Rethink hydrologic modeling framework with AI integrating multi-processes across scales

Authors

Eugene Yan, Emil Constantinescu, Rajkumar Kettimuthu, Kenneth Kemner, Cheng Wang, Romit Maulik, Park, Seongha Park (Argonne); Praveen Kumar and Maria L. Chu (UIUC).

Focal Area(s)

The proposed work is to develop a new hydrologic modeling framework that leverages vast amount earth and human system observational data, AI technologies, and data-driven information flows to improve predictability of hydrologic system that involves hydrologic, terrestrial, and biogeochemical processes and their interactions with the human and atmosphere systems.

Science Challenge

The predictability of the current earth system modeling is hampered by some critical scientific gaps, including the difficulty of capturing processes and subgrid-processes across scales, mismatch of data and model resolutions, inconsistency of system and subsystem complexities, and lack of coupling with the human system. A proposed novel hydrologic modeling framework will identify a set of AI technologies to construct hybrid models to translate across spatiotemporal scales and complexities and address resolution mismatches, incorporate data-driven causal inference and learning to explore interactions and feedbacks among processes, and develop coupling with the human system by leveraging large amount of earth and human system data. We expect that the new modeling framework will significantly improve the predictability of coupled hydrologic, terrestrial, and biogeochemical processes and outcomes.

Rationale

The current modeling of the earth system has several critical limitations that affect predictability. These include the difficulty of incorporating detailed terrestrial, hydrologic, biogeochemical, groundwater, and human processes due to process complexity and scale mismatches, resulting poor model capacity of simulating multiple dimensions of connectivity in terms of how water, sediment, nutrients, and biota move across critical zones [1-3], biogeochemical process from reactions at the pore scale to its global effects (e.g. microbially-driven methanogenesis that governs release of methane accounting for ~20% of human-induced radiative forcing), and missing effects of groundwater system. To integrate those processes, we need to include additional modeling components that could facilitate translations between spatiotemporal scales, identify interactions and/or feedbacks across different processes, and integrate subgrid-processes seamlessly into the existing earth system modeling. We identify a set of machine learning technologies and causal inference methods that leverage vast amount data to assist in integration of various processes across temporal and spatial scales.

Furthermore, human-induced dynamics (e.g., economic, social, political) with contrasting geographical teleconnections are commonly represented through land cover changes that rarely involve mechanistic descriptions of the human decision processes as well as their feedback interactions. These models are typically applied at coarse resolutions and ignore the influence of critical land management activities on natural processes and the micro-to-regional climate association with fine-resolution factors [4-9]. Besides, in the current earth system modeling, a wide range of other human activities affecting the response of the natural system are usually excluded from the representation of natural processes, resulting poor predictability because they aggregate to affect global processes [4]. With an explosive

DOI: https://doi.org/10.2172/1769773
DOE BER, AI4ESP 2021 #AI4ESP1093 (https://www.ai4esp.org)

growth of data from a plethora of sensors measuring states, fluxes, and time/space-integrated variables, including unconventional autonomous sensors from various sectors and citizen science observations, a data-driven causal network learning process can be leveraged to construct agent-based models to evaluate interactions of the hydrological-human system and predict effects of perturbation from human activities on the hydrologic system.

Narrative

We proposed three types of data-driven analysis tools: (1) machine learning (ML) techniques, (2) ML with causal inference, and (3) a hybrid of ML and agent-based models (Figure 1).

1. *ML methods to reduce model bias and facilitate translation of the feedbacks between processes across scales.*

Our approaches will focus on (1) ML techniques for predicting model bias and quantifying uncertainty, (2) ML model (or hybrid) to replace simulation of empirical or semi-empirical processes and an emulator to replace simulation of computational expansive sub-processes or dynamics with evolving states (e.g., vegetation model), and (3) ML algorithms for downscaling and upscaling data and translating feedbacks across scales and processes.

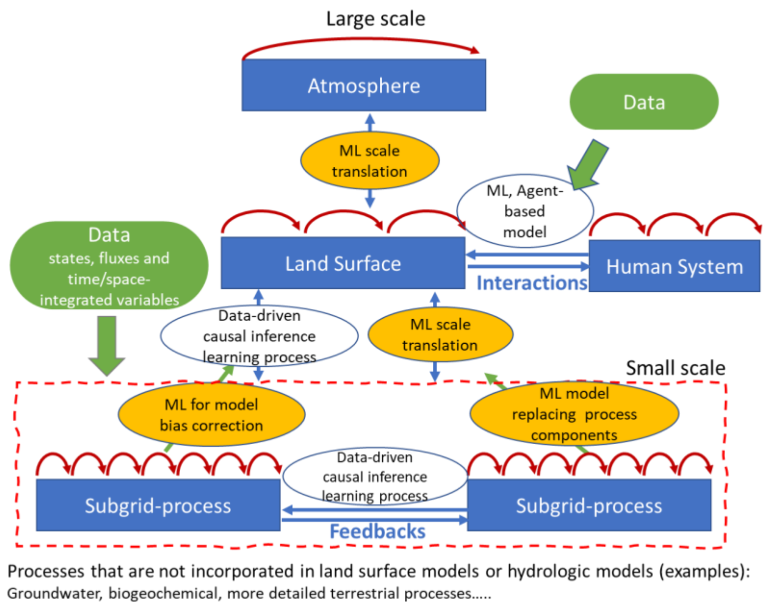

Figure 1. A new AI-blanded hydrologic modeling framework

An example of a data-driven model using random forest and support vector machine regression is used in a recent study [10] to evaluate model bias using observations (e.g soil moisture, baseflow) without parameterization (Figure 2). To improve the estimation of bias interval, the data driven model can incoporate the input for physics-based model, model predictions, model residual based on observations, and input data that cannot be readily taken into consideration in the physics-based model (e.g. dynamic landscape changes using satelite data, seasonal water extraction from unconventional sector data, and database from tagged aqutic species). The predicted bias intervals can be used to improve model output and the remaining errors can be quantified as model uncertainty. We can apply this method to predict model bias for any other subgrid-processes (e.g. surface hydrologic, terrestrial processes using soil moisture data to

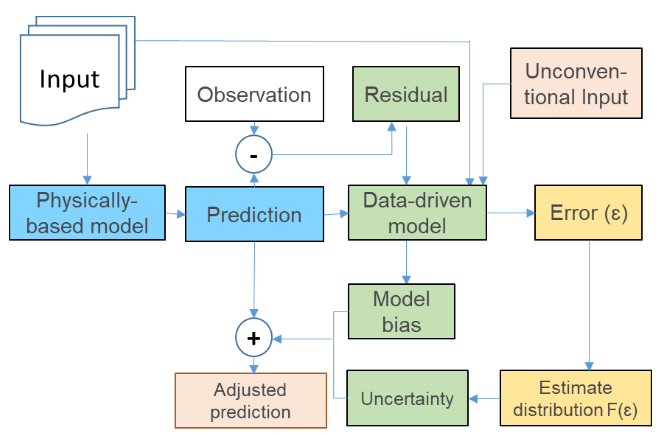

Figure 2. Framework of data-driven model to identify model bias and quantify remaining errors as model uncertainty

improve prediction of intermedia “state”). However, for subgrid-processes that are empirical in nature and when observations are available, several machine learning algorithms can be tested to improve the empirical models. Furthermore, an emulator can be developed to replace expansive computational subgrid-processes and dynamic processes with evolving states (e.g., dynamic vegetation model).

2. Data-driven causal inference learning process to identify feedbacks among processes

We propose to adopt causal complex network analysis to identify feedbacks among the process. In standard approaches, nodes are defined as the time series at different grid locations, links are typically based on correlations between the grid point time series, and the node degree quantifies the number of processes linked to a node. Causal network comparison metrics can be utilized for a causal evaluation of physical models. As an example, the PC algorithm [11] starts with a fully connected graph and test for the removal of a link between two variables iteratively to assess causal directions for contemporaneous links. Greedy equivalence search, on the other hand, starts with an empty graph and iteratively adds edges. The statistical criterion for removing or adding an edge can either be a conditional independence test or a properly defined score function that quantifies the likelihood of a particular graph structure given the data [12]. The causal inference has been recently improved via learning process with advanced machine learning approaches addresses a wide range of independence and dependence types [13]. The potential applications for the earth system modeling have been discussed by a co-author of the white paper [14] and others [15].

3. *Hybrid model of ML, information flow, and agent-based model for evaluating interactions of hydrological-human systems and predicting effects of perturbation from human activities on a hydrologic system*.

Agent based modeling (ABM) uses the concept of an ‘agent’ (a computational entity that represents a real-world actor) to simulate behaviors and interactions of decision-making entities, including feedbacks between human and environmental processes, physical constraints, and the different spatiotemporal scales in which these dynamics unfold. ABM has been used to build a number of socio-hydrological models in the past [16-23]. Though ABM is a promising approach, a key challenge in ABMs is in identifying the appropriate level of abstraction and defining rules to govern the agent’s behaviors in respons to changes in environmental system. The model must contain enough detail such that the key dynamics of the system can be analyzed. With the lack of formal process theories for human-hydrologic interactions, there is skepticism in models designed to integrate complexity in human decision making with complexity in interactions with hydrologic system. However, there is a fast growth of natural and human system data including remote sensing from a few metres to hundreds of kilometres above Earth, in situ observations (increasingly from autonomous sensors) at and below the surface, and unconvetional sector, social and econmic data which are further being complemented by citizen science observations. The recent advanced deep learning methods are capable to levaerage those data to identify causal interactions between human and hydrologic system across spatiotemporal scales, which can be used to develop hidden mechanisms that regulate the human behaviors and responses for an ABM model. Deep Reinforcement Learning (RL) architectures have been shown to achieve human level performance in complex tasks such as video games. It has also been used to study societal issues such as the Prisoner’s Dilemma. In recent work [24], researchers have combined RL and ABM to address the self-organizing dynamics of social segregation. We propose to combine RL and ABM to improve predictability of hyorologic modeling.

Rethink hydrologic modeling framework with AI integrating multi-processes across scales

Suggested Partners/Experts (Optional)

James Evans at University of Chicago

References (Optional)

[1} Goss, H., 2020, Next steps for the critical zone, *Eos, 101,* https://doi.org/10.1029/2020EO149358. Published on 24 September 2020.

[2] Kumar, P., E. Herndon, and D. D. Richter, 2020, Critical agents of change at Earth's surface, *Eos, 101,* https://doi.org/10.1029/2020EO149750. Published on 29 September 2020.

{3} Praveen Kumar, Phong V.V. Le, A.N. Thanos Papanicolaou, Bruce L. Rhoads, Alison M. Anders, Andrew Stumpf, Christopher G. Wilson, E. Arthur Bettis, Neal Blair, Adam S. Ward, Timothy Filley, Henry Lin, Laura Keefer, Donald A. Keefer, Yu-Feng Lin, Marian Muste, Todd V. Royer, Efi Foufoula-Georgiou, Patrick Belmont, 2018, Critical transition in critical zone of intensively managed landscapes, *Anthropocen*e, 22: 10-19.

[4] Robinson, D. T., Di Vittorio, A., Alexander, P., Arneth, A., Barton, C. M., Brown, D. G., Kettner, A., Lemmen, C., O'Neill, B. C., Janssen, M., Pugh, T. A. M., Rabin, S. S., Rounsevell, M., Syvitski, J. P., Ullah, I., and Verburg, P. H., 2018, Modelling feedbacks between human and natural processes in the land system, Earth Syst. Dynam., 9, 895–914, https://doi.org/10.5194/esd-9-895-2018.

[5] Robinson, D. T., Brown, D. G., and Currie, W. S., 2009, Modelling carbon storage in highly fragmented and human dominated landscapes: linking land-cover patterns and ecosystem models, Ecol. Model., 220, 1325–1338,

[6] Smith, B., Prentice, I. C., and Sykes, M. T., 2001, Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space, Global Ecol. Biogeogr., 10, 621–637.

[7] Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S., 2014, Implications of incorporating N ycling and N limitations on primary production in an individual-based dynamic vegetation model, Biogeosciences, 11, 2027–2054, https://doi.org/10.5194/bg-11-2027-2014.

[8] Lawrence, D. M., K. W. Oleson, M. G. Flanner, P. E. Thornton, S. C. Swenson, P. J. Lawrence, and S. G. Slater, 2011, Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. Journal of Advances in Modeling Earth Systems. 3(1):27.

[9] Coops, N. C. and Waring, R. H., 2001, The use of multiscale remote sensing imagery to derive regional estimates of forest growth capacity using 3-PGS, Remote Sens. Environ., 75, 324–334.

[10] Xu, T., & Valocchi, A. J., 2015, A Bayesian approach to improved calibration and prediction of groundwater models with structural error. *Water Resources Research*, *51*(11), 9290-9311. https://doi.org/10.1002/2015WR017912

[11] Spirtes, P. & Glymour, C., 1991, An algorithm for fast recovery of sparse causal graphs. *Soc. Sci. Comput. Rev.* **9**, 62–72.

[12] Runge, J., Bathiany, S., Bollt, E. *et al.*, 2019, Inferring causation from time series in Earth system sciences. *Nat Commun* **10,** 2553.

[13] Runge, J., 2018, Causal network reconstruction from time series: from theoretical assumptions to practical estimation. *Chaos Interdiscip. J. Nonlinear Sci.* **28**, 075310.

[14] Praveen Kumar, Hoshin V. Gupta, 2020, Debates—Does Information Theory Provide a New Paradigm for Earth Science?, Water Resources Research, 10.1029/2019WR026398, **56**, 2.

[15] Nowack, P., Runge, J., Eyring, V. *et al., 2020,* Causal networks for climate model evaluation and constrained projections. *Nat Commun* **11,** 1415. https://doi.org/10.1038/s41467-020-15195-y

[16] Walker, W.E.; Loucks, D.P.; Carr, G. Social responses to water management decisions. Environ. Process. 2015, 2, 485–509.

[17] Dziubanski, D.; Franz, K.J.; Gutowski, W. Linking economic and social factors to peak flows in an agricultural watershed using socio-hydrologic modeling. Hydrol. Earth Syst. Sci. Discuss. 2019, 1–55.

[18] Noël, P.H.; Cai, X. On the role of individuals in models of coupled human and natural systems: Lessons from a case study in the Republican River Basin. Environ. Model. Softw. 2017, 92, 1–16.

[19] Tesfatsion, L.; Rehmann, C.R.; Cardoso, D.S.; Jie, Y.; Gutowski, W.J. An agent-based platform for the study of watersheds as coupled natural and human systems. Environ. Model. Softw. 2017, 89, 40–60.

[20] Pouladi, P.; Afshar, A.; Afshar, M.H.; Molajou, A.; Farahmand, H. Agent-based socio-hydrological modeling for restoration of Urmia Lake: Application of theory of planned behavior. J. Hydrol. 2019, 576, 736–748.

[21] Pouladi, P.; Afshar, A.; Molajou, A.; Afshar, M.H. Socio-hydrological framework for investigating farmers' activities affecting the shrinkage of Urmia Lake; hybrid data mining and agent-based modelling. Hydrol. Sci. J. 2020, 65, 1249–1261.

[22] Aghaie, V.; Alizadeh, H.; Afshar, A. Agent-Based hydro-economic modelling for analysis of groundwater-based irrigation Water Market mechanisms. Agric. Water Manag. 2020, 234, 106140.

[23] Robinson, D. T., D. G. Brown, D. C. Parker, P. Schreinemachers, M. A. Janssen, M. Huigen, H. Wittmer, N. Gotts, P. Promburom, E. Irwin, T. Berger, F. Gatzweiler, and C. Barnaud. 2007. Comparison of empirical methods for building agent-based models of land and resource use. Journal of Land Use Science 2(1):31-55.

[24] Sert, E., Bar-Yam, Y. & Morales, A.J. Segregation dynamics with reinforcement learning and agent based modeling. Sci Rep 10, 11771 (2020). https://doi.org/10.1038/s41598-020-68447-8