. Within FATES, the transpiration rate is mainly determined by leaf-level photosynthesis, competition as determined by growth and mortality, and hydraulic function changes under water stress. FATES allocates photosynthetic carbon to storage and different tissues, such as leaf, root, and stem based on the allometry of different plant species. Mortality within FATES is simulated by carbon starvation caused by depletion of carbon storage

DOI: https://doi.org/10.2172/1769768
DOE BER, AI4ESP 2021 #AI4ESP1147 (https://www.ai4esp.org)

and hydraulic function failure caused by embolism via a hydrodynamic model [8]. Comprehensive sensitivity analysis has been conducted to understand the relative importance of a large number of parameters (>80) [9].

The next generation demographic vegetation models provide great opportunities to advance our understanding of vegetation feedback to climate; however, we are faced with a new challenge. These models are much more complex and computationally expensive. Thus, it is really difficult to calibrate the model with heterogeneous sources of data. So far, scientists have been using two approaches: 1) empirical tuning based on expert knowledge of the model; and 2) traditional statistical estimation based on emulators. While these two approaches are useful for current applications, they are subjected to key limitations. The empirical tuning is not efficient as it might be only targeted for one data source. The emulator has been mainly focused on the relationship between model inputs and outputs. Due to the larger number of parameters and complexity of models, it is only feasible to build emulators with a constrained and limited number of key parameters based on traditional kernel methods[1]. The reason why traditional machine learning (ML) approaches are not able to build efficient emulators is because these approaches typically treat the model as a black box and do not consider the causality structure within the model. Therefore, it is difficult to map a large number of parameters to the model output spaces.

The recent advancement of ML, especially the Long Short-Term Memory (LSTM) [10,11], deep Convolutional Autoencoders (CA)[12], and Recursive Neural Networks (RNN) [13], provides a great potential to build alternative predictive models. However, because these models are purely data driven without considering the physical and biological processes, they cannot give us good scientific reasoning and might fail when applied to out-of-distribution data. In summary, we have no efficient process-based modeling capability to forecast vegetation dynamics that can be easily calibrated against heterogeneous sources of data.

Narrative

We call for a *process-based dynamic vegetation recursive neural network (FATES-RNN)* learned from the biological and physical knowledge within DOE-sponsored Functionally Assembled Terrestrial Simulator (FATES). If successful, this neural network will transform our capability of vegetation modeling to efficiently fuse heterogeneous data and process-based understanding. FATES-RNN will be built to mimic FATES process structure. Specifically, we aim to first divide the model into subcomponents with corresponding climate forcing (D(t)), inputs (Y_{0i} (t)), parameters ($X_{i,j}$) and outputs ($Y_{i,j}$), depending on the model structure. For example, the photosynthetic component of the model will include 1) climate forcing of temperature and prediction; 2) inputs of soil moisture; 3) parameters of the maximum carboxylation rate at 25 ^{o}C ($V_{c,max25}$) and stomata slope; and 4) outputs of the gross primary production (GPP) and net primary production (NPP). The carbon allocation component will use predicted GPP combined with plant allometry parameters to determine plant stem diameter growth. This iteration can be repeated until the last layer of model output, $Y_{o.}$ In the case of FATES, Y_o could include the total ecosystem biomass. The initial condition Y_0 is a subset of model outputs, which are updated at each time step. For each recurrent cell, there are two types of layers: 1) LSTM layers, which has been widely used in Earth science and has the advantage to learn long-term temporal dependencies[14]; 2) dense layers. LSTM layers will be used for outputs that are dependent on previous time steps (e.g., the structure wood accumulation) and dense layers will be used to learn structures that are not time-dependent (only dependent on the parameters and the current climate forcing). The complexity of the

networks (such as the depth and width) are hyper-parameters that need to be tuned for a better performance.

The key advantage of using FATES process structure to determine the recurrent layers is that when FATES-RNN is learned from FATES simulations, it will learn the intrinsic physical and biological knowledge embedded through simulations. In the model learning process, the loss function will be determined by comprehensive model outputs that include many intermediate variables not typically part of the model outputs. Thus, the model output routine needs to be modified to make sure all the necessary intermediate variables will be exported for learning. A second advantage of FATES-RNN is that supplemental training samples can be generated for target subcomponents from target subroutines and thus, a smaller number of training samples is needed to capture the relationships between model parameters and outputs. A third advantage is that we can use FATES-RNN to accurately estimate the derivative of different model outputs to model parameters, thus allowing efficient model calibration against multiple sources of data. Finally, FATES-RNN is able to capture all the causal relationships and is much better for scientific hypothesis testing related to vegetation changes that is linked to regional and global water cycles.

FATES is coupled to the land component of DOE's Exascale Energy Earth System Model (ELM) through a common interface of water, carbon and nutrients. To make FATES-RNN applicable across different climate conditions and vegetation types, FATES-RNN needs to be trained against broad climate and soil conditions from arctic to tropics, parameter values ranges from global observations (e.g., TRY database[15]), and future climate projections. In the end, FATES-RNN can be coupled to ELM to predict vegetation dynamics under current and future climate conditions. The model will be calibrated against portions of observations of vegetation status and validated against other portions that are not used for calibration.

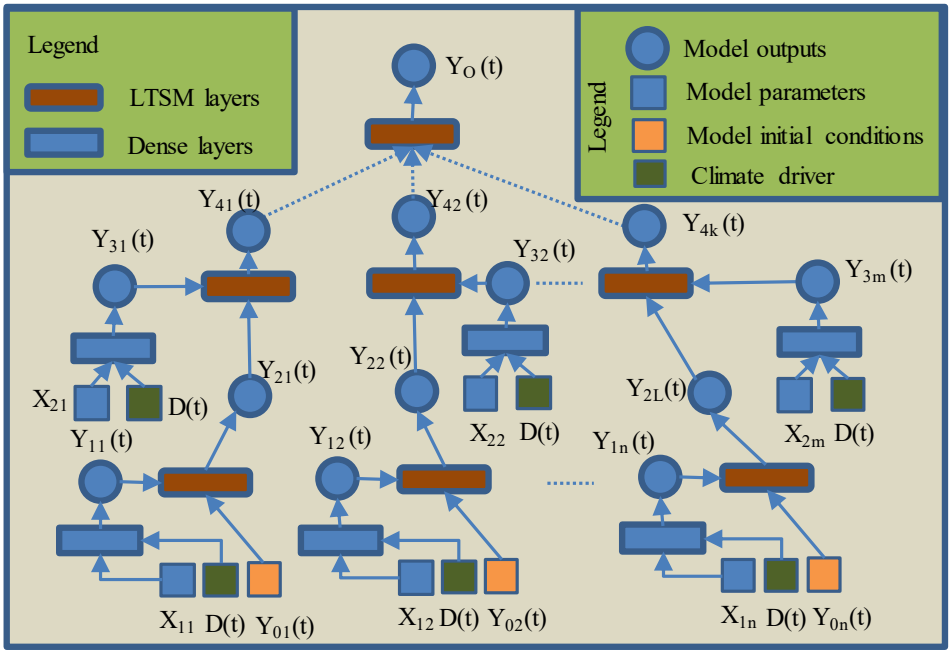

Figure 1: Framework of one recurrent cell that explicitly considers relevant parameters (x_{ij}) for different outputs of model (Y_{ij}).

In summary, this white paper calls for 1) a computationally efficient process-based neural network dynamic vegetation model that accommodates the physical and biological processes within FATES; and 2) A computationally efficient model calibration approach using this process-based neural network emulator using heterogeneous data sources.

References

1. Massoud, E. C. Emulation of environmental models using polynomial chaos expansion. *Environmental Modelling & Software* **111**, 421–431 (2019).
2. Flato, G. M. Earth system models: an overview: Earth system models. *Wiley Interdiscip. Rev. Clim. Change* **2**, 783–800 (2011).
3. Lemordant, L., Gentine, P., Swann, A. S., Cook, B. I. & Scheff, J. Critical impact of vegetation physiology on the continental hydrologic cycle in response to increasing CO2. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 4093–4098 (2018).
4. Xu, C. *et al.* Increasing impacts of extreme droughts on vegetation productivity under climate change. *Nat. Clim. Chang.* **9**, 948–953 (2019).
5. Fisher, R. A. *et al.* Vegetation demographics in Earth System Models: A review of progress and priorities. *Glob. Chang. Biol.* **24**, 35–54 (2018).
6. Moorcroft, P. R., Hurtt, G. C. & Pacala, S. W. A method for scaling vegetation dynamics: The ecosystem demography model (ed). *Ecol. Monogr.* **71**, 557–586 (2001).
7. Fisher, R. A. *et al.* Taking off the training wheels: the properties of a dynamic vegetation model without climate envelopes, CLM4.5(ED). *Geosci. Model Dev.* **8**, 3593–3619 (2015).
8. Christoffersen, B. O. *et al.* Linking hydraulic traits to tropical forest function in a size-structured and trait-driven model (TFS v.1-Hydro). *Geosci. Model Dev.* **9**, 4227–4255 (2016).
9. Massoud, E. C. *et al.* Identification of key parameters controlling demographically structured vegetation dynamics in a land surface model: CLM4.5(FATES). *Geoscientific Model Development* vol. 12 4133–4164 (2019).
10. Kratzert, F., Klotz, D., Brenner, C., Schulz, K. & Herrnegger, M. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **22**, 6005–6022

(2018).

11. Fang, K. & Shen, C. Near-Real-Time Forecast of Satellite-Based Soil Moisture Using Long Short-Term Memory with an Adaptive Data Integration Kernel. *J. Hydrometeorol.* **21**, 399–413 (2020).
12. Lee, K. & Carlberg, K. T. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *J. Comput. Phys.* **404**, 108973 (2020).
13. Mishra, A. K. & Desai, V. R. Drought forecasting using feed-forward recursive neural network. *Ecol. Modell.* **198**, 127–138 (2006).
14. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
15. Fraser, L. H. TRY-A plant trait database of databases. *Global change biology* vol. 26 189–190 (2020).