Title
Dominant Controls on Preferential Flow and Their Implications for Future Soil Water Fluxes

Journal
Earth's Future, 14(7)

ISSN
2328-4277

Authors
Li, Bonan
Sprenger, Matthias
Araki, Ryoko
et al.

Publication Date
2026-07-01

DOI
10.1029/2026ef008296

ADVANCING EARTH AND SPACE SCIENCES

Earth's Future

RESEARCH ARTICLE

10.1029/2026EF008296

Key Points:

- The effect of precipitation peak intensity on soil water velocities declined with increasing precipitation intensity
- Antecedent soil moisture failed to predict preferential flow (PF), contrasting the high predictive power of sand content
- Climate predictions suggest that soil water velocities through PF paths will increase ~15% by 2099

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

B. Li,
libon@oregonstate.edu

Citation:

Li, B., Sprenger, M., Araki, R., Keen, R., Mayernik, C. M., Rudisill, W., et al. (2026). Dominant controls on preferential flow and their implications for future soil water fluxes. *Earth's Future*, *14*, e2026EF008296. https://doi.org/10.1029/2026EF008296

Received 12 FEB 2026
Accepted 7 JUL 2026

Author Contributions:

Conceptualization: Bonan Li, Matthias Sprenger, Ryoko Araki, Rachel Keen, Hoori Ajami, Octavia Crompton, Daniel Giménez, Jannis Groh, Daniel Hirmas, Aaron N. Koop, Nitin Singh, Inge Wiekenkamp, Briana M. Wyatt, Tianfang Xu, Pamela L. Sullivan
Data curation: Bonan Li, Ryoko Araki, William Rudisill
Formal analysis: Bonan Li, Ryoko Araki
Funding acquisition: Matthias Sprenger, Pamela L. Sullivan

Dominant Controls on Preferential Flow and Their Implications for Future Soil Water Fluxes

Bonan Li[1], Matthias Sprenger[2,3], Ryoko Araki[4,5], Rachel Keen[6], Caitlin M. Mayernik[1], William Rudisill[7], Hoori Ajami[8], Octavia Crompton[9], Daniel Giménez[10], Jannis Groh[11,12], Daniel Hirmas[13], Aaron N. Koop[14], Nitin Singh[15], Inge Wiekenkamp[16], Briana M. Wyatt[17], Tianfang Xu[18], and Pamela L. Sullivan[1]

[1]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA, [2]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [3]Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, USA, [4]Department of Geography, San Diego State University, San Diego, CA, USA, [5]Department of Geography, University of California, Santa Barbara, CA, USA, [6]Department of Ecology and Evolutionary Biology and Kansas Biological Survey & Center for Ecological Research, University of Kansas, Lawrence, KS, USA, [7]Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [8]Department of Environmental Sciences, University of California Riverside, Riverside, CA, USA, [9]Hydrology and Remote Sensing Laboratory, USDA Agricultural Research Service, Beltsville, MD, USA, [10]Department of Environmental Sciences, Rutgers University, New Brunswick, NJ, USA, [11]Agrosphere Institute (IBG-3), Forschungszentrum Jülich (FZJ), Jülich, Germany, [12]Isotope Biogeochemistry and Gas Fluxes, Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany, [13]Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, USA, [14]Aaron N. Koop Soils Research, Newton, KS, USA, [15]Department of Crop, Soil & Environmental Sciences, Auburn University, Auburn, AL, USA, [16]GFZ Helmholtz Centre for Geosciences, Potsdam, Germany, [17]Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA, [18]School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA

Abstract

Soil water flow, particularly preferential flow (PF), is a critical control on hydrological and biogeochemical processes, including groundwater recharge, contaminant transport, and carbon cycling. However, it remains challenging to predict PF occurrence across large environmental gradients. Here, we developed a deep learning (DL) model to estimate event-scale soil water flow velocity and the probability of PF occurrence using high-frequency soil moisture and precipitation data from 33 sites across the National Ecological Observatory Network. The model demonstrated high skill in predicting the binary occurrence of PF (91% F1-score; 85% accuracy) but the performance was limited in predicting soil water velocity ($R^2 = 0.31$). We found that precipitation characteristics (duration, volume, and intensity) were the most important predictors for soil water velocity. Among the non-precipitation event variables, sand content showed relatively high predictive skill, though differences among non-event climate variables were generally modest. Lower sand content was associated with increased predicted soil water velocity, a finding that highlights the role of soil structure in producing more non-uniform flow, which contrasts with traditional uniform flow models. Projecting a reduced DL model under both moderate and high-emissions future climate scenarios (2060–2099 Representative Concentration Pathways 4.5 and 8.5), we found ~7.3% increase under RCP4.5 and ~15% under RCP8.5 of soil water velocities compared to the historical simulation, while modeled likelihood of PF changed little. These findings suggest climate change is not making PF more frequent, but it is making existing PF pathways more efficient with important consequences for associated nutrient and contaminant transport under climate change.

Plain Language Summary

Water movement in soil is critical for water quality. While often modeled as a uniform flow process, in reality water moves rapidly through cracks and burrows in what is called "preferential flow" (PF), which limits natural filtration and can transport pollutants. We developed a deep learning model, trained on data from 33 U.S. sites, to predict when and how fast this PF occurs based on precipitation, soil, and climate data. The model showed that precipitation characteristics (duration, intensity, volume) were the most important predictors of PF. Lower soil sand content/higher clay content was associated with faster water flow, likely due to clay soils forming aggregates and cracks that water moves through rather than infiltrating uniformly. Further analyses based on climate projections suggest that the speed at which PF occurs will become more rapid under future climate scenarios compared to historical simulation. This highlights the need to represent PF in soil water models when assessing future water quality.

Methodology: Bonan Li, Matthias Sprenger, Ryoko Araki, Tianfang Xu, Pamela L. Sullivan
Resources: Matthias Sprenger
Supervision: Matthias Sprenger, Pamela L. Sullivan
Writing – original draft: Bonan Li, Matthias Sprenger, Ryoko Araki, Pamela L. Sullivan
Writing – review & editing: Bonan Li, Matthias Sprenger, Ryoko Araki, Rachel Keen, Caitlin M. Mayernik, William Rudisill, Hoori Ajami, Octavia Crompton, Daniel Giménez, Jannis Groh, Daniel Hirmas, Aaron N. Koop, Nitin Singh, Inge Wiekenkamp, Briana M. Wyatt, Tianfang Xu, Pamela L. Sullivan

1. Introduction

Preferential flow (PF) is a commonly observed phenomenon where water rapidly moves through relatively few distinct pathways in the soil, bypassing much of the bulk soil matrix. PF typically occurs in quick response to significant precipitation events. PF influences water balance partitioning, affecting streamflow generation, groundwater recharge, and the water stored in soil available for evapotranspiration (ET). PF governs water transit times (Sprenger et al., 2019), thereby impacting the transport of nutrients and pollutants in the subsurface (Radolinski et al., 2022), which in turn affects ecosystem and agricultural productivity by influencing root zone water storage and solute dynamics. These PF paths reduce the ability of soils to filter new water before it recharges groundwater and directly or indirectly diminishes many ecosystem services provided by soils (Clothier et al., 2008). Compared to water fluxes through the bulk soil matrix, PF pathways wet and drain more rapidly and frequently, creating highly heterogeneous moisture conditions in soil. These dynamics can generate biogeochemical "hot spots" and "hot moments" (Bernhardt et al., 2017; Franklin et al., 2021; McClain et al., 2003), where transient wetting triggers microbial activity and nutrient and carbon-rich water is transported from the near surface to different depths. Despite these advances in our understanding of individual components of PF dynamics, the mechanisms driving the formation of PF pathways and its occurrence remain one of the most challenging processes to constrain (K. J. Beven, 2018). As a consequence, this limits our ability to discern its potential responses to climate change and the cascading implications for subsurface biogeochemical processing and transport. Sensor-based approaches have been used in watersheds around the world to understand the occurrence and drivers of PF (Allaire et al., 2009; Nimmo et al., 2025). Yet synthesis across sites or ecoregions has been limited due to inconsistencies in sensor types, frequency of collection, and the availability of contextual information on potential drivers (Sprenger et al., 2021). The research presented here leverages a sensor-based approach across sites using a deep learning (DL) model to estimate soil water flow velocity and the probability of PF occurrence based on dynamic (event-based) variables.

A recurring theme across PF studies is the contrasting influence of event-based (e.g., precipitation intensity) versus site-based drivers (e.g., mean annual precipitation) on PF occurrence and soil water flow velocity. Most studies report that increasing, event-based precipitation intensity—up to a certain threshold—enhances the likelihood of PF (Hu et al., 2019; Lin & Zhou, 2008; Radolinski et al., 2021). In contrast, the role of antecedent soil moisture (at the onset of an event), another event-based variable, is more variable, with studies showing positive, negative, or negligible effects on PF likelihood (Jarvis, 2007). Both precipitation intensity and soil moisture variability are projected to increase in many regions due to climatic change (C. Li et al., 2021; Prein et al., 2017) highlighting the need for better insights into the complex interaction of these climatic drivers with PF occurrence. Recent advances also highlight the sensitivity of PF pathways (e.g., macropore flow) to dynamic biotic influences such as root activity (Guthrie et al., 2025; Jarecke et al., 2024; Koop et al., 2023; Sullivan et al., 2022) and soil fauna including earthworms (Rutgers et al., 2016; Van Schaik et al., 2014). Meanwhile, relatively static soil properties such as clay content and structural development frequently emerge as strong predictors of PF, with higher clay content and well-developed structure often associated with increased PF potential (Demand et al., 2019; B. Li et al., 2025). These findings suggest that the propagation of PF across soil horizons (e.g., topsoil A horizons vs. subsoil B horizons) is likely influenced by the degree of pedogenesis, including processes such as clay illuviation. Indeed, a continental-scale study of macroporosity—pores that facilitate PF—found that the accumulation of complexed clay and soil organic C increases macroporosity, particularly in subsoils, as these horizons are often less C saturated and retain more available mineral and clay surfaces for additional organic C to stabilize, thereby influencing pore development (Koop et al., 2023). This highlights the potential for subsoil characteristics to modulate PF dynamics in ways that are distinct from surface horizons.

PF occurrence is shaped by complex interactions among multiple variables, making it challenging for traditional physically-based modeling approaches to capture these dynamics. However, machine learning methods offer a powerful alternative, capable of uncovering patterns in large, multidimensional data sets (Massoud et al., 2023). For example, a recent application of random forest models to explain PF occurrence across 40 sites spanning 17 ecoregions identified meteorological conditions, such as precipitation intensity and soil moisture variability, as key predictors (B. Li et al., 2025). Notably, distinct thresholds for individual variables were observed across sites, underscoring the context-dependent nature of PF. However, this binary classification approach (i.e., PF occurrence as yes/no) limits the ability to explore continuous soil water flow velocity responses, thereby constraining insights into the mechanisms driving PF behavior. To overcome these limitations, we develop a DL model to estimate both soil water flow velocity and the probability of PF occurrence at the event scale. DL models are particularly well-suited for this task because they can ingest large volumes of event-based data and uncover complex, non-linear relationships between inputs and outputs, without the need for extensive feature engineering. Our approach tests three key hypotheses:

1. Event-based variables (e.g., precipitation intensity, antecedent soil moisture) are more informative for predicting soil water flow velocities than site-based variables (e.g., mean annual precipitation).
2. Differences in soil properties between topsoil (<35 cm) and subsoil (35–100 cm) will influence soil water flow velocities.
3. Increased precipitation intensity under future climate scenarios will drive faster soil water velocities.

To test these hypotheses, we leverage data from the U.S. National Ecological Observatory Network (NEON), a U.S. National Science Foundation program. NEON provides high-frequency soil moisture data from 33 sites across 16 ecoregions in the continental United States and Puerto Rico. At each site, soil moisture data are collected at various depths, recording data every 1 min since approximately 2017. We isolate all precipitation events that triggered a measurable soil moisture response (>0.01 cm^3/cm^3) and use the soil water flow velocities to train and validate the DL model. We further trained the DL model to predict PF based on observed storm events with precipitation and ancillary environmental variables as input features. The methodological consistency and temporal resolution of NEON's data set make it uniquely suited for training and validating DL models aimed at advancing our understanding of soil water flow velocity and PF dynamics across diverse environmental conditions.

2. Methods

2.1. NEON Data

We used high-frequency volumetric soil moisture and precipitation data, along with soil properties from 33 of the 40 NEON terrestrial sites (due to data quality limitations) across 16 ecoregions in the contiguous US and Puerto Rico to assess the extent of event-based drivers on PF. Each site provided data over a period of 4–7 years. Soil moisture measurements were collected using Sentek - EnviroSCAN TriSCAN soil moisture sensors (0.01% change repeatability) at up to eight depths ranging from 6 to 200 cm in five soil plots around the tower infrastructure, recorded at 1-min intervals (NEON, 2023c). For this study, we followed soil moisture processing approaches that are outlined in detail in B. Li et al. (2025). We filtered the soil moisture data where internal quality flags were raised or where the moisture exceeded the maximum porosity of the entire soil profile across all soil pits. The soil moisture data were resampled to a 10-min interval using the median value within that time period, with missing values assigned to the resampled data sequence if the 1-min interval data contained fewer than seven good quality measurements (NEON, 2023c). Gaps of less than 6 hours in the resampled soil moisture data were filled using linear interpolation.

We used both primary and secondary precipitation data from these sites (NEON, 2023a). Primary precipitation data was recorded every 5 min and includes measurements by a weighing gauge (manufacturer-specified precision: 0.05 mm) located within 0.5 km of the tower infrastructure at all NEON core sites. Secondary precipitation data was recorded at 1-min intervals and includes measurements collected at the tower top (above the canopy, 8–60 m) by a tipping bucket with a precision of 0.001 mm across all NEON sites. We included both primary and secondary precipitation data sets in our study based on their availability, preferring secondary data where both types were available due to its finer temporal resolution. Primary precipitation data were filtered using internal final quality flags, while secondary data were filtered using a combination of plausible value range flags and science review flags raised by NEON scientists due to known unfavorable conditions. The filtered precipitation data were aggregated to 10-min intervals, with missing values assigned as zeros if the original data at 1-min or 5-min intervals had fewer than five good quality measurements passing NEON's QA/QC (Kesting & Allen, 2024; NEON, 2023a).

2.2. Calculating Soil Water Velocity and PF

Precipitation event was defined as any precipitation initiated (precipitation onset) by a non-zero 10-min precipitation pulse. Any additional precipitation occurring within the following 6 hr was considered part of the same event. To avoid excessively long events, a maximum duration of 5 days was imposed. For each precipitation event, we derived the soil water sensor event response velocity, here called soil water velocity, which is widely used to identify PF (Lin & Zhou, 2008; Nimmo et al., 2025). For each precipitation event at a given soil moisture sensor depth, we calculated the time lag from the precipitation onset to the detected soil moisture response onset, and we refer to this lag as the event response time. The soil moisture response onset was determined following Hirmas et al. (2025) and B. Li et al. (2025), where the onset of the soil moisture response was defined as the timestamp at which the increase in soil moisture was greater or equal to 0.01 cm^3/cm^3. We note that the onset of soil moisture response identified might be slightly delayed from the first physical response by this threshold approach. However, it can reduce the risk of identifying sensor noise as hydrological responses. Soil water velocity is then calculated as the ratio of soil moisture sensing depth to the event response time. In accordance with Hardie et al. (2011), such velocity represents the water flow "within a portion of the soil that occurs within the sphere of influence of the soil moisture sensors."

In this study, PF is not identified using the non-sequential approach, in which PF is inferred when a deeper soil sensor responds earlier than the shallower sensor, resulting in a zero/negative response time between two soil sensors. Instead, we focused on the velocity threshold approach, where PF (0 for not a PF event; 1 for PF event) was inferred when the soil water velocity exceeded the maximum estimated saturated hydraulic conductivity (K_{sat}) value, determined using the pedotransfer function model Rosetta3 (Y. Zhang & Schaap, 2017) based on soil texture data, for the given soil thickness above each sensor (details in B. Li et al., 2025). We focus on sensors between 10 and 100 cm to reduce uncertainties from very shallow and deep sensors.

2.3. Event and Static Variables for Predicting Soil Water Velocity

2.3.1. NEON PF Predictor Variables

Soil texture, precipitation, and soil moisture data were obtained from the NEON data. Soil organic carbon measurements and rooting depth information were not included in this analysis, as insufficient information was available for each sensor. Site specific soil texture was determined using the <2 mm soil fraction using the wet and dry sieving (sand fractions) and sedimentation (silt and clay fractions) in accordance with the USDA NRCS Soil Survey Laboratory Methods Manual (Soil Science Division Staff, 2017) published by NEON (NEON, 2023b, 2024b). For each precipitation event, we calculated the precipitation peak intensity, duration, and sum. Additionally, three metrics were derived from the soil moisture data for each event: (a) the antecedent soil moisture trend (Antecedent SM trend), defined as the magnitude of the slope of the relationship between soil moisture at the onset of the precipitation event and soil moisture one day prior to the event for each sensor and depth; (b) the average soil moisture one day prior to the event onset (Average SM (1d history)) measured by the individual sensors; (c) the mean surface soil moisture (Mean surface SM (1d history)), measured by the upper most soil moisture sensor one day before the precipitation event.

While our study sites span 16 ecoregions with diverse soil and climate characteristics, a broader range of drivers at a finer temporal scale might allow for better understanding and predictions. Another limitation is that the NEON soil texture data is not perfectly aligned with the soil moisture sensor depth and locations.

2.3.2. Satellite Derived Data

The mean annual temperature (MAT, °C) and mean annual precipitation sum (MAP, mm/y) were obtained from NEON metadata (NEON, 2024a), and aridity index from Fiorella et al. (2021). To derive Standardized Precipitation-Evapotranspiration Index (SPEI), we used 3-hourly air temperature and precipitation data from Global Land Data Assimilation System (GLDAS) Noah Land Surface Model L4 Version 2.1 at 0.25° for 2010–2022 (Beaudoing et al., 2020; Rodell et al., 2004). Potential evapotranspiration (PET) was calculated using Hamon's method (Vremec et al., 2024), which requires air temperature as primary input. The 3-month SPEI (SPEI3) was then computed using code provided by Vonk (2024) driven by GLDAS-derived Hamon PET and GLDAS precipitation.

For each precipitation event, we extracted leaf area index (LAI, MCD15A3H, Myneni et al., 2015) at 500-m resolution from MODIS with 4-day temporal resolution. We assigned the most recent LAI values to each precipitation event. The gross primary productivity (GPP) was obtained from MOD17A2HGF Version 6.1 (Running & Zhao, 2021). Similar to LAI, we assigned the most recent GPP value to each precipitation event. Consequently, these variables could differ among events when the temporal spacing between precipitation events exceeds the temporal resolution of the corresponding GPP and LAI products. Event-based antecedent conditions were determined from the actual ET and land surface temperature (T) 24 hr prior to each precipitation event using the GLDAS and designated as ET (1d history) and Surface T (1d history), respectively.

2.4. Future Climate Scenarios

We incorporated future climatic conditions using IM3/HyperFACETS Thermodynamic Global Warming Simulation Data sets at 12-km resolution (Jones et al., 2022, 2023). The data set provides climate projections under Representative Concentration Pathways (RCP) scenarios grouped by "cooler" or "hotter" models relative to the multi-model mean climate sensitivity. We used historical (1980–2019) and end-of-century (2060–2099) projections for both the least extreme scenario (RCP 4.5, with cooler models) and the most extreme scenario (RCP 8.5, with hotter models). For each NEON site, climate variables were extracted from the nearest IM3/HyperFACETS grid cell.

Precipitation events were delineated from the hourly historical (1980–2019) and future (2060–2099) precipitation time series of the two scenarios using the same event separation criteria as applied to NEON observations (B. Li et al., 2025). For each delineated event, precipitation sum, peak intensity, and duration were recorded. We also calculated historical and scenario-specific site climate variables, including MAT, MAP, SPEI3, and aridity. The MAT and MAP were calculated as the multi-year annual averages. Daily PET was calculated using the Hamon method (Vremec et al., 2024) based on temperature and latitude, and the aridity index was estimated as the PET: MAP ratio. We derived the SPEI3 from precipitation and Hamon PET using the same procedure applied to the NEON observational data set. Similarly, seasonal values of these variables were calculated as multi-year seasonal averages/sum for winter, DJF; spring, MAM; summer, JJA; fall, SON for both historical and future scenarios (Figure S3 in Supporting Information S1).

We estimated soil water velocities and PF probabilities for historical and future scenarios by applying a reduced DL model to the corresponding event level precipitation characteristics and site level climate variables (Section 2.5; Table S1 in Supporting Information S1). The reduced model was used because several observational predictors were not directly available or reliably projected based on future scenarios. In this set-up, soil structure and properties, which affect K_{sat}, remain relatively static over time, and remain unchanged between the historical and future periods. However, we acknowledge that for example, extreme conditions (Robinson et al., 2016; Sullivan et al., 2022) or land-use (Hofmeister et al., 2025; Schlüter et al., 2022), that are not represented in this projection framework, could affect the soil structure.

We also acknowledge that the DL models are generally limited in their ability to extrapolate beyond the range of their training data distributions. However, our training set covers a wide range of hydroclimatic conditions, and RCP 4.5 and 8.5 projections may include out-of-distribution inputs that carry inherent uncertainties (Figure S5 in Supporting Information S1). Therefore, we interpret the projected soil water velocity and PF probability as scenario-based estimates rather than deterministic forecasts.

2.5. Deep Learning Model

To predict soil water velocity and the probability of PF occurrence, we developed the DL model adapted from Residual Neural Network (ResNet) architecture (He et al., 2015). This architecture incorporates skip connections, which link earlier layers directly to later ones, helping to mitigate vanishing gradient issues in deeper networks and improving training stability. The model was designed as a multi-output network with two prediction targets: soil water velocity and the probability of PF occurrence of a precipitation event.

The model consists of two ResNet blocks (excluding the input layer), each comprising a feature normalization layer, a non-linear activation layer (exponential linear unit; ELU), a dropout layer, a fully connected (FC) layer, and a normalization layer (Figure 1). The feature normalization layer centers each input feature within a range to improve model training stability. The ELU activation function introduces non-linearities that enable the model to capture nonlinearities among features. The dropout layer serves as a regularization method by randomly deactivating a fraction of neurons during training to reduce overfitting. The FC layer transforms inputs to outputs through learnable parameters (weights), allowing the model to capture relationships across features. The sigmoid function maps the output from the FC layer to a value between 0 and 1 as the probability of PF occurrence, which is used to represent the probability (i.e., confidence) that a precipitation event initiates PF occurrence.

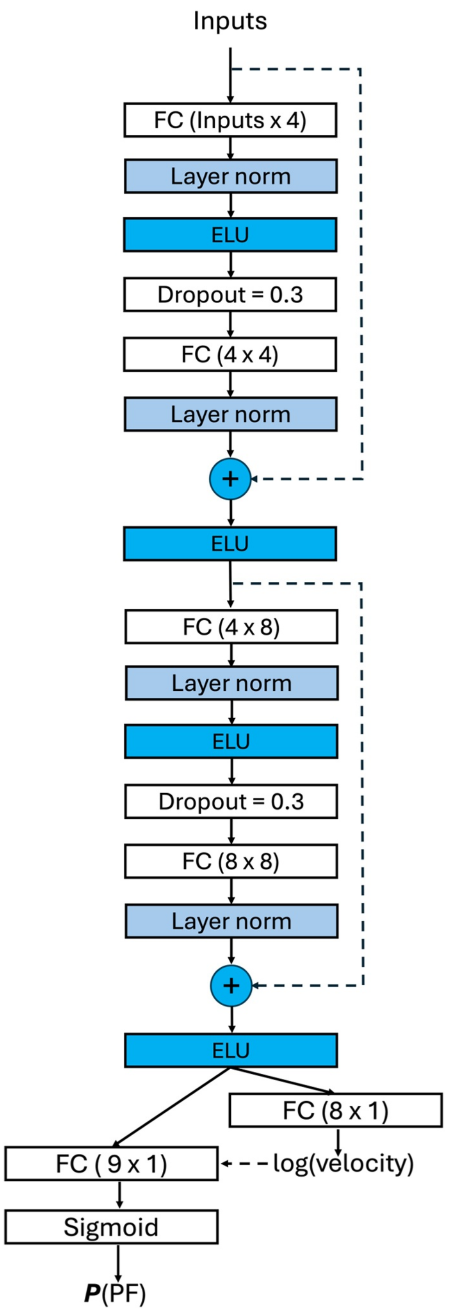

Figure 1. Structure of the deep learning model for soil water velocity at log scale. Note: FC = fully connected layer, ELU = exponential linear unit, Layer norm = feature normalization layer, and Dropout = fraction of the model's connections that are randomly turned off to avoid overfitting.

We developed two versions of the model: a full model for interpreting PF controls under observed conditions and a reduced model for analyzing changes between historical and future soil water velocity and probability of PF at event scale. The full model includes the complete set of predictors (Table S1 in Supporting Information S1): precipitation peak intensity, precipitation sum, precipitation duration, sensor depth, sand content, clay content, porosity, aridity index, MAT, MAP, GPP, LAI, SPEI3, ET (1d history), Surface T (1d history), Average SM (1d history), Antecedent SM trend, Mean surface SM (1d history). The reduced model included only predictors that could be consistently derived for the historical and future scenarios. These predictors were precipitation peak intensity, precipitation duration, sensor depth, precipitation sum, sand content, clay content, porosity, aridity index, MAT, MAP, and SPEI3 (Table S1 in Supporting Information S1). Both full and reduced models were trained on NEON observational data sets. The reduced model allowed us to apply the same predictor set to the historical and future scenarios while avoiding the use of predictors that were unavailable or highly uncertain under future scenarios. The reduced model yielded broadly comparable performance to the full model, suggesting that model reduction has limited effect on historical to future comparisons.

After quality filtering, our observational data set (full predictors and reduced predictors) consisted of 15554 event-sensor samples from 586 sensors at 33 NEON sites. Each event-sensor sample represents the response of one soil moisture sensor at a given depth to one precipitation event. To ensure a rigorous assessment of our model's ability to generalize to future conditions and to prevent data leakage, the data set was partitioned chronologically at each sensor location. This resulted in training (∼70%, 11498), validation (∼15%, 1837), and testing (∼15%, 2219) sets that maintained the temporal sequence of observations. The validation set provided a basis for model hyperparameter tuning (e.g., learning rate, epochs) during model development, while the testing set served as held-out data for model performance evaluation. Sensors with fewer than 20 data points were directly assigned to the training data set. We standardized the training data set to have a mean of zero and a standard deviation of one to facilitate faster convergence of the DL model and prevent large-scale features from dominating the learning process. The same parameters were also used to transform the validation and testing data, respectively.

We passed the training data sets up to 100 times (epochs) to train the model and a mini-batch size of 128 (update the model parameters every 128 samples) using Adam optimizer (Kingma & Ba, 2017). The model was optimized using a multi-task loss function that combined regression loss from soil water velocity and classification loss for PF occurrence as:

$$\mathcal{L}_{total} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| - \frac{1}{N}\sum_{i=1}^{N}[c_i \log(\hat{p}_i) + (1 - c_i)\log(1 - \hat{p}_i)] \quad (1)$$

where N is the number of samples in the mini-batch (here 128), $\hat{y}_i$ and y_i are the predicted and observed soil water velocity in log scale, c_i is the binary PF label (0 for non-PF, 1 for PF), $\hat{p}_i$ represents the predicted PF probability.

We used an initial learning rate of 0.01 and the learning rate was reduced by a factor of 0.8 (i.e., multiplied by 0.8) whenever the validation loss failed to reduce by at least 0.1% for 5 consecutive epochs. We set a minimum learning rate of 0.005 to allow the model to continue updating with smaller and more stable steps. To reduce the risk of overfitting, we applied early stopping based on the validation loss. Training was terminated if the validation loss did not decrease by at least 0.01 for 20 consecutive epochs. The final model was selected as the

10.1029/2026EF008296

checkpoint with the lowest validation loss before early stopping was triggered (Figure S6 in Supporting Information S1).

We trained 50 independent models using bootstrap-resampled training data set with the same model architecture and optimization procedures for both full and reduced models. The final predictions of soil water velocity and PF probability were calculated as the average of the 50 model ensembles. The ensemble based approach reduces the dependence on a single model realization and provides a more robust basis for predictions and interpretations.

To interpret the model predictions of factors affecting soil water velocity variation, we calculated SHAP values using the GradientSHAP (Lundberg & Lee, 2017) method. Unlike global feature importance metrics (e.g., from Random Forest), SHAP provides locally consistent explanations by averaging output gradients with respect to inputs over a baseline (training data). Positive SHAP values indicate that a feature increases the predicted soil water velocity relative to the baseline, whereas negative SHAP values indicate a decrease. Mean absolute SHAP values were used to summarize the overall magnitude of each feature's contribution, while SHAP dependence plots were used to evaluate the direction. Small differences between mean absolute SHAP values should not be interpreted as evidence that one variable is more important than the other, as SHAP values could be shared among strongly correlated predictors. We also note that direct single-variable analysis, such as correlation between soil water velocity and individual predictors, can provide insights into potential PF drivers. However, these analyses often fail to disentangle the confounding effects of other variables compared to SHAP analysis. To look at spatial patterns in PF and factors that influence it across contiguous US and Puerto Rico, we grouped NEON sites according to climate regions as described by Karl and Koss (1984): Western US (NEON ecoregions: 12, 13, 14, 15, 16, and 17) Southeast (NEON ecoregions: 3, 4, 8, 11), Central North (NEON ecoregions: 5, 6, 9, and 10), and Northeast (NEON ecoregions: 1, 2, 7).

3. Results and Discussion

3.1. Model Performance

The trained DL model reproduced the broad distribution of the observed soil water velocities (Figure 2), but its ability to reproduce the exact soil water velocities was still limited. The R^2 values of soil water velocity predictions were 0.36, 0.35, and 0.31 for training, validation, and testing data sets, respectively (Figure 2). The relatively small difference between training and validation performance suggests that strong overfitting was not evident (Figure S6 in Supporting Information S1). In contrast, the model showed strong skills in predicting binary classification of PF with precision, recall, accuracy, and F1-score all greater than 84%, with a low occurrence of false positive and false negative values (Figure 2).

Our model had a tendency to underrepresent the extreme lowest and highest soil water velocities and to overestimate the occurrence of the average soil water velocities. This is not surprising given the majority of the derived velocity data the model was trained on fell near this average. As a result, the model was successful at identifying when PF was likely to occur but was less precise at estimating the exact velocity of the flow, which is reflected in the smoothed-out distribution of the velocities.

Although our model showed a limited performance in predicting the exact soil water velocity, it captured the broad distributional structure of the observed soil water velocities. Our result is meaningful because soil water velocity associated with PF is a continuous response variable and is rarely predicted directly across large environmental gradients. Most of the previous PF analyses focused on binary PF occurrence at limited scale, whereas our approach extends this framework by estimating both whether PF occurs and how rapidly water moves. Thus, our study provides an important step toward identifying broad controls on PF-associated water movement across diverse soil and climate conditions at event level.

Our model's performance indicates a fundamental difference between predicting a continuous hydrological value and classifying a hydrological state. Although predicting the exact magnitude of the soil water velocity remains challenging due to limitations in data quantity and quality, as well as inherent uncertainties in identifying PF (i.e., delineations of precipitation events, soil moisture onsets, and peak detections etc.), the occurrence of PF itself, as defined here, tends to be more predictable. This is because the definition of PF for the velocity threshold method primarily relies on the comparisons between derived soil water velocity and K_{sat} values based on Rosetta (B. Li et al., 2025). Therefore, the classification results remain correct as long as the computed flow velocity is larger than the K_{sat} threshold that defines PF.

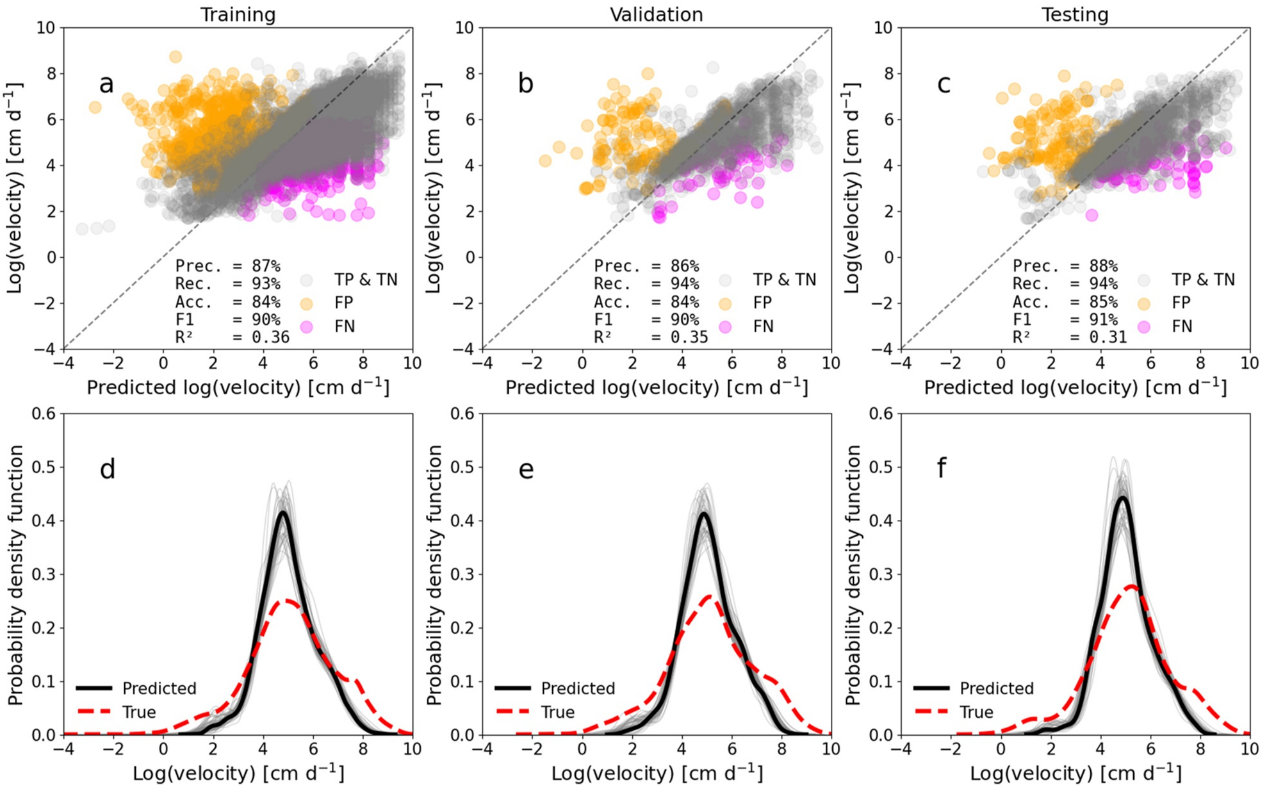

Figure 2. Model predicted versus soil moisture data derived soil water velocity across all sensors for each precipitation event observed at all 33 national ecological observatory network sites. TP and TN = true positives and true negatives, FP and FN = false positives and negatives, respectively. Results are shown for the training (a, d), the validation (b, e), and testing (c, f) data. Subplots (a–c) represent the comparison of the modeling results with the data-derived results. Subplots (d–f) show probability density functions of the model predictions (solid) and soil moisture data derived velocities (dashed). Note that velocity is log-scaled. The color code shows whether the deep learning derived classification into preferential flow or not is a true positive (TP, grey), a true negative (TN, grey), a false positive (FP, orange), or a false negative (FN, pink). Prec = precision; Rec = recall; Acc = accuracy; F1 = the combination of precision and recall; R^2 = coefficient of determination. Each dot in (a–c) represents the ensemble mean of 50 models. The grey lines in (d–f) represent the distribution of modeled velocity of 50 individual models.

It's important to note that this approach might mask the local scale heterogeneity, which creates additional uncertainty in velocity estimations. Also, the lack of key parameters on root characteristics (Kang et al., 2023), such as density and diameter, may partly explain why some flow velocities and PF events are not accurately modeled, as root systems are critical in shaping bio-pore networks (Lu et al., 2020), which directly influence water movement.

3.2. Event Based Precipitation Characteristics Most Strongly Influence PF

To infer key controls on PF and soil water velocity derived from DL model, we assessed mean absolute SHAP values of each model variable (Figure 3). Here, the "control" refers to predictors that strongly influence model-estimated soil water velocity and PF probability across the event-sensor data set. This distinction is important because PF responses emerge from interacting precipitation, texture, climate, vegetation status, and antecedent conditions, many of which covary across sites.

Among all evaluated event-based, site-based climate, and vegetation metrics, precipitation duration emerged as the single most dominant control, followed closely by precipitation sum (volume) and precipitation peak intensity. This hierarchy of feature importance remains remarkably consistent across both the topsoil (10–35 cm) and subsoil (35–100 cm) soil moisture data, indicating that the precipitation characteristics overwrites soil characteristics in regulating soil water velocities.

We observed highly non-linear relationships between precipitation event characteristics and soil water velocities, with clear thresholds in the feature contributions evident in the SHAP dependence plots (Figures 4a and 4b). Specifically, low precipitation sums did not lead to sufficient infiltration flux and therefore decreased the

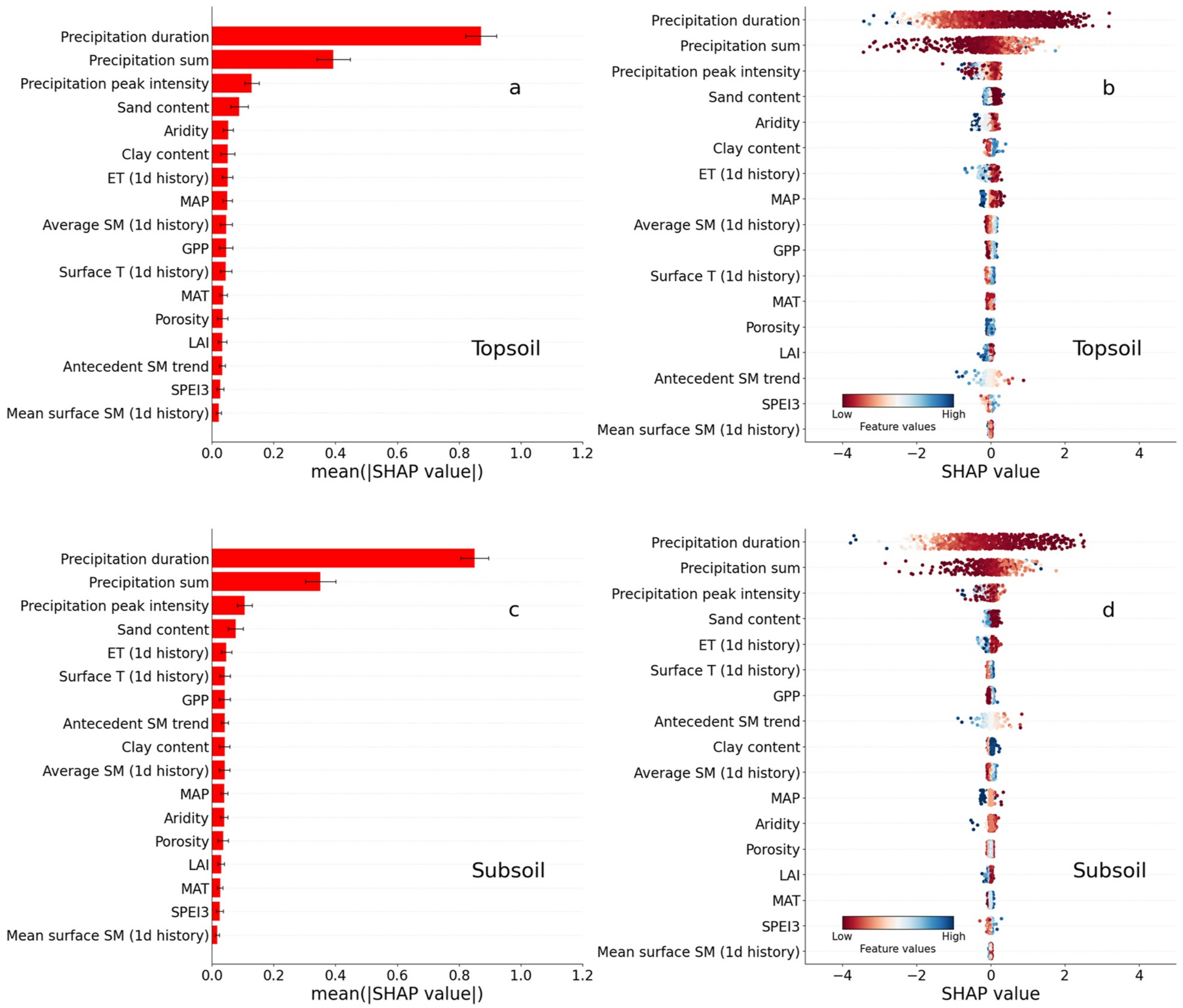

Figure 3. Mean absolute SHAP values of each considered variable explaining the soil water velocity for the topsoil (a) and subsoil (c), respectively. SHAP values of each considered variable for explaining the soil water velocity with color-coded feature values (red = low and blue = high) for topsoil (b) and subsoil (d), respectively. Note: the error bars in (a) and (c) represent the standard deviation of the SHAP values of 50 models. Each point in (b) and (d) represents the averaged SHAP value of 50 models for that sample.

predicted soil water velocity. However, for precipitation events with volumes greater than ~20 mm there was a clear positive effect on the model's prediction of soil water velocities (Figure 4a). An increase of both PF occurrence and the associated soil water velocities with increasing precipitation sum were observed in several previous studies from catchments in Europe and China (Demand et al., 2019; M. Li et al., 2020; Wiekenkamp et al., 2016) and our analysis confirms that the precipitation amount is generally a strong driver of soil water velocities. The sharp threshold behavior captured at the sum of 20 mm in precipitation highlights the volumetric requirement necessary to initiate PF. This threshold suggested that minor precipitation volumes are absorbed by the soil's matrix leading to dominance of uniform flow. Once an event delivers more than 20 mm of water, the likelihood of PF initiation increases substantially due to processes outlined by Nimmo (2016). For precipitation peak intensity, less intense precipitation (<30 mm/hr) had a net positive contribution in predicting soil water velocities (Figure 4b). This may occur because when precipitation intensity exceeds the soil matrix's infiltration capacity, water preferentially enters macropores with less hydraulic resistance rather than the fine-pore network. Although these macropores usually account for a small soil volume (<5%), they function as PF pathways which allow exceptionally high subsurface water flow (K. Beven & Germann, 1981). During very intense precipitation events (>38 mm/hr), the soil's infiltration capacity may be exceeded, leading to surface runoff or ponding and

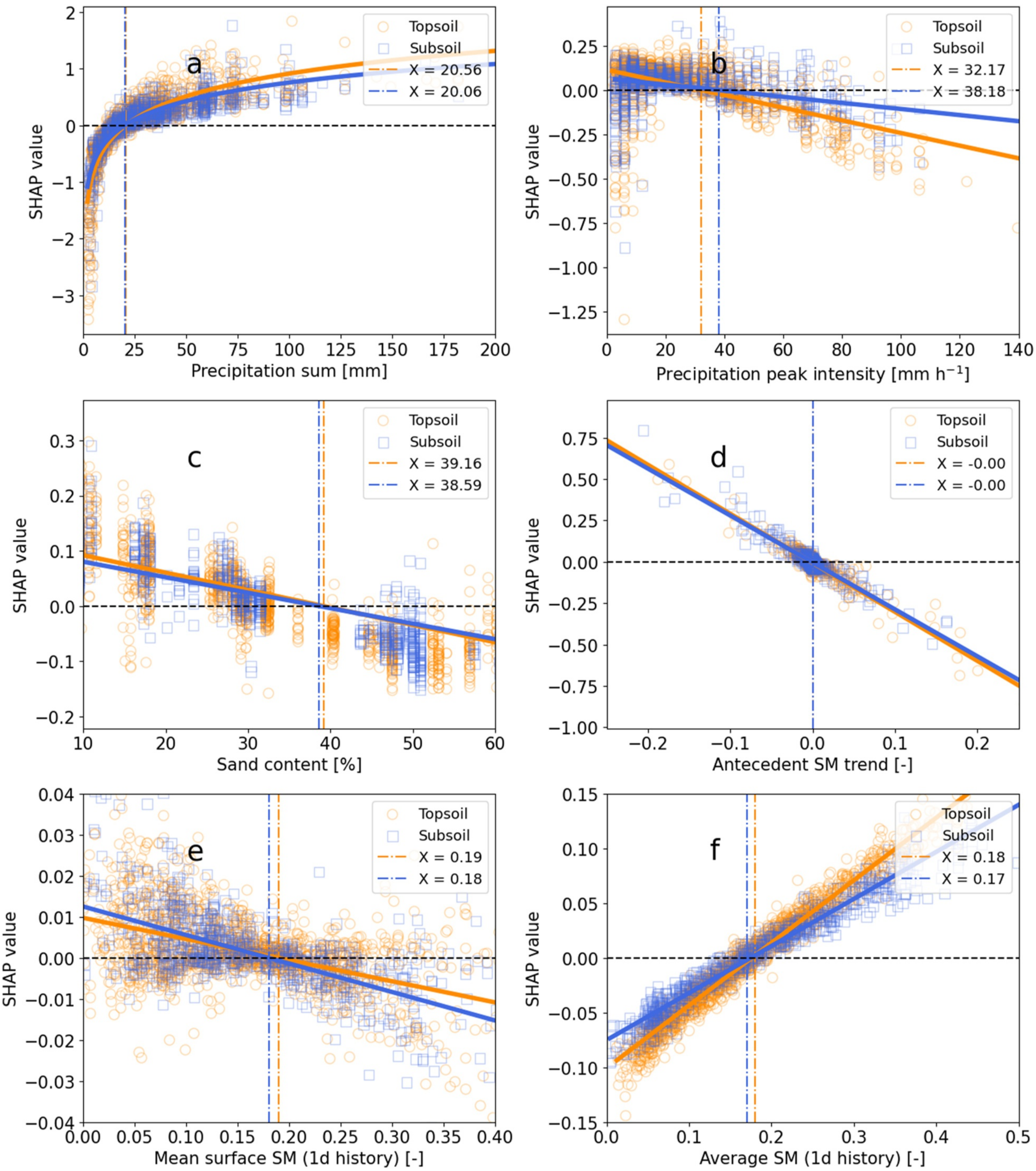

Figure 4. SHAP values for the soil water velocity detected at sensors between 10 and 100 cm as a function of precipitation sum (a), precipitation peak intensity (b), sand content (c), antecedent soil moisture trend (d), mean surface soil moisture (e), and antecedent soil moisture content at the sensor depth (f). Relationships are shown for the topsoil (10–35 cm, orange) and subsoil (35–100 cm, blue), respectively. The lines do not necessarily show significant trends, but are added for visualization. Vertical dashed lines represent the X-value for the intercept of the trend lines with a SHAP value = 0. All other variables are shown in Figure S9 in Supporting Information S1. Each point in these subpanels represents the averaged SHAP value of 50 models.

resulting in a negative effect on the model's prediction of soil water velocities. These trends of limited to no increase in PF occurrence or soil water velocity during high precipitation intensities have been reported by prior studies (Hu et al., 2019; Radolinski et al., 2021; Weiler & Naef, 2003). Our DL model results reveal that this pattern holds across a wide range of ecological and climatic conditions. The effect of precipitation duration on soil water velocities reflects how precipitation sum and intensity drive PF. Relatively short (<8 hr) precipitation events have a positive effect on soil water velocities (Figure S9k in Supporting Information S1), because they are most likely to result in initiation of fast flow paths, while long lasting precipitation events are often of lower intensity and, therefore more likely to get absorbed by the soil's matrix.

There was very little difference between topsoil and subsoil with regard to the variables that emerged as dominant features, as evident in the overall ranking of feature importances (Figure 3), and the directionality of variables and SHAP values in the partial dependence plots (Figure 4).

3.3. Higher Sand Content and Antecedent Soil Moisture Modulate Soil Water Flow

Sand content showed a comparable SHAP value to precipitation peak intensity and was the most important variable not related to precipitation characteristics (Figure 3). This suggests that soil texture provides crucial information for the model in explaining variation in soil water velocity across sites and depths. The linear trend indicated that decreasing sand content is associated with an increased predicted soil water velocity (Figure 4c). At sand contents below 40% there is a positive effect on soil water velocity, but at higher sand contents the soil water velocities are generally lower. This relationship is contrary to the common description of soil water flow based on Darcy-Richards conceptualization of a uniform matrix flow in which a low sand content (i.e., high clay content in general) would lead to reduced soil water velocities. Thus, our observations of soil water velocities do not reflect uniformly infiltrating wetting fronts, but rather the higher end of the soil water velocity distribution that reflects non-uniform heterogeneous bypass flow in response to precipitation events.

Clay content exhibits a threshold-like effect, with generally low SHAP values observed when clay content is below 30% (Figure S9 in Supporting Information S1). However, at clay contents above 30% there is a clear positive effect on water flow velocities. Such a positive effect of high clay contents on PF has long been observed in field experiments (Bouma, 1991; Flury et al., 1994; Grant et al., 2019) and we show that the spatial variability of PF and soil water velocities across the US is to a large degree due to soil texture, confirming earlier work that flow velocity in clay textured soil can be just as large as in non-structured sands due to the role of macropores (Jarvis et al., 2013). While the topsoil has generally lower clay content than the subsoil, the trend in SHAP values is similar for top- and subsoil water velocities.

The three antecedent soil moisture variables showed relatively low mean absolute SHAP values (Figure 3). However, we can nevertheless learn about the potential influence of moisture state on PF processes. The antecedent surface soil moisture had a positive effect on velocities for drier conditions (Figure 4e), which can be related to some degree to soil cracking (Dos Santos et al., 2016; Tang et al., 2021) and water repellency (Gimbel et al., 2016; Ritsema et al., 1993). However, wet antecedent conditions reduce predicted velocity, because infiltration capacity was likely already reached at high moisture conditions and hydraulic gradients were smaller.

Interestingly, the antecedent soil moisture content of the individual sensors showed the opposite trend as the topsoil antecedent moisture, because wetter conditions at the individual sensor depths generally had a positive effect on the predictions of soil water velocities (Figure 4f). Thus, at the subsurface level, wetter initial conditions enhance vertical connectivity and flow making it more likely for the soil moisture sensors to sense an increase in soil moisture to a precipitation event earlier and thus detect PF. Tracer-based studies have shown that water stored in the soil matrix can be mobilized and contribute to PF flows (Klaus et al., 2013; McDonnell, 1990), leading to higher flow velocities at higher moisture conditions. Various studies came to different conclusions about antecedent soil moisture with some studies showing an increase in PF occurrence with higher antecedent soil moisture across the upper 60 cm soil (e.g., Hu et al., 2019), while other studies found more PF under drier antecedent moisture conditions (e.g., Demir et al., 2022).

For antecedent moisture trends, which represent whether the soil was generally drying (negative slope) or wetting up (positive slope) at the depth of the soil moisture sensors right before a precipitation event, an interesting relationship emerged in which drying conditions contributed positively to the predicted soil water velocities, while wetting conditions tended to contribute negatively (Figure 4f). Thus, soils that were drying led to formation of PF paths when rewetting from infiltrating precipitation began. Soils that were wetting up prior to precipitation events, were less likely to experience PF or rapid soil water velocities. These patterns are expected in soils that contain some amount of 2:1 clays, which shrink and swell with changes in moisture. After a rainfall, large cracks in the soil can close quickly at the surface, making a previously open pathway for PF no longer available during future rain events (Stewart et al., 2016).

3.4. Other Factors Affecting the Water Flow Velocities

We found the mean absolute SHAP values of the remaining site-based hydroclimate state (e.g., aridity, MAT, MAP) and dynamics (e.g., ET, surface temp., SPEI3) variables, soil porosity, and vegetation variables (e.g., GPP, LAI) were generally of low importance and relatively similar to each other (Figure 3). We therefore refrain from interpreting small differences in their rankings as strong evidence, but these variables likely represent background environmental conditions that help explain why similar precipitation events can produce different soil water velocity responses across sites and depths. For example, at higher temperatures, both the land surface temperature (event-based, 1-day history) and the air temperature (MAT; static, annual average) contribute positively to the prediction of soil water velocities. While the site's air temperature has a positive effect on soil water velocities at MAT > 11°C, the surface temperature has a positive effect at >15°C (Figure S9 in Supporting Information S1). Importantly, MAT and land surface temperatures tend to be higher at sites where precipitation peak intensities are higher (Figure S8 in Supporting Information S1), which might explain part of the relative importance. The clear pattern of SHAP values for land surface temperatures could reflect an influence of soil evaporation and potentially hydrophobicity due to drying, which can lead to non-uniform flow (e.g., finger flow) pattern (Gimbel et al., 2016).

The soil porosity has generally a low impact on soil water velocities according to our model results and there is a wide spread in SHAP values (Figure S9i in Supporting Information S1). However, the model identifies a slight negative relationship with soil water flow velocity, which suggests that soils with higher total porosity are often dominated by high micro- or meso-porosity within the soil matrix, which exerts strong capillary suction (K. Beven & Germann, 1982; Hajnos et al., 2006; Nimmo et al., 2023). This high matrix suction actively abstracts infiltrating water laterally out of PF paths, reducing the bypass flow. Conversely, lower matrix porosity, when paired with stable structural macropores, prevents this lateral abstraction, focusing water fluxes into macropores.

The effect of vegetation on soil water velocity is represented in our model in two ways: the soil water velocity increased with increasing in GPP, which indicates that higher productivity accelerates velocities. More biomass likely increases root biopores leading to more macropore flow and root exudates can increase soil aggregate stability (Hamza et al., 2026). Soil water velocities were lower with higher leaf area, likely due to increased canopy interception that dampens infiltration intensity via throughfall during precipitation events and leads to overall lower infiltration volumes.

Generally, lower aridity favors the prediction of higher soil water velocities as indicated by a negative relationship with the SHAP values (Figure S9g in Supporting Information S1). In drier climate conditions, incoming precipitation might replenish the soil matrix water before PF pathways are activated, whereas lower aridity with sufficient moisture may favor hydrological connectivity, thereby faster soil water velocities.

When examining the most influential non-precipitation factors for the study sites grouped into four climate regions, a distinct pattern emerges (Figure 5): Soil texture and GPP contribute more to predicting water velocities in the Eastern US, while sand and clay content were most important in the central part of the US, and moisture-related variables (aridity and MAP) play a more critical role in the Western US. Thus, in the drier Western US, a lack of precipitation events and the strong evaporative demand overwrites the effects of soil texture or vegetation effects we saw east of the Rocky Mountains.

3.5. Water Flow Velocities Increase in a Future Climate Across CONUS

Under future climate scenarios, soil water velocities are predicted to increase by ∼7% for the RCP 4.5 (cooler) and increase by 15% for the RCP 8.5 (hotter) scenario (Table S2 in Supporting Information S1). The distribution of predicted soil water velocities shifted toward a greater density of higher water velocities under both scenarios (Figures 6a–6e). The RCP 8.5 hotter scenario resulted in the most pronounced increases in densities of high soil water velocity that remained fairly consistent across annual and seasonal timeframes. Of particular interest are results for the summer (JJA) season (Figure 6d), when soil water velocities under the RCP 8.5 hotter showed a wider range with both much higher and lower velocities than the historical and the RCP 4.5 cooler scenario. This distinctive behavior results from the wide range of projected changes in precipitation frequency and peak intensity during the JJA season under the RCP 8.5 hotter scenario, spanning from +100% to −50% (Figures S3 and S4 in

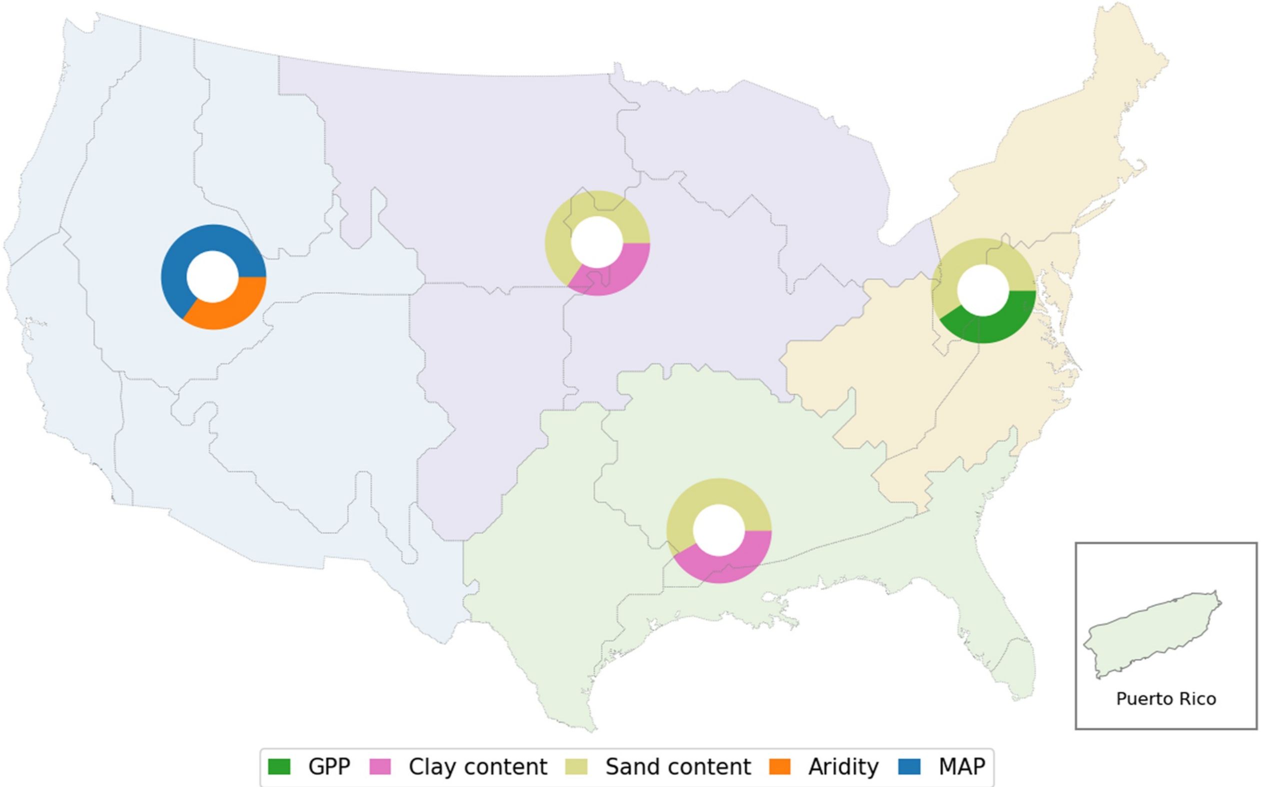

Figure 5. The most important non-precipitation features (absolute SHAP values) explaining soil water velocities for four different regions. Note that for this analysis, we excluded precipitation event characteristics that are generally the most important features (see Figure 3) and we do not distinguish between soil depths. Different colors reflect climate regions grouped according to Karl and Koss (1984).

Supporting Information S1). We consistently observe an increase in soil water velocities across all regions of the US (Figure 7), but the lowest increase in the Southeast, where subtropical rainfall conditions are already favorable for high soil moisture velocities as previously shown by B. Li et al. (2025). Our findings of increased soil water velocities add to predictions of an intensified water cycle due to increased soil microporosity (Hirmas et al., 2025).

Interestingly, we do not see much of a change of the projected probability of PF occurrence for any season (Figures 6f–6j) and neither for any of the considered regions (blue bars in Figure 7). The projections indicate a consistent, but slight (1%–3%) decline in PF occurrence on average across both climate scenarios and for all seasons. While the model could predict PF occurrence relatively well (Figure 2), there is large scatter in the relationship between predictions of soil water velocity and likelihood of PF (Figure S10 in Supporting Information S1); this suggests the model is missing key variables to better constrain soil water velocity, like soil structure, root effects, or organic matter. In addition, PF is inferred as soil water velocities exceeding a threshold of an expected matrix flow velocity (Hardie et al., 2011; Nimmo et al., 2025). Although soil water velocity and PF occurrence probability are linked, they may not change in a one to one manner, as a projected increase in soil water velocity would not necessarily exceed the threshold for that soil to lead to PF occurrence. Since both precipitation sum and intensity are projected in the climate scenarios to increase at all considered sites (Figures S1 and S2 in Supporting Information S1), the distribution of velocities is likely to shift upward toward higher soil water velocities (Figure 6), but this increase happens mostly for events that result in high soil water velocities that are already classified as PF (Figure S10 in Supporting Information S1). Changes at the lower soil water velocity distribution are relatively minor and thus the threshold from matrix flow dominated to PF is less likely to be reached.

In summary, the climate projections across all ecoregions in the contiguous US show a clear and strong increase in soil water flow rates, yet PF occurrence will probably not become more frequent. This means that water flow

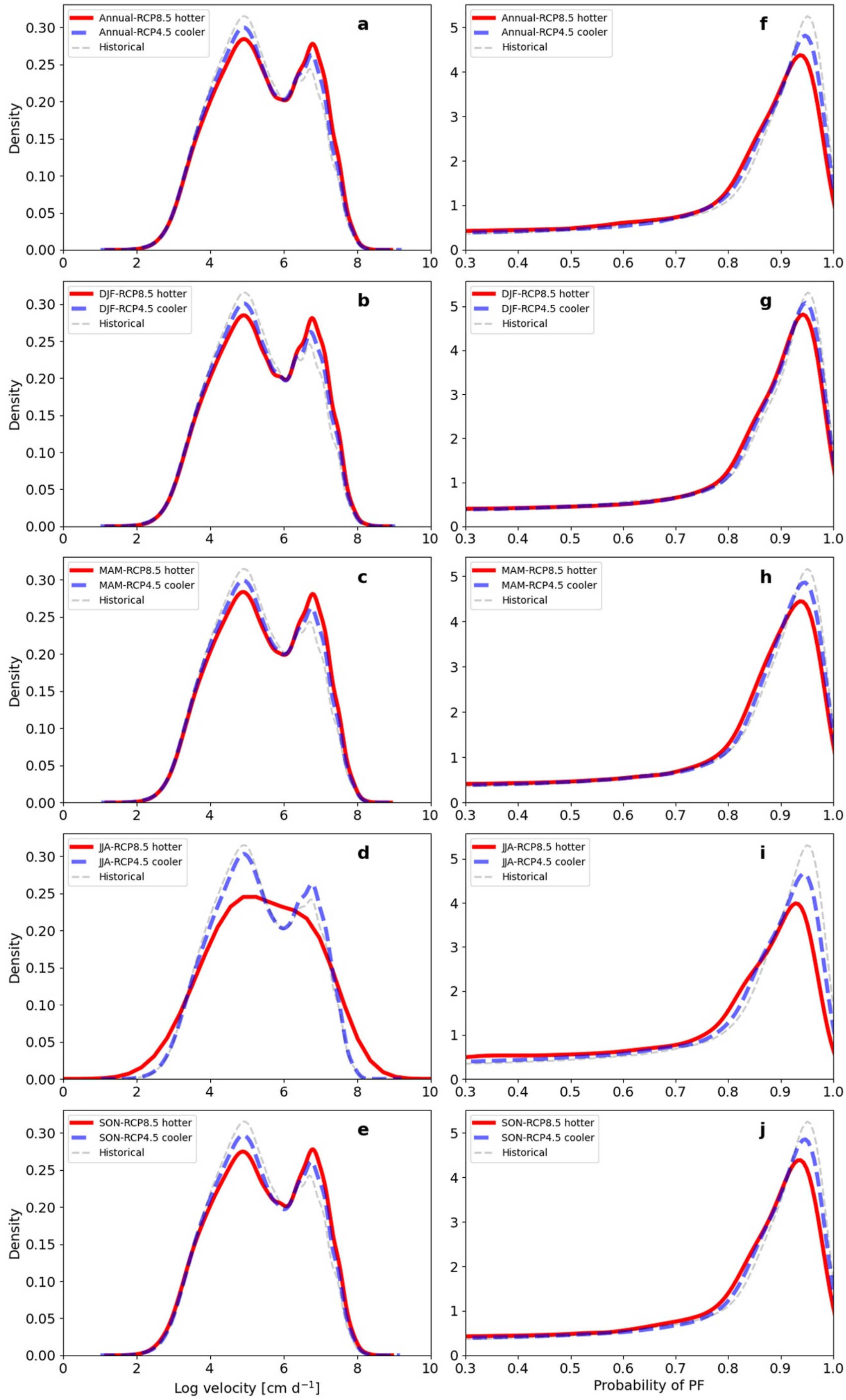

Figure 6. Comparison between the historical and future scenarios of soil water velocities (left panels a to e) and probability of preferential flow (PF) (right panels f to j) at annual (a and f) and seasonal (b–e and g–j) scale. The simulated probability density distribution of the log velocity and PF probability for historical (1980–2019) conditions are shown in grey dashed lines and future predictions (2060–2099) for Representative Concentration Pathways (RCP) 8.5 hotter and RCP 4.5 cooler are shown in red and blue lines, respectively. All precipitation events are included in "Annual" (a and f) and simulations by season are parsed as follows: winter, "DJF" (b and g); spring, "MAM" (c and h); summer, "JJA" (d and i) and fall, "SON" (e and j). Seasonal simulations are calculated by replacing mean annual temperature with seasonal mean temperature (Figure S3e in Supporting Information S1), MAP with seasonal mean precipitation (Figure S3f in Supporting Information S1), aridity with seasonal mean aridity (Figure S3g in Supporting Information S1), while precipitation event information remains the same as "Annual."

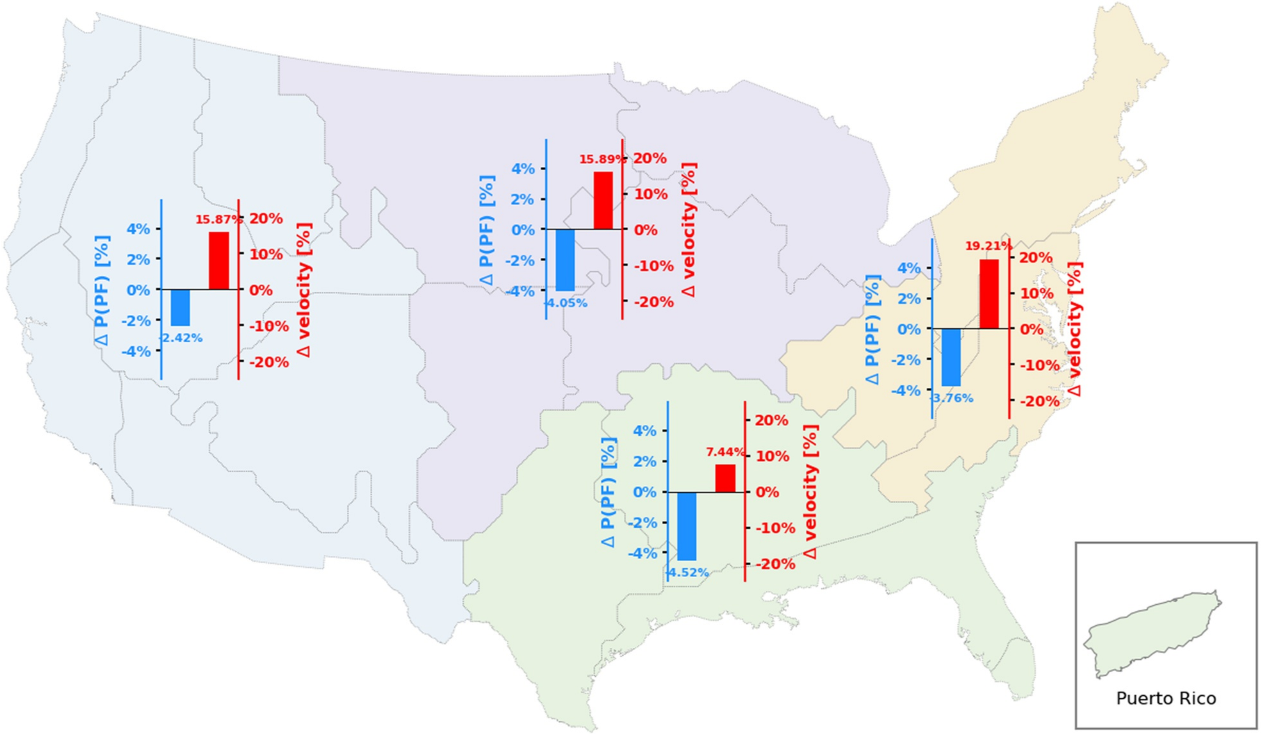

Figure 7. Relative change [%] of the probability of preferential flow occurrence (blue bars) and changes in soil water velocity (red bars) of the four climate regions between historical (1980–2019) and end of century (2060–2099) period based on the Representative Concentration Pathways (RCP) 8.5 hotter climate scenario. See Supporting Information S1 for seasonal changes and the results for the RCP 4.5 cooler scenario (Figure S11 in Supporting Information S1). The outlines show National Ecological Observatory Network ecoregions and the different hues reflect climate regions grouped according to Karl and Koss (1984).

associated with PF occurrence will become faster, which will likely increase the share of soil water bypassing the root zone storage, potentially increasing groundwater recharge as well as nutrient and contaminant transport (Vereecken et al., 2022) and biogeochemical cycling (Franklin et al., 2021). PF pathways occupy only a small percentage of the soil volume, yet currently they can account for over 70% of soil water flow (Watson & Luxmoore, 1986; X. Zhang et al., 2019) and often contain elevated concentrations of soil organic C and nutrients relative to the surrounding matrix. While PF can mobilize substantial quantities of water and C, the fate of C within these pathways is tightly coupled to the broader moisture status and microbial function of soils (Schimel, 2018; Singh et al., 2023; Vargas et al., 2018). Microbial decomposition rates peak under intermediate moisture conditions where microbial activity is optimized (Davidson et al., 2012). Whereas, the effects of dry antecedent conditions on C remains less clear. Under dry conditions, PF could re-activate reaction sites by delivering water and C for microbial processing (Homyak et al., 2018) and remove processed C. Alternatively, PF following dry soil conditions could promote rapid C transport with limited time for microbial processing and possible cell lysis (Schimel, 2018; Singh et al., 2023), decreasing the amount of processed C at depth. Specifically, rapid re-wetting facilitated by PF can cause lysis of microbial cells (Griffiths & Birch, 1961), decreasing the potential for C processing prior to export yet potentially mobilizing released metabolites. Further, high soil water velocities could promote bypass of the root water storage zone where C cycling is most active. Thus, higher soil water velocities through pathways of PF combined with shifts in antecedent soil moisture may exert an outsized influence on whether PF enhances or bypasses microbial decomposition, ultimately shaping C retention and loss in terrestrial ecosystems (Crowther et al., 2019; Schimel, 2018) and informing nutrient and contaminant transport more broadly.

It is important to note that soil texture and porosity were not dynamically updated in the velocity and probability predictions for historical and future scenarios. We also acknowledge that soil structure, organic matter cycling, and root growth can interact dynamically and influence long-term soil water balance (Jarvis et al., 2024). However, we must assume that soil texture and porosity do not evolve between historical and future scenarios in our reduced-model framework due to a lack of data. Therefore, it might be more appropriate to interpret the

10.1029/2026EF008296

projected changes in velocity and PF probability as the response to precipitation and hydroclimatic changes under static soil-structural conditions, rather than a fully coupled land-surface projection.

4. Conclusion

This study used a DL modeling approach to derive key predictive controls of PF and soil water velocity and make projections for future climate scenarios. Generally, the DL model was very successful at predicting PF occurrence, while predicting event-based soil water velocities remained challenging. Not all event-based variables were more informative (i.e., soil moisture related) for predicting soil water velocities than static variables. Somewhat expected, the precipitation characteristics were by far the major drivers of soil water velocity variation in time and space. Thus, as hypothesized, event-based precipitation variables (volume, duration, intensity) were more informative for predicting soil water velocities than site-based variables (e.g., mean annual precipitation), but contrary to our hypothesis antecedent soil moisture characteristics did not play a major role. However, soil texture, especially sand content, was an important variable and showed that decreasing sand content (increased clay content) results in an increase in soil water velocities. This pattern is contrary to uniform subsurface water flow and thus does not support the usual assumption in the conceptualization of subsurface flow in most hydrological, biogeochemical, and Earth System models.

Contrary to our hypothesis, we did not find evidence for major differences in dominant predictive controls of soil water velocity between topsoil (10–35 cm) and subsoil (35–100 cm). While topsoils tended to have lower clay content, the feature importance as well as the relationships between the various predictors and soil water velocities were consistent for both depth classes.

A major motivation of our study was to assess the influence of a future climate with increased precipitation intensity and this exercise supported our hypothesis that we should expect higher soil water velocities for a wide range of climate scenarios with precipitation intensity being an important aspect. However, also the general increase in volume of future precipitation events is a major driver for soil water velocities during end-of-the century rainfall events compared to historical conditions. While our model predicted soil water velocities associated with PF to increase by about 15% for precipitation events at the end of the century, there was no increase in the likelihood of PF occurring. This implies that, while the frequency of PF will not increase, the water flow through PF paths may increase across all considered study sites. As a consequence, we will expect a stronger potential for recharge of the groundwater, reduced root zone storage, and a greater likelihood of nutrient and contamination transport to the aquifers. Models that do not represent these processes are therefore likely to underestimate their effects in simulations of current conditions, and even more so when predictions of future conditions are made.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The soil moisture and precipitation data sets are available at NEON (2023c) at https://data.neonscience.org/data-products/DP1.00094.001 and NEON (2023a) at https://data.neonscience.org/data-products/DP1.00006.001, respectively. The NEON site meta data is available at NEON (2024a) at https://www.neonscience.org/field-sites/explore-field-sites. The NEON aridity information is available in Fiorella et al. (2021). The soil physical information is obtained from NEON (2023b, 2024b) at https://data.neonscience.org/data-products/DP1.00096.001 and https://data.neonscience.org/data-products/DP1.10047.001, respectively. The Global Land Data Assimilation System (GLDAS) Noah Land Surface Model L4 3-hourly 0.25×0.25 Version 2.1 data is available in Beaudoing et al. (2020) and Rodell et al. (2004) at https://ldas.gsfc.nasa.gov/gldas/. The IM3/HyperFACETS Thermodynamic Global Warming Simulation Data sets at 12-km resolution is available in Jones et al. (2022, 2023) at https://tgw-data.msdlive.org/. The MODIS Leaf area index (LAI, MCD15A3H) and MODIS gross primary productivity (GPP, MOD17A2HGF Version 6.1) data products are available in Myneni et al. (2015) at https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MCD15A3H and Running and Zhao (2021) at https://www.earthdata.nasa.gov/data/catalog/lpcloud-mod17a2hgf-061, respectively. The software used to calculate

PET is available in Vremec et al. (2024) at https://pypi.org/project/pyet/. The software used to calculate SPEI3 is available in Vonk (2024) at https://doi.org/10.5281/ZENODO.10816741.

Acknowledgments

This work was conducted as part of the AI4PF Working Group (Using a network of networks for high-frequency multi-depth soil moisture observations to infer spatial and temporal drivers of subsurface preferential flow) supported by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) and the John Wesley Powell Center for Analysis and Synthesis, funded by the U.S. Geological Survey. MS was supported by the U.S. Department of Energy Office of Science under contract DE-AC02-05CH11231 as part of Lawrence Berkeley National Laboratory Watershed Function Science Focus Area. PLS acknowledges support from the U.S. National Science Foundation (EAR-2012796, EAR-2121694). JG acknowledges the support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project no. 460817082). We acknowledge the data provided by National Ecological Observatory Network (NEON), a program sponsored by the U.S. National Science Foundation and operated by Battelle. The manuscript was made open access thanks to funds from CUAHSI.

References

Allaire, S. E., Roulier, S., & Cessna, A. J. (2009). Quantifying preferential flow in soils: A review of different techniques. *Journal of Hydrology*, *378*(1–2), 179–204. https://doi.org/10.1016/j.jhydrol.2009.08.013

Beaudoing, H., & Rodell, M., & NASA/GSFC/HSL. (2020). GLDAS Noah land surface model L4 3 hourly 0.25 x 0.25 degree, version 2.1 [Dataset]. *NASA Goddard Earth Sciences Data and Information Services Center*. https://doi.org/10.5067/E7TYRXPJKWOQ

Bernhardt, E. S., Blaszczak, J. R., Ficken, C. D., Fork, M. L., Kaiser, K. E., & Seybold, E. C. (2017). Control points in ecosystems: Moving beyond the hot spot hot moment concept. *Ecosystems*, *20*(4), 665–682. https://doi.org/10.1007/s10021-016-0103-y

Beven, K., & Germann, P. (1981). Water flow in soil macropores II. A combined flow model. *Journal of Soil Science*, *32*(1), 15–29. https://doi.org/10.1111/j.1365-2389.1981.tb01682.x

Beven, K., & Germann, P. (1982). Macropores and water flow in soils. *Water Resources Research*, *18*(5), 1311–1325. https://doi.org/10.1029/WR018i005p01311

Beven, K. J. (2018). On hypothesis testing in hydrology: Why falsification of models is still a really good idea. *WIREs Water*, *5*(3), e1278. https://doi.org/10.1002/wat2.1278

Bouma, J. (1991). Influence of soil macroporosity on environmental quality. *Advances in Agronomy*, *46*, 1–37. https://doi.org/10.1016/S0065-2113(08)60577-5

Clothier, B. E., Green, S. R., & Deurer, M. (2008). Preferential flow and transport in soil: Progress and prognosis. *European Journal of Soil Science*, *59*(1), 2–13. https://doi.org/10.1111/j.1365-2389.2007.00991.x

Crowther, T. W., Van Den Hoogen, J., Wan, J., Mayes, M. A., Keiser, A. D., Mo, L., et al. (2019). The global soil community and its influence on biogeochemistry. *Science*, *365*(6455), eaav0550. https://doi.org/10.1126/science.aav0550

Davidson, E. A., Samanta, S., Caramori, S. S., & Savage, K. (2012). The dual Arrhenius and Michaelis–Menten kinetics model for decomposition of soil organic matter at hourly to seasonal time scales. *Global Change Biology*, *18*(1), 371–384. https://doi.org/10.1111/j.1365-2486.2011.02546.x

Demand, D., Blume, T., & Weiler, M. (2019). Spatio-temporal relevance and controls of preferential flow at the landscape scale. *Hydrology and Earth System Sciences*, *23*(11), 4869–4889. https://doi.org/10.5194/hess-23-4869-2019

Demir, G., Michalzik, B., Filipzik, J., Metzger, J. C., & Hildebrandt, A. (2022). Spatial variation of grassland canopy affects soil wetting patterns and preferential flow. *Hydrological Processes*, *36*(12), e14760. https://doi.org/10.1002/hyp.14760

Dos Santos, J. C. N., De Andrade, E. M., Guerreiro, M. J. S., Medeiros, P. H. A., De Queiroz Palácio, H. A., & De Araújo Neto, J. R. (2016). Effect of dry spells and soil cracking on runoff generation in a semiarid micro watershed under land use change. *Journal of Hydrology*, *541*, 1057–1066. https://doi.org/10.1016/j.jhydrol.2016.08.016

Fiorella, R. P., Good, S. P., Allen, S. T., Guo, J. S., Still, C. J., Noone, D. C., et al. (2021). Calibration strategies for detecting macroscale patterns in NEON atmospheric carbon isotope observations. *Journal of Geophysical Research: Biogeosciences*, *126*(3), e2020JG005862. https://doi.org/10.1029/2020JG005862

Flury, M., Flühler, H., Jury, W. A., & Leuenberger, J. (1994). Susceptibility of soils to preferential flow of water: A field study. *Water Resources Research*, *30*(7), 1945–1954. https://doi.org/10.1029/94WR00871

Franklin, S. M., Kravchenko, A. N., Vargas, R., Vasilas, B., Fuhrmann, J. J., & Jin, Y. (2021). The unexplored role of preferential flow in soil carbon dynamics. *Soil Biology and Biochemistry*, *161*(February), 108398. https://doi.org/10.1016/j.soilbio.2021.108398

Gimbel, K. F., Puhlmann, H., & Weiler, M. (2016). Does drought alter hydrological functions in forest soils? *Hydrology and Earth System Sciences*, *20*(3), 1301–1317. https://doi.org/10.5194/hess-20-1301-2016

Grant, K. N., Macrae, M. L., & Ali, G. A. (2019). Differences in preferential flow with antecedent moisture conditions and soil texture: Implications for subsurface P transport. *Hydrological Processes*, *33*(15), 2068–2079. https://doi.org/10.1002/hyp.13454

Griffiths, E., & Birch, H. F. (1961). Microbiological changes in freshly moistened soil. *Nature*, *189*(4762), 424. https://doi.org/10.1038/189424a0

Guthrie, A., Hirmas, D., Sullivan, P. L., Li, L., Ajami, H., Singha, K., et al. (2025). Deep root loss and regeneration in the Anthropocene drive continental-scale changes in deep soil structure. *Earth's Future*, *13*(9), e2025EF006326. https://doi.org/10.1029/2025EF006326

Hajnos, M., Lipiec, J., Świeboda, R., Sokołowska, Z., & Witkowska-Walczak, B. (2006). Complete characterization of pore size distribution of tilled and orchard soil using water retention curve, Mercury Porosimetry, nitrogen adsorption, and water desorption methods. *Geoderma*, *135*, 307–314. https://doi.org/10.1016/j.geoderma.2006.01.010

Hamza, A., Karčauskienė, D., Tahir, M. A., Mockevičienė, I., Noor-Us-Sabah, Repšienė, R., & Skuodienė, R. (2026). Root-soil interactions a key driver of aggregate formation and stability: A trait-based and mechanistic review. *Plant and Soil*, *521*(2), 1037–1066. https://doi.org/10.1007/s11104-026-08453-3

Hardie, M. A., Cotching, W. E., Doyle, R. B., Holz, G., Lisson, S., & Mattern, K. (2011). Effect of antecedent soil moisture on preferential flow in a texture-contrast soil. *Journal of Hydrology*, *398*(3–4), 191–201. https://doi.org/10.1016/j.jhydrol.2010.12.008

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv*. https://doi.org/10.48550/ARXIV.1512.03385

Hirmas, D. R., Ajami, H., Sena, M. G., Zhang, X., Cao, X., Li, B., et al. (2025). Predicting soil interpedal macroporosity and hydraulic conductivity dynamics: A model for integrating laser-scanned profile imagery with soil moisture sensor data. *Water Resources Research*, *61*(9), e2024WR039836. https://doi.org/10.1029/2024WR039836

Hofmeister, K. L., Feitl, G., Tackett, K., Fleury, L., & Nave, L. E. (2025). Land-use effects on soil organic matter and related soil properties in a mixed agricultural–forest landscape of central Wisconsin, USA. *Soil Science Society of America Journal*, *89*(1), e70009. https://doi.org/10.1002/saj2.70009

Homyak, P. M., Blankinship, J. C., Slessarev, E. W., Schaeffer, S. M., Manzoni, S., & Schimel, J. P. (2018). Effects of altered dry season length and plant inputs on soluble soil carbon. *Ecology*, *99*(10), 2348–2362. https://doi.org/10.1002/ecy.2473

Hu, H., Wen, J., Peng, Z., Tian, F., Tie, Q., Lu, Y., & Khan, M. Y. A. (2019). High-frequency monitoring of the occurrence of preferential flow on hillslopes and its relationship with rainfall features, soil moisture and landscape. *Hydrological Sciences Journal*, *64*(11), 1385–1396. https://doi.org/10.1080/02626667.2019.1638513

Jarecke, K. M., Bladon, K. D., Meinzer, F. C., & Wondzell, S. M. (2024). Impact of rainfall and vapor pressure deficit on latewood growth and water stress in Douglas-fir in a Mediterranean climate. *Forest Ecology and Management*, *551*, 121529. https://doi.org/10.1016/j.foreco.2023.121529

Jarvis, N., Coucheney, E., Lewan, E., Klöffel, T., Meurer, K. H. E., Keller, T., & Larsbo, M. (2024). Interactions between soil structure dynamics, hydrological processes, and organic matter cycling: A new soil-crop model. *European Journal of Soil Science*, *75*(2), e13455. https://doi.org/10.1111/ejss.13455

Jarvis, N., Koestel, J., Messing, I., Moeys, J., & Lindahl, A. (2013). Influence of soil, land use and climatic factors on the hydraulic conductivity of soil. *Hydrology and Earth System Sciences*, *17*(12), 5185–5195. https://doi.org/10.5194/hess-17-5185-2013

Jarvis, N. J. (2007). A review of non-equilibrium water flow and solute transport in soil macropores: Principles, controlling factors and consequences for water quality. *European Journal of Soil Science*, *58*(3), 523–546. https://doi.org/10.1111/j.1365-2389.2007.00915.x

Jones, A. D., Rastogi, D., Vahmani, P., Stansfield, A., Reed, K., Thurber, T., et al. (2022). IM3/HyperFACETS thermodynamic global warming (TGW) simulation datasets [Dataset]. *MultiSector Dynamics–Living, Intuitive, Value-Adding, Environment*. https://doi.org/10.57931/1885756

Jones, A. D., Rastogi, D., Vahmani, P., Stansfield, A. M., Reed, K. A., Thurber, T., et al. (2023). Continental United States climate projections based on thermodynamic modification of historical weather. *Scientific Data*, *10*(1), 664. https://doi.org/10.1038/s41597-023-02485-5

Kang, W., Tian, J., Reemt Bogena, H., Lai, Y., Xue, D., & He, C. (2023). Soil moisture observations and machine learning reveal preferential flow mechanisms in the Qilian Mountains. *Geoderma*, *438*, 116626. https://doi.org/10.1016/j.geoderma.2023.116626

Karl, T., & Koss, W. J. (1984). *Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983*. National Climatic Data Center. Retrieved from https://repository.library.noaa.gov/view/noaa/10238

Kesting, H. M., & Allen, S. T. (2024). The roles of winter versus summer precipitation in supplying evapotranspiration across US ecoregions: A cross-catchment end-member-splitting analysis. *Geophysical Research Letters*, *51*(12), e2024GL109112. https://doi.org/10.1029/2024GL109112

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. *arXiv*. https://doi.org/10.48550/arXiv.1412.6980

Klaus, J., Zehe, E., Elsner, M., Külls, C., & McDonnell, J. J. (2013). Macropore flow of old water revisited: Experimental insights from a tile-drained hillslope. *Hydrology and Earth System Sciences*, *17*(1), 103–118. https://doi.org/10.5194/hess-17-103-2013

Koop, A. N., Hirmas, D. R., Billings, S. A., Li, L., Cueva, A., Zhang, X., et al. (2023). Is macroporosity controlled by complexed clay and soil organic carbon? *Geoderma*, *437*, 116565. https://doi.org/10.1016/j.geoderma.2023.116565

Li, B., Sprenger, M., Wyatt, B. M., Giménez, D., Hirmas, D. R., Ajami, H., et al. (2025). Ubiquity and causes of soil water preferential flow across 17 ecoregions. *Geophysical Research Letters*, *52*(19), e2025GL118045. https://doi.org/10.1029/2025GL118045

Li, C., Zwiers, F., Zhang, X., Li, G., Sun, Y., & Wehner, M. (2021). Changes in annual extremes of daily temperature and precipitation in CMIP6 models. *Journal of Climate*, *34*(9), 3441–3460. https://doi.org/10.1175/JCLI-D-19-1013.1

Li, M., Yao, J., & Cheng, J. (2020). Study on the preferential flow characteristics under different precipitation amounts in simian Mountain grassland of China. *Water*, *12*(12), 3489. https://doi.org/10.3390/w12123489

Lin, H., & Zhou, X. (2008). Evidence of subsurface preferential flow using soil hydrologic monitoring in the Shale Hills catchment. *European Journal of Soil Science*, *59*(1), 34–49. https://doi.org/10.1111/j.1365-2389.2007.00988.x

Lu, J., Zhang, Q., Werner, A. D., Li, Y., Jiang, S., & Tan, Z. (2020). Root-induced changes of soil hydraulic properties – A review. *Journal of Hydrology*, *589*, 125203. https://doi.org/10.1016/j.jhydrol.2020.125203

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *arXiv*. https://doi.org/10.48550/ARXIV.1705.07874

Massoud, E. C., Hoffman, F., Shi, Z., Tang, J., Alhajjar, E., Barnes, M., et al. (2023). Perspectives on artificial intelligence for predictions in ecohydrology. *Artificial Intelligence for the Earth Systems*, *2*(4), e230005. https://doi.org/10.1175/AIES-D-23-0005.1

McClain, M. E., Boyer, E. W., Dent, C. L., Gergel, S. E., Grimm, N. B., Groffman, P. M., et al. (2003). Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. *Ecosystems*, *6*(4), 301–312. https://doi.org/10.1007/s10021-003-0161-9

McDonnell, J. J. (1990). A rationale for old water discharge through macropores in a steep, humid catchment. *Water Resources Research*, *26*(11), 2821–2832. https://doi.org/10.1029/WR026i011p02821

Myneni, R., Knyazikhin, Y., & Park, T. (2015). MCD15A3H MODIS/terra+aqua leaf area Index/FPAR 4-day L4 global 500m SIN grid V006 [Dataset]. *NASA Land Processes Distributed Active Archive Center*. https://doi.org/10.5067/MODIS/MCD15A3H.006

NEON. (2023a). Precipitation (DP1.00006.001) [Dataset]. *National Ecological Observatory Network (NEON)*. https://doi.org/10.48443/DXDR-W376

NEON. (2023b). Soil physical and chemical properties, Megapit (DP1.00096.001) [Dataset]. *National Ecological Observatory Network (NEON)*. https://doi.org/10.48443/T70Z-NP08

NEON. (2023c). Soil water content and water salinity (DP1.00094.001) [Dataset]. *National Ecological Observatory Network (NEON)*. https://doi.org/10.48443/7EXD-N727

NEON. (2024a). Explore field sites [Dataset]. Retrieved from https://www.neonscience.org/field-sites/explore-field-sites

NEON. (2024b). Soil physical and chemical properties, distributed initial characterization (DP1.10047.001) [Dataset]. *National Ecological Observatory Network (NEON)*. https://doi.org/10.48443/r517-kg13

Nimmo, J. R. (2016). Quantitative framework for preferential flow initiation and partitioning. *Vadose Zone Journal*, *15*(2), 1–12. https://doi.org/10.2136/vzj2015.05.0079

Nimmo, J. R., Katuwal, S., & Lucas, M. (2023). Porosity and pore-size distribution. In *Encyclopedia of soils in the environment* (pp. 16–24). Elsevier. https://doi.org/10.1016/B978-0-12-822974-3.00139-7

Nimmo, J. R., Wiekenkamp, I., Araki, R., Groh, J., Singh, N. K., Crompton, O., et al. (2025). Identifying preferential flow from soil moisture time series: Review of methodologies. *Vadose Zone Journal*, *24*(2), e70017. https://doi.org/10.1002/vzj2.70017

Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P., & Holland, G. J. (2017). The future intensification of hourly precipitation extremes. *Nature Climate Change*, *7*(1), 48–52. https://doi.org/10.1038/nclimate3168

Radolinski, J., Le, H., Hilaire, S. S., Xia, K., Scott, D., & Stewart, R. D. (2022). A spectrum of preferential flow alters solute mobility in soils. *Scientific Reports*, *12*, 1–11. https://doi.org/10.1038/s41598-022-08241-w

Radolinski, J., Pangle, L., Klaus, J., & Stewart, R. D. (2021). Testing the two water worlds hypothesis under variable preferential flow conditions. *Hydrological Processes*, *35*(6), 1–14. https://doi.org/10.1002/hyp.14252

Ritsema, C. J., Dekker, L. W., Hendrickx, J. M. H., & Hamminga, W. (1993). Preferential flow mechanism in a water repellent sandy soil. *Water Resources Research*, *29*(7), 2183–2193. https://doi.org/10.1029/93WR00394

Robinson, D. A., Jones, S. B., Lebron, I., Reinsch, S., Domínguez, M. T., Smith, A. R., et al. (2016). Experimental evidence for drought induced alternative stable states of soil moisture. *Scientific Reports*, *6*(1), 20018. https://doi.org/10.1038/srep20018

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., et al. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, *85*(3), 381–394. https://doi.org/10.1175/BAMS-85-3-381

Running, S., & Zhao, M. (2021). MODIS/terra gross primary productivity gap-filled 8-Day L4 global 500m SIN grid V061 [Dataset]. *NASA Land Processes Distributed Active Archive Center*. https://doi.org/10.5067/MODIS/MOD17A2HGF.061

Rutgers, M., Orgiazzi, A., Gardi, C., Römbke, J., Jänsch, S., Keith, A. M., et al. (2016). Mapping earthworm communities in Europe. *Applied Soil Ecology*, *97*, 98–111. https://doi.org/10.1016/j.apsoil.2015.08.015

Schimel, J. P. (2018). Life in dry soils: Effects of drought on soil microbial communities and processes. *Annual Review of Ecology, Evolution, and Systematics*, *49*(1), 409–432. https://doi.org/10.1146/annurev-ecolsys-110617-062614

Schlüter, S., Roussety, T., Rohe, L., Guliyev, V., Blagodatskaya, E., & Reitz, T. (2022). Land use impact on carbon mineralization in well aerated soils is mainly explained by variations of particulate organic matter rather than of soil structure. *SOIL*, *8*(1), 253–267. https://doi.org/10.5194/soil-8-253-2022

Singh, S., Mayes, M. A., Kivlin, S. N., & Jagadamma, S. (2023). How the Birch effect differs in mechanisms and magnitudes due to soil texture. *Soil Biology and Biochemistry*, *179*, 108973. https://doi.org/10.1016/j.soilbio.2023.108973

Soil Science Division Staff. (2017). *Soil survey manual*. United States Department of Agriculture.

Sprenger, M., Saia, S., Dennedy-Frank, P., & Woodburn, E. (2021). Preferential flow in subsurface hydrology: From a century of denial to a decade of addressing it via ML? (no. AI4ESP1125, 1769765). https://doi.org/10.2172/1769765

Sprenger, M., Stumpp, C., Weiler, M., Aeschbach, W., Allen, S. T., Benettin, P., et al. (2019). The demographics of water: A review of water ages in the critical zone. *Reviews of Geophysics*, *57*(3), 800–834. https://doi.org/10.1029/2018RG000633

Stewart, R. D., Rupp, D. E., Abou Najm, M. R., & Selker, J. S. (2016). A unified model for soil shrinkage, subsidence, and cracking. *Vadose Zone Journal*, *15*(3), 1–15. https://doi.org/10.2136/vzj2015.11.0146

Sullivan, P. L., Billings, S. A., Hirmas, D., Li, L., Zhang, X., Ziegler, S., et al. (2022). Embracing the dynamic nature of soil structure: A paradigm illuminating the role of life in critical zones of the Anthropocene. *Earth-Science Reviews*, *225*, 103873. https://doi.org/10.1016/j.earscirev.2021.103873

Tang, C.-S., Zhu, C., Cheng, Q., Zeng, H., Xu, J.-J., Tian, B.-G., & Shi, B. (2021). Desiccation cracking of soils: A review of investigation approaches, underlying mechanisms, and influencing factors. *Earth-Science Reviews*, *216*, 103586. https://doi.org/10.1016/j.earscirev.2021.103586

Van Schaik, L., Palm, J., Klaus, J., Zehe, E., & Schröder, B. (2014). Linking spatial earthworm distribution to macropore numbers and hydrological effectiveness. *Ecohydrology*, *7*(2), 401–408. https://doi.org/10.1002/eco.1358

Vargas, R., Sánchez-Cañete, P. E., Serrano-Ortiz, P., Curiel Yuste, J., Domingo, F., López-Ballesteros, A., & Oyonarte, C. (2018). Hot-moments of soil CO2 efflux in a water-limited grassland. *Soil Systems*, *2*(3), 47. https://doi.org/10.3390/soilsystems2030047

Vereecken, H., Amelung, W., Bauke, S. L., Bogena, H., Brüggemann, N., Montzka, C., et al. (2022). Soil hydrology in the Earth system. *Nature Reviews Earth and Environment*, *3*(9), 573–587. https://doi.org/10.1038/s43017-022-00324-6

Vonk, M. A. (2024). SPEI: A simple python package to calculate and visualize drought indices (version v0.4.0) [Software]. *Zenodo*. https://doi.org/10.5281/ZENODO.10816741

Vremec, M., Collenteur, R. A., & Birk, S. (2024). PyEt v1.3.1: A python package for the estimation of potential evapotranspiration. *Geoscientific Model Development*, *17*(18), 7083–7103. https://doi.org/10.5194/gmd-17-7083-2024

Watson, K. W., & Luxmoore, R. J. (1986). Estimating macroporosity in a forest watershed by use of a tension infiltrometer. *Soil Science Society of America Journal*, *50*(3), 578–582. https://doi.org/10.2136/sssaj1986.03615995005000030007x

Weiler, M., & Naef, F. (2003). An experimental tracer study of the role of macropores in infiltration in grassland soils. *Hydrological Processes*, *17*(2), 477–493. https://doi.org/10.1002/Hyp.1136

Wiekenkamp, I., Huisman, J. A., Bogena, H. R., Lin, H. S., & Vereecken, H. (2016). Spatial and temporal occurrence of preferential flow in a forested headwater catchment. *Journal of Hydrology*, *534*, 139–149. https://doi.org/10.1016/j.jhydrol.2015.12.050

Zhang, X., Zhu, J., Wendroth, O., Matocha, C., & Edwards, D. (2019). Effect of macroporosity on pedotransfer function estimates at the field scale. *Vadose Zone Journal*, *18*(1), 1–15. https://doi.org/10.2136/vzj2018.08.0151

Zhang, Y., & Schaap, M. G. (2017). Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *Journal of Hydrology*, *547*, 39–53. https://doi.org/10.1016/j.jhydrol.2017.01.004