Science and Technology Infusion Climate Bulletin
NOAA's National Weather Service
44th NOAA Annual Climate Diagnostics and Prediction Workshop
Durham, NC, 22-24 October 2019

Physically Informed Machine Learning for Hydrological Modeling Under Climate Nonstationarity

Grey S. Nearing[1, 2, 3], Craig S. Pelissier[3, 4, 5], Frederik Kratzert[6], Daniel Klotz[6], Hoshin V. Gupta[7], Jonathan M. Frame[2], and Alden K. Sampson[1]

[1]*Upstream Tech, Public Benefit Corporation*
[2]*Department of Geological Sciences, University of Alabama, Tuscaloosa, AL*
[3]*Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Catonsville, MD*
[4]*NASA Center for Climate Simulation, NASA Goddard Space Flight Center, Greenbelt, MD*
[5]*Science Systems and Applications Inc., Greenbelt, MD*
[6]*LIT AI Lab & Institute for Machine Learning, Johannes Kepler University, Linz, Austria*
[7]*Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ*

1. Introduction

There is an understanding in the hydrological sciences community that physical realism is necessary for providing hydrological forecasts under changing conditions (Blöschl *et al.*, 2019; Clark *et al.*, 2016; Milly *et al.*, 2008). At present, however, machine learning (ML) generally provides the best estimates of most hydrological states and fluxes, even in extrapolation (*e.g.*, Best *et al.*, 2015; Kratzert *et al.*, 2019a,b; Nearing *et al.*, 2018). A notable example of this was provided by Kratzert *et al.* (2019a), who showed that Long Short Term Memory networks (LSTMs) produce, on average, better predictions in basins that did not supply training data (effectively *ungaged* basins) than a conceptual model well-calibrated to gauge data in individual basins (gaged basins). This is significant in that the 2003-2012 decadal problem of the International Association of Hydrological Sciences (IAHS) was `*Prediction in Ungauged Basins*' (PUB) (Hrachowitz *et al.*, 2013). Prior to Kratzert *et al.* (2019a), best practices for PUB required extensive catchment-specific investment (Blöschl, 2016), which is infeasible at large scales (*e.g.*, regional, continental, global).

The purpose of this talk is to suggest ways that hydrological forecasting efforts could begin to bridge the gap between the reliability we typically associate with models based on physical understanding and the reality that data-driven models out-perform process-driven models. We give examples of two ways to approach this problem: (i) explainable ML and (ii) physics-informed ML.

2. Explainable ML under Climate Nonstationarity

The first project discussed in this talk is an example of constructing and deconstructing a deep learning model to gain better understanding of how it learns to organize information. We did this in the context of predicting streamflow in catchments under dynamic climate. We used LSTMs because they are conceptually similar to traditional dynamical systems models, in that they have a memory state that is updated in time through a set of input-state-output relationships.

Our LSTM took three types of inputs, which are described in more detail in Table 1 by Kratzert *et al.* (2019a): (1) daily meteorological forcings, (2) static catchment attributes related to soil, vegetation, geology, and topography parameters, and (3) dynamic climate statistics related to annual precipitation, potential evaporation, temperature, aridity, *etc*. Dynamic climate indexes were calculated on the 365 days just previous to each daily streamflow prediction. Static catchment attributes and climate statistics were input to an embedding network consisting of three fully connected layers with 35, 35, and 30 nodes, respectively. The 30 outputs from this embedding network at each timestep were concatenated with daily meteorological forcings and input to the LSTM.

Correspondence to: Grey S. Nearing, Department of Geological Sciences, University of Alabama, Tuscaloosa, AL;
E-mail: gsnearing@ua.edu

The resulting LSTM was trained using the NCAR CAMELS data set (Addor *et al.*, 2017; Newman *et al.*, 2015) with 15-year training (1981-1995) and test (1996-2010) periods. This LSTM was trained and tested to develop a single continental-scale model using all of the data from 447 catchments over CONUS. We used the same hyperparameters, training procedure, and benchmark models as Kratzert *et al.* (2019b), and additionally benchmarked against two LSTMs with static climate indexes: one using climate indexes calculated over the training period and one using climate indexes calculated over the test period.

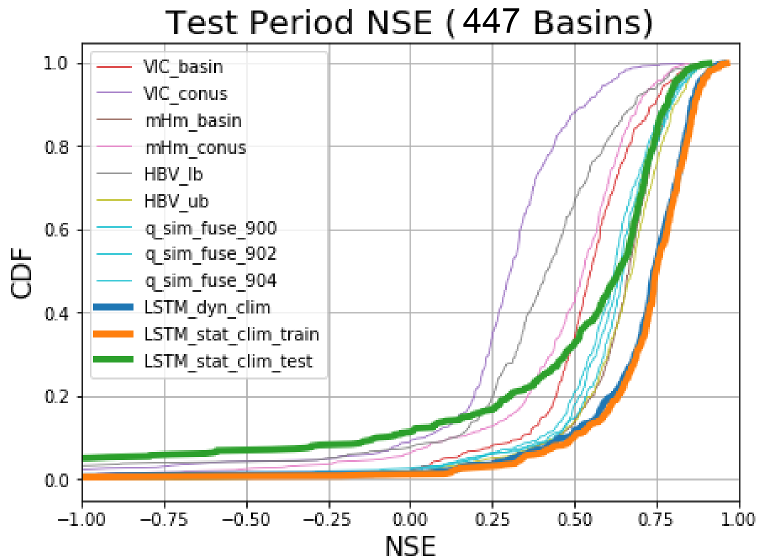

Fig. 1 NSE CDFs over 447 CAMELS basins of several hydrology models. The dynamic climate LSTM from Fig. 2 is in blue. All curves are for the test period.

Cumulative density functions of the Nash Sutcliffe Efficiencies (NSE) obtained for the 447 separate catchments over the test period are shown in Fig. 1. LSTMs performed better than all of the standard hydrology models, except for the LSTM that used static climate indexes calculated over the test period (green line). This demonstrates that we cannot simply change the climate attributes in a basin (*e.g.*, as climate in that basin evolves) if the model is trained assuming static climate in each basin.

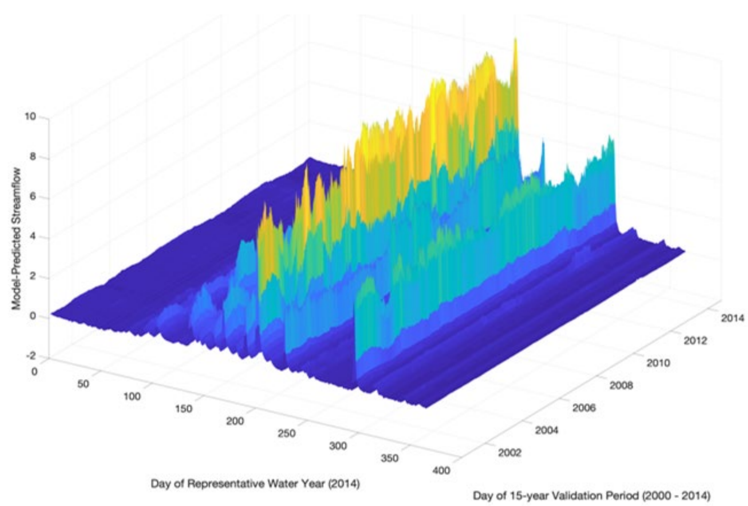

Fig. 2 Hydrographs in a single catchment due to putting the 2014 (representative) water-year forcings through a trained LSTM with different climate indexes. The only source of variability along the y-axis (Day of 15-year Validation Period) are the annual climate indexes. Notice that the peak flows in this catchment (around DOY 180) are highly responsive to background climate.

Figure 2 shows that the dynamic-climate LSTM gives very different hydrographs as climate changes. This figure was generated by putting the meteorological forcings from a single water year (Oct 2014 – Nov 2015) through the LSTM with the background climate from different years (2000–2014). The only thing that is variable in these runs is the climate index. Notice, for example, that the peak flow around day 180 of the water year changes drastically under different climates. Figures 1 and 2 taken together show that we must directly account (during model training) for catchment-specific variable climate.

The reason that the LSTM is able to extrapolate to different catchments and to different climates is that it effectively 'sees everything'. A single model trained on all available data from all catchments has a wide variety of training experience to draw from, and thereby extrapolate into, as the conditions in any given catchment change. Kratzert *et al.* (2019b) showed that the LSTM learned a representation of catchment similarity and used this to model rainfall-runoff behavior in similar catchments using shared parts of the network structure.

We can do the same thing under changing climate. Figure 3 is a still-frame from a movie that reduces the portion of the LSTM gate structure that reacts to the embedding network (which is variable only due to climate indexes) down to first principal components. Proximity in this space indicates catchment similarity, and as local climates change, individual catchments move in this space and become more or less hydro-climatically 'similar' to others.

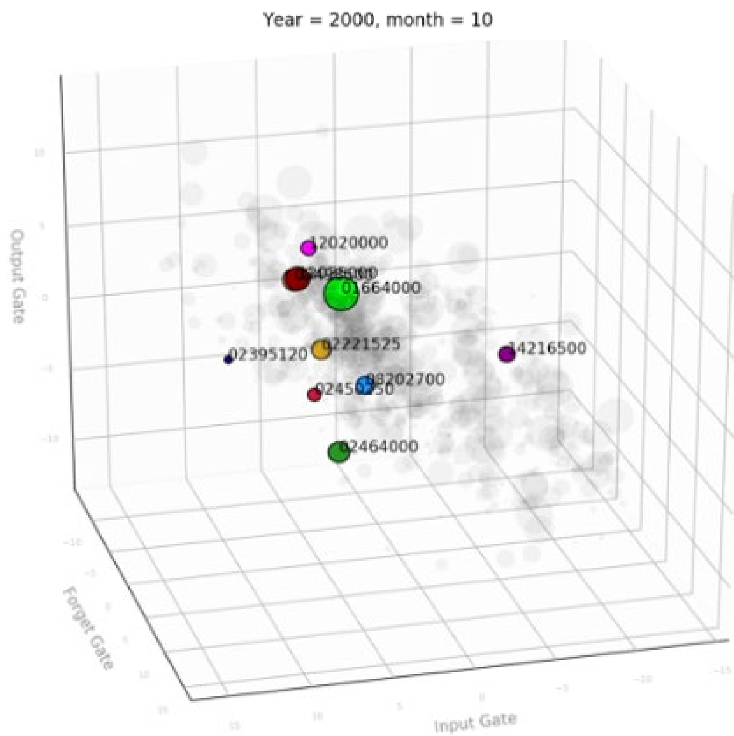

Fig. 3 A low-dimensional representation of catchment similarity as learned by the LSTM based on static catchment attributes and annual climate indexes. Ten of 447 CAMELS catchments are highlighted in color. Basins with close proximity in this space are treated as having similar rainfall-runoff processes by the LSTM. This is one time slice in the test period, and this learned representation of catchment similarity will evolve in time as local climate changes in each catchment. The LSTM is effectively interpolating in this space to make new predictions under dynamic conditions.

3. An example of combining process modeling with machine learning

There are several possible ways to combine physics with ML. This talk presents an example of integrating an ML kernel into a fully developed process-based model. We used Noah-MP, which is the land surface component of the NWS National Water Model.

A basic approach to integrating ML into a process model was described by Ghahramani and Roweis (1998), whereby the ML model is trained on the analysis states resulting from data assimilation (*e.g.*, by a Kalman-type filter). We can generalize that idea as follows. Suppose that we have a dynamical systems model that solves a set of PDEs:

$$\frac{dX}{dt} = f(X, U, \theta)$$

where X are model states, U are boundary conditions, θ are model parameters, and f(·) is the total divergence. A discrete-time solution is:

$$X_t = f^*(X_{t-1}, U_t, \theta).$$

We can then augment the state transition function f* (·) with a ML component as follows:

$$X_t = f^*(X_{t-1}, U_t, \theta) + g(X_{t-1}, U_t, \theta)$$

where $g(\cdot)$ is an ML model. The challenge is to train $g(\cdot)$ given that we don't have direct access to all of the system states. Nearing & Gupta (2015) used an ensemble Kalman filter to derive analysis states from a calibrated rainfall runoff model, and then trained a Gaussian process (Williams & Rasmussen, 2006) to use as $g(\cdot)$.

Here we applied a similar technique to the soil moisture state of Noah-MP using data from 11 FluxNet towers globally. We used only FluxNet sites and years of data having almost complete ½-hourly surface soil moisture data records, so that data assimilation was not necessary. Figure 4 shows the improvement to surface soil moisture estimates over the standalone Noah-MP model in individual data-years at each of the 10 FluxNet sites. These results were obtained using k-fold cross-validation, so that only out-of-sample data is shown in this figure. Prior to training and testing the ML integration, Noah-MP was calibrated against soil moisture data at each site so that the error being corrected here is model structural error.

Almost all out-of-sample simulations were improved as compared to a calibrated Noah-MP model, indicating potential to improve systematic structural errors in Noah-MP by integrating an ML component.

*Acknowledgements***.** Grey Nearing was supported in part by a UCAR COMET Cooperative Project between the University of Alabama and the National Water Center and in part by a grant from the NASA Terrestrial Hydrology Program. Frederik Kratzert was supported by a Google Faculty Research Award.

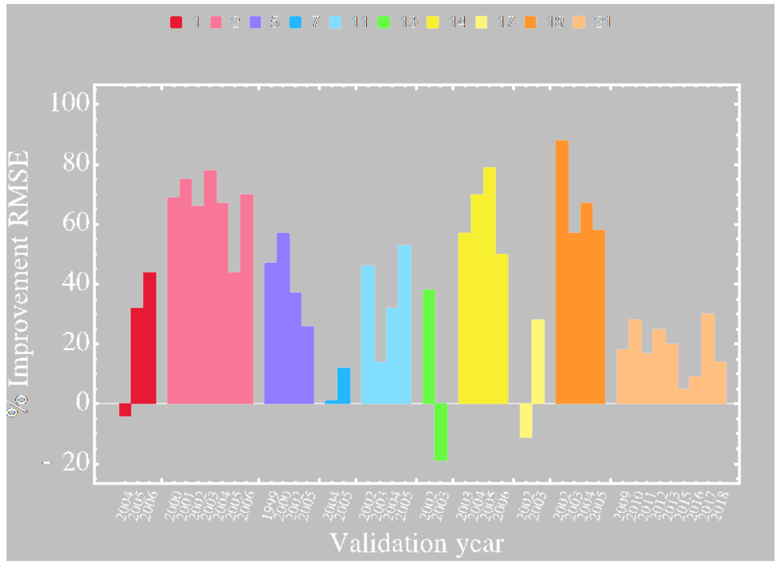

Fig. 4 Improvements to out-of-sample surface soil moisture at 10 FluxNet sites (site numbers are indexes into the sites used for the Protocol for Analysis of Land Surface Models; Abramowitz 2012)

References

Abramowitz, G., 2012: Towards a public, standardized, diagnostic benchmarking system for land surface models. *Geosci. Model Dev.*, **5** (3), 819-827.

Addor, N., A. J. Newman, N. Mizukami, and M. P. Clark, 2017: The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrol. Earth Syst. Sci. (HESS)*, **21** (10), 5293-5313.

Best, M. J., G. Abramowitz, H. R. Johnson, A. J. Pitman, G. Balsamo, A. Boone, and Coauthors, 2015: The Plumbing of Land Surface Models: Benchmarking Model Performance. *J. Hydrometeorol.*, **16** (3), 1425-1442.

Blöschl, G., 2016: Predictions in ungauged basins – where do we stand. *Proc. IAHS*, **373**, 57-60.

Blöschl, G., M. F. Bierkens, A. Chambel, C. Cudennec, G. Destouni, A. Fiori, and Coauthors, 2019: Twenty-three unsolved problems in hydrology (UPH)–a community perspective. *Hydrolog. Sci. J.*, **64** (10), 1141-1158.

Clark, M. P., R. L. Wilby, E. D. Gutmann, J. A. Vano, S. Gangopadhyay, A. W. Wood, and Coauthors, 2016: Characterizing uncertainty of the hydrologic impacts of climate change. *Current Climate Change Reports*, **2** (2), 55-64.

Ghahramani, Z., and S. T. Roweis, 1998: Learning nonlinear dynamical systems using an EM algorithm. Paper presented at the Advances in Neural Information Processing Systems Conference. [Available online at https://papers.nips.cc/paper/1594-learning-nonlinear-dynamical-systems-using-an-em-algorithm]

Hrachowitz, M., H. Savenije, G. Blöschl, J. McDonnell, M. Sivapalan, J. Pomeroy, and Coauthors, 2013: A decade of Predictions in Ungauged Basins (PUB) - A review. *Hydrolog. Sci. J.*, **58** (6), 1198-1255.

Kratzert, F., D. Klotz, M. Herrnegger, A. K. Sampson, S. Hochreiter, and G. S. Nearing, 2019a: Towards improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resour. Res.*, **55**, 11,344-11,354.

Kratzert, F., D. Klotz, G. Shalev, G. Klambauer, S. Hochreiter, and G. Nearing, 2019b: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrol. Earth Syst. Sci. (HESS)*, **23** (12), 5089-5110.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, 2008: Stationarity is dead: Whither water management? *Science*, **319** (5863), 573-574.

Nearing, G. S., and H. V. Gupta, 2015: The quantity and quality of information in hydrologic models. *Water Resour. Res.*, **51** (1), 524-538.

Nearing, G. S., B. R. Ruddell, M. P. Clark, B. Nijssen, and C. D. Peters-Lidard, 2018: Benchmarking and Process Diagnostics of Land Models. *J. Hydrometeorol.*, **19** (11), 1835-1852.

Newman, A., M. Clark, K. Sampson, A. Wood, L. Hay, A. Bock, and Coauthors, 2015: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrol. Earth Syst. Sci. (HESS)*, **19** (1), 209-223.

Williams, C. K., and C. E. Rasmussen, 2006: *Gaussian processes for machine learning.* Vol. 2. MIT press, Cambridge, MA.