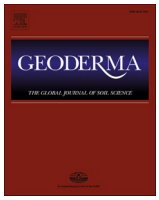
Geoderma 438 (2023) 116626

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma



Soil moisture observations and machine learning reveal preferential flow mechanisms in the Qilian Mountains

Weiming Kang [a,d], Jie Tian [a,d,*], Heye Reemt Bogena [b], Yao Lai [a,d], Dongxiang Xue [a,d], Chansheng He [a,c,d]

[a] Key Laboratory of West China's Environmental System (Ministry of Education), College of Earth and Environmental Sciences, Lanzhou University, Lanzhou, Gansu 730000, China
[b] Agrosphere Institute (IBG-3), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
[c] Department of Geography, Environment, and Tourism, Western Michigan University, Kalamazoo, MI 49008, USA
[d] Dayekou Hydrological Process Observation and Research Station, Lanzhou University, Lanzhou, Gansu 730000, China

ARTICLE INFO

Handling Editor: Morgan Cristine L.S.

Keywords:
Preferential Flow
Machine learning
Occurrence pattern
Mountainous areas

ABSTRACT

The complexity of the spatial distribution and temporal occurrence of preferential flow (PF) makes it challenging to understand the mechanisms of PF. This study aims to identify the spatial and temporal patterns of PF occurrence using machine learning (Classification and Regression Trees and Random Forests) in the Qilian Mountains, Northwest China. Our results show that detected PF events transport much more rainfall down to the subsoil than non-PF events. Different vegetation types exhibit variations in the main soil layers where PF occurs, which is closely related to the distribution of roots. The PF proportion varies significantly both vertically and horizontally. Based on the Random Forests, we found that the spatial distribution of the PF proportion is mainly controlled by the saturated hydraulic conductivity and residual soil moisture, which cannot be identified by conventional correlation analysis methods. With these soil properties, the spatial distribution of the PF proportion can be estimated with reasonable performance. Using the Classification and Regression Trees method, we identified the temporal occurrence pattern of the PF for different vegetation types and all observation stations. Results indicate that the dominant factors controlling the temporal occurrence of the PF varied for different vegetation types. The thresholds at which these factors initiate the PF also varied. Finally, we found that the PF occurs particularly under wet conditions (except for hydrophobic soils), under denser vegetation, and under conditions of high rainfall amount and intensity, regardless of vegetation type. Our study confirms that both site factors (e.g., soil properties and vegetation) and temporal factors (e.g., initial soil moisture and rainfall characteristics) control the occurrence of the PF in mountainous regions such as the Qilian Mountains and that the Classification and Regression Trees has great potential to study the temporal occurrence of the PF.

1. Introduction

Preferential flow (PF) refers to the phenomenon whereby a fluid bypasses most of the matrix and chooses a preferred path to pass through a porous medium at a faster rate (Flury et al., 1994; Lin, 2010; Guo and Lin, 2018). Due to its relatively fast transport rate (relative to piston flow), it has an important effect on the distribution of water in the soil (Ritsema and Dekker, 1994), root uptake (Schwärzel et al., 2009), and groundwater recharge (Ireson and Butler, 2011). Especially in arid and semi-arid areas, where precipitation is scarce and potential evapotranspiration is intense (Cao et al., 2011), the growth of vegetation is highly dependent on deep soil moisture and groundwater (Loheide and Booth, 2011; Orellana et al., 2012; Yang et al., 2022). And the PF helps to transfer more water to deeper soils (Gazis and Feng, 2004), which is used for groundwater recharge and vegetation consumption. Despite the increasing attention and research on this topic, its complex mechanisms hinder further progress in understanding and modeling the PF (Guo and Lin, 2018).

Previous studies have shown that the occurrence of PF is controlled by temporal and spatial factors (Guo and Lin, 2018). The spatial factors controlling the occurrence of PF mainly include soil properties, topography, and vegetation (Guo and Lin, 2018; Demand et al., 2019; Tang

* Corresponding author at: NO. 222 Tianshui Road (South), Chengguan District, Lanzhou Gansu 730000, China.
E-mail address: tianjie@lzu.edu.cn (J. Tian).

https://doi.org/10.1016/j.geoderma.2023.116626
Received 28 February 2023; Received in revised form 28 July 2023; Accepted 30 July 2023
Available online 8 August 2023

et al., 2020). However, the main spatial factors controlling the occurrence of PF vary in different regions and scales (Liu and Lin, 2015; Wiekenkamp et al., 2016; Demand et al., 2019), and even the direction of influence of some soil properties on the PF varied in different regions (Koestel and Jorda, 2014; Larsbo et al., 2014, 2016). Wiekenkamp et al. (2016) found that the spatial occurrence of the PF could not be explained by watershed-scale topographic or soil-specific controls, and there was no significant relationship between the proportion of the PF occurrence and spatial factors. However, some other studies found that soil texture, topography, and land cover significantly influenced the occurrence of PF (Demand et al., 2019; Tang et al., 2020). Liu and Lin (2015) found that the control of topography on the PF occurrence was amplified when the scale was expanded from hillslope to watershed scale. In addition, some soil properties related to soil porosity (e.g., soil bulk density, and saturated hydraulic conductivity (K_S)) have a complex effect on the PF. Intuitively, higher soil porosity enhances the PF since macropores provide an important pathway for the PF (Mossadeghi-Björklund et al., 2016), but high soil porosity implies a larger surface area between the pores and the matrix, which is detrimental to the occurrence of the PF (Jarvis, 2007; Larsbo et al., 2016). These interacting spatial factors have different sensitivities and directions of influence on the occurrence of PF (Nimmo, 2020).

Temporal factors controlling the occurrence of PF include rainfall characteristics and initial soil moisture (SM_{int}) (Wiekenkamp et al., 2016; Guo and Lin, 2018; Demand et al., 2019). Wet soils, high rainfall amounts, and intensity are generally considered conditions conducive to the occurrence of PF (Liu and Lin, 2015; Wiekenkamp et al., 2016; Demand et al., 2019). However, the control of temporal factors on PF occurrence is also complex. Liu and Lin (2015) found that the occurrence of the PF requires rainfall to be above a threshold on the valley floor and swales, but the PF can occur directly on hilltops. Similar rainfall thresholds were found by Wiekenkamp et al. (2016). In addition, the influence of SM_{int} on the PF also varies across different study areas. In hydrophobic soils, dry soil favors the occurrence of PF, but PF in other areas is positively correlated with SM_{int} (Guo and Lin, 2018; Demand et al., 2019; Tang et al., 2020).

Current models of the PF lag in empirical understanding due to the complexity of the PF control factors and the specificity of control effects (Jarvis et al., 2016; Guo and Lin, 2018). Moreover, current research on the PF (identification of the PF using soil moisture observation) mainly focused on humid mountains or hills (e.g., Graham and Lin, 2011; Liu and Lin, 2015; Wiekenkamp et al., 2016; Demand et al., 2019; Guo et al., 2019; Tang et al., 2020). Few studies have investigated the occurrence and control of PF in cold, mountainous areas (Li et al., 2013; Hu et al., 2016). It is important to explore mechanisms for the control and occurrence of the PF in different environments to improve the understanding of the PF and to construct predictive models. In addition, due to the interaction of control factors in the natural environment and the relationship between them and the PF is highly nonlinear (Liu and Lin, 2015; Guo and Lin, 2018), and it is very difficult for traditional statistical methods to deal with this complex issue. In the past decades, Classification and Regression Trees (CART) and Random Forests (RF) have been widely used in hydrology-related research due to their powerful ability to handle complex nonlinear problems, such as water quality analysis (Li et al., 2019), flood prediction (Choubin et al., 2019), and determination of relationships between different soil properties (pedotransfer functions) (Koestel and Jorda, 2014; Lai et al., 2022a; Palladino et al., 2022). The CART provides a conceptual framework for automatic model selection that is not only easy to interpret, and monotonic changes in explanatory variables do not affect the model structure (Genuer and Poggi, 2020). The RF is suitable for dealing with high-dimensional cases like PF occurrences that have complex control factors (the number of variables is much greater than the number of observations) (Biau, 2012; Genuer and Poggi, 2020).

Our previous preliminary experiment explored the occurrence of PF under the typical land covers in mountain areas (Kang et al., 2022). We were able to identify soil properties, vegetation, and rainfall to affect the occurrence of PF. However, due to the limitations of the research method, the following questions remain unanswered: 1) What are the spatial and temporal patterns of PF occurrences at a large scale? 2) Can we develop a model to predict the spatial pattern of PF occurrences in large-scale mountainous areas? 3) Can we identify the mechanisms that trigger the occurrence of the PF, e.g., under what precedent soil conditions and rainfall does the PF occur (the temporal patterns of the PF occurrence)? This study is the first to try to find answers to the above questions based on a large-scale, long-term in-situ soil moisture observation network in high and cold mountainous areas using the latest machine learning techniques.

2. Data and methods

2.1. Study area

This study was conducted in the upper reaches of the Heihe River Basin, the second-largest inland river watershed (or terminal lake) in Northwest China (Cheng et al., 2014). It is located in the Qilian Mountains (97°29′-101°32′E, 37°43′-39°39′N) on the northern margin of the Qinghai-Tibet Plateau and has an area of over 27×10^3 km^2 (Fig. 1). The study area is in elevation from 1700 to 5600 m above sea level. The average annual temperature ranges from −3 to 7 °C and the annual precipitation ranges from 200 to 700 mm, with most of the rainfall occurring in the summer (65% of total rainfall between June and August) (Geng et al., 2014; Zhang et al., 2016), and the precipitation shows a decreasing trend from the eastern region to the western region (Geng et al., 2017). Due to the strong vertical differences in temperature and precipitation, the soils and vegetation show strong spatial heterogeneity and vertical zonation in the study area. The landscapes include glaciers, cold deserts, alpine meadows, shrub meadows, forests, grassland, and desert grassland from high to low elevation (Lu et al., 2017). The main vegetation consists of forests, shrubs, meadows, grasslands, and sparse vegetation (Tian et al., 2017). Under the influence of zonal differences in vegetation and temperature, the main soil types (FAO World Reference Base (WRB)) in the west are Phaeozem and Podzol, while Mountain grassland soil, Chernozem and Leptosol soils predominate in the southeast. The main soil textures in the area include silt loam, sandy loam, and silt (USDA classification), and the loam and sandy loam soils are mainly located in the upper mountainous areas, while the silt soils are mainly located in the upper river valleys (Lu et al., 2017).

2.2. Data

2.2.1. Soil moisture network

In this study, we analyzed soil moisture measured with 30 min intervals from 2014 to 2019 at 32 stations established in the upper reaches of the Heihe River Basin (Fig. 1). These stations cover different soil, vegetation, and elevation zones of the study area (Zhang et al., 2017a; Tian et al., 2019). Soil moisture and temperature sensors (5TE sensors, Decagon Devices Inc., Pullman, USA) were installed at each station at depths of 5, 15, 25, 40, and 60 cm. Soil samples were also collected at each depth and used to analyze the soil properties (including bulk density, soil porosity, K_S, soil organic content (SOC), soil–water characteristic curve (n, α, saturated soil moisture (SM_s), and residual soil moisture (SM_r) are the parameters of the soil–water characteristic curves (van Genuchten model) (van Genuchten, 1980)), and soil texture), see Tian et al. (2017) for more details. Tian et al. (2023) demonstrated the spatial distribution characteristics of major soil properties in the study area. The clay shows a decreasing trend from eastern to western (similar to the changes in rainfall and vegetation), while the bulk density increases from eastern to western. The K_S, n, SM_s, SM_r, silt, and sand are higher in the central region than in the west and east, while the α shows an opposite variation, with the central region lower than the west and east.

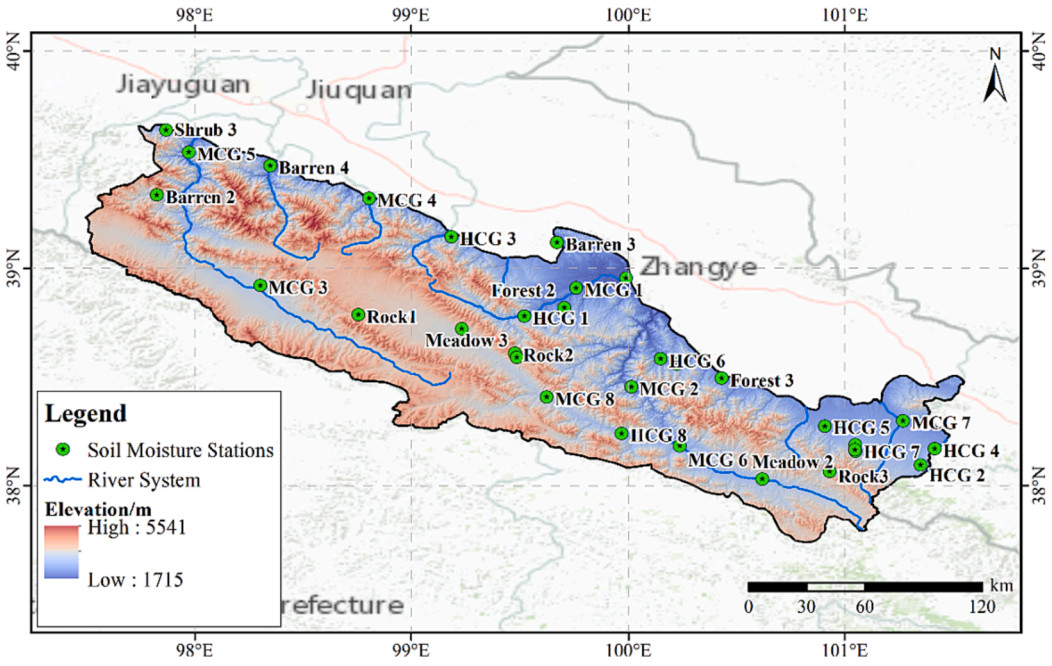


Fig. 1. Location of the study area and the distribution of the soil moisture stations. HCG and MCG represent high-coverage grassland and medium-coverage grassland, respectively.

In addition, we calculated the field capacity, wilting point, effective porosity, and total porosity. The field capacity and wilting point are the soil moisture at soil water pressure potential of −330 kPa and −1500 kPa, respectively (can be calculated from soil-water characteristic curves). The effective porosity is the difference between the total porosity and field capacity (Rawls et al., 1998), while the total porosity is calculated using the bulk density (Hao et al., 2008).

$$TP = 1 - \frac{BD}{PD} \tag{1}$$

where *BD* is the bulk density; *PD* is the soil particle density, generally taken as the average value of 2.65 g/cm^3.

To avoid the impact of data quality on subsequent analyses, checks of data credibility and consistency over time were performed according to the data quality control methods (Dorigo et al., 2013; Wiekenkamp et al., 2016): 1) excluding soil moisture data during the seasonal freezing periods based on soil temperature and soil moisture dynamics during the freeze–thaw cycles (Dorigo et al., 2013; Yang et al., 2017); 2) removing outliers (e.g., values outside the 1–90 vol% range (Wiekenkamp et al., 2016) and unreasonable fluctuations) (Tian et al., 2019); 3) excluding unreliable data due to instrumentation problems (e.g. insufficient battery power) by visual data inspection; 4) retaining only the periods when all five layers of soil moisture meet quality control. The periods in which soil moisture at each station met the above criteria are shown in Fig. S1. All stations except the MCG 2 met these criteria for more than 50% of the time period, and 78% of the stations met these criteria for more than 70% of the time period throughout the study period.

It is worth noting that the sensors of the Rock 1, Rock 2, and Rock 3 (Fig. 1) were installed in gravel, which might make the measured soil moisture inaccurate. In addition, it was not possible to determine the soil properties at these stations due to the inability to sample with ring knives (Zhao et al., 2020), so these three stations were excluded from the subsequent analysis.

2.2.2. Rainfall and normalized difference vegetation index (NDVI)

We did not deploy ground-based meteorological observation stations for rainfall observations at these locations due to budget constraints. Therefore, to analyze the impact of rainfall on the PF occurrence, we extracted rainfall data from 2014 to 2018 for these stations from the widely-used reanalysis dataset (China Meteorological Forcing Dataset, CMFD) (time resolution is 3 h, spatial resolution is 0.1° × 0.1°) (Yang et al., 2010; He et al., 2020). CMFD has been widely used in China (Meng et al., 2021; Zhang et al., 2021), and it is the relatively high-accuracy rainfall dataset in the Qilian Mountains (Lai et al., 2022b). Using the CMFD, we calculated the average rainfall (P_{mean}) for each station for the annual growing season (from May to October each year).

We also extracted the NDVI during the growing season from 2014 to 2019 from the surface vegetation index data (time resolution is 8 days, spatial resolution is 30 m × 30 m; MODIS (250 m) and Landsat (30 m) time series data were fused by the Gap Filling and Savitzky-Golay method) of the Qinghai-Tibetan Plateau (Cao et al., 2022; Chen et al., 2021) and used their maximum values as the NDVI of the corresponding stations. In the classification trees, the NDVI at the corresponding time of the infiltration event was used as the explanatory variable. Both the rainfall data and NDVI were from the National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/en/).

2.3. Hypothetical control mechanisms for the PF

In order to gain a preliminary understanding of the PF in the study area, we integrated knowledge from some in-situ observation stations and concepts from the literature to propose possible PF control mechanisms (Liu and Lin, 2015; Wiekenkamp et al., 2016; Guo and Lin, 2018; Demand et al., 2019; Kang et al., 2022). The main spatially controlling factors for PF occurrence are soil properties, vegetation, and topographic features. In particular, the effect of topography on the control of PF occurrence is more important at the watershed scale. Soil properties that control the PF occurrence are mainly macropore and hydrophobicity, and their related soil properties (e.g., *SOC*, bulk density, K_S, etc.). Temporal controls on the PF occurrence are mainly the SM_{int} and rainfall characteristics, and the higher rainfall, the wetter soil (as opposed to hydrophobic soils), and the more likely the PF will occur.

2.4. Infiltration event

2.4.1. Determination of the infiltration event

According to the definition of infiltration events by Tian et al. (2019) and Kang et al. (2022), the starting and ending times of infiltration events were determined for each station as follows (Fig. 2a): 1) selecting the time series of continuous increase in soil moisture, defining the time of the start of the soil moisture increase as the start time of the infiltration event (SM_{int}) and the time of the end of the soil moisture increase as the end of the infiltration event (SM_{end}); 2) combining infiltration events that occurred within 6 h into the same event; 3) excluding events with the soil moisture increase less than 1% (SM_{end} - SM_{int} less than 1 vol %) (Saito et al., 2013; Wiekenkamp et al., 2016; Demand et al., 2019). The identification of infiltration events was performed automatically using a dedicated Matlab script. The total number of selected infiltration events for these stations is shown in Fig. S2. In order to characterize the soil moisture change process for infiltration events, we defined the following quantitative indexes (Fig. 2a) (Lozano-Parra et al., 2016; Tian et al., 2019):

The initial soil saturation (S_{int}) was calculated by:

$$S_{int} = \frac{SM_{int} - SM_{min}}{SM_{max} - SM_{min}} \tag{2}$$

where SM_{max} and SM_{min} are the maximum and minimum values of the recorded soil moisture, respectively.

Initial drying time (DT_{int}, hour):

$$DT_{int} = Startingtime_j - Endingtime_{j-1} \tag{3}$$

where j is the serial number of the infiltration event.

Maximum variation or slope of the soil moisture wetting curve (S_{max}) (vol. %/hour) (Lozano-Parra et al., 2016):

$$S_{max} = max\left(\frac{SM_{t+\Delta t} - SM_t}{\Delta t}\right) \times 2 \tag{4}$$

where SM_t is the soil moisture value in the time t, and Δt is the variation of the time in the measurement interval, which is 30 min.

Soil water storage increment (SWS, cm):

$$SWS = d_l \times \sum_{t=Startingtime}^{Endingtime} \Delta SM_{t,l}^{j} \times 0.01 \tag{5}$$

with

$$\Delta SM_{t,l}^{j} = \begin{cases} \Delta SM_{t,l}^{j}, \Delta SM_{t,l}^{j} > 0 \\ 0, \ \Delta SM_{t,l}^{j} < 0 \end{cases} \tag{6}$$

where $\Delta SM_{t,l}^{j} = SM_{t+\Delta t,l}^{j} - SM_{t,l}^{j}$; $\Delta SM_{t,l}^{j}$ is the change of soil moisture for the j^{th} infiltration event in the l layer (vol. %); $\Delta SM_{t,l}^{j}$ is multiplied by 0.01 to convert its units to cm^3/cm^3; d_l is the thickness of the l^{th} layer of soil.

2.4.2. Event classification

These events meeting the described quality criteria can be classified into three categories based on the starting time of the infiltration event at two adjacent layers (Liu and Lin, 2015; Wiekenkamp et al., 2016; Demand et al., 2019):

I. Not classifiable (NC): Events where the upper soil moisture (5 cm) responds (SM_{end} - $SM_{int} \geq 1$ vol%) but the lower soil moisture does not (SM_{end} - SM_{int} less than 1 vol%) (Demand et al., 2019) (e.g., the third layer of soil moisture responds but the fourth layer of soil moisture does not respond, in Fig. 2b);

II. PF: Events where at least one depth sensor detects an out-of-sequence soil moisture response or both layers respond simultaneously (e.g., the third soil layer shows a soil moisture response before the second layer, in Fig. 2b);

III. Sequential flow (SF): Events followed the expected sequence of soil moisture responses with depth (e.g., the second soil layer shows a soil moisture response before the third layer).

According to the above principles, we can use the soil moisture of two adjacent layers to detect whether PF was detected between them.

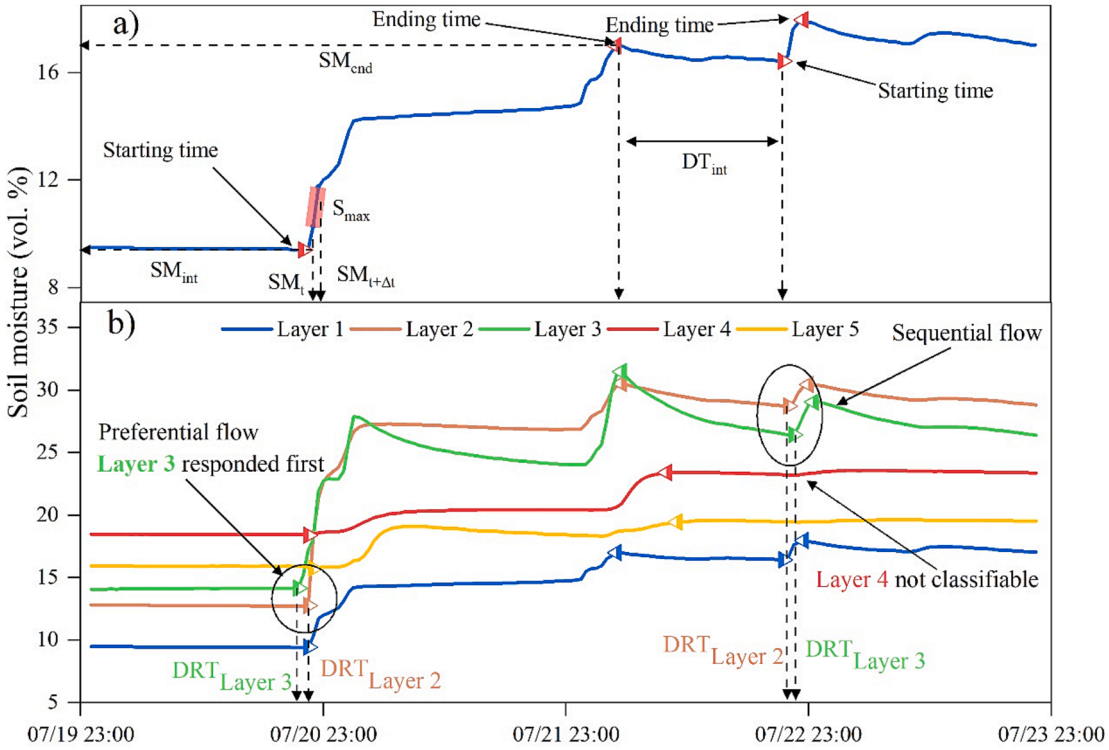


Fig. 2. Example of determining infiltration event and schematic diagram of variables SM_{end}, SM_{int}, S_{max}, and DT_{int} (**a**), and schematic diagram of the PF, sequential flow (SF), and not classifiable (NC) events (**b**).

We can classify four types of response scenarios (Layer 1–2, Layer 2–3, Layer 3–4, and Layer 4–5) (Kang et al., 2022).

In addition, besides the event classification for the adjacent layers, we defined the event classification for the whole soil profile using the following rules:

I. NC: only the first layer (5 cm) of soil moisture showed a response;

II. PF: at least one soil layer was detected with an out-of-sequence response;

III. SF: sequential response at all soil layers.

The PF proportion was calculated from the ratio of PF events to total infiltration events (total number of infiltration events in the first soil layer (5 cm)). In the subsequent analysis, except for the special description of the PF proportion of the different layers, all other PF proportions refer to the PF proportions of the profiles.

The relative proportion of PF in each layer is the ratio of the PF events detected in that layer to the number of PF events in the entire profile. For example, the PF relative proportion of soil Layer 2–3 can be calculated by:

$$PF_{L2-3} = \frac{N_{PF,L2-3}}{N_{PF,profile}} \times 100 \quad (7)$$

where $N_{PF,\ L2-3}$ is the number of PF events detected between layer 2 (15 cm) and layer 3 (25 cm); $N_{PF,\ profile}$ is the number of PF events detected for the entire profile.

2.5. Methods of analysis

2.5.1. Classification and regression trees

Classification and Regression Trees (CART) refers to a statistical method for constructing tree predictors (also called decision trees) for both regression and classification problems (Breiman et al., 1984). CART is an upside-down tree (the root is at the top). The leaves of the tree are nodes without descendants, and the other nodes of the tree are non-terminal nodes. Also, each non-terminal node has two child nodes; therefore, the tree is a binary tree. Nonterminal nodes distinguish between two child nodes with a judgment condition, marking the leaves with a class label or the value of a response variable. Building a CART is a two-step process. Firstly, a maximal tree is constructed using recursive binary splitting, and the second step, called pruning, builds a sequence of optimal subtrees pruned from the maximal tree sufficiently, using the complexity parameter (*CP*) (Rothwell et al., 2008; Genuer and Poggi, 2020). The "rpart" package (Therneau and Atkinson, 2018) and the "rpart.plot" package (Milborrow, 2018) in R were used to construct the classification tree in this study.

In order to determine the occurrence pattern of the PF, we constructed classification trees using vegetation (NDVI), previous soil conditions (SM_{int} and DT_{int}), and water input characteristics (rainfall amount and intensity) as explanatory variables, and the type of infiltration events (the PF or Non-preferential flow (NPF, including NC and SF)) as the response variable. Given the absence of in situ rainfall observations in the soil moisture network, we used surface soil moisture (5 cm) dynamics (*SWS* and S_{max}) instead of rainfall characteristics (rainfall amount and intensity) (Zhu et al., 2014; Glaser et al., 2019). We constructed classification trees (maximal tree) with events for each of the three typical vegetation stations (since there were few PF events at the bare land and meadow, only three vegetation types, forest, high cover grassland (HCG), and medium cover grassland (MCG), were selected) and for all the stations (excluding Rock stations) (Table 1). In order to test the reliability of the classification tree, we randomly selected 2/3 of the infiltration events to construct the classification tree and used the remaining events to verify the classification tree. The number of infiltration events for different vegetation types and all stations is shown in Table 1. The error calculation formula for the classification tree is as follows:

Table 1
The stations correspond to different vegetation types and the number of infiltration events in the training and validation sets.

LUC	Stations	Number of infiltration events	
		Training	Validation
Forest*	Shrub 1, Shrub 3, Forest 1, Forest 2, Forest 3	228	113
HCG	HCG 1, HCG 2, HCG 3, HCG 4, HCG 5, HCG 6, HCG 7, HCG 8	511	255
MCG	MCG 1, MCG 2, MCG 3, MCG 4, MCG 5, MCG 6, MCG 7, MCG 8	432	215
ALL	Excluding Rock 1, 2 and 3	1681	839

Note: * In order to have enough events to construct a representative classification tree, we merged the forest and shrub stations. Shrub 2 was not added to the subsequent analysis because the temporal factor of the station had little influence on whether the PF occurred.

$$err = \frac{1}{n}\sum_{i=1}^{n} 1_{Y_i \neq T(X_i)} \quad (8)$$

where n is the number of infiltration events; Y is the response variable (PF or NPF); T is the constructed classification tree, X is the prediction variable, and $T(X)$ denotes the type of event predicted by the classification tree.

Pruning is the second step of the CART algorithm. Pruning is a model selection process with the idea of finding the best tree between two extremes: satisfying the allowed prediction error while minimizing the complexity (i.e., the number of nodes) (Genuer and Poggi, 2020). The subsequent analysis is based on the pruned tree.

Once the tree is given, it is easy to use it to predict the type of infiltration event. Simply start at the root and determine in turn whether the explanatory variables meet the conditions of the nonterminal node, and if so, go to the left node, and if not, go to the right (Genuer and Poggi, 2020). By sequential judgments, the unique path from the root to the leaves is obtained, and the type of infiltration event (i.e., PF or NPF) can be determined.

2.5.2. Random Forest

Random Forest (RF) is a collection of un-pruned CART trees. Since individual trees are randomly perturbed, the forest benefits from a more extensive exploration of the space of all possible tree predictors, which always results in better predictive performance (Therneau and Atkinson, 2018). The importance calculation and the selection of variables in this study were implemented through the "VSURF" package in R (Genuer et al., 2015; Genuer and Poggi, 2020). The steps are as follows:

I. Ranking and preliminary elimination. Ranking the variables by decreasing importance, and eliminating the variables with low importance.

II. Selecting variables for interpretation. Starting with the model with only the most important variables and ending with the model involving all the previously selected important variables, the average of the Out-Of-Bag error for these models is calculated, and finally, the model variables that lead to the lowest Out-Of-Bag error are selected.

III. Selecting variables for prediction. From the variables selected for interpretation, a sequence of models is constructed by sequentially introducing the variables in increasing order of importance and iteratively testing them. The variables of the last model are finally selected.

The Out-Of-Bag error (*OOB error*) and importance of variables (*VI*) were calculated as follows:

$$OOBerror = \frac{1}{n}\sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2 \quad (9)$$

$$VI(X^j) = \frac{1}{q}\sum_{k=1}^{q} \left(\widetilde{OOBerror_k^j} - OOBerror_k \right) \quad (10)$$

where n is the number of samples; Y is the response variable; $\widehat{Y}$ is the

corresponding predicted value; X is the explanatory variables; q is the number of trees constructed; $OOBerror_k$ is the *OOB error* of trees k; $\widetilde{OOBerror}_k^j$ is the error of the trees k after perturbation of X^j (Randomly permute the values of variable X^j).

2.5.3. Regression and statistical analysis

The RF was used to determine the main control factors of the PF, but the exact relationship between these main control factors and the proportion of the PF occurrence is not determined by the method because the RF is a non-parametric method (Gao et al., 2018). Cftool, an application in MATLAB (R2022a, The MathWorks), was used to establish the empirical formula for predicting PF in this study. We used the coefficient of determination (R^2), adjusted coefficient of determination (Adjusted R^2), and root mean square error (*RMSE*) to assess the predictive power of the empirical formula.

$$R^2 = \frac{\sum (\widehat{Y} - \overline{Y})^2}{\sum (Y - \overline{Y})^2} \tag{11}$$

$$Adjusted R^2 = 1 - \frac{\frac{\sum (Y-\widehat{Y})^2}{n-2}}{\frac{\sum (Y-\overline{Y})^2}{n-1}} \tag{12}$$

$$RMSE = \sqrt{\frac{\sum (Y - \widehat{Y})^2}{n}} \tag{13}$$

where $\overline{Y}$ is the means of measured values.

3. Results

3.1. Variation of PF proportion with depth

Fig. 3 shows the relative proportions of PF detected at different depths for the 32 soil moisture stations. The relative proportion of detected PF events gradually decreases with increasing depth, but at Layer 4–5, median of the PF relative proportions is higher than at Layer 3–4. The *SWS* proportions of PF events (Because it is not possible to distinguish between the amount of water transported by SF and PF, the *SWS* of PF events here is the result of the combined effect of SF and PF.) are also calculated (Fig. 4a), and their distribution (median value) at different depths is similar to the distribution of the relative proportions

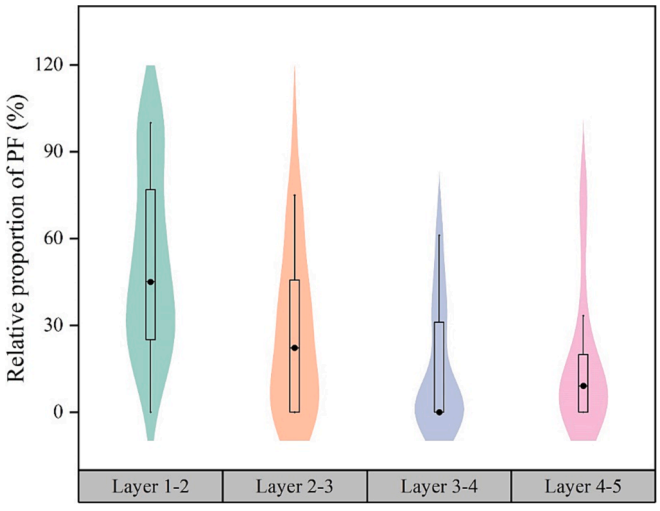


Fig. 3. The relative proportions of PF were detected at different soil depths. The dots and horizontal lines in the plot indicate the median of the relative proportions, the boxes indicate the 25–75% range, and the vertical lines indicate the 5th and 95th quartiles.

of PF and is also larger for Layer 4–5. For individual events, the mean *SWS* (SWS_{mean}) of the detected PF events is greater than the SF events, and the SWS_{mean} in deep soils (Layer 3–4 and Layer 4–5) is also larger than in shallow soils (Layer 1–2 and Layer 2–3) (Fig. 4b). In deep soils, the water transport of the PF events is much larger than that of the SF event, while in shallow soils, they are close to each other.

We also analyzed in detail the relative proportions of PF in different soil horizons at all stations (Fig. 5). The results showed that the distribution characteristics of PF in the soil horizons of different vegetation types were different. At Barren, PF was mainly concentrated above the second soil layer (15 cm) (Layer 1–2); at MCG, PF was mainly concentrated above the third soil layer (25 cm) (Layer 1–2 and 2–3); at HCG, PF was concentrated above the fourth soil layer (40 cm) (Layer 1–2, 2–3, and 3–4); at Meadow, Shrub, and Forest, PF was concentrated primarily above the fifth soil layer (60 cm) (Layer 1–2, 2–3, 3–4, and 4–5) (Fig. 5). These distribution characteristics correlate with the distribution of roots at each station. As vegetation cover increased and roots penetrated deeper into the soil, the thickness of soil where PF was detected increased accordingly.

3.2. Differences in the PF proportion between stations

Fig. 6 shows the proportion of PF occurrence at different stations. In general, the proportion of the PF gradually increases from Barren land to Meadow and grassland (MCG and HCG), and then to Shrub and Forest. However, at some stations, the proportion of the PF is much higher than at other stations with the same vegetation types. This difference implies that the occurrence of the PF is not only controlled by vegetation type but also influenced by other factors. The proportions of the PF, SF, and NC events at these stations are shown in Fig. 7. Most of the stations are dominated by the SF events (65.6%, 21 of the 32 stations) and the NC events (21.9%, 7 of the 32 stations), with a small proportion of the PF events (12.5 %, 4 of the 32 stations). Nonetheless, at some stations in the eastern and central parts of the study area, the proportion of PF events is higher.

3.3. The spatial control of PF occurrence

In order to explore the factors controlling the occurrence of the PF, we analyzed the correlations of 18 spatial attributes, such as soil texture, soil hydraulic properties, topography, and vegetation, with the PF proportions. In general, the correlations are low, and significant correlations are only found between the *SOC* and NDVI, and PF proportions (P less than 0.05). The proportion of the PF shows a slight increase trend with increasing the *SOC* (Pearson $R = 0.41$), and with increasing the NDVI (Pearson $R = 0.38$). The relationship between the PF proportion and other factors, such as soil properties and topography, is not clear (P greater than 0.05). Obviously, the relationships between these spatial factors and the PF proportions are not simply linear, and their relationships need to be further analyzed using other statistical methods.

3.4. The spatial estimation of the PF occurrence at the station-scale

We further analyzed the relationship between the control factors and the PF to develop empirical relationships, which may be useful for hydrological simulations. Given the complexity and intercorrelation of the PF control factors, we explored the relative influence of these control factors on the PF based on the VSURF. The result shows that the importance of variables of K_S, SM_r, *SOC*, α, *Clay*, and NDVI is higher than the other factors (Fig. 8), while K_S and SM_r are the final chosen interpretation and prediction variables. We used multiple nonlinear regression to establish equations for predicting the PF proportions using the K_S, SM_r, *SOC*, and NDVI (there are significant correlations between the *SOC* and NDVI and the proportion of the PF occurrence) as prediction variables, respectively (Table 3). We found that the prediction equation built using the SM_r has the best prediction performance ($R^2 = 0.43$*)

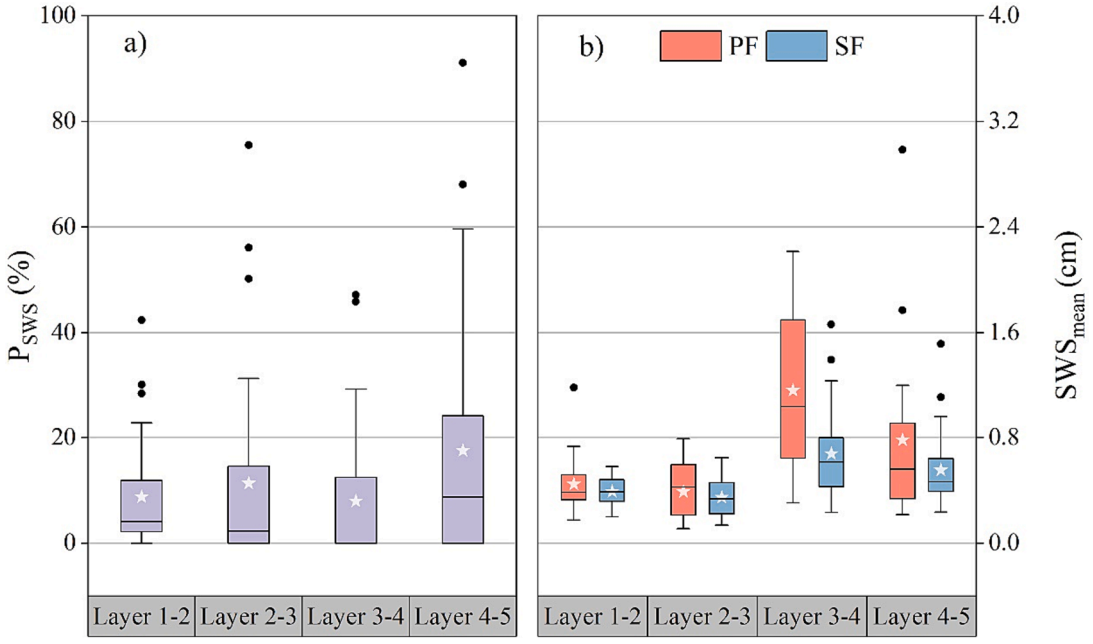


Fig. 4. The proportion of cumulative soil water storage increment (*SWS*) of PF events to cumulative *SWS* of all events (**a**); Mean *SWS* (SWS_{mean}) for individual PF and SF events (**b**). The pentagram and the horizontal black line in the plot indicate the mean and median, respectively. *SWS* here is the incremental water storage in the lower soil layer. For example, the *SWS* of Layer 1–2 is the incremental water storage in the second soil layer (15 cm). We assumed that any increase in water storage in the second soil layer is transported from soil water in the first layer to the second soil layer through PF or SF events.

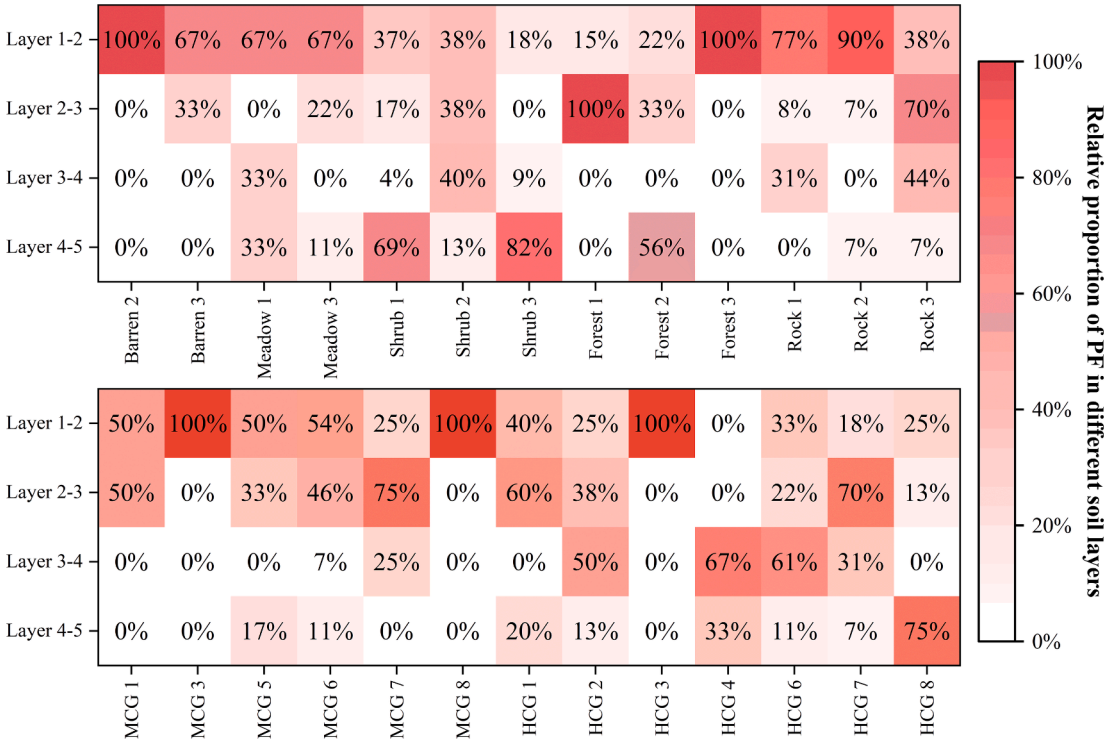


Fig. 5. Relative proportions of PF for different soil layers at each station. Because no PF events were detected at Barren 1 and 4, MCG 2 and 4, HCG 5, and Meadow 3, these stations are not shown in the figure.

instead of the prediction equation built by *SOC* ($R^2 = 0.22$*) or NDVI ($R^2 = 0.15$). In addition, the predictive performances of the equations were established using the K_S and SM_r, *SOC* and NDVI, and the SM_r and *SOC* (SM_r and *SOC* are the two variables with the best prediction performance using multiple nonlinear regression) as prediction variables, respectively (Table 3). We found that the equations built by the NDVI and *SOC* have the worst prediction performance ($R^2 = 0.37$), while the other two equations had approximately similar performances ($R^2 = 0.49$* and 0.52*). Obviously, the K_S and SM_r play an important role in the prediction performance despite no significant correlation being found between them and the proportion of the PF. This may be because the K_S and SM_r are influenced by other soil properties (such as bulk density,

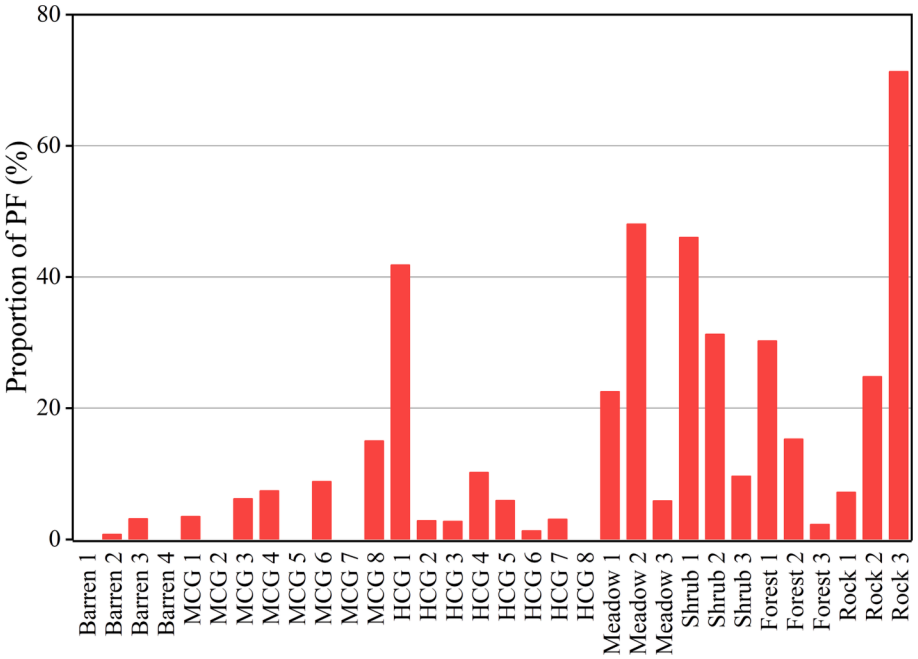


Fig. 6. The proportion of the PF events at different stations.

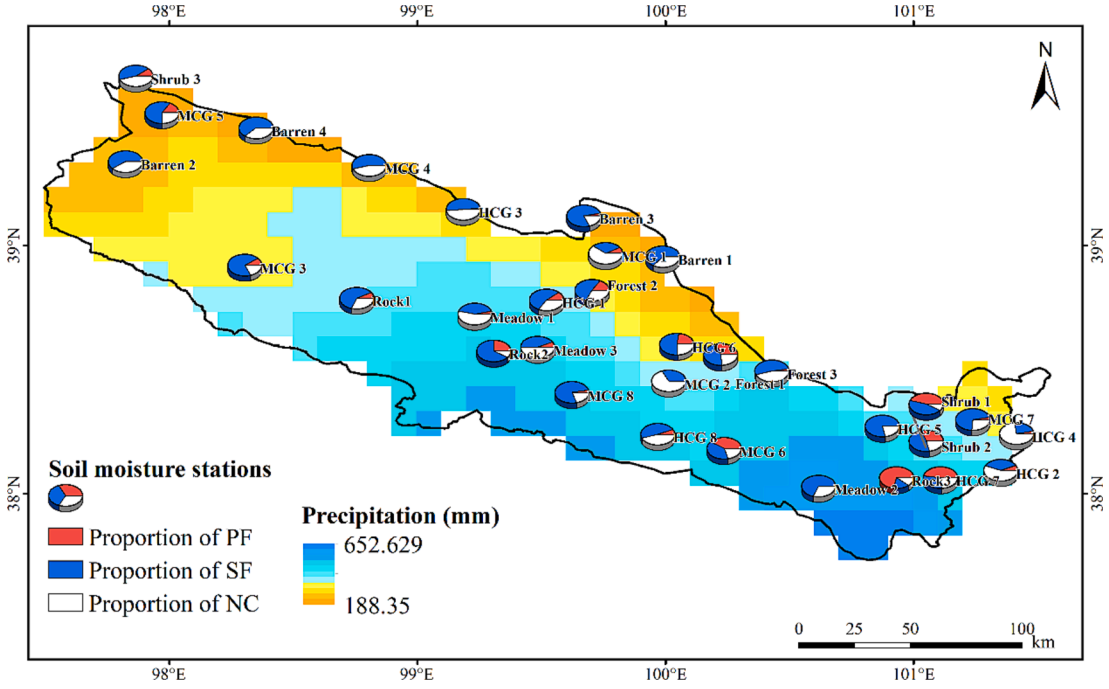


Fig. 7. The spatial distribution of annual mean precipitation (2014–2018, CMFD) in the study area and soil moisture stations, and the proportions of the PF, SF, and NC at each station.

SOC, and effective porosity) and are indicators that can represent the overall soil hydraulic characteristics. Since the prediction performance of the equations built with the K_S and SM_r (Adjusted $R^2 = 0.39$; *RMSE* = 11.65%) and with the SM_r and *SOC* (Adjusted $R^2 = 0.39$; *RMSE* = 11.64%) are comparable, we choose the relatively simple equation (Eq. (14) ($R^2 = 0.49$* ; Adjusted $R^2 = 0.39$; *RMSE* = 11.65%) as the final prediction equation (Fig. S3). There are limitations to this kind of prediction that only uses spatial factors such as soil properties because spatial factors only provide flow paths for the occurrence of the PF, and temporal factors controlling the occurrence of the PF will be analyzed in Section 3.5.

$$PF = 4.21 \times K_s - 24.59 \times K_s \times SM_r + 7734 \times SM_r{}^2 - 956.5 \times SM_r + 28.31 \quad (14)$$

3.5. Temporal control of the PF occurrence at the event scale

Table 4 shows the importance of variables of each explanatory variable (tree before pruning) for different vegetation types. S_{int} is the most important factor controlling whether the PF occurs in Forest, MCG, and all stations, while S_{max} is the most important factor in HCG. However,

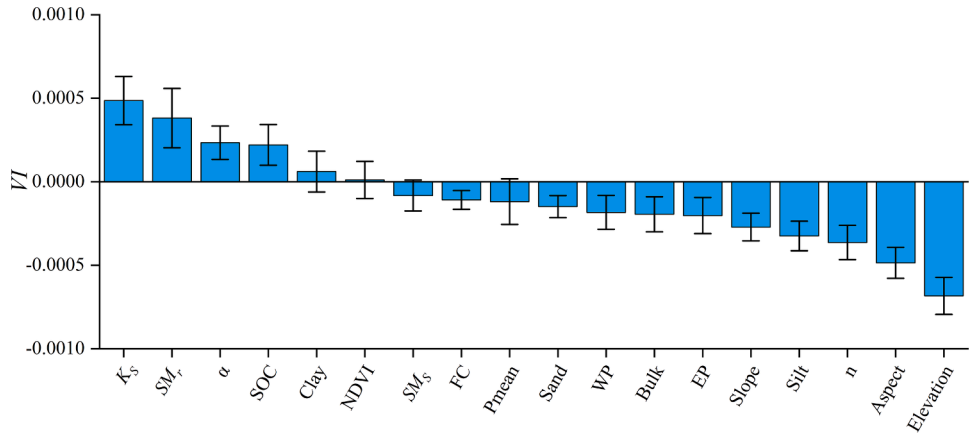


Fig. 8. Importance of variables (*VI*) ranked in descending order for the PF proportion. Due to missing soil properties data for some stations, only 25 stations were used for importance ranking. The *FC*, *WP*, *Bulk*, and *EP* refer to field capacity, wilting point, bulk density, and effective porosity, respectively.

Table 2
Pearson and Spearman correlation coefficients between spatial factors (soil properties, vegetation, topography, elevation, precipitation) and the PF proportions. (R = correlation coefficient, significance: * = 0.05).

	Sand (%)	Clay (%)	Silt (%)	K_S (m/d)	α	n	SM_s (m^3/m^3)	SM_r (m^3/m^3)	Bulk Density (g/cm^3)
Pearson R	−0.15	0.16	0.14	0.11	0.19	−0.07	0.11	0.06	−0.23
Spearman R	0.01	0.34	−0.01	0.30	0.33	0.01	0.11	0.06	−0.23
	SOC (g/100 g)	Effective Porosity	Field Capacity (m^3/m^3)	Wilting Point (m^3/m^3)	Elevation (m)	Slope (°)	Aspect (°)	NDVI	P_{mean} (mm)
Pearson R	**0.41***	0.25	−0.03	0.12	0.28	−0.06	−0.03	**0.38***	0.24
Spearman R	0.36	0.24	−0.13	0.00	0.29	0.04	−0.09	0.22	0.17

Table 3
Statistical parameters of equations obtained by multiple nonlinear regression. Only the spatial attributes and the PF proportions of the 25 stations used in VSURF were used to build these equations.

Variables	Regression equation**	R^2	Adjusted R^2	*RMSE* (%)
K_S	$-1.357 \times K_S^2 + 10.73 \times K_S + 0.4368$	0.16	0.08	14.28
SM_r	$8215 \times SM_r^2$-$1064 \times SM_r + 36.94$	0.43*	0.37	11.80
NDVI	$-21.36 \times \text{NDVI}^2 + 46.1 \times$ NDVI-5.518	0.15	0.07	14.38
SOC	$0.6822 \times SOC^2$-$3.174 \times SOC +$ 9.78	0.22*	0.15	13.78
K_S and SM_r	$4.214 \times K_S$-$24.59 \times K_S \times SM_r +$ $7734 \times SM_r^2$-$956.5 \times SM_r + 28.31$	0.49*	0.39	11.65
NDVI and *SOC*	$-77.81 \times \text{NDVI}^2 + 81.9 \times$ NDVI $+6.205 \times$ NDVI $\times SOC + 0.6968$ $\times SOC^2$-$7.552 \times SOC$-6.328	0.36	0.20	13.36
SM_r and *SOC*	$6601 \times SM_r^2$-$778.5 \times SM_r$-3.961 $\times SM_r \times SOC + 0.6038 \times SOC^2$-$3.053 \times SOC + 26.81$	0.52*	0.39	11.64

Note: * The multiple regression equation was significant. ** Only one form of the equation with the best simulation results was chosen here.

Table 4
Importance of variables of input variables for stations with different vegetation and all stations.

LUC	Importance of explanatory variables*				
	S_{int}	*SWS*	NDVI	S_{max}	DT_{int}
Forest	**33.343**	28.386	18.118	16.53	12.276
HCG	23.632	30.305	19.118	**34.983**	23.659
MCG	**26.403**	14.445	16.286	20.375	19.390
All	**112.83**	87.359	77.274	84.267	75.948

Note: * The indicators of each event are calculated using the surface soil moisture (5 cm). Bold numbers represent the variable importance values corresponding to the variables with the highest importance in that vegetation type.

Table 5
The normalized complexity parameter, number of splits, and training and validation errors of the different decision trees for different land covers.

LUC	Complexity Parameter	Number of Splits	Training Errors	Validation Errors
Forest	0.029	6	0.129	0.204
HCG	0.027	5	0.077	0.106
MCG	0.024	6	0.074	0.114
All	0.010	7	0.105	0.111

the factors of the lowest importance of variables vary among vegetation types, in Forest and all stations are DT_{int}, HCG is NDVI, and MCG is *SWS* (Table 4).

Fig. S4-6 shows the results of classification trees after pruning for the three vegetation types. All four trees show good classification results with the validation errors and training errors less than 0.15 (except the validation error of Forest is 0.204) (Table 5). Therefore, we can determine the occurrence pattern of the PF in different vegetation types by going through the classification trees.

1) In Forest, the PF occurs when the NDVI is above 0.59, the rainfall exceeds a certain threshold (corresponding to the *SWS* greater than 0.31 cm), and the S_{int} is greater than 0.57. In addition, when the S_{int} is less than 0.57, but rainfall intensity exceeds a certain threshold value (corresponding to the S_{max} greater than 1.2 vol. %/h), and the S_{int} is greater than 0.15, there is also a high susceptibility to the PF (Fig. S4);

2) In HCG, the PF occurs more frequently when rainfall intensity exceeds a certain threshold value (corresponding to the S_{max} greater than 9 vol. %/h), and the S_{int} is greater than 0.23 (Fig. S5);

3) In MCG, the PF occurs when the S_{int} is greater than 0.47, the rainfall amount is greater than a certain threshold value (corresponding to the *SWS* greater than 0.21 cm), the DT_{int} is less than 113 h, the NDVI is greater than 0.23, and rainfall intensity is less than a certain threshold value (corresponding to the S_{max} less than 1.5 vol. %/h). In addition, the PF is more likely to occur when rainfall intensity exceeds a certain threshold value (corresponding to the S_{max} greater than 1.5 vol. %/h), but rainfall intensity exceeds a certain threshold value (corresponding to a value of S_{max} greater than 2.2 vol. %/h) (Fig. S6).

In summary, we found various thresholds of these factors for different vegetation types to control the PF occurrences (Table 6), for example, the threshold for *SWS* in Forest is 0.31 cm but 0.21 cm in MCG; the threshold for S_{int} is 0.57 in Forest, 0.23 in HCG, and 0.47 in MCG. However, in all of these vegetation types, the PF is generally more likely to occur when the soil is wetter (larger S_{int}), and the S_{max} and *SWS* are larger. In addition, the PF is also more likely to occur in Forest with better vegetation (larger NDVI). But the PF events also occur at some of the nodes with smaller S_{int} (e.g., the 6th terminal nodes of Forest (Fig. S4)).

Finally, a classification tree was constructed using all stations (except Rock) in order to identify patterns of the PF occurrence for the entire study area (Fig. 9). The PF is more likely to occur under the following three conditions (Fig. 9): 1) S_{max} greater than 11 vol. %/h, NDVI greater than 0.22 and S_{max} greater than 21 vol. %/h; 2) S_{max} less than 21 vol. %/h, but *SWS* less than 1.7 cm, S_{int} less than 0.57 and DT_{int} greater than 74 h; 3) DT_{int} less than 74 h and S_{max} greater than 16 vol. %/h. Obviously, this classification tree identifies the most complex conditions for the occurrence of the PF, as it includes all infiltration events. However, the conditions for the occurrence of PF vary among the in-situ observation stations.

4. Discussion

4.1. The significance of the PF for ecohydrology

Numerous studies have shown that PF can enhance the deeper and faster infiltration of rainfall into soils (Guo et al., 2018; Worthington, 2019). We came to a similar conclusion that although PF events account for only a small fraction (less than 20%) of the infiltration events at most of the stations (about 75%) (Fig. 7), they play a crucial role in the recharge of soil moisture in the study area. Our results suggested that the PF events detected at depth can provide more water to the soil, compared to SF events (Fig. 4b). Some previous studies have shown that evaporation affects shallow soil moisture, while soil moisture that infiltrates through the PF pathway to the deeper layers can be retained for a longer period to provide water for plant growth (Gazis and Feng, 2004; Jarvis et al., 2016; Guo et al., 2019). This finding reinforces the significance of PF on soil water dynamics and water resources for plant growth, particularly in arid areas.

4.2. The spatial distribution of PF

We find that the relative proportion of PF detected decreases with depth but increases at Layer 4–5 (Fig. 3). As the soil depth increases, the root and macroporosity tends to decrease. This can lead to a decrease in the number of PF events detected (Liu et al., 2007; Wang et al., 2020). In contrast, the increased number of PF events at Layer 4–5 may be due to the high gravel content in the deeper soil layers (Yang et al., 2020). Both roots and gravels are important factors controlling the occurrence of PF (Johnson and Lehmann, 2006; Qin et al., 2015; Zhao et al., 2020). The proportion of PF also varies significantly between stations. The spatial distribution of PF proportions and annual precipitation (Fig. 7) and some soil properties (e.g., *Clay*) (Tian et al., 2023) in the study area are similar, both gradually increasing from west to east and from north to south. However, the variability of PF at different depths and different stations also depends on the control of spatial and temporal factors discussed below (Guo et al., 2018; Demand et al., 2019).

Table 6
The threshold for the temporal factors for initiating the PF.

	S_{int}	*SWS* (cm)	NDVI	S_{max} (vol. %/h)	DT_{int} (h)
Forest	0.15	0.31	0.59	1.2	–
HCG	0.23	–	–	9	–
MCG	0.47	0.21	0.23	2.2	113

Note: - The pruned classification tree does not have this variable.

4.3. The spatial control of PF proportion

The previous study found that the soil properties significantly related to the proportion of PF occurrence in the study area are the bulk density, SM_r, K_S, and *SOC* (Kang et al., 2022). However, when upscaling to the catchment scale, only significant linear relationships remain between the *SOC* and PF (Table 2). In addition, a significant relationship was found between the NDVI and the PF at some stations (Kang et al., 2022) and the entire catchment scales (Table 2), which implies an important control of vegetation on the PF occurrence. Interestingly, although no significant linear relationship was found between the K_S and SM_r and the proportion of the PF occurrence at the catchment scale, they are the dominant factors controlling the PF occurrence identified using the VSURF method (Fig. 8) (This is in contrast to our hypothesis, which did not find the control of topographic factors on the occurrence of PF.). The results of the multiple nonlinear regression also proved that SM_r and K_S simulated the proportion of the PF occurrence better than *SOC* and NDVI (Table 3). This implies that the RF method can find the intrinsic effects of prediction variables (Koestel and Jorda, 2014; Lai et al., 2022a) on the response variables that cannot be obtained through traditional methods, such as significant analysis.

The most important factor controlling the proportion of PF occurrence is K_S (Fig. 8). It is worth noting that in the previous study, a positive correlation between the K_S and the PF occurrence proportion was found (Kang et al., 2022), but our results show that the relationship between the PF proportions and K_S is nonlinear (Table 2 and 3). Previous studies have found that PF is more likely to occur in soils that contain only a few continuous but poorly interconnected large pores (Jarvis, 2007; Bianchi et al., 2011). This is attributed to the fact that a well-connected network of small and macropores increases the lateral convection and diffusion of water between pores and soil matrix during infiltration while decreasing vertical transport (Jarvis, 2007; Koestel et al., 2012). Many soil properties influenced by porosity have this contradictory effect on the PF, such as the bulk density and K_S (Mossadeghi-Björklund et al., 2016; Nimmo, 2020). Larsbo et al. (2014) proposed a conceptual model to explain this phenomenon, as pore connectivity in the soil decreases, the PF intensity initially increases and then decreases.

We established an empirical relationship (Eq. (14) between the PF and soil properties (K_S and SM_r). This relationship, along with the spatial distribution of K_S and SM_r, was used to characterize the spatial distribution of the proportion of PF occurrence across the study area. In addition, the equation can also be used to determine the proportion of PF occurrence in similar regions around the globe, providing a preliminary understanding of the PF characteristics of the region.

Previous studies have also established empirical relationships between spatial factors and the proportion of PF occurrence, but the main factors controlling the occurrence of PF varied greatly in different studies (Liu and Lin, 2015; Gao et al., 2018). The empirical relationships we established can only provide a limited simulation of the proportion of the PF occurrence ($R^2 = 0.49$*) since only 50–66% of the variability of the PF is related to spatial factors such as soil properties (van Schaik, 2009). Spatial factors provide the pathway for the occurrence of the PF, while temporal factors determine whether the PF occurs at each station (Guo and Lin, 2018).

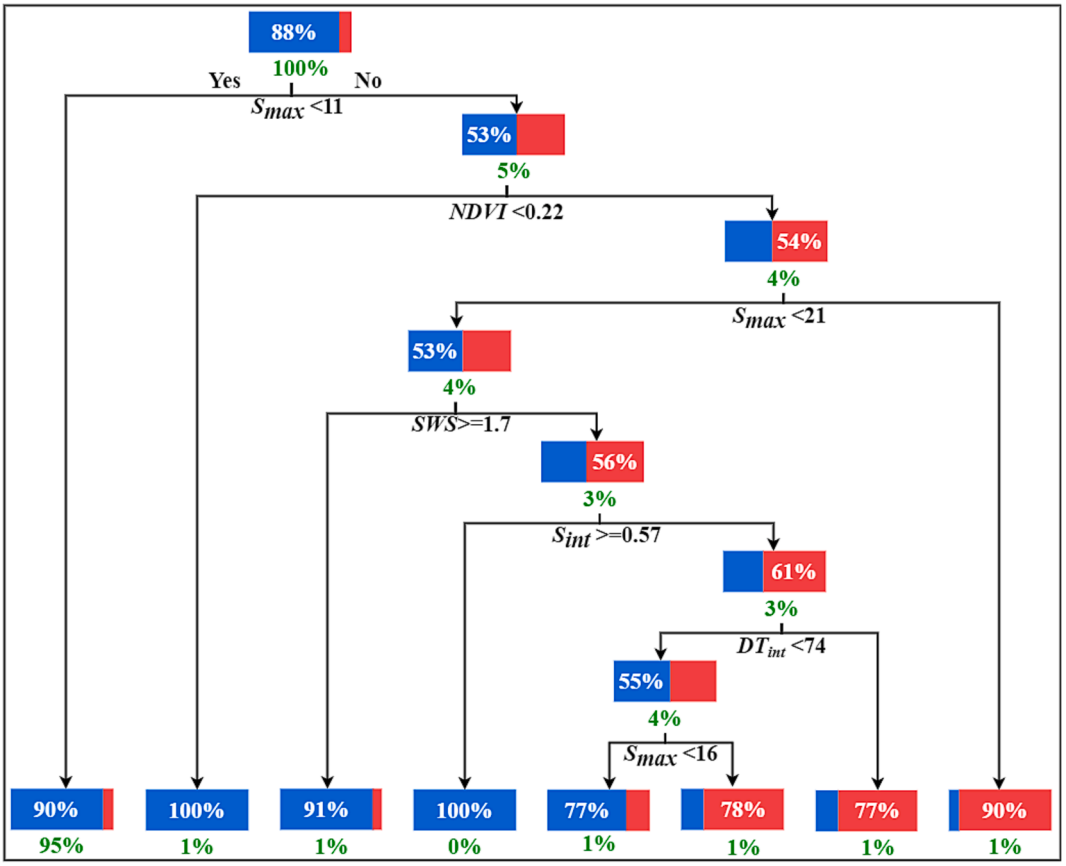


Fig. 9. Classification tree of NPF/PF for all stations (except Rock) generated by CART. The red and blue columns are the nodes of the tree and represent PF events and NPF events, respectively. The size of the different colored columns and the number in the column indicate the proportion of the corresponding type of events in that node. The green number below the node indicates the proportion of observations for that node to the total number of events.

4.4. Temporal control of the PF occurrence

Previous studies have found that rainfall characteristics and SM_{int} together determine whether the PF occurs (Guo and Lin, 2018). However, the mechanisms controlling the occurrence of PF are complex due to the interplay of these factors. The present studies mainly focused on the correlation between the PF and these control factors (e.g., Liu and Lin, 2015; Wiekenkamp et al., 2016; Demand et al., 2019), and it's hard to determine the occurrence pattern of the PF. However, using the CART we can determine the mechanism of the PF occurrence, i.e., under which circumstances the PF occurs.

We found that sufficient moisture input (*SWS* and S_{max} greater than thresholds) is required for the PF to occur, and the threshold varies across vegetation types (Table 6). It may imply that the PF is initiated by different rainfall amounts and intensities in the different vegetation types, which is similar to the previous findings. Liu and Lin (2015) found that rainfall amount needs to exceed a threshold for the PF to occur at valley floors and swales, but the PF can occur directly at hilltops. Wiekenkamp et al. (2016) found that the effect of rainfall on PF is governed by SM_{int}, especially for rainfall events over 25 mm, where the drier the soil, the more likely PF is to occur. A similar phenomenon is found in this study at the Forest, where the probability of PF occurring is high when *SWS* is greater than 0.31 cm, but S_{int} is less than 0.57 (Figs. S4, 6th terminal node). In addition, the NDVI is not only related to the proportion of the PF occurrence as a spatial factor (differences in NDVI between stations) (Table 2) but also as a temporal factor (changes in vegetation) controlling the occurrence of the PF in Forest (Fig. S4 and Table 4).

We also constructed the occurrence pattern of the PF for the whole region using the infiltration events at all stations (Fig. 9). The PF occurs under conditions in which rainfall intensity and amount exceed certain thresholds and the initial soil wetting which is similar to previous studies (Liu and Lin, 2015; Wiekenkamp et al., 2016; Demand et al., 2019). The specific thresholds for the occurrence of the PF can be obtained using the CART, and although limited by the absence of rainfall observations, this method appears to have great potential for the determination of the PF occurrence patterns. In conclusion, spatial and temporal factors together control the occurrence of PF in the study area.

4.5. Limitations and future considerations

Our previous study has found that vegetation, rainfall, soil properties, and SM_{int} influence the occurrence of PF. Specifically, low bulk density, low SM_r, high K_S, and high *SOC* favored the occurrence of PF; low S_{int} favored the occurrence of the PF in areas with high sand content (probably due to hydrophobicity), while high S_{int} favored the occurrence of the PF in areas with low sand content (Kang et al., 2022). However, the model that predicts the spatial pattern of the PF and the mechanism triggering the temporal occurrence of the PF remains unavailable. In this study, the occurrence pattern of the PF and the prediction equation of the proportion of the PF occurrence are determined using the CART, RF, and multiple nonlinear regression based on the infiltration events at a large scale using 29 observation stations.

However, the detection of the PF at distinct depths in this study has some limitations. It should be noted that detecting the PF at a specific depth does not guarantee that PF is taking place at that depth. Instead, detecting an out-of-sequence response at a certain layer indicates the presence of a PF path above a certain layer. For instance, the detection of PF in Layer 3–4 may be attributed to the existence of a PF path between Layer 3–4, leading to a synchronized response of soil moisture in both

layer 3 and 4. However, it is also possible that a PF path exists between Layer 2–4 or 1–4, which can also lead to the out-of-sequence response at Layer 3–4. While the change in the relative proportion of PF (Fig. 3) with depth may not accurately represent the vertical distribution of the PF path, it does serve as a reference for identifying its vertical variation. As far as we know, the vertical distribution of macropores can only be assessed using techniques such as staining experiments or X-ray computed tomography of soil columns (Anderson et al., 1990; Katuwal et al., 2015; Zhang et al., 2017b; Zhang et al., 2018). These invasive methods can only obtain the occurrence pathway of PF at the sampling moment. Instead, by analyzing the PF detected at varying depths, it is possible to obtain long-term dynamic characteristics of the pathways in which the PF occurs. Moreover, this does not impact the assessment of whether the PF is occurring in the actual profile.

In addition, the data sources in this study have some limitations. Since no in situ rainfall observation is available, some indicators calculated from soil moisture dynamics are used instead of the rainfall characteristics to construct the PF occurrence pattern, which may influence the computed rainfall thresholds. Previous studies have found a monotonic relationship between rainfall characteristics and soil moisture dynamics (Zhu et al., 2014, 2021; Yang et al., 2018; Glaser et al., 2019), and the advantage of the CART approach is that monotonic changes in the explanatory variables do not affect the structure of the final PF occurrence pattern (De'ath and Fabricius, 2020; Genuer and Poggi, 2020). However, it is undeniable that there are variations in the relationship between rainfall characteristics and soil moisture dynamics at different stations, and therefore there are large deviations between the thresholds obtained for PF initiation and the true thresholds. We also attempted to use the available rainfall datasets, but the current reanalysis datasets (e.g., CMFD) are not effective in estimating rainfall intensity in mountainous areas (Lai et al., 2022b), and the estimation error of rainfall intensity, in particular, is large (*RMSE* greater than 0.5 mm/3h, and $R^2 \leq 0.025$) (Fig. S7).

In addition, limited by the harsh mountainous environment, the measurement interval of soil moisture data was 30 min, which is coarser for identifying the PF (e.g., Hardie et al., 2010; Liu and Lin, 2015; Wiekenkamp et al., 2016; Demand et al., 2019), may cause some uncertainty in the identification of the PF. Finally, due to non-human factors such as instrument failure caused by animals (yak, rat, etc.), data gaps exist in the time series of soil moisture available at each station (Fig. S1). Therefore, the number of infiltration events identified at each station varies (Fig. S2) (mainly because rainfall varies greatly from station to station), which may also contribute to some uncertainty. However, the stations used in this study include the typical soils, vegetation, and elevation zones across the high and cold mountainous areas (2.7×10^4 km^2). The soil moisture at these stations was monitored over 6 years (2014–2019), and the missing soil moisture data at most stations are only a small fraction of the data series (Fig. S1). Previous studies have concluded that 1 year of continuous monitoring is sufficient to determine the proportion of PF occurrence, and greater than 3 years is sufficient to determine the controlling factors for PF (Graham and Lin, 2011; Liu and Lin, 2015). Thus, the uncertainty due to differences in soil moisture series is small and can be ignored. The large-scale and long-term soil moisture observations make the recognized patterns and mechanisms of the PF occurrence more representative of the study area and robust than the other previous studies (e.g., Liu and Lin, 2015; Wiekenkamp et al., 2016; Demand et al., 2019). Thus, the datasets used in this study help to extend the methods and the results to other similar high and cold mountainous regions.

Results indicate that these methods are significant for determining the occurrence mechanism of PF. On the one hand, RF can obtain the intrinsic influence of some factors on the proportion of the PF occurrence, which is not available by traditional methods (Koestel and Jorda, 2014) (Fig. 8); on the other hand, the occurrence pattern of the PF constructed by CART is not available by other existing methods (Fig. S4-6, Fig. 9 and Table 6). To further enhance the understanding of the PF occurrence mechanisms, we need to collect soil moisture datasets from observation networks in different climates, soils, topography, and vegetation. We need also to use more accurate rainfall data to construct the temporal patterns of the PF occurrence. Moreover, since these methods utilize some common environmental parameters (such as K_S, SM_r, rainfall characteristics, and SM_{int}), it is plausible to couple the PF prediction models and occurrence patterns with hydrological models to improve the accuracy of hydrological modeling.

5. Conclusion

Our study identified a large number of PF infiltration events in the high and cold mountainous areas of the Qilian Mountains, based on profile soil moisture data from a long-term monitoring network. By combining these long-term observations with the machine learning methods of RF and CART, we revealed the robust distribution and spatiotemporal control mechanisms of PF in mountainous areas. The main findings of our study are as follows:

- The PF varied considerably in the mountainous areas, and both site factors (e.g., soil properties and vegetation) and dynamic factors (e.g., SM_{int} and rainfall characteristics) control the occurrence of the PF at large-scale mountainous areas.
- As soil depth increases, the relative proportion of PF events detected tends to decrease. With increasing vegetation cover, the dominant soil layer in which PF was detected increased.
- Based on the RF, the study identified the K_S and SM_r as the main local factors controlling the spatial distribution of the PF, which can't be identified using the conventional methods of correlation analysis. Furthermore, an equation to predict the spatial distribution of PF proportion was developed with reasonable accuracy in mountainous regions.
- Based on the CART, the PF temporal occurrence patterns were established for different vegetation types and the whole catchment in the Qilian Mountains, and it was found that the PF is initiated when the temporal control factors exceed certain thresholds, with SM_{int}, rainfall amount, rainfall intensity being the main factors.

Our study provides new insights into the mechanism of the PF by comprehensively analyzing its spatiotemporal controlling factors, which can be useful for eco-hydrological studies in large-scale mountainous areas. Additionally, we demonstrate the great potential of combining in-situ soil moisture sensor networks with machine learning techniques, such as the CART and RF, to explore the complex mechanism of the PF.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This project is partially funded by the National Natural Science Foundation of China (Grants: 42101022, 42030501, and 91125010), and the Scherer Endowment Fund of the Department of Geography, Western Michigan University. We are grateful to the members of the Center for Dryland Water Resources Research and Watershed Science at Lanzhou University for their hard field work in installing instruments, retrieving data, and collecting and analyzing soil samples in the alpine mountains since 2012.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2023.116626.

References

Anderson, S.H., Peyton, R.L., Gantzer, C., 1990. Evaluation of Constructed and Natural Soil Macropores Using X-Ray Computed-Tomography. Geoderma. 46 (1–3), 13–29. https://doi.org/10.1016/0016-7061(90)90004-S.

Bianchi, M., Zheng, C., Wilson, C., Tick, G.R., Liu, G., Gorelick, S.M., 2011. Spatial connectivity in a highly heterogeneous aquifer: From cores to preferential flow paths. Water Resour. Res. 47 (5) https://doi.org/10.1029/2009wr008966.

Biau, G., 2012. Analysis of a random forests model. J. Mach. Learn. Res. 13 (38), 1063–1095.

Cao, S., Chen, L., Shankman, D., Wang, C., Wang, X., Zhang, H., 2011. Excessive reliance on afforestation in China's arid and semi-arid regions: Lessons in ecological restoration. Earth-Sci. Rev. 104 (4), 240–245.

Cao, R., Xu, Z., Chen, Y., Chen, J., Shen, M., 2022. Reconstructing High-Spatiotemporal-Resolution (30 m and 8-Days) NDVI Time-Series Data for the Qinghai-Tibetan Plateau from 2000–2020. Remote Sens. 14 (15) https://doi.org/10.3390/rs14153648.

Chen, Y., Cao, R., Chen, J., Liu, L., Matsushita, B., 2021. A practical approach to reconstruct high-quality Landsat NDVI time-series data by gap filling and the Savitzky-Golay filter. ISPRS J. Photogramm. 180, 174–190. https://doi.org/10.1016/j.isprsjprs.2021.08.015.

Cheng, G. et al., 2014. Integrated study of the water-ecosystem-economy in the Heihe River Basin. Natl. Sci. Rev. 1(3), 413-428. 10.1093/nsr/nwu017.

Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., Mosavi, A., 2019. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Sci. Total Environ. 651, 2087–2096.

De'ath, G., Fabricius, K.E., 2020. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology. 81 (11), 3178–3192. https://doi.org/10.2307/177409.

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

Dorigo, W.A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A.D., Zamojski, D., Cordes, C., Wagner, W., Drusch, M., 2013. Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network. Vadose Zone J. 12 (3) https://doi.org/10.2136/vzj2012.0097.

Flury, M., Flühler, H., Jury, W., Leuenberger, J., 1994. Susceptibility of soils to preferential flow of water: A field study. Water Resour. Res. 30 (7), 1945–1954. https://doi.org/10.1029/94wr00871.

Gao, M., Li, H.Y., Liu, D.F., Tang, J.Y., Chen, X.Y., Chen, X., Blöschl, G., Ruby Leung, L., 2018. Identifying the dominant controls on macropore flow velocity in soils: A meta-analysis. J. Hydrol. 567, 590–604.

Gazis, C., Feng, X., 2004. A stable isotope study of soil water: evidence for mixing and preferential flow paths. Geoderma. 119 (1–2), 97–111. https://doi.org/10.1016/s0016-7061(03)00243-x.

Geng, H., Pan, B., Huang, B., Cao, B., Gao, H., 2017. The spatial distribution of precipitation and topography in the Qilian Shan Mountains, northeastern Tibetan Plateau. Geomorphology. 297, 43–54. https://doi.org/10.1016/j.geomorph.2017.08.050.

Geng, X., Wang, X., Yan, H., Zhang, Q., Jin, G., 2014. Land Use/Land Cover Change Induced Impacts on Water Supply Service in the Upper Reach of Heihe River Basin. Sustainability. 7 (1), 366–383. https://doi.org/10.3390/su7010366.

Genuer, R., Poggi, J. M., Tuleau-Malot, C., 2015. VSURF: An R package for variable selection using random forests. The R Journal. 7(2). 10.32614/RJ-2015-018.

Genuer, R., Poggi, J.M., 2020. Random Forests with R. Use R! Springer International Publishing.

Glaser, B., Jackisch, C., Hopp, L., Klaus, J., 2019. How Meaningful are Plot-Scale Observations and Simulations of Preferential Flow for Catchment Models? Vadose Zone J. 18 (1), 1–18. https://doi.org/10.2136/vzj2018.08.0146.

Graham, C.B., Lin, H.S., 2011. Controls and Frequency of Preferential Flow Occurrence: A 175-Event Analysis. Vadose Zone J. 10 (3), 816–831. https://doi.org/10.2136/vzj2010.0119.

Guo, L., Fan, B., Zhang, J., Lin, H., 2018. Occurrence of subsurface lateral flow in the Shale Hills Catchment indicated by a soil water mass balance method. Eur. J. Soil Sci. 69 (5), 771–786. https://doi.org/10.1111/ejss.12701.

Guo, L., Lin, H., 2018. Addressing Two Bottlenecks to Advance the Understanding of Preferential Flow in Soils. Adv. Agron. 61–117.

Guo, L.i., Lin, H., Fan, B., Nyquist, J., Toran, L., Mount, G.J., 2019. Preferential flow through shallow fractured bedrock and a 3D fill-and-spill model of hillslope subsurface hydrology. J. Hydrol. 576, 430–442.

Hao, X., Ball, B. C., Culley, J. L., Carter, M. R., Parkin, G. W. 2008. Soil density and porosity. In Carter M. R. & Gregorich E. G. (Eds.), Soil sampling and methods of analysis. CRC Press, Taylor & Francis: Boca Raton, FL, USA, vol. 57, pp. 743-760.

Hardie, M.A., Cotching, W.E., Doyle, R.B., Holz, G., Lisson, S., Mattern, K., 2010. Effect of antecedent soil moisture on preferential flow in a texture-contrast soil. J. Hydrol. 398 (3-4), 191–201.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., Li, X., 2020. The first high-resolution meteorological forcing dataset for land process studies over China. Sci. Data. 7 (1) https://doi.org/10.1038/s41597-020-0369-y.

Hu, X., Li, Z.C., Li, X.Y., Liu, L.Y., 2016. Quantification of soil macropores under alpine vegetation using computed tomography in the Qinghai Lake Watershed, NE Qinghai-Tibet Plateau. Geoderma. 264, 244–251. https://doi.org/10.1016/j.geoderma.2015.11.001.

Ireson, A.M., Butler, A.P., 2011. Controls on preferential recharge to Chalk aquifers. J. Hydrol. 398 (1–2), 109–123. https://doi.org/10.1016/j.jhydrol.2010.12.015.

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. Eur. J. Soil Sci. 58 (3), 523–546. https://doi.org/10.1111/j.1365-2389.2007.00915.x.

Jarvis, N., Koestel, J., Larsbo, M., 2016. Understanding Preferential Flow in the Vadose Zone: Recent Advances and Future Prospects. Vadose Zone J. 15 (12) https://doi.org/10.2136/vzj2016.09.0075.

Kang, W., Tian, J., Lai, Y., Xu, S., Gao, C., Hong, W., Zhou, Y., Pei, L., He, C., 2022. Occurrence and controls of preferential flow in the upper stream of the Heihe River Basin, Northwest China. J. Hydrol. 607, 127528.

Katuwal, S., Moldrup, P., Lamandé, M., Tuller, M., de Jonge, L.W., 2015. Effects of CT Number Derived Matrix Density on Preferential Flow and Transport in a Macroporous Agricultural Soil. Vadose Zone J. 14 (7) https://doi.org/10.2136/vzj2015.01.0002.

Koestel, J., Jorda, H., 2014. What determines the strength of preferential transport in undisturbed soil under steady-state flow? Geoderma. 217–218, 144–160. https://doi.org/10.1016/j.geoderma.2013.11.009.

Koestel, J.K., Moeys, J., Jarvis, N.J., 2012. Meta-analysis of the effects of soil properties, site factors and experimental conditions on solute transport. Hydrol. Earth Syst. Sci. 16 (6), 1647–1665. https://doi.org/10.5194/hess-16-1647-2012.

Lai, X., Liu, Y.a., Li, L., Zhu, Q., Liao, K., 2022a. Spatial variation of global surface soil rock fragment content and its roles on hydrological and ecological patterns. Catena. 208, 105752.

Lai, Y., Tian, J., Kang, W., Gao, C., Hong, W., He, C., 2022b. Rainfall estimation from surface soil moisture using SM2RAIN in cold mountainous areas. J. Hydrol. 606, 127430.

Larsbo, M., Koestel, J., Jarvis, N., 2014. Relations between macropore network characteristics and the degree of preferential solute transport. Hydrol. Earth Syst. Sci. 18 (12), 5255–5269. https://doi.org/10.5194/hess-18-5255-2014.

Larsbo, M., Koestel, J., Kätterer, T., Jarvis, N., 2016. Preferential Transport in Macropores is Reduced by Soil Organic Carbon. Vadose Zone J. 15 (9) https://doi.org/10.2136/vzj2016.03.0021.

Li, X.Y., Hu, X., Zhang, Z.H., Peng, H.Y., Zhang, S.Y., Li, G.Y., Li, L., Ma, Y.J., 2013. Shrub Hydropedology: Preferential Water Availability to Deep Soil Layer. Vadose Zone J. 12 (4) https://doi.org/10.2136/vzj2013.01.0006.

Li, Y., Khan, M.Y.A., Jiang, Y., Tian, F., Liao, W., Fu, S., He, C., 2019. CART and PSO+KNN algorithms to estimate the impact of water level change on water quality in Poyang Lake, China. Arab. J. Geosci. 12 (9) https://doi.org/10.1007/s12517-019-4350-z.

Lin, H., 2010. Linking principles of soil formation and flow regimes. J. Hydrol. 393 (1–2), 3–19. https://doi.org/10.1016/j.jhydrol.2010.02.013.

Liu, H., Lin, H., 2015. Frequency and Control of Subsurface Preferential Flow: From Pedon to Catchment Scales. Soil Sci. Soc. Am. J. 79 (2), 362–377. https://doi.org/10.2136/sssaj2014.08.0330.

Liu, H., Zhao, W., He, Z., Zhang, L., 2007. Stochastic modelling of soil moisture dynamics in a grassland of Qilian Mountain at point scale. Sci. China Ser. D-Earth Sci 50 (12), 1844–1856. https://doi.org/10.1007/s11430-007-0128-3.

Loheide, S.P., Booth, E.G., 2011. Effects of changing channel morphology on vegetation, groundwater, and soil moisture regimes in groundwater-dependent ecosystems. Geomorphology. 126 (3–4), 364–376. https://doi.org/10.1016/j.geomorph.2010.04.016.

Lozano-Parra, J., van Schaik, N.L.M.B., Schnabel, S., Gómez-Gutiérrez, Á., 2016. Soil moisture dynamics at high temporal resolution in a semiarid Mediterranean watershed with scattered tree cover. Hydrol. Process. 30 (8), 1155–1170. https://doi.org/10.1002/hyp.10694.

Lu, L., Liu, C., Li, X., Ran, Y., 2017. Mapping the Soil Texture in the Heihe River Basin Based on Fuzzy Logic and Data Fusion. Sustainability. 9 (7) https://doi.org/10.3390/su9071246.

Meng, C., Mo, X., Liu, S., Hu, S., 2021. Extensive evaluation of IMERG precipitation for both liquid and solid in Yellow River source region. Atmos. Res. 256, 105570.

Milborrow, S., 2018. rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'. R package version 3.1.0. https://CRAN.R-project.org/package=rpart.plot.

Mossadeghi-Björklund, M., Arvidsson, J., Keller, T., Koestel, J., Lamandé, M., Larsbo, M., Jarvis, N., 2016. Effects of subsoil compaction on hydraulic properties and preferential flow in a Swedish clay soil. Soil Till. Res. 156, 91–98.

Nimmo, J.R., 2020. The processes of preferential flow in the unsaturated zone. Soil Sci. Soc. Am. J. 85 (1), 1–27. https://doi.org/10.1002/saj2.20143.

Orellana, F., Verma, P., Loheide, S.P., Daly, E., 2012. Monitoring and modeling water-vegetation interactions in groundwater-dependent ecosystems. Rev. Geophys. 50 (3) https://doi.org/10.1029/2011RG000383.

Palladino, M., Romano, N., Pasolli, E., Nasta, P., 2022. Developing pedotransfer functions for predicting soil bulk density in Campania. Geoderma. 412, 115726.

Qin, Y., Yi, S., Chen, J., Ren, S., Ding, Y., 2015. Effects of gravel on soil and vegetation properties of alpine grassland on the Qinghai-Tibetan plateau. Ecol. Eng. 74, 351–355. https://doi.org/10.1016/j.ecoleng.2014.10.008.

Rawls, W.J., Giménez, D., Grossman, R., 1998. Use of soil texture, bulk density, and slope of the water retention curve to predict saturated hydraulic conductivity. Transactions of the ASAE. 41(4), 983-988. 10.13031/2013.17270.

Ritsema, C.J., Dekker, L.W., 1994. How water moves in a water repellent sandy soil: 2. Dynamics of fingered flow. Water Resour. Res. 30 (9), 2519–2531. https://doi.org/10.1029/94wr00750.

Rothwell, J.J., Futter, M.N., Dise, N.B., 2008. A classification and regression tree model of controls on dissolved inorganic nitrogen leaching from European forests. Environ. Pollut. 156 (2), 544–552. https://doi.org/10.1016/j.envpol.2008.01.007.

Saito, T., Fujimaki, H., Yasuda, H., Inosako, K., Inoue, M., 2013. Calibration of Temperature Effect on Dielectric Probes Using Time Series Field Data. Vadose Zone J. 12 (2) https://doi.org/10.2136/vzj2012.0184.

Schwärzel, K., Menzer, A., Clausnitzer, F., Spank, U., Häntzschel, J., Grünwald, T., Köstner, B., Bernhofer, C., Feger, K.H., 2009. Soil water content measurements deliver reliable estimates of water fluxes: A comparative study in a beech and a spruce stand in the Tharandt forest (Saxony, Germany). Agric. For. Meteorol. 149 (11), 1994–2006.

Tang, Q., Duncan, J.M., Guo, L.i., Lin, H., Xiao, D., Eissenstat, D.M., 2020. On the controls of preferential flow in soils of different hillslope position and lithological origin. Hydrol. Process. 34 (22), 4295–4306.

Therneau, T., Atkinson, B., 2018. rpart: Recursive Partitioning and Regression Trees. R package version, 4.1-13. https://CRAN.R-project.org/package=rpart.

Tian, J., Zhang, B., Wang, X., He, C., 2023. In situ observations of soil hydraulic properties and soil moisture in a high, cold mountainous area of the northeastern Qinghai-Tibet Plateau. Sci. China Earth Sci. 10.1007/s11430-022-1120-5.

Tian, J., Zhang, B., He, C., Yang, L., 2017. Variability in Soil Hydraulic Conductivity and Soil Hydrological Response Under Different Land Covers in the Mountainous Area of the Heihe River Watershed, Northwest China. Land Degrad. Develop. 28 (4), 1437–1449. https://doi.org/10.1002/ldr.2665.

Tian, J., Zhang, B., He, C., Han, Z., Bogena, H.R., Huisman, J.A., 2019. Dynamic response patterns of profile soil moisture wetting events under different land covers in the Mountainous area of the Heihe River Watershed, Northwest China. Agric. For. Meteorol. 271, 225–239.

Van Genuchten, M.T., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. Soil Sci. Soc. Am. J. 44 (5), 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x.

van Schaik, N.L.M.B., 2009. Spatial variability of infiltration patterns related to site characteristics in a semi-arid watershed. Catena. 78 (1), 36–47. https://doi.org/10.1016/j.catena.2009.02.017.

Wang, R., Dong, Z., Zhou, Z., Wang, N., Xue, Z., Cao, L., 2020. Effect of vegetation patchiness on the subsurface water distribution in abandoned farmland of the Loess Plateau. China. Sci. Total Environ. 746, 141416.

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. J. Hydrol. 534, 139–149. https://doi.org/10.1016/j.jhydrol.2015.12.050.

Worthington, S.R.H., 2019. How preferential flow delivers pre-event groundwater rapidly to streams. Hydrol. Process. 33 (17), 2373–2380. https://doi.org/10.1002/hyp.13520.

Yang, Y., Chen, R.S., Song, Y.X., Han, C.T., Liu, Z.W., Liu, J.F., 2020. Spatial variability of soil hydraulic conductivity and runoff generation types in a small mountainous catchment. J. Mt. Sci. 17 (11), 2724–2741.

Yang, K., He, J., Tang, W., Qin, J., Cheng, C.C.K., 2010. On downward shortwave and longwave radiations over high altitude regions: Observation and modeling in the Tibetan Plateau. Agric. For. Meteorol. 150 (1), 38–46. https://doi.org/10.1016/j.agrformet.2009.08.004.

Yang, J., He, Z., Du, J., Chen, L., Zhu, X., Lin, P., Li, J., 2017. Soil water variability as a function of precipitation, temperature, and vegetation: a case study in the semiarid mountain region of China. Environ. Earth Sci. 76 (5) https://doi.org/10.1007/s12665-017-6521-0.

Yang, L., Zhang, H., Chen, L., 2018. Identification on threshold and efficiency of rainfall replenishment to soil water in semi-arid loess hilly areas. Sci. China Earth Sci. 61 (3), 292–301. https://doi.org/10.1007/s11430-017-9140-0.

Zhang, L., He, C., Zhang, M., 2017a. Multi-Scale Evaluation of the SMAP Product Using Sparse In-Situ Network over a High Mountainous Watershed, Northwest China. Remote Sens. 9 (11) https://doi.org/10.3390/rs9111111.

Zhang, A., Liu, W., Yin, Z., Fu, G., Zheng, C., 2016. How Will Climate Change Affect the Water Availability in the Heihe River Basin, Northwest China? J. Hydrometeorol. 17 (5), 1517–1542. https://doi.org/10.1175/jhm-d-15-0058.1.

Zhang, G., Nan, Z., Zhao, L., Liang, Y., Cheng, G., 2021. Qinghai-Tibet Plateau wetting reduces permafrost thermal responses to climate warming. Earth Planet. SC. Lett. 562, 116858.

Zhang, Y., Zhao, W., Fu, L., 2017b. Soil macropore characteristics following conversion of native desert soils to irrigated croplands in a desert-oasis ecotone, Northwest China. Soil Till. Res. 168, 176–186. https://doi.org/10.1016/j.still.2017.01.004.

Zhang, Y., Zhao, W., He, J., Fu, L., 2018. Soil Susceptibility to Macropore Flow Across a Desert-Oasis Ecotone of the Hexi Corridor, Northwest China. Water Resour. Res. 54 (2), 1281–1294. https://doi.org/10.1002/2017wr021462.

Zhao, S.Y., Jia, Y.W., Gong, J.G., Niu, C.W., Su, H.D., Gan, Y.D., Liu, H., 2020. Spatial Variability of Preferential Flow and Infiltration Redistribution along a Rocky-Mountain Hillslope, Northern China. Water. 12 (4), 1102.

Zhu, Q., Nie, X., Zhou, X., Liao, K., Li, H., 2014. Soil moisture response to rainfall at different topographic positions along a mixed land-use hillslope. Catena. 119, 61–70. https://doi.org/10.1016/j.catena.2014.03.010.

Zhu, P., Zhang, G., Wang, H., Zhang, B., Liu, Y., 2021. Soil moisture variations in response to precipitation properties and plant communities on steep gully slope on the Loess Plateau. Agric. Water Manage. 256, 107086.