Towards Trustworthy and Interpretable Deep Learning-assisted Ecohydrological Models

Authors

Peishi Jiang[1], Xingyuan Chen[1], Maruti K. Mudunuru[1], Praveen Kumar[2], Pin Shuai[1], Kyongho Son[1], Alexander Sun[3]
[1]Pacific Northwest National Laboratory, Richland, WA 99352, USA
[2]University of Illinois at Urbana-Champaign, Champaign, IL 61801, USA
[3]The University of Texas at Austin, Austin, TX 78712, USA

Focal Area(s)

Insights gleaned from complex data (both observed and simulated) using AI, big data analytics, and other advanced methods, including explainable AI and physics- or knowledge-guided AI.

Science Challenge

The transformational science question we plan to address is: *How do we leverage in-situ observations and simulations from process-based ecohydrological model to construct interpretable and trustworthy deep learning (DL) models for improved reliability of prediction of quantities of interest (QoIs) under hydro-climatic extremes?*

Rationale

Research needs or gaps: Watershed responses are driven by atmospheric forcing through land-atmosphere coupling as well as integrated surface and subsurface ecohydrological processes. Under global warming, they are leading to more extreme events[1] such as floods or droughts, as well as variability shift of high frequency events[2]. In order to understand how increase in hydro-climatic extremes or altered variability impact watershed dynamics, a variety of ecohydrological models can be employed to simulate the watershed responses as QoIs (e.g., runoff, nutrient loading)[3,4], with in-situ observations used for calibrating the models. Simulations from the calibrated process-based model provide new physical insights and further guide follow-up site sampling/measurement activities, leading to the Model-Experimental Coupling (ModEx) approach.

Nevertheless, the traditional ModEx approach using process-based models requires numerous realizations for either calibration or uncertainty quantification, which would be computationally expensive, if not unaffordable, for a large-scale watershed simulation. Such computational demand further inhibits predictions to understand watershed dynamics under hydro-climatic extremes or disturbances. To address this problem, deep learning (DL)-based emulator can be an ideal alternative to process-based model for providing fast simulations[5], thereby speeding up the ModEx life cycle. However, using DL techniques in ModEx poses the following challenges:

1. The first challenge is associated with the **trustworthiness** of the DL-based emulator[6,7]. That is, we need an emulator that is able to capture the dynamical interdependencies of the corresponding process-based model. A trustworthy emulator can not only provide accurate predictions but also facilitate identifying unrepresented dynamics using observations in the ModEx life cycle. However, the traditional way of training DL models[8] usually does not explicitly account for the interactions between inputs and outputs, and thus the trustworthiness of the DL model is not guaranteed.
2. The second challenge is associated with the **interpretability** of the DL-based emulator[9]. That is, once the model is trained, we must interpret each individual prediction in order to understand how the input features (i.e., atmospheric forcing/model parameters) impact the predictions (i.e., simulated QoIs). The interpretation will help address issues such as to what extent a flooding event contributes to the downstream nutrient loading, thereby guiding the new sampling/measurement activities in the ModEx life cycle. However, the black box nature of DL models masks interpretable dependencies, and thus prevents such predictive understanding on simulated QoIs.

DOI: https://doi.org/10.2172/1769787
DOE BER, AI4ESP 2021 #AI4ESP1065 (https://www.ai4esp.org)

3. The third challenge is associated with the ability in **capturing and understanding extreme events** using the DL-based emulator. This challenge arises as a result of the contradiction between the non-stationary characteristic of the extreme events and the static nature of most DL models[8] (i.e., DL model is usually only faithful to the training data). Moreover, the uncertainty associated with observed forcing further hinders accurate predictions under hydro-climatic extremes using DL-based emulator[10]. How to address the extrapolation issue of DL models as well as account for the uncertainty of the observed forcings remains an open question.

The proposed approach to address the challenges: We **propose a DL-driven system** to systematically leverage simulations from an ecohydrological model as well as the in-situ observations (see Figure 1), aiming to generate *fast, trustworthy, and interpretable* simulations for QoIs. The system consists of two DL models: (1) an emulator to simulate QoIs given observed forcing and model parameters and (2) an inverse mapping to estimate model parameters from observed QoIs. Developing and using the system includes two steps: Step 1 – training the DL-driven system using ensemble simulations from the process-based model and Step 2 – employing the system to gain predictive understanding in the ModEx life cycle. In this two-step workflow, the three challenges are addressed in the following ways:

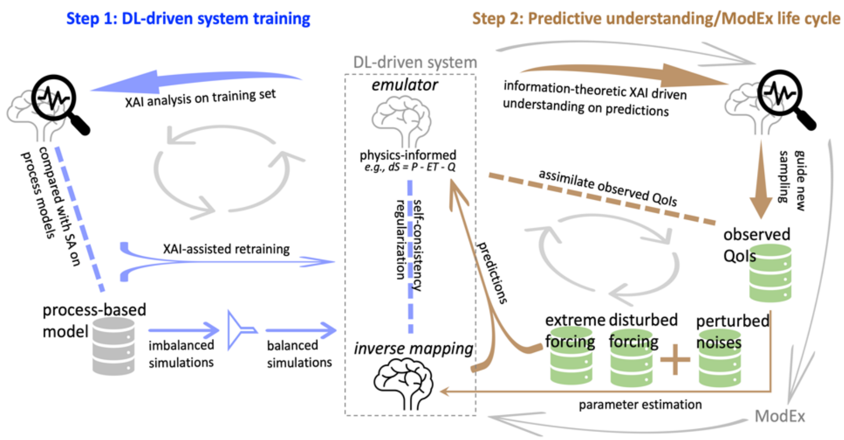

Figure 1 Workflow for DL-driven system training and predictive understanding under ModEx life cycle.

1A. Trustworthiness of the DL-based emulator will be enhanced during DL model training in Step 1 from two perspectives. First, domain knowledge will be incorporated in the DL-based emulator. This can be achieved either using physics-informed machine learning (PIML)[11] to constrain the emulator with physical laws or using a specific DL model (e.g., graph neural net[12]) to bring in better prior knowledge (e.g., river network) in the model inputs. Second, the discrepancy between the emulator and the process-based model regarding the input-output interactions will be reduced by evaluating the faithfulness of each prediction from the emulator and retraining the model to overcome those 'unfaithful' predictions. The faithfulness of an emulator prediction is defined as to what extent an emulator prediction is able to capture a process-based model prediction regarding the contribution from each input to each predicted output.

2A. Interpretability of the DL-based emulator will be obtained using explainable artificial intelligence (XAI)[9] to provide the local interpretability at each prediction instance in Step 2. The benefits of the XAI analysis are two folds. First, these XAI-driven insights are particular helpful to evaluate how the QoIs at different locations in the watershed responds to hydro-climatic extreme events such as drought or flood. Second, the XAI analysis can guide follow-up sampling/measurement activities, which then provides more observed QoIs, thereby forming a new ModEx life cycle.

3A. Prediction on noisy forcing data under extremes will be improved by leveraging the following efforts in both steps. First, in Step 1, a preprocessing[13] will be performed to increase the underrepresented extreme data in the imbalanced training data before training the DL models. Second, during the training, the use of PIML will facilitate the extrapolation ability of the emulator to extreme scenarios. Third, the uncertainty impact of the observed forcing data will be delineated using a perturbation-based XAI approach[9] by adding noises to the forcing data in Step 2.

Data needs: Both ensemble simulations and observations are needed to train, validate, and improve the DL-driven system. The simulations will be obtained from ensemble runs of ecohydrological process-based models (e.g., ATS or SWAT) by varying model forcing and parameters. The observations include in-situ data collected at gauge stations (e.g., USGS-NWIS), spatial surface/subsurface properties (e.g., NASA-SMAP and USDA-SSURGO), and atmospheric conditions (e.g., NASA-NLDAS).

The expected benefits: By developing this DL-driven system, we can have transformative advances in the ModEx life cycle. The justification of our approach is as follows: (1) the developed DL-based emulator, calibrated using a corresponding DL inverse mapping, is expected to generate fast and trustworthy emulations of QoIs; (2) the enhanced DL performance in extrapolation benefits the predictions under hydro-climatic extreme conditions; and (3) the interpretability of the DL emulator provides predictive understanding on the QoIs of the watershed system.

Narrative

Scientific and technical description of the opportunities: The proposed approach opens up unique opportunities in leveraging recent efforts in AI community and developing new tools in earth science to address the following technical barriers. The **first** barrier is the difficulty in ensuring the consistency between the DL-based emulator and inverse mapping during training in Step 1. The fact that training the emulator and mapping are usually independent can lead to the inconsistency between the forward mapping from parameters to outputs and the corresponding inverse mapping. As a result, recent advance in cyclic-consistency training[14] is promising in guaranteeing the self-consistency between the two models. The **second** barrier is the difficulty in formulating the performance metrics for DL model training. To ensure the self-consistency as well as the, the loss function includes the following additive metrics: (1) a general performance metric (e.g,. mean squared error); (2) a physics-constraints metric adopted from laws (e.g., the water mass balance and the water energy balance[15]); and (3) a self-consistency constraint metric. The **third** barrier is the difficulty in updating the DL-driven system based on the observed QoIs in the ModEx life cycle of Step 2. That is, how do we assimilate the observed QoIs to update the weights and biases used in the emulator and inverse mapping while still keeping the self-consistency between the two? Recent development in integrating DL with data assimilation[16] is promising in addressing this assimilation problem. The **fourth** barrier is the difficulty in unraveling multivariate interactions among atmospheric forcing and predicted QoIs using XAI in Step 2. While the traditional XAI approachs[17,18] interpret the input-output relation in a pairwise manner, a predictive understanding on a complex watershed system needs a more sophisticated approach that can unravel multivariate interactions (e.g., what are the most *k* responsive downstream outlets when storms happen at upstream for a period of time). To address this barrier, recent progresses of information theory (IT)[19,20] shows the promise of using IT in delineating nonlinear dynamics among multiple variables.

Activities that will advance the science: Oriented at providing interpretability and enhancing performance of DL-driven modeling in earth science, the proposed approach will advance both earth science and any AI-related communities. First, the DL-driven system enables a new ModEx life cycle. Indeed, the emulator can replace the physical modeling to provide efficient and accurate simulations for identifying new measurement locations, e.g., through a corresponding XAI analysis in spotting the 'hot' or sensitive locations regarding the responses of QoIs. Second, the novel XAI-assisted training will benefit the general AI community in enhancing DL learning by leveraging ecohydrologic model simulations. Third, the proposed model-agnostic information-theoretic XAI approach will be a new tool for opening up any DL 'black-box' by assessing the multivariate interactions in the model.

References

1. Myhre G, Alterskjær K, Stjern CW, et al. Frequency of extreme precipitation increases extensively with event rareness under global warming. *Sci Rep*. 2019;9(1):2-11. doi:10.1038/s41598-019-52277-4
2. Roque-Malo S, Kumar P. Patterns of change in high frequency precipitation variability over North America. *Sci Rep*. 2017;7(1):1-12. doi:10.1038/s41598-017-10827-8
3. Shuai P, Chen X, Song X, et al. Dam Operations and Subsurface Hydrogeology Control Dynamics of Hydrologic Exchange Flows in a Regulated River Reach. *Water Resour Res*. 2019;55(4):2593-2612. doi:10.1029/2018WR024193
4. Wang Y, Bian J, Zhao Y, Tang J, Jia Z. Assessment of future climate change impacts on nonpoint source pollution in snowmelt period for a cold area using SWAT. *Sci Rep*. 2018;8(1):1-13. doi:10.1038/s41598-018-20818-y
5. Hutson M. AI shortcuts speed up simulations by billions of times. *Science*. 2020;367(6479):728. doi:10.1126/science.367.6479.728
6. Floridi L. Establishing the rules for building trustworthy AI. *Nat Mach Intell*. 2019;1(June):261-262. doi:10.1038/s42256-019-0055-y
7. Thiebes S, Lins S, Sunyaev A. Trustworthy artificial intelligence. 2020.
8. Goodfellow I. Deep Learning.
9. Samek W, Montavon G, Vedaldi A, Hansen LK, Muller K-R. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Vol 11700.; 2019. doi:10.1007/978-3-030-28954-6
10. Loquercio A, Segu M, Scaramuzza D. for Uncertainty Estimation in Deep Learning. 2020:1-10.
11. Raissi M, Perdikaris P, Karniadakis GE. Physics informed deep learning (Part II): Data-driven discovery of nonlinear partial differential equations. *arXiv*. 2017;(Part I):1-22.
12. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *5th Int Conf Learn Represent ICLR 2017 - Conf Track Proc*. 2017:1-14.
13. Branco P, Ribeiro RP, Torgo L, Krawczyk B, Moniz N. SMOGN: a Pre-processing Approach for Imbalanced Regression. *Proc Mach Learn Res*. 2017;74:36-50.
14. Anirudh R, Thiagarajan JJ, Bremer PT, Spears BK. Improved surrogates in inertial confinement fusion with manifold and cycle consistencies. *Proc Natl Acad Sci U S A*. 2020;117(18):9741-9746. doi:10.1073/pnas.1916634117
15. Jia X, Willard J, Karpatne A, et al. Physics-Guided Machine Learning for Scientific Discovery: An Application in Simulating Lake Temperature Profiles. 2020:1-25. http://arxiv.org/abs/2001.11086.
16. Brajard J, Carrassi A, Bocquet M, et al. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations : a case study with the Lorenz 96 model.
17. Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One*. 2015;10(7):1-46. doi:10.1371/journal.pone.0130140
18. Ribeiro MT, Singh S, Guestrin C. “Why should i trust you?” Explaining the predictions of any classifier. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min*. 2016;13-17-Augu:1135-1144. doi:10.1145/2939672.2939778
19. Jiang P, Kumar P. Information transfer from causal history in complex system dynamics. *Phys Rev E*. 2019;99(1):1-13. doi:10.1103/PhysRevE.99.012306
20. Goodwell AE, Jiang P, Ruddell BL, Kumar P. Debates—Does Information Theory Provide a New Paradigm for Earth Science? Causality, Interaction, and Feedback. *Water Resour Res*. 2020;56(2):1-12. doi:10.1029/2019WR024940