Preferential flow in subsurface hydrology: From a century of denial to a decade of addressing it via ML?

Matthias Sprenger[1], Sheila M. Saia[2], James P. Dennedy Frank[1], Erica Woodburn[1],
[1]Earth & Environmental Sciences Area, Lawrence Berkeley National Laboratory
[2]Dept. of Biological and Agricultural Engineering, North Carolina State University

Focal Areas:

This white paper highlights the potential of using machine learning (ML) to gain new insights from large, high-frequency Earth science observatory network datasets (focal area 3) and better represent complex subsurface processes via ML in Earth System Models (ESMs, focal area 2).

Science Challenge:

This white paper outlines the potential for machine learning models to improve understanding and implementation of subsurface hydrologic processes in ESMs. There is currently limited understanding about where (e.g., soil texture, topography, land cover) and when (e.g., rainfall intensity, antecedent wetness) preferential flow (PF), defined as rapid subsurface bypass flow via soil macropores, occurs. While such heterogeneous flow through soils and bedrock especially affects water and solute exports from catchments during extreme events (e.g., during high intensity rainfall or rainfall after prolonged drought), the representation of PF in large-scale land-surface models is currently not considered. ML models have great potential to explore drivers of PF from global soil moisture databases and emulate these complex hydrologic processes in ESMs.

Rationale:

Hydrologic processes of a highly heterogeneous subsurface strongly impacts biogeochemical cycling (Li et al., 2021). Though the first scientific description of PF dates back over a century (Schumacher, 1864), when and where heterogeneous subsurface flow occurs is still poorly understood-- and is not implemented in ESMs. Instead, uniform flow, often described by the Buckingham-Richards equation originating from Darcy's law, remains the most common approach to modeling subsurface flows in soils (Beven, 2018). This approach is the one most often used to model subsurface flows in ESMs (e.g., Swenson et al. 2019, Fig. 1).

There are two main reasons why PF is not universally included in hydrologic models such as ESMs. One reason is a lack of understanding under which conditions PF occurs. This is true in terms of physical characteristics in the landscape, such as soil texture, vegetation cover, slope or aspect, (i.e., variability in space) and also hydro-meteorological conditions triggering PF, such as rainfall or snowmelt intensity, precipitation amount, antecedent wetness (i.e., variability in time). While several studies have looked at drivers of PF using field observations from one or few sites (e.g., Lin & Zhou, 2008; Demand et al. 2019), the uniqueness of place (Beven, 2000) prevents a generalizable process understanding from individual study sites.

Due to these challenges, it is not surprising that when and where PF occurs was one of the twenty-three recently identified unresolved problems in hydrology (Bloeschl, 2019) and that the quantification and prediction of PF was recognized as one of five key challenges in modeling soil processes (Vereecken et al., 2016).

DOI: https://doi.org/10.2172/1769765
DOE BER, AI4ESP 2021 #AI4ESP1125 (https://www.ai4esp.org)

The second reason why PF is not universally included in hydrologic models is a limited understanding of how to integrate complex PF dynamics into ESMs. ML models will help conceptualize complex PF processes across heterogeneous landscapes and subsurface environments in a computationally efficient manner as outlined in the next section.
So far, applications of ML models to soil science applications have been limited to remote sensing products, like the Soil Moisture Active Passive (SMAP) mission (Fang et al., 2017; Shen, 2018). These studies only include the top few centimeters of the soil profile; thus, they do not address subsurface flow processes. In this white paper, novel ML models that incorporate soil moisture time series to overcome the limitations of remote sensing products are discussed.

Narrative:

The outlined two current limitations for including PF in ESMs could be tackled in two steps that involve ML models (Fig. 1): First, a ML algorithm can be applied on field observations of hydro-meteoric time series (Fig. 2.I) and site characteristics of locations (Fig. 2.II) where high-frequency soil moisture time series are available (e.g., International Soil Moisture Network, Doriogo et al., 2021). The prior two can serve as dynamic (x^d) and static (x^s) attributes in Entity-Aware Long Short-Term Memory networks (EA-LSTM, Fig. 2.IV), as described by Kratzert et al. (2019) for hydrograph modelling for catchments. In the case of PF, the target would not be a simulation of soil moisture, but relating the static and dynamic attributes to occurrence of PF (Fig. 2.III arrows indicate PF). Occurrence of PF can be derived from non-sequential soil moisture sensor response: PF is recognized when a deeper moisture probe responds earlier to rainfall than a shallower probe (Lin & Zhou, 2008; Demand et al., 2019). The time-variant contribution of the dynamic features to the PF occurrence can be derived from cell states of the EA-LSTM to gain novel process understanding (e.g., impact of hydro-meteorological extremes like high intensity rainfall during the wet season in Fig. 2.Va related to the state of one LSTM cell in Fig. 2.Vb). The role that static features play for PF occurrence can then be inferred from clusters grouped via Uniform Manifold Approximation and Projection (UMAP, McInnes et al., 2020, Fig. 2.VI). EA-LSTM was shown to allow for interpreting the effect of individual static characteristics (Kratzert et al., 2019), which would be very helpful in determining where PF plays an important role.

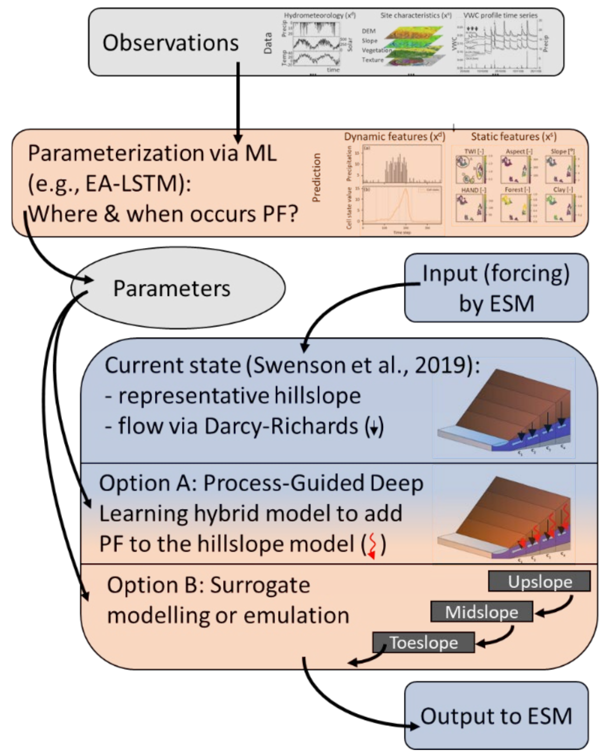

Figure 1 Connecting the ML derived process understanding of preferential flow from observations (see Figure 2) to the hillslope model in an Earth System Model (e.g., GLM in Swenson et al., 2019). Grey boxes indicate data and parameters. Figure inspired by Reichstein et al. (2019).

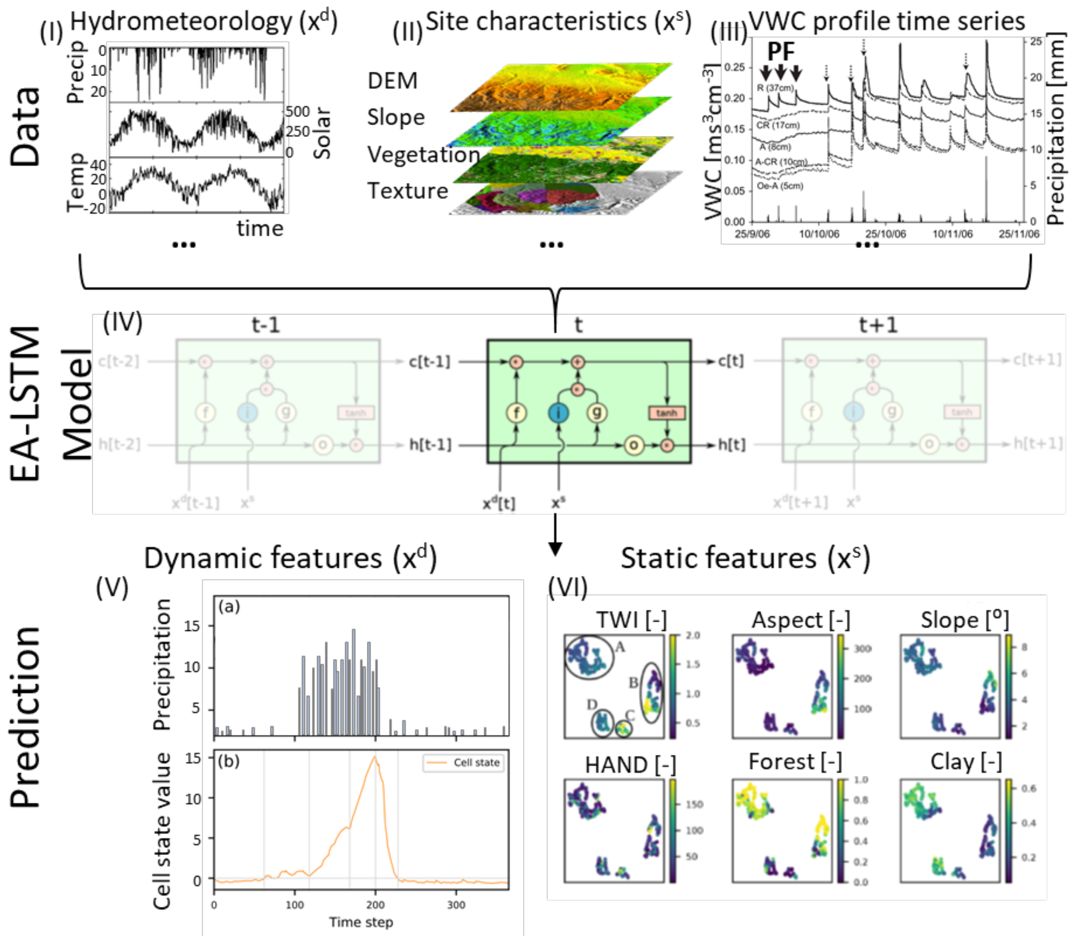

Figure 2 Machine learning model (i.e., Entity-Aware Long Short-Term Memory networks, EA-LSTM; Kratzert et al., 2019) to derive dynamic (x^d) and static (x^s) attributes that explain occurrence of preferential flow. Both, x^d and x^s are inputs into the EA-LSTM. Target for the ML model is the occurrence of preferential flow (PF) derived from high-frequency soil moisture data. After the model is trained and tested to accurately predict PF, x^d and x^s can be extracted and clustered via Uniform Manifold Approximation and Projection (UMAP, McInnes et al. 2020) to explain spatio-temporal drivers of PF. Graphic includes adapted subplots from Zhi et al. (2021), https://en.wikiversity.org/wiki/File:GIS_Layers.png, Lin & Zhou (2008), Kratzert et al. (2019), and Kratzert et al. (2018).

In a second step, the novel insights about where and when PF occurs can be incorporated into ESMs, which recently adapted the representative hillslope concept that allows accounting for intra-hillslope connectivity (e.g., in the Community Land Model; Swenson et al., 2019). One option could be to couple the physically-based hillslope flow model to a ML model via Process-Guided Deep Learning hybrid model (Read et al., 2019). This way, the ML would be embedded in the ESM accounting for constraints from the water mass and energy balances. Another option would be to replace the subsurface hydrology model with a surrogate model or an emulation based on knowledge gained from the EA-LSTM in combination with training the model to stream discharge data. Such approaches, replacing computationally expensive, physics-based submodels in ESMs have gained momentum in the Earth Science community (see Reichstein et al., 2019).

Suggested Partners/Experts:

This AI4ESP white paper is the outcome of a breakout session organized as part of the Early Career Critical Zone Network-of-networks Workshop during the 2020 AGU Fall Meeting. The network has several experts on subsurface and hillslope hydrology that could contribute to the process understanding with ML.

Xiaojing Fu, Caltech Division of Engineering and Applied Science, California Institute of Technology
Chaopeng Shen, Pennsylvania State University, Civil and Environmental Engineering
Bill Riley, Earth & Environmental Sciences Area, Lawrence Berkeley National Laboratory
Frederik Kratzert, Johannes Kepler University, Linz
Sean C. Swenson, Climate and Global Dynamics Laboratory, National Center for Atmospheric Research

Data access and FAIR standards

The International Soil Moisture Network operates with the FAIR Data Principles (Dorigo et al., 2021).

References

Beven: Uniqueness of place and process representations in hydrological modelling, Hydrol. Earth Syst. Sci., 4(2), 2000.

Beven: A Century of Denial: Preferential and Nonequilibrium Water Flow in Soils, 1864-1984, Vadose Zone J., 17(1), 2018.

Blöschl, Bierkens, Zhang, et al.: Twenty-three Unsolved Problems in Hydrology (UPH) – a community perspective, Hydrol. Sci. Journal, 64(10), 2019.

Demand, Blume, & Weiler: Spatio-temporal relevance and controls of preferential flow at the landscape scale, Hydrol. Earth Syst. Sci., 23(11), 2019.

Dorigo, Himmelbauer, Aberer, et al.: The International Soil Moisture Network: serving Earth system science for over a decade, Hydrol. Earth Syst. Sci. Discuss., 2021.

Fang, Shen, Kifer, & Yang: Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental U.S. Using a Deep Learning Neural Network, Geophys. Res. Lett., 44(21), 2017.

Read, Jia, Willard, et al.: Process-Guided Deep Learning Predictions of Lake Water Temperature, Water Resour. Res., 55(11), 2019.

McInnes, Healy, & Melville: UMAP: Uniform manifold approximation and projection for dimension reduction, arXiv, 2020.

Kratzert, Klotz, Brenner, Schulz, & Herrnegger: Rainfall – runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., (22), 2018.

Kratzert, Klotz, Shalev, Klambauer, Hochreiter, and Nearing: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrol. Earth Syst. Sci., 23(12), 2019.

Li, Sullivan, Benettin, et al.: Toward catchment hydro-biogeochemical theories, Wiley Interdiscip. Rev. Water, 8(1), 2021.

Lin & Zhou: Evidence of subsurface preferential flow using soil hydrologic monitoring in the Shale Hills catchment, Eur. J. Soil Sci., 59(1), 2008.

Reichstein, Camps-Valls, Stevens, Jung, Denzler, Carvalhais, & Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566(7743), 2019.
Schumacher: Die Physik des Bodens, Wiegandt & Hempel, Berlin., 1864.
Shen: A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists, Water Resour. Res., 54(11), 2018.
Swenson, Clark, Fan, Lawrence, & Perket: Representing Intrahillslope Lateral Subsurface Flow in the Community Land Model, J. Adv. Model. Earth Syst., 11(12), 2019.
Vereecken, Schnepf, Hopmans, et al.: Modeling Soil Processes: Review, Key challenges and New Perspectives, Vadose Zone J., 15(15), 2016.