Interpretable Deep Learning for the Earth System with Fractal Nets

Carolyn Begeman, Fluid Dynamics and Solid Mechanics, Los Alamos National Laboratory
Marian Anghel, Computer, Computational and Statistical Sciences, Los Alamos National Laboratory
Ishanu Chattopadhyay, Department of Medicine, University of Chicago

Focal Area 3: Explainable AI

Our confidence in the projections made by Earth System Models (ESMs) depends on understanding them to be, in some important respects, faithful representations of the Earth system. Here we present an "explainable Artificial Intelligence (AI)" method that allows us to uncover the dynamical structure of the observed and modeled Earth system, discover hidden links across wide spatiotemporal scales, target model development efforts at poorly-represented dynamics, and optimize observed or modeled data collection to maximize predictive information.

Science Challenge

Dynamical system science for the Earth system poses unique challenges given the large degree of internal climate variability. Thus, tools that help us understand how ESMs succeed and fail at representing these dynamics are crucial, particularly in relation to the observed system. Furthermore, the computational and memory constraints on ESM data output motivate *in situ* analysis of ESM dynamics, including automatic detection of dynamical shifts. Also, of key importance are procedures that leverage ESMs to optimize observational campaigns for improving process representation, reducing structural uncertainty and improving model skill.

Rationale

Classical approaches to the study of Earth system dynamics leverage correlation and linear response functions, limiting our mechanistic insights into the underlying dynamics. Recently deep learning approaches with neural networks (NN) have demonstrated the ability to characterize nonlinear relationships between many variables. However, key problems remain with NN inference, including a general lack of transparency and a limited ability to directly model stochastic phenomena. Furthermore, learning temporal relationships between spatially-separated variables (e.g., climate teleconnections) with a NN *in situ* requires much data communication - a limitation even for emerging HPC architectures. Thus, offline NN applications to ESM data output are most commonly employed for dynamical studies, but predictive information is lost as a result of temporal averaging of simulation output.

We propose a lightweight approach, archetypically distinct from NNs, to constructing a dynamical representation of the Earth system from data streams with any spatial distribution of any variable of interest. This approach is capable of representing nonlinear relationships between variables at any temporal scale and is robust even for noisy, chaotic systems. It can be deployed directly within ESMs at the model's spatial and temporal resolution because it is computationally efficient and minimizes data communication. Furthermore, the approach may be identically deployed in both observed and modeled domains, allowing both for direct dynamical comparisons as well as mutual information flows regarding the spatial locations, timescales, and variables that hold the most predictive power for a quantity of interest.

DOI: https://doi.org/10.2172/1769730
DOE BER, AI4ESP 2021 #AI4ESP1010 (https://www.ai4esp.org)

Narrative

To understand our approach, consider deep learning and the NN paradigm as conceptually distinct; we have shown that **it is possible to learn "deep" associations without invoking the ubiquitous NN strategy** of global optimization via backpropagation of the loss gradient[1]. Our key insight is tied to self-similar structures arising in ergodic stochastic processes which take values over a finite alphabet. If such a process has a well-defined set of dynamical states, self-similarity implies that future trajectories from each distinct visit of the same state are statistically indistinguishable. Uncovering hidden dynamical states is thus equivalent to recovering fractal structures in observed data. Such an approach is capable of learning complex behaviors with long-range persistence over space and time[2].

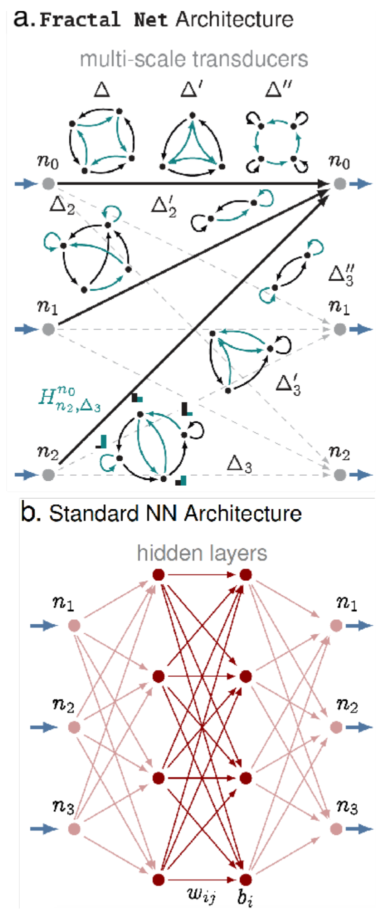

Figure 1. Architecture of (a) Fractal Nets (FN) vs. (b) Neural Networks (NN). n_x are the input and modeled data streams. For the FN in (a), models with Δ correspond to XPFSAs relating two n's at a time delay of Δ steps. Each arrow corresponds to one or more locally learned activation functions, one for each inferred time scale. Thus $H^{n_0}_{n_2,\Delta_3}$ denotes an inferred XPFSA predicting the effect of n_0 on n_2 at a time delay of Δ_3 steps. Note that the FN discovers connectivity, and is aware of the direction of influence, and explicitly models the time scale(s) over which influence transpires.

To uncover this emergent self-similar structure, we employ none of the NN features of fixed activation functions, user-specified loss functions, and global optimization via back-propagation. Instead, we distill local models, which are assembled into a predictive network which we call the Fractal Net (FN) (Fig. 1). Each elemental unit is a probabilistic finite state automaton (PFSA) or a PFSA for cross-dependencies (XPFSA). These models have a finite set of states; the number of states, the state transition map, and the output event probabilities are all inferred from data without prior constraints using efficient algorithms. FN's dynamical links are distinct from correlations due to their directionality, corresponding to the flow of predictive information. By quantifying the predictive influence of one variable on another, its directionality, and its timescales, this approach has unique interpretability lacking in other deep learning approaches (Fig. 2).

We demonstrate how this approach might be applied to answer an open scientific question for Earth system predictability: **How will the likelihood of high precipitation events (90th percentile) in the southwestern U.S. change in the next 30 years?** Informed by domain knowledge, we can use the following input time series to build a predictive FN for these events:

- Sea Surface Temperature (SST) at 10 regions in the Pacific Ocean
- Aerosol concentrations at 4 ARM sites in the continental U.S.
- Land Surface Temperature (LST) averaged over 20 regions in the U.S.
- Precipitation from 50 weather stations across the Western U.S.
- Area burned by wildfires, by state.

We can also construct a similar FN from simulated data. A direct comparison of dynamical linkages at multiple temporal scales can be made for the data streams that are reasonably analogous for the observed and modeled systems. The comparison between observed and modeled causal network might offer the following (fabricated) information:

- The dynamical link between Pacific SSTs and Southwest LSTs is weaker in the model than in observations at sub-seasonal timescales, suggesting an inadequate modeled teleconnection.
- There is a compensating too-strong model link between aerosol concentrations and precipitation, which had inflated our confidence in model skill for Southwest precipitation.

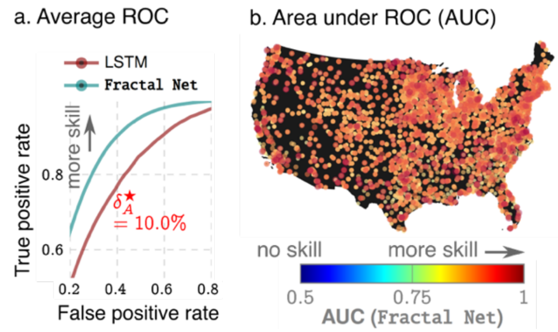

Figure 2. Seven-day predictive skill of Fractal Nets (FN) for extreme weather events (precipitation and cold/snow events). (a) Receiver Operator Curve (ROC). δ_A^* denotes the average increase in the Area Under the ROC (AUC) for FN over the Long Short-Term Memory (LSTM) recurrent Neural Network model. (b) AUC at 2000 airport-based weather stations for FN.

This kind of dynamical information can be used to guide model development efforts, as demonstrated in these bulleted examples. FNs may also reveal hidden dynamical links in the Earth system that motivate further study. The dynamical information that FNs offer is fundamentally connected with predictability on multiple timescales, information that is difficult to achieve in other ways, as we explained above. By analyzing a large number of data streams with this approach, there are several exciting applications:

(a) **Detect changes in ESM dynamics over time** by detecting changes in FN coefficients, indicating changes in the strength of dynamical links. This could reveal when certain feedbacks become important or when a threshold is crossed. When constructed *in situ*, the FN incorporates synoptic-scale variability that is lost with time-averaged output.

(b) **Optimize data output for predictive power**. We could write an algorithm to automatically output those data streams (a variable at a model grid cell) that have the highest influence coefficients in relation to a quantity of interest (e.g., 90th percentile precipitation in the Southwest). This data could be used to inform more intensive studies of the dynamics of these events or could be used as the training data for ML surrogates.

(c) **Optimize site choice for observations**. For example, an algorithm could start with a set of "seed" candidate ARM sites, compute their FN coefficients in relation to 90th percentile precipitation according to the ESM, keep the sites with coefficients in the 75th percentile, spawn new candidate sites with proximity to the top performers and repeat.

This dynamical systems approach is revolutionary both for its ease of deployment in requiring no prior physics knowledge, its expressiveness in comparison to current approaches, and its interpretability by identifying not only dynamical linkages but also their directionality and the timescales on which they operate.

References

1. Chattopadhyay, I. and Lipson, H. 2013. Abductive learning of quantized stochastic processes with probabilistic finite automata. Philos. Trans. R. Soc. A, Vol. 371(1984), pp 20110543.
2. Chattopadhyay, I., Huang, Y., Evans, J. 2020. Deep Learning Without Neural Networks: Fractal-nets for Rare Event Modeling. Preprint. doi:10.21203/rs.3.rs-86045/v1