Process oriented insights from interpretable machine learning - what influences flood generating processes?

L. Stein[1], M. P. Clark[2,3], W. J. M. Knoben[2], F. Pianosi[1], R. Woods[1]

[1]Department of Civil Engineering, University of Bristol, Bristol, UK
[2]University of Saskatchewan Coldwater Lab, Canmore, Alberta T1W 3G1, Canada
[3]Centre for Hydrology, University of Saskatchewan, Saskatoon, Saskatchewan S7N 1K2, Canada

Key Points:

- Flood generating processes are mostly influenced by climate attributes.
- Influential attributes vary between different processes and climates.
- Mix of flood generating processes can be predicted for ungauged catchments with high accuracy.

Corresponding author: Lina Stein, `lina.stein@bristol.ac.uk`

Abstract

Hydroclimatic flood generating processes, such as excess rain, short rain, long rain, snowmelt and rain-on-snow, underpin our understanding of flood behaviour. Knowledge about flood generating processes helps to improve modelling decisions, flood frequency analysis, estimation of climate change impact on floods, etc. Yet, not much is known about how climate and catchment attributes influence the distribution of flood generating processes. With this study we aim to offer a comprehensive and structured approach to close this knowledge gap. We employ a large sample approach (671 catchment in the conterminous United States) and test attribute influence on flood processes with two complementary approaches: firstly, a data-based approach which compares attribute probability distributions of different flood processes, and secondly, a random forest model in combination with an interpretable machine learning approach (accumulated local effects). This machine learning technique is new to hydrology, and it overcomes a significant obstacle in many statistical methods, the confounding effect of correlated catchment attributes. As expected, we find climate attributes (fraction of snow, aridity, precipitation seasonality and mean precipitation) to be most influential on flood process distribution. However, attribute influence varies both with process and climate type. We also find that flood processes can be predicted for ungauged catchments with relatively high accuracy (R2 between 0.45 and 0.9). The implication of these findings is that flood processes should be taken into account for future climate change impact studies, as impact will vary between processes.

1 Introduction

Flood processes influence flood behaviour (Gaál et al., 2012; Fischer et al., 2016; Keller et al., 2018; Merz & Blöschl, 2005; Tarasova et al., 2019). Possible hydroclimatic flood processes include excess rain, short rain, long rain, snowmelt and rain-on-snow. The need to classify these processes has long been recognised and several studies have developed flood classification approaches (e.g. Berghuijs et al., 2016, 2019; Blöschl et al., 2017; Diezig & Weingartner, 2007; Merz & Blöschl, 2003; Sikorska et al., 2015; Stein et al., 2019; Tarasova et al., 2020). However, very few of those studies look into how catchment and climate characteristics influence flood generating processes (Merz & Blöschl, 2003; Stein et al., 2019).

Being able to predict which flood generating processes might occur in a catchment is relevant for many applications. For hydrological model development it is important to know which process representations must be included (Clark et al., 2016); for model evaluation it can help to evaluate model outputs in the sense of getting the right results for the right reasons (Kirchner, 2006). Moreover, knowing which catchment characteristics are relevant for processes in various areas might improve the choice of donor catchments for predictions in ungauged catchments through regionalization (Rosbjerg et al., 2013). Furthermore, climate change can drive changes in flood process, which may affect flood magnitude (Blöschl et al., 2017, 2019). Knowing the temporal and spatial distribution of processes can potentially inform or explain changes in flood characteristics.

Based on existing literature we can formulate several hypotheses regarding which climate and catchment attributes we expect to influence the mix of flood generating processes. In the following section we will describe the studies that inform the hypotheses which are summarised in Table 1.

1.1 Flood hypotheses - What do we expect?

Deciding which processes in a catchment generate flood events depends on two factors: the availability of the flood producing input, and how the catchment stores and transmits water. Here we briefly outline some of the impacts of climate and catchment attributes on flood generation.

Climate and Weather: The availability of the input is dependent both on climate and weather. Precipitation and temperature distribution influence snowpack accumulation. Locations with winter precipitation and winter temperatures continuously below zero during the winter months can accumulate a snowpack that will not melt until the spring or summer. Sudden increases in temperature (Ward, 1978) or the melting of the snowpack through precipitation creates either snowmelt or rain/snow floods. In catchments with winter temperatures fluctuating around freezing, rain/snow events can also occur during the winter. Southern Germany for example often experiences floods in late December caused by combinations of rain and snow (Sui & Koehler, 2001).

For floods generated by short rain, long rain or excess rainfall, the availability of input is dependent on rainfall and evapotranspiration distribution. As the name implies short rainfall floods occur after short intense periods of rainfall that exceed infiltration capacity or quickly saturated the catchment (Merz & Blöschl, 2003). Arid regions can be more prone to this type of flood since convective thunderstorms are a common precipitation input. The distinction between excess rainfall and long rainfall is based on antecedent conditions (Stein et al., 2019) and therefore dependent on precipitation versus evapotranspiration seasonality. The seasonality is out of phase if precipitation maximum is in the winter (summer) with an evapotranspiration maximum in the summer (winter). This means precipitation maximum falls into a time where drying of the soil is minimum leading to saturated conditions. This increases the chances of excess rainfall floods. With in phase seasonality, precipitation maximum and evapotranspiration maximum aligns, leading to drier conditions. Heavy multi-day rainfall is needed to saturated catchment storage before runoff is increasing.

In catchments with only one input type, for example under continuously saturated conditions with no snow, flood generating process is independent of catchment characteristics. Catchment characteristics might influence flood characteristics but not generating process. However, in catchments that receive various inputs, the catchment storage and transmission behaviour will heavily influence which process generates the highest flows. Catchment attributes that increase runoff influence flood magnitude. Attributes that influence time of concentration can equally be deciding between processes. A short time of concentration means an immediate reaction to precipitation input. This would make short rainfall more likely than long rainfall (Blöschl et al., 2013). However, a short time of concentration can also be reached through prior saturation of the catchment (Acreman & Holden, 2013; Ward, 1978). If this is necessary, it will mean the catchment is more prone to excess rainfall floods.

Snowmelt and the interaction of rainfall and snowmelt is dependent on snowpack conditions. These depend both on climate conditions as well as weather conditions during snowpack accumulation and melting season. Rainfall retention capacity of a catchment varies depending on the snowpack conditions (Singh et al., 1998). This influences reaction of the snow pack to rainfall, thus increasing or decreasing the chance of a rain-on-snow flood. This kind of flood is strongly dependent both on antecedent condition of the snowpack and the rainfall producing weather system (Marks et al., 1998, 2001; Musselman et al., 2018; Sui & Koehler, 2001)

Slope: The influence of slope varies between different flood processes. Chang et al. (2014) find steep catchments transport meltwater more quickly to the stream especially in combination with thin soils. Yet, in steep catchments melting conditions are reached gradually along the elevation gradient, as temperature varies with elevation. For catchments in the plains there is little or no temperature gradient and melting conditions are reached simultaneously over the whole area. This can cause large snowmelt flood peaks (Ward, 1978). In regard to rainfall induced floods, slope can influence transit time, with steeper catchments transporting water more quickly to the outlet (Tetzlaff et al., 2009). Subsequently, Tetzlaff et al. (2009) found transit times in flatter terrain in temperate regions to be longer. However, this was only the case in areas with permeable soils. Slope can also be an influential attribute as a proxy for soil thickness, where steeper slopes can also have thinner soils and therefore less storage and quicker transmission (Pitlick, 1994). Additionally, catchments with steeper slope might be more prone to flash floods, where increased erosion promotes efficient drainage systems. These in turn are going to contribute to flash floods (Gaál et al., 2012; Weingartner et al., 2003). Despite these findings, in a global flood frequency study Smith et al. (2015) did not find slope to be a good predictor for the shape of the flood frequency curve. Similarly, Pitlick (1994) demonstrate that flood magnitude does not vary with catchment slope in their study region. The effect of slope on processes is therefore still under debate.

Area: Smith et al. (2015) find area to be a good predictor for flood magnitude in a flood frequency approach. It is important to note, though, that this performance varies with climate. In arid regions, area alone was not a good predictor. This was confirmed by Tooth (2000) who find rainfall variability has a stronger effect on flood magnitude.

Short rainfall events might be more common in smaller catchments due to two effects. Firstly, area affects time of concentration with smaller catchments having a shorter time of concentration. Secondly, small catchments can be covered in its entirety by high intensity convective storms. A larger catchment might only be partially covered with rainfall amounts too small to cause a flood (Weingartner et al., 2003). For very large catchments Ward (1978) mention that snowmelt is the most likely flood producing process as it is the only input that can occur across the whole large area at the same time. In arid regions, area has been found to be less influential for flood magnitude, as rainfall variability has a stronger effect (Tooth, 2000). Pitlick (1994) similarly found no increase flash flood potential for larger catchments.

Shape: Catchment shape influences flood peak shape (Ward, 1978). In a round catchment with simultaneous input everywhere the flood waves from different parts of the catchment will overlap at the outlet with high peak flows as a result. This effect will be strongest when storm duration is the same as catchment time of concentration (Viglione & Blöschl, 2009; Blöschl et al., 2013). There are exceptions though. Elongated catchments can receive very high peak flow if a storm cells moves along the catchment toward the outlet. Again the flood waves will overlap and cause high peak flow.

Soils: Soils in addition to geology and topography contribute to storage capacity of a catchment. A high storage capacity requires larger input volumes before runoff occurs (Merz & Blöschl, 2003). Once storage capacity is exceeded floods can reach larger magnitudes which is visible as a step-change in the flood frequency curve (Rogger et al., 2012). Wood et al. (1990) found that soil properties are most relevant for floods of small magnitude while rainfall properties are more relevant for larger magnitude floods. Wetlands similarly contribute to the storage capacity of a catchment. A wetland's effect on downstream flood characteristics depends largely on the saturation state. Once saturated most rainfall contributes immediately to runoff (Acreman & Holden, 2013; Bullock & Acreman, 2003; McCartney et al., 1998). Catchments with larger storage capacity would therefore be more likely to flood after wet antecedent conditions (excess rainfall floods).

Elevation: Merz and Blöschl (2003) found that in Austria flood process and time of occurrence changes with elevation. The change agrees with availability of input, e.g. flash floods occur only in the summer months and snowmelt floods in the spring. The higher the elevation the later in the spring snowmelt floods occur. Elevation is directly related to temperature and precipitation.

Geology: Geology shapes topography. The way the drainage network will form depends on geology as well as climate, vegetation and soils. Large subsurface storage dampens flood response. This leads to less erosion and more soil development thus again increasing storage (Rosbjerg et al., 2013).

Vegetation: The influence of vegetation, in particular forest vegetation, deforestation and reforestation, has been discussed in depth in the literature. While some large scale studies find forests to have an effect on magnitude and frequency (Bradshaw et al., 2007), other based on extensive literature research, disagree (Bruijnzeel, 2004; Calder & Aylward, 2006). Some of this disagreement is due to scale, with smaller catchments and smaller flood magnitude more influenced by vegetation (Calder & Aylward, 2006; van Dijk et al., 2009). In regard to flood processes, snowmelt floods have been shown to be influenced by coniferous trees, which intercept snowfall and increase sublimation rates (Storck et al., 2002). We can hypothesise that vegetation decreases quick runoff since it both increases surface roughness as well as soil infiltration capacity(Lull & Reinhart, 1972) . The effect of land-use on floods is stronger in smaller catchments (Calder & Aylward, 2006) and for smaller flood magnitudes (van Dijk et al., 2009). Vegetation is an important influence on runoff behaviour in semi-arid and arid regions as it increases infiltration capacity (Osterkamp & Friedman, 2000; Shafer et al., 2007).

1.2 Aims and Objectives

The brief review above demonstrates that the majority of studies do not in fact evaluate the influence of catchment characteristics on flood generating processes but only on other flood characteristics (runoff, magnitude, duration...). We can only infer hypotheses for the effect on flood generating processes, as we have done in Table 1, while a comprehensive, data-based, comparative study to test the influence of catchment characteristics on flood generating processes is still missing. To this end, the aim of this study is to evaluate which assumptions and prior findings hold true when tested on a large sample of catchments across several climates. We hypothesise that climate attributes will be very influential, especially the seasonal availability of flood producing input and saturation conditions. Catchment attributes that influence the storage behaviour of the catchment will likely have an effect as well (Merz & Blöschl, 2009).

We have previously developed and tested the first global event-based flood classification methodology (Stein et al., 2019). In this study we use that methodology to classify flood generating processes for a large sample of catchments across several climates. We then explore the influence of catchment attributes on generating processes. Finding influential attributes in correlated datasets is challenging, as the correlation among attributes might obscure findings (Dormann et al., 2013). We therefore use two approaches that complement each other and allow interpretation despite collinearity. The first is a data based approach that evaluates the influence of each attribute individually. The second approach uses a random forest model and an interpretable machine learning method (called accumulated local affects (Apley & Zhu, 2016)) which is unbiased toward correlated predictors (Molnar et al., 2018). This is, to our knowledge, the first application of accumulated local effects in a hydrological study.

2 Methodology

In order to understand the influence of catchment and climate attributes on flood generating processes, we propose the following steps. We will use the CAMELS dataset (Catchment Attributes and MEteorology for Large-sample Studies) (Section 2.1) by Newman et al. (2015) and Addor et al. (2017). We first classify flood generating processes of flood events using a recent methodology (Stein et al., 2019) (Section 2.2). Then, for three distinct climate groups (Section 2.3) the influence of catchment attributes is determined using both a data-based approach (probability density comparison, Section 2.4.1) and a machine learning approach (Accumulated Local Effects applied to random forest models, Sections 2.4.2 and 2.4.3).

Table 1. Literature informed hypotheses on which catchment attributes influence flood generating processes. ↑ (↓) indicates that an increase (decrease) in that attribute indicates an increased influence of the attribute on the process. − indicates that the attribute has not been found influential. Study location indicates where the study took place, with a dash (−) indicating that location was not specified.

Process	Catchment attribute	Study location	Reference
Excess rain	Vegetation ↑	United States	Lull and Reinhart (1972)
	Subsurface Storage ↑	−	Rosbjerg et al. (2013)
	Soil Storage ↑	Austria	Merz and Blöschl (2003)
	Wetlands ↑	−− Zimbabwe	Acreman and Holden (2013); Bullock and Acreman (2003); McCartney et al. (1998)
	Precipitation peak winter	Great Britain	Institute of Hydrology (IoH) (1999)
	Mean annual precipitation ↑	Austria	Merz and Blöschl (2009)
	Precipitation intensity ↓	−	Rosbjerg et al. (2013)
Short rain	Vegetation ↓ − ↓	United States, United States, United States	Osterkamp and Friedman (2000); Pitlick (1994); Shafer et al. (2007)
	Subsurface Storage ↓	−	Rosbjerg et al. (2013)
	Soil Storage ↓	Austria	Merz and Blöschl (2003)
	Round catchment shape	−	Ward (1978)
	Area ↓ −	United States ,United States, Switzerland	O'Connor and Costa (2004); Pitlick (1994); Weingartner et al. (2003)
	Slope ↑ −	United States, Austria, Global, −, temperate regions	Costa (1987); Gaál et al. (2012); Smith et al. (2015); Ward (1978); Tetzlaff et al. (2009)
	Precipitation intensity ↑	United States	Pitlick (1994)
	Aridity ↑	−, Global	French and Miller (2011); Stein et al. (2019)
	Mean annual precipitation ↓	Austria	Merz and Blöschl (2009)
Long rain	Subsurface Storage ↑	Austria	Rosbjerg et al. (2013)
	Soil Storage ↑	Austria	Merz and Blöschl (2003)
	Area ↑	Switzerland	Weingartner et al. (2003)
	Precipitation peak summer		
Snowmelt	Vegetation ↓	−, United States	D. Miller (1964); Storck et al. (2002)
	Soil Storage ↓	United States	Chang et al. (2014)
	Area ↑	United States	Ward (1978)
	Slope ↑↓	United States, −	Chang et al. (2014); Ward (1978)
Rain/Snow	Vegetation ↓	United States, United States, United States	Marks et al. (1998, 2001); D. Miller (1964)
	Winter temperature	Austria	Singh et al. (1998)
	Precipitation peak summer	Germany	Sui and Koehler (2001)
	Elevation ↑	Germany, Germany, United States	Sui and Koehler (2001); Freudiger et al. (2014); Musselman et al. (2018)
	Precipitation intensity	United States	Musselman et al. (2018)

2.1 Data

We used the publicly available CAMELS dataset which combines hydro-climatological data (Newman et al., 2015) with catchment attributes (Addor et al., 2017) for 671 catchments in the contiguous United States. The daily data covers a time period from 1980 to 2014. Newman et al. (2015) selected these catchments specifically to have minimal human influence. The majority are therefore small headwater catchments. For the flood event classification daily observed streamflow data and Daymet meteorological forcing data (precipitation, temperature) were used. For the soil moisture routine (Stein et al., 2019) available water capacity of the soil is a necessary variable. We used data from the Gridded National Soil Survey Geographic (gNATSGO) Database for the Conterminous United States (Soil Survey Staff, 2019), because this parameter is not available in CAMELS. Newman et al. (2015) provide daily actual evapotranspiration values from the Sacramento Soil Moisture Accounting Model. Addor et al. (2017) extended the CAMELS dataset by Newman et al. (2015) to include catchment attributes in six thematic groups: topography, climate, soil, vegetation, geology and streamflow indices. These combine continuous and categorical attributes. Examples of continuous attributes are mean annual precipitation or fraction of the catchment covered by forest. Categorical attributes include, for example, dominant land cover and geologic class. Detailed descriptions and definitions for each attribute can be found in Tables 1-6 in Addor et al. (2017). We additionally calculated catchment shape as represented by the elongation ratio Schumm (1956). A value closer to 1 indicates a round catchment, a value closer to zero a long catchment.

2.2 Flood process classification

Flood events were identified using a peaks-over-threshold approach. It identifies the highest independent streamflow peaks in the time series. The number of peaks varies depending on the threshold which can be set to find a certain number of peaks per year. The R function "findPeaks" from the package "quantmod" (Ryan & Ulrich, 2019) was used to identify all peak streamflow days. Only independent flood peaks are kept for further analysis. For any flood peak identified by "quantmod" to be independent from another, the time difference between both peaks has to be larger than the mean rising time calculated from 5 'clean hydrographs' (Cunnane, 1979), which we take here as the 5 highest peaks with a large time difference to the previous peak. An additional independence criterion is that a trough between two peaks needs to be less than 2/3 of the first peak (Cunnane, 1979). Two subsets of peaks with different magnitudes were identified: One with an average of one event per year (larger peaks) and one with an average of three (smaller peaks) events per year to compare if a difference in magnitude has an effect. We include this option because several studies indicate that land use or storage capacity are more influential for floods of smaller magnitude (Rogger et al., 2012; van Dijk et al., 2009; Wood et al., 1990). If there are more events than the threshold, the smallest peak events are removed.

We classified the identified flood events in each catchment into one of five hydro-climatological generating processes (Stein et al., 2019): excess rainfall, short rainfall, long rainfall, snowmelt and a combination of rain and snowmelt. A decision tree evaluates hydro-climatic conditions in a 7-day time period before any flood event. It uses the date of the flood event and hydroclimatological input data, as well as soil moisture and snowmelt estimates obtained from a simple lumped model routine run at a daily time step (Stein et al., 2019). Critical temperature for snowfall and melt was set to 1° C (Jennings et al., 2018). The thresholds of the tree are based on the hydro-climatological time series of each catchment. This methodology allows us to classify a large sample of flood events across various climatic regions without prior knowledge about dominant flood generating processes for each catchment. The tree is structured to first evaluate if snowmelt and rainfall occur simultaneously. This would be classified as rain/snow floods. In a next step it checks if snowmelt was higher than the threshold, which would indicate a snowmelt flood event. Then, if neither was the case, soil moisture state in combination with higher than mean weekly rainfall is evaluated to determine if the flood event was an excess rainfall flood. If that was not the case it evaluates whether the thresholds for long rainfall and then short rainfall are crossed. If no process could be identified, the class "other" will be assigned. Events classified as other will not be considered in this analysis. For an in-depth description and evaluation of this methodology please refer to Stein et al. (2019).

2.3 Climate type definition

Climatic catchment attributes are influential on catchment flow behaviour (Addor et al., 2018; Berghuijs et al., 2014; Jehn et al., 2020; Knoben et al., 2018). In regard to flood process distribution Berghuijs et al. (2016) note the influence of aridity on distribution of excess rainfall and short/long rainfall floods. Based on availability of generating processes, there will be very few or no snowmelt or rain/snow floods in catchments with small or zero fraction of precipitation falling as snow. Since the importance of theses two attributes, aridity and fraction of snow, is already known we can split the dataset into different climate types to evaluate the interaction of these attributes with others. We want to determine whether the importance of other catchment attributes varies between the different climate types. The CAMELS data is well suited to answer this as the catchments lie within various different climatic regions. Based on two climatic indices from the CAMELS dataset (Addor et al., 2017) the catchments were separated into three different groups: wet, dry, and snow influenced catchments. Wet catchments were defined as catchments with an aridity index < 1. Potential evapotranspiration in those catchments is lower than precipitation (i.e. energy-limited catchments). Dry catchment have an aridity index > 1 respectively with mean potential evapotranspiration larger than mean precipitation (i.e. water-limited catchments). All catchments with a fraction of precipitation falling as snow higher than 20 % were designated as snow catchments, regardless of their aridity. The flood process classification shows that his thresholds delivers largely similar numbers of catchments for all climate types while grouping catchments with snowmelt flood contributions together (see Figure 5a). The distribution of catchments for each climate type is depicted in Figure 1.

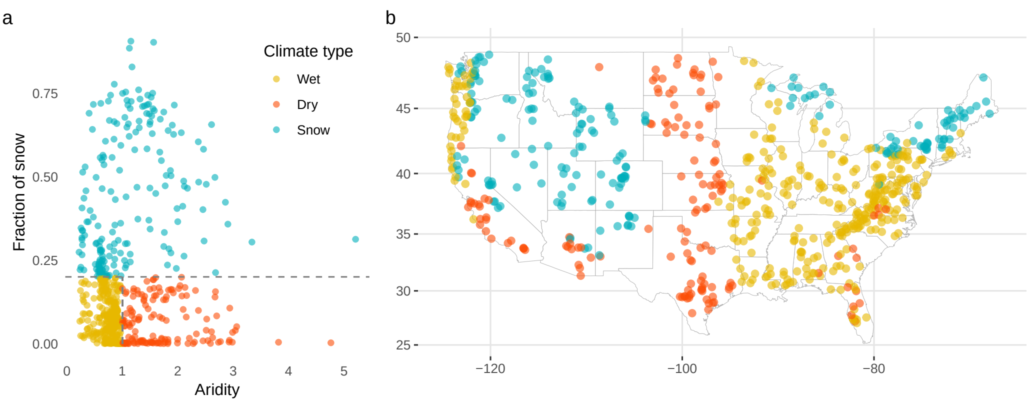

Figure 1. a: Classification of 671 CAMELS catchments into three climate types wet, dry and snow based on aridity and fraction of snow. Climate type thresholds are indicated through dashed lines. Aridity and fraction of snow taken from Addor et al. (2017). b: Spatial distribution of the three climate types for the CAMELS catchments.

2.4 Attribute influence estimation

We employ two methods to evaluate which attributes influence the flood process distribution. Both methods are described in detail further below. The distribution comparison (Section 2.4.1) draws its results directly from the data. A single attribute's influence on each flood-generating process is directly evaluated. However, the method struggles with unequal sample sizes, which is the case particularly in the "wet" climate type. The second approach compares importance between different attributes using a random forest model (Section 2.4.2) to which we apply an interpretable machine learning method (Section 2.4.3). This method, called accumulated local effects, is not biased towards collinearity in the data and less susceptible to differences in sample size. However, its reliability depends on the performance of the model to which it is applied (random forest model in our case).

The first approach depicts the influence of a single attribute on the different flood processes within catchments. It allows a comparison of influence between different processes and climates. In contrast, the second approach compares the influence of all attributes on the spatial distribution of a single process. This then allows us to compare the influence between attributes, but only for a single process.

2.4.1 Probability distribution comparison

We first evaluate the influence of each continuous attribute on each flood generating process. We stratify this analysis by climate type (Section 2.3) because it is plausible that the influence of attributes varies with environmental setting. For this we apply a comparative hydrology approach (Falkenmark et al., 1989; Gaál et al., 2012). To assess the influence of one attribute on one process we compare two attribute distributions with each other sampled across all catchments within one climate type. Each catchment can contribute the same attribute multiple times depending on the number of events. We compare the empirical cumulative distribution function (ecdf) of the attribute, sampled from all catchments with events of that process, with the ecdf associated with all events (independent of process). If the two distributions differ, we infer that this attribute influences the occurrence of that process (Gaál et al., 2012; Merz et al., 2006; Pianosi & Wagener, 2015). E.g. say a catchment had 15 excess rain events and the mean annual rainfall attribute in that catchment were 400 mm per year. Then this catchment would contribute the value 400 mm 15 times to the specific process distribution. Whereas, the next catchment has 10 excess rainfall events and a mean annual rainfall of 350 mm per year. It would contribute the value 350 mm 10 times to the same distribution. Similar methods have been applied by Merz et al. (2006) to study runoff coefficients and by Gaál et al. (2012) to evaluate flood duration.

To make the distributions comparable, all attribute values are normalised (min-max-normalisation). To summarise the divergence between the two distribution functions, we calculate the mean difference between 100 values along the ecdf curve for each process and the curve for all events. The resulting value may be either positive or negative. Figure 2c illustrates that a negative (positive) value indicates an increased occurrence of the process for smaller (larger) values of the attribute. Figure 2d displays how the mean difference between each process curve and the full range curve translates into a single metric. We use cumulative density functions instead of probability density functions as they can be calculate without any prior parameter assumptions (Pianosi & Wagener, 2015).

We chose this approach over a correlation based analysis since a simple correlation analysis would only be able to determine linear relationships between attribute and flood process. The comparison of ecdf curves instead is able to indicate both linear and non-linear relationships by taking into account variations across the whole attribute space. Although rank correlation would be able to give similar results as the curve summary statistics, the comparison over the whole curve additionally allows a visual interpretation of influential attribute ranges (supplemental information).

A drawback of this approach is that the distribution functions are sensitive to unequal sample size and to small samples (e.g. the overall number of snowmelt and rain/snow flood events in dry catchments is small). If one sample is much larger than the others, it dominates the comparison distribution (e.g. there are much more excess rainfall events in wet catchments than any other process). A small sample size may lead to a possibly inaccurate approximation of the real distribution function (an example in Figure 2b is the distribution of snowmelt events in dry catchments). For this reason more weight should be given to distributions based on a larger sample size. In Figure 2d this is taken into account by adjusting the point size according to the sample size. Another limitation is that only continuous variables can be analysed in this way. Lastly, a limitation of the applied summary statistic is that attributes that reverse their influence (i.e., the ecdf curves cross one another) would be summarised to zero. We visually checked all curves and there is only one case (influence of mean annual precipitation on rain/snow floods in "snow" climate) where this occurs.

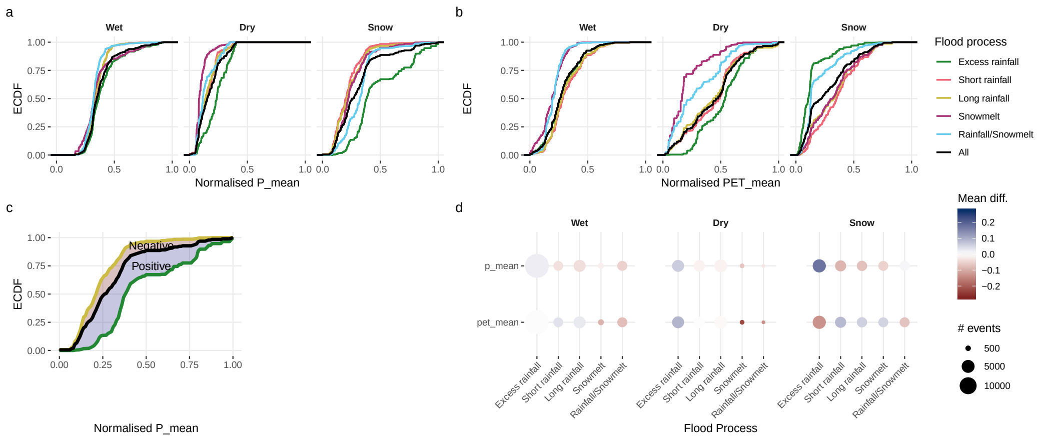

Figure 2. Example figure to explain distribution comparison and difference between distribution value. a: empirical distribution functions of normalised mean annual rainfall for each flood generating process compared to all events (black). b: empirical distribution functions of normalised mean annual potential evapotranspiration for each flood generating process compared to all events (black). c: Example distribution differences - the coloured space indicates the difference between the long rainfall events and all events (red, difference between distributions is negative) and difference between excess rainfall events and all events (blue, difference between distributions is positive). d: Summary statistic - Mean difference between distribution value for both example attributes. Colour indicates the direction and strength of difference. The number of events that contribute to a distribution is indicated through point size.

2.4.2 Random forests

A random forest is a machine learning model approach, where an algorithm creates and combines multiple regression trees (Breiman, 2001) . Addor et al. (2018) use a random forest model to predict hydrologic signatures in space, using the CAMELS dataset. They list the benefits of random forest models as allowing multiple predictors, being able to incorporate nonlinear relationships, flexibility, a reduced risk of data overfitting compared to individual regression trees, interpretability, and computational efficiency.

We use a random forest model in two ways: (1) As a model that we can interpret using interpretable machine learning; and (2) to predict flood process distributions from catchment characteristics. The latter can demonstrate both that catchment attributes influence process distribution, and that it is possible to predict flood process distribution in ungauged catchments. Prediction accuracy is evaluated using 10-fold cross-validation (see e.g. Addor et al., 2018). Random forest models tend to overfit on training data and cross validation gives a better evaluation of prediction accuracy than performance evaluation based on training data (Dormann et al., 2013). Therefore, the dataset is split into ten equal-sized samples. Ten random forest models are trained with nine parts of the data and evaluated on the respective tenth part. This way prediction accuracy can be evaluated for all catchments.

2.4.3 Accumulated local effects applied to random forest

To interpret a random forest model, Addor et al. (2018) refer to the possibility of determining the influence of an attribute on the outcome through variable importance (the increase in error when a predictor is shuffled). However, this metric is unsuitable for datasets with correlated features (Toloşi & Lengauer, 2011; Degenhardt et al., 2019; Dormann et al., 2013) such as the CAMELS dataset. Jehn et al. (2020) demonstrate high correlations between various attributes of the CAMELS dataset. An alternative to variable importance are interpretable machine learning approaches. One method that is particularly suitable to give an unbiased result despite collinearity, are accumulated local effects plots (Apley & Zhu, 2016). Accumulated local effects (ALE) plots improve the application of more commonly used partial dependence plots (Anchang et al., 2020; Friedman, 2001; Molnar, 2019). After a model was fit to the data, ALE plots evaluate the change in model prediction over a small interval of an input variable. Interval size is determined by quantiles in the distribution (Molnar, 2019). For all observed data points in that interval, differences in prediction between the interval boundaries are calculated. This way the change in the variable of interest (local effect) is recorded, disregarding any effect correlation with other variables might have. The local effects for each boundary are accumulated into a curve and centred around zero. Example accumulated local effects curves are displayed in Figure 3a (black lines) for mean precipitation, mean potential evapotranspiration and water fraction in the soil. Any divergence from zero reveals an influence of the attribute on the prediction outcome. Blue bars in Figure 3a indicate the divergence which is taken as influence. For an in depth explanation of ALE plots we refer the reader to Apley and Zhu (2016).

Accumulated local effects are a relatively new method. They have proven their applicability in several fields, for example in ecology (e.g. Anchang et al., 2020; Brown et al., 2020). One limitation is that accumulated local effects evaluate the reaction of a model to changes in an attribute. Results are not directly based on data. It can therefore be assumed that accumulated local effects calculated on a model with low performance will yield less reliable results. We recognise but were unable to quantify this limitation when interpreting the results. Another limitation is that ALE plots do not give reliable results in attribute ranges with scarce data (Molnar, 2019). Interval size over which the accumulated local effects is calculated is not regular but instead is based on an equal number of observations per interval. In unevenly distributed data this can lead to large interval sizes. In the CAMELS dataset that is the case for the attributes fraction of top 1.5 m considered water (Water Fraction) and organic (Organic Fraction) as well as carbonated rocks fraction (Carbonate rocks fraction). Figure 3 demonstrates how the unevenly distributed fraction of water in the soil data (Figure 3b) translates into only two intervals, one a zero and one at 10 (Figure 3a, blue bars).

To summarise the influence an attribute has into one number (instead of a curve), we calculate the mean absolute values of the accumulated local effects (bars in Figure 3a). This value is comparable between attributes of the same model, but not between different models. Therefore the summarised values are normalised (min-max-normalisation) for comparability. In the example given mean evapotranspiration would rank as most influential with a value of 0.93, followed by mean precipitation at 0.84. Due to the uneven distribution water fraction would still have a relatively high importance at 0.78.

Although random forest models and accumulated local effects can interpret categorical variables, we decided to not include them. This way the two methods analyse the same catchment attributes. The random forest model has been implemented using the 'randomForest' package in R (Liaw & Wiener, 2002) and the accumulated local effects were calculated using the package 'iml' (Molnar et al., 2018).

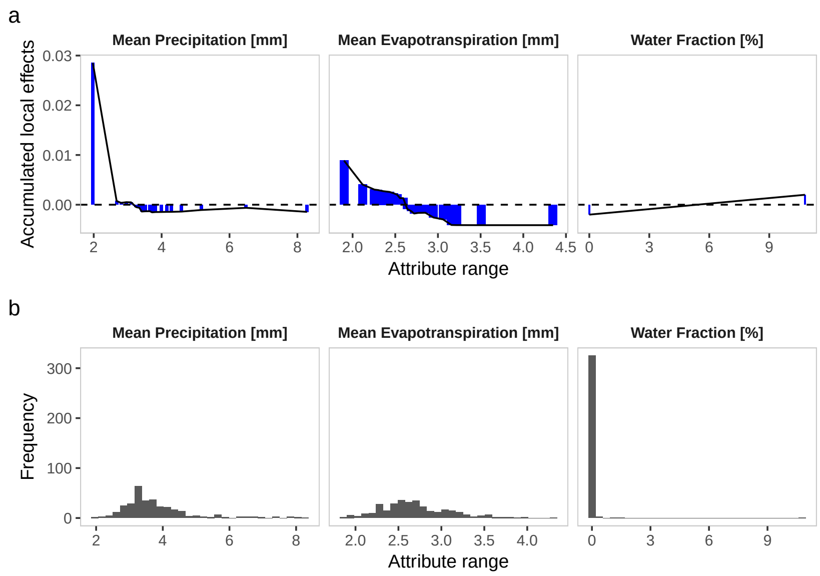

Figure 3. Example figure accumulated local effects plot and its limitation. a: Accumulated local effects plot for predictions of snowmelt floods in wet climate catchments. The dashed line is the zero line. Blue bars indicate interval locations identified by the ALE algorithm. Their divergence from zero is calculated and the mean is taken as a summary value. b: Data distribution for each of the example attributes.

3 Results

3.1 Event classification

Figure 4 illustrates the contribution for each flood generating process in each catchment. Excess rainfall floods are most common in the eastern and north-western United States. Short rainfall floods occur most often in the western United States. Snowmelt floods are most common in the western-central United States where the Rocky Mountains are. In the north-eastern United States rain/snow floods are common. Long rainfall floods are most common in the great plains area in the central US. Out of all (61,764) identified flood events the majority of events are excess rainfall floods (Figure 5b). In wet climates excess rainfall floods occur in every catchment (Figure 5a). In drier regions short and long rainfall events are more common, with fewer or no events classified as excess rainfall. The combination of rainfall and snowmelt rarely occurs, but several snowmelt floods were identified. Catchments with the climate type snow accordingly classify more events as snowmelt and rainfall/snowmelt. Several catchments with large percentages of snowmelt floods also classify large contributions from short rainfall/long rainfall events (Figure 5a).

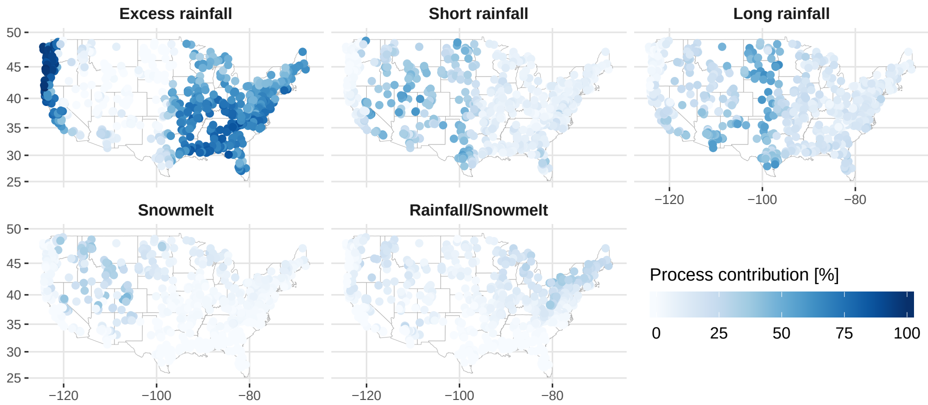

Figure 4. Contribution in percent for each flood generating process across the CAMELS catchment dataset. Flood events are defined as peak-over-threshold with an average of 3 events per year.

3.2 Distribution comparison

The distribution of each catchment attribute for each process was compared with the distribution of each attribute across all processes. The more different the distribution the more influential an attribute is for that specific process. This difference is measured by taking the difference between the empirical distribution functions for the specific process and across all events. The results are detailed in Figure 6a. The plotted empirical distributions functions are shown in the supplement (see Figures S3-S7).

From the distribution comparison (Figure 6a, read by row) we learn that in wet catchments (P>PET) catchment and climate attributes influence the mix of flood processes only marginally. Attribute distribution does not differ widely between the different processes. Excess rainfall is only slightly influenced by precipitation seasonality and mean precipitation. The other processes see a minor influence by further climate attributes. The only noticeable exception is the positive influence of difference in green vegetation fraction on snowmelt. We can therefore conclude that, of the attributes we have considered, only the two attributes, aridity and fraction of snow, that created the climate type influence the distribution of processes. This is confirmed by the difference in distributions between the three climate types demonstrated by Figure 5a.

In drier catchments the difference in attribute distributions are stronger. Excess rainfall floods increase with higher precipitation and potential evapotranspiration and decrease with precipitation seasonality, e.g. with a precipitation maximum in the summer. Increased vegetation (fraction of forest, green vegetation fraction and leaf area index) similarly increase contribution from excess rainfall floods and decrease other occurrences of other processes. Snowmelt floods most decrease with increasing potential evapotranspiration and increase with seasonality indicated both by precipitation seasonality and difference in green vegetation fraction.

The strongest differences in distribution can be seen in snowy catchments. Elevation decreases excess rainfall floods and rainfall/snowmelt floods and increases short rainfall/long rainfall and snowmelt floods. Several climatic attributes have a strong effect as well. Vegetation attributes have the strongest effect in Figure 6a in snow-dominated

Figure 5. a: Contribution in percent for each flood generating process across the CAMELS catchment dataset shown per catchment. Flood events are defined as peak-over-threshold with an average of 3 events per year. The gap in the snow climate type is due to all flood events in that catchment being classified as 'other' (see Stein et al. (2019)). Catchments are sorted by their catchment ID (Addor et al., 2017) which approximates spatial proximity and an ordering from East to West. b: Overview of number of events.

catchments. Similarly to drier climates we can see with increasing vegetation an increase in excess rainfall and rainfall/snowmelt floods and a decrease in short rain/long rain and snowmelt floods. The same methodology applied to larger floods (peaks-over-threshold with one event per year) yields similar results (see Figure S2 in the supplement).

3.3 Attribute influence using accumulated local effects

The summarised accumulated local affects (sALE) are shown in Figure 6b. In contrast to Figure 6a the values here are standardised for each process/climate. Values are not comparable between processes but between attributes for each process (i.e. read the figure by column). Therefore, for each process it can be assessed how the ranking of attributes changes between climates.

Precipitation seasonality and fraction of snow are ranked influential on excess rainfall floods in wet climates. The process contribution from excess rainfall is influenced by fraction of snow, since more rainfall/snowmelt floods decrease the contribution by excess rainfall (see Figure 5a). In dry climates aridity and mean annual precipitation are important as well as precipitation seasonality. Fraction of snow is here less prominent. Climatic attributes in snow-influenced catchments on the other hand do not influence contribution of excess rainfall floods. Instead, elevation is the most relevant attribute for the spatial distribution of excess rainfall floods.

The distribution of short rainfall floods is not well predicted in wet catchments (R^2 0.45, Figure 7). Any conclusion here are therefore less reliable. However, in dry catchments aridity is most dominant for predicting this type of event, whereas in snow-dominated catchments elevation is dominant.

This is in contrast to long rainfall floods. While aridity is influential in wet climatic conditions, in dry climates precipitation seasonality and mean annual rainfall are more influential than aridity. In snow-dominated climates elevation is similarly important.

The spatial distribution of snowmelt induced flood events under wet climatic conditions is influenced by several catchment attributes. Mostly climate attributes such as fraction of snow and mean annual precipitation and potential evapotranspiration, but the difference in green vegetation fraction influences prediction as well. Water fraction, which refers to the top 1.5 m of soil marked as water in the soil database (STATSGO) (Addor et al., 2017), shows as relevant as well, although this will be due to skewness of the data. In a dry climate only fraction of snow is influential and in snow-dominated catchments elevation and fraction of snow dominate.

The contribution of events caused by a combination of rainfall and snowmelt seems to be differently influenced by catchment attributes than sole snowmelt events. In wet and dry climate catchment fraction of snow is the most important attribute. However, in snow-dominated catchments average duration of dry periods seems to be most influential.

3.4 Predictions in space using random forest

Random forest was used as an unsupervised learning model to predict the distribution of each flood generating process and for each climate. The results of a ten-fold cross validation are presented in Figure 7. It demonstrates that prediction accuracy varies with process and climate. For all processes, higher observed contributions are slightly underestimated and low ones slightly overrated. For all processes there are few outliers. Most occur in snow-dominated catchments. From Figure 7b we can see that prediction accuracy using all attributes is lowest for short rainfall events in wet climates (R^2 0.45) and highest for excess rainfall in snowy climate (R^2 0.92). Except for rainfall/snowmelt floods, prediction accuracy is always lowest in dry climates.

4 Discussion

4.1 Influential catchment attributes

A combined interpretation of the two methods takes the direction of influence (positive/negative) from the distribution comparison in Figure 6a. The accumulated local effects (Figure 6b) then confirm if that attribute is influential in comparison to other attributes. In addition to that a comparison between climates is possible using the distribution comparison (Figure 6a) as well. We interpret the combined results in regard to the hypotheses formulated in Section 1.1 and Table 1.

In dry catchments (P<PET) precipitation seasonality has a slight negative influence on excess rainfall floods. Higher precipitation seasonality values indicate a precipitation peak in summer/warm season and lower a peak in winter/cold season (Addor et al., 2017; Woods, 2009). In catchments with a precipitation peak in winter we therefore see more excess rainfall floods. The colder temperatures prohibit the drying out of soils during peak rainfall leading to saturated conditions. Archer (1981) found for the humid catchments in Great Britain, that soil moisture deficits in the summer prevent flooding despite rainfall events with high intensity. Instead, flooding is more common in the winter, when soils are saturated. We can conclude that in catchments where precipitation peak coincides with lower temperatures (and thus lower evaporation), excess rainfall floods would be even more likely. The effect of precipitation seasonality on excess rainfall floods can be seen for wet and particularly for dry catchments. This is confirmed by the ac-

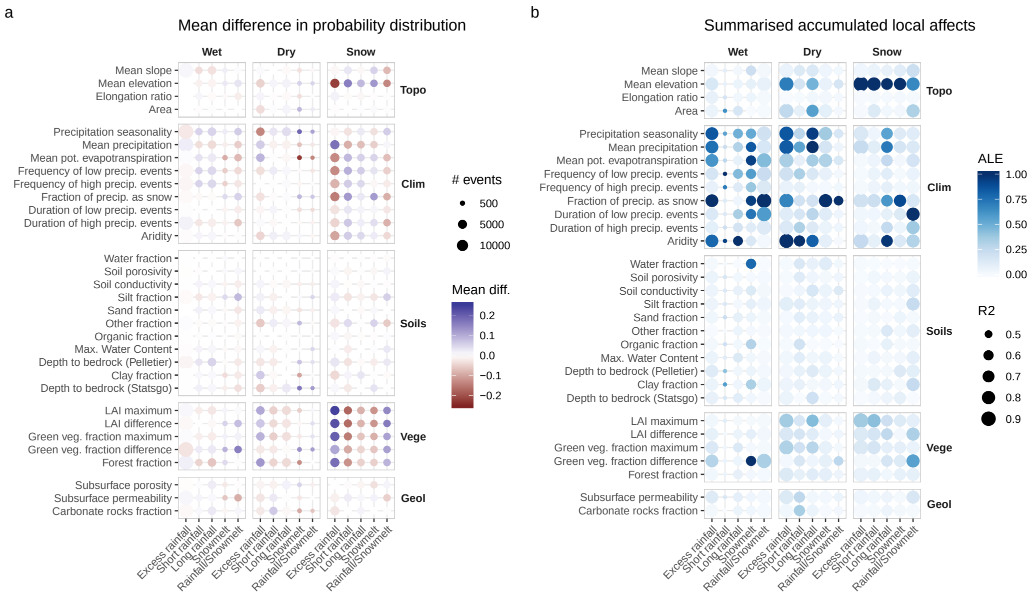

Figure 6. Attribute influence on process distribution. **A**: Mean difference between the empirical distribution function (ECDF) of the attributes for a single process and for all events. The larger the absolute value the more different the two ECDF's. Size of points give an overview of how many events contributed to the distribution. Colour indicates the direction of influence. Blue values point at a decrease of the process for smaller values of the attribute. Red values at an increase for smaller values of the attribute. Horizontal breaks point at different groups of catchment characteristics. **B**: Summarised accumulated local effects. For each climate type and flood process the accumulated local effects for all attributes were calculated. The point colour shows the mean absolute values for each accumulated local effects curve. Higher values indicate increased importance. Values were normalised for each climate/process to enable comparability between climates/processes. Point sizes represent cross-validation R2 prediction accuracy for the random forest model.

cumulated local effects. Although influence in the distribution comparison is minor, ALE confirms that it is strongest in comparison with the other attributes. In snow catchments precipitation seasonality is less influential. In these places, a precipitation peak in winter will instead contribute towards snowpack (Woods, 2009) and not cause floods immediately.

Catchments in snow dominated climate (more than 20 % of precipitation falling as snow) show the strongest influence of catchment attributes on process distribution. Despite both methods pointing to an influence of elevation as an attribute, it is difficult to distinguish elevation from various other catchment attributes (Dingman, 1981; Merz & Blöschl, 2009), some of which are not included in the analysis. All high elevation catchments in the conterminous United States are located in the Rocky Mountains, Sierra Nevadas and in the Appalachian Mountains. Mountainous catchments are steeper, have a higher fraction of bare soils, are smaller, receive more precipitation and due to a temperature gradient have a higher fraction of snow (Wohl, 2013). Elevation thus affects various aspects of flow behaviour (Dingman, 1981). We can therefore not conclusively decide if elevation is influential because it mimics a combination of attributes (fraction of snow, slope, catchment area...) or if it is a proxy for attributes that are either not measured at all (drainage density, infiltration capacity), or measured with a high uncertainty (soil char-

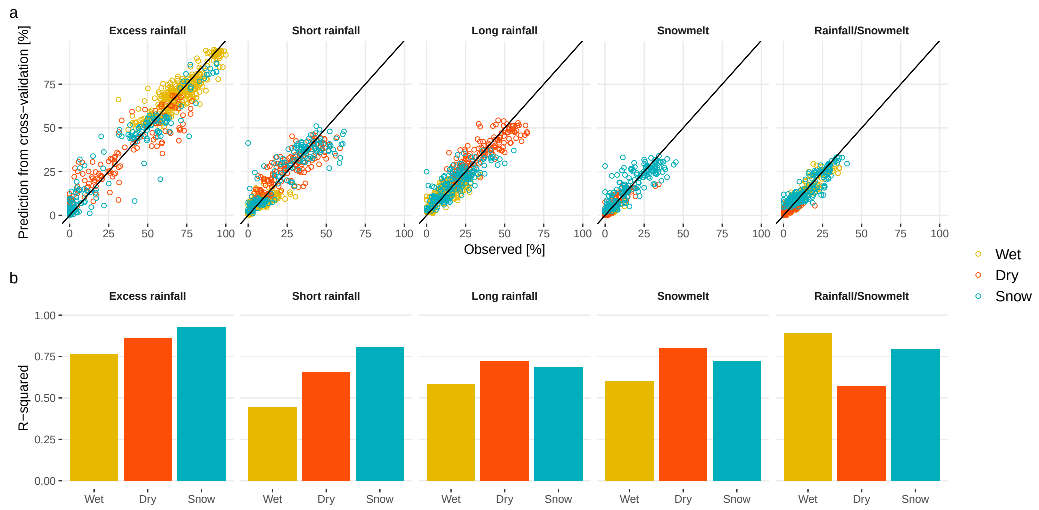

Figure 7. Random forest cross-validation results. For each climate type and flood processes a separate random forest was trained and validated through cross-validation. a: Validation results in comparison to the observed classification. A black line indicates the perfect fit. b: R-Squared between cross-validation and observed values.

acteristics). The interaction between attributes is outside the scope of this paper (with an exception for aridity and fraction of snow which define the different climate types). Therefore, we can only take elevation as a proxy for mountainous catchments, indicating that in mountainous catchments flood generating processes are more likely to be short rainfall floods and snowmelt than excess rainfall.

If and how forest and vegetation in general affect flood characteristics is widely debated (Bradshaw et al., 2007; Bruijnzeel, 2004; Calder & Aylward, 2006) and varies for different processes (Table 1). However, several studies showed that runoff processes can be influenced by land use especially in arid/semi-arid and snow-influenced areas (Lull & Reinhart, 1972; Osterkamp & Friedman, 2000; Pariente, 2002; Storck et al., 2002; Shafer et al., 2007; Zhang et al., 2011). The results from the distribution comparison agree with findings in the literature. The comparison shows a stronger influence of vegetation on excess rainfall floods (positive) and short/long rain floods (negative) with increasing vegetation compared to wet catchments. Shafer et al. (2007) notes for desert areas that vegetation increases infiltration capacity of the soil. Additionally, in arid to semi-arid areas in Israel shrubs will locally increase soil water retention (Pariente, 2002), this will reduce quick runoff leading to less short rain floods. Zhang et al. (2011) describe for the sub-humid east Qinghai–Tibet Plateau that forest vegetation in comparison to shrubs increase water retention of the soil. (Merz & Blöschl, 2003) describe for Austria, that an increased water retention requires larger rainfall amounts or previous saturation to cause flood sized runoff events. This explains why vegetation that increases water retention increases excess rainfall floods (and decreases short rainfall/long rainfall floods).

However, the distribution comparison approach is sensitive to correlated attributes. Figure S1 in the supplemental material reveals that vegetation attributes are correlated to climate attributes. The correlation is strongest for snow dominated catchments. The effect we are seeing could therefore just be due to correlation and only climate not vegetation attributes influence flood process distribution. Yet, the accumulated local effects approach which is unbiased to correlated features, sees a minor influence of vegetation as well. So does vegetation play a role or not? The answer is, while vegetation attributes do have some influence on flood processes, the influence is small if compared to climate attributes (which is the result the accumulated local effects show). This has been noted for other flow behaviour as well. Jehn et al. (2020) cluster the CAMELS catchments by hydrological signature. They notice that clustering is most strongly shaped by climate but that vegetation and soil information play a role as well. Similar conclusions have been reached by Berghuijs et al. (2014) for similarity in a seasonal water balance. Based on their experience it stands to reason that in study areas with very similar climate, vegetation will determine mix of flood generating processes.

In contrast to vegetation we did not find the topographic attributes area, slope and shape to be influential. This contrast with previous studies (Table 1). An explanation for that is that topographic attributes like slope and area influence flood magnitude (Jehn et al., 2020), but not necessarily flood processes. Additionally, the size of CAMELS catchments might not be large enough for area to have an effect particularly on snowmelt floods, which are more prevalent in very large catchments (Ward, 1978).

4.2 Predictions in space using random forest

In addition to evaluating attribute influence, we were able to show that a random forest model is able to predict the spatial distribution of each flood generating process. The accuracy of the prediction varies between climates, especially in a wet climate, several processes are not as well predicted. A possible explanation might be that with excess rainfall being the most common process in these regions, any other processes can be related less to catchment or climate attributes and more to extreme weather events. Stein et al. (2019) highlighted that in the southeastern United States, several catchments have a different flood generating process for the most extreme flood event in the time series. These single event contributions from different processes are difficult for a random forest model to predict based on stationary input attributes. Therefore, while the overarching prediction accuracy might be high, the possible uncertainty of extreme flood generating processes should be kept in mind.

4.3 Limitations

We recognise that environmental data is prone to uncertainties. Especially soil data relies on uncertain interpolation of point measurements over space and depth Addor et al. (2017); Merz and Blöschl (2009); D. A. Miller et al. (1998). This uncertainty might be a possible explanation for having found little influence of soil attributes on flood processes, despite the influence of soil on storage capacity (Section 1.1, Table 1). Similar uncertainties can be found in large scale geology data sets, especially since the CAMELS dataset uses information from global geology datasets (Addor et al., 2017). The evaluated attributes were all taken from the CAMELS dataset (Addor et al., 2017) as a consolidated source. Further studies might want to take additional and non-stationary catchment attributes into account. Possible suggestions for additional stationary attributes are drainage density, wetland area, slope aspect and urbanised areas. Possible suggestions for non-stationary attributes are: forest cover, leaf area index, green vegetation fraction, annual precipitation and annual fraction of snow. Furthermore, the dataset includes mostly small headwater catchments. It is possible that the conclusions might change if larger catchments are taken into account (FAO, 2002).

5 Conclusion

We employed a data-based approach (comparing empirical distribution functions) and an interpretable machine learning approach (random forest model combined with accumulated local effects) to evaluate which catchment characteristics influence flood generating processes. This is the first application of accumulated local effects in a hydrological study. We were able to demonstrate that the two approaches complemented each other. The combined interpretation of both results allowed us to detect limitations and advantages of each method. This resulted in a more complete picture. With an increasing use of machine learning approaches in hydrology we recommend that the hydrologic community make use of interpretable machine learning approaches to improve the transferability of results.

In regard to flood generating processes we found that climatic attributes, such as fraction of snow, aridity, precipitation seasonality and mean precipitation have the strongest influence within the catchment and within space. In comparison, vegetation plays a minor role. This confirmed previous findings that flow behaviour across climates is most strongly influenced by climate attributes. In snow influenced catchments, elevation as a proxy for one or more attributes is influential in predicting flood processes across space. Neither of the methods we used found soil and geologic attributes to be influential. This might be due to limitations in data quality or attribute selection for both groups.

With the available catchment attribute information the mix of flood generating processes can be predicted with relatively high accuracy. A prediction of processes for ungauged catchments is therefore possible, although climate dependent uncertainties should be taken into account.

Further studies are necessary to evaluate the implication of these findings in regard to changes in climate and land use. Changes in flood magnitude and frequency have been observed, yet direction and magnitude of the trends are not homogeneous (Blöschl et al., 2019; Gudmundsson et al., 2019; Mallakpour & Villarini, 2015; Sharma et al., 2018; Wasko & Nathan, 2019). The results of this study can give an indication why: not all flood processes are influenced by the same climate attributes and the influence varies between different climates.

Acknowledgments

The Catchement Attributes and Meteorlogy for Large-sample Studies (CAMELS) dataset (Addor et al., 2017; Newman et al., 2015) is freely available at `https://ral.ucar.edu/solutions/products/camels`. The National Soil Geographic Database (NATSGO) used to calculate available soil water storage (Soil Survey Staff, 2019) is freely accessible at `https://nrcs.app.box.com/v/soils`. Processed NATSGO data are provided in a table in the supporting information. The code needed to replicate this work can be found here: `http://doi.org/10.5281/zenodo.3941515`.

This work was funded as part of the Water Informatics Science and Engineering Centre for Doctoral Training (WISE CDT) under a grant from the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/L016214/1. We thank Maria Xenochristou for her helpful advice on random forest and interpretable machine learning.

References

Acreman, M., & Holden, J. (2013, 10). How Wetlands Affect Floods. *Wetlands*, *33*(5), 773–786. Retrieved from `http://link.springer.com/10.1007/s13157-013-0473-2` doi: 10.1007/s13157-013-0473-2

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018, 11). A Ranking of Hydrological Signatures Based on Their Predictability in Space. *Water Resources Research*, *54*(11), 8792–8812. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2018WR022606` doi: 10.1029/2018WR022606

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hy-*

drology and Earth System Sciences Discussions. doi: 10.5194/hess-2017-169

Anchang, J. Y., Prihodko, L., Ji, W., Kumar, S. S., Ross, C. W., Yu, Q., ... Hanan, N. P. (2020, 1). Toward Operational Mapping of Woody Canopy Cover in Tropical Savannas Using Google Earth Engine. *Frontiers in Environmental Science*, *8*, 4. doi: 10.3389/fenvs.2020.00004

Apley, D. W., & Zhu, J. (2016, 12). Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. Retrieved from `http://arxiv.org/abs/1612.08468`

Archer, D. R. (1981, 3). Seasonality of flooding and the assessment of seasonal flood risk. *Proceedings of the Institution of Civil Engineers*, *71*(4), 1023–1035. Retrieved from `http://www.icevirtuallibrary.com/doi/abs/10.1680/iicep.1981.1753`

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., & Kirchner, J. W. (2019, 5). The relative importance of different flood-generating mechanisms across Europe. *Water Resources Research*, 2019WR024841. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2019WR024841` doi: 10.1029/2019WR024841

Berghuijs, W. R., Sivapalan, M., Woods, R. A., & Savenije, H. H. G. (2014, 3). Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. *Water Resources Research*, *50*(7), 5638–5661. Retrieved from `http://onlinelibrary.wiley.com/doi/10.1002/2014WR015692/full`

Berghuijs, W. R., Woods, R. A., Hutton, C. J., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. *Geophysical Research Letters*. doi: 10.1002/2016GL068070

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., ... Chirico, G. B. (2017). Changing climate shifts of European floods. *Science*, *357*(6351), 588–590. Retrieved from `http://science.sciencemag.org/content/357/6351/588.abstract` doi: 10.1126/science.aan2506

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., ... Živković, N. (2019, 9). Changing climate both increases and decreases European river floods. *Nature*, *573*(7772), 108–111. Retrieved from `http://www.nature.com/articles/s41586-019-1495-6` doi: 10.1038/s41586-019-1495-6

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. H. G. (2013). *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales.* doi: 10.1017/CBO9781139235761

Bradshaw, C. J., Sodhi, N. S., Peh, K. S., & Brook, B. W. (2007, 11). Global evidence that deforestation amplifies flood risk and severity in the developing world. *Global Change Biology*, *13*(11), 2379–2395. Retrieved from `http://doi.wiley.com/10.1111/j.1365-2486.2007.01446.x` doi: 10.1111/j.1365-2486.2007.01446.x

Breiman, L. (2001, 10). Random forests. *Machine Learning*, *45*(1), 5–32. doi: 10.1023/A:1010933404324

Brown, S. C., Wells, K., Roy-Dufresne, E., Campbell, S., Cooke, B., Cox, T., & Fordham, D. A. (2020, 1). Models of spatiotemporal variation in rabbit abundance reveal management hotspots for an invasive species. *Ecological Applications*, eap.2083. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/eap.2083` doi: 10.1002/eap.2083

Bruijnzeel, L. A. (2004, 9). Hydrological functions of tropical forests: Not seeing the soil for the trees? In *Agriculture, ecosystems and environment* (Vol. 104, pp. 185–228). Elsevier. doi: 10.1016/j.agee.2004.01.015

Bullock, A., & Acreman, M. (2003). The role of wetlands in the hydrological cycle. *Hydrology and Earth System Sciences*, *7*(3), 358–389. Retrieved from `http://www.hydrol-earth-syst-sci.net/7/358/2003/` doi: 10.5194/hess-7

-358-2003
Calder, I. R., & Aylward, B. (2006, 3). Forest and Floods. *Water International*, *31*(1), 87–99. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/02508060608691918 doi: 10.1080/02508060608691918
Chang, H., Johnson, G., Hinkley, T., & Jung, I.-W. (2014, 4). Spatial analysis of annual runoff ratios and their variability across the contiguous U.S. *Journal of Hydrology*, *511*, 387–402. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022169414000882#b0255 doi: 10.1016/J.JHYDROL.2014.01.066
Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C., Jackson, B. M., ... Ceola, S. (2016, 3). Improving the theoretical underpinnings of process-based hydrologic models. *Water Resources Research*, *52*(3), 2350–2365. Retrieved from http://doi.wiley.com/10.1002/2015WR017910 doi: 10.1002/2015WR017910
Costa, J. E. (1987, 9). Hydraulics and basin morphometry of the largest flash floods in the conterminous United States. *Journal of Hydrology*, *93*(3-4), 313–338. Retrieved from https://www.sciencedirect.com/science/article/pii/0022169487901028 doi: 10.1016/0022-1694(87)90102-8
Cunnane, C. (1979, 4). A note on the Poisson assumption in partial duration series models. *Water Resources Research*, *15*(2), 489–494. Retrieved from http://doi.wiley.com/10.1029/WR015i002p00489 doi: 10.1029/WR015i002p00489
Degenhardt, F., Seifert, S., & Szymczak, S. (2019, 3). Evaluation of variable selection methods for random forests and omics data sets. *Briefings in Bioinformatics*, *20*(2), 492–503. Retrieved from https://academic.oup.com/bib/article/20/2/492/4554516 doi: 10.1093/bib/bbx124
Diezig, R., & Weingartner, R. (2007). Hochwasserprozesstypen in der Schweiz. *Wasser und Abfall*, *4*(1), 18–26. Retrieved from https://boris.unibe.ch/id/eprint/25512
Dingman, S. L. (1981, 12). Elevation: a major influence on the hydrology of New Hampshire and Vermont, USA / L'altitude exerce une influence importante sur l'hydrologie du New Hampshire et du Vermont, Etats-Unis. *Hydrological Sciences Bulletin*, *26*(4), 399–413. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/02626668109490904 doi: 10.1080/02626668109490904
Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Lautenbach, S. (2013, 1). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. Retrieved from http://doi.wiley.com/10.1111/j.1600-0587.2012.07348.x doi: 10.1111/j.1600-0587.2012.07348.x
Falkenmark, M., Chapman, T., & others. (1989). *Comparative hydrology: An ecological approach to land and water resources.* The Unesco Press. Retrieved from http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src=google&base=REPIDISCA&lang=p&nextAction=lnk&exprSearch=97638&indexSearch=ID
FAO. (2002). *Land-water linkages in rural watersheds–Electronicworkshop.Conclusions and Recommendations.* Food and Agriculture Organization of the United Nations.
Fischer, S., Schumann, A., & Schulte, M. (2016, 7). Characterisation of seasonal flood types according to timescales in mixed probability distributions. *Journal of Hydrology*, *539*, 38–56. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022169416302712
French, R. H., & Miller, J. J. (2011). *Flood hazard identification and mitigation in semi- and arid environments.* World Scientific Publishing Co. doi: 10.1142/8175
Freudiger, D., Kohn, I., Stahl, K., & Weiler, M. (2014, 7). Large-scale analysis

of changing frequencies of rain-on-snow events with flood-generation potential. *Hydrology and Earth System Sciences*, *18*(7), 2695–2709. Retrieved from https://www.hydrol-earth-syst-sci.net/18/2695/2014/ doi: 10.5194/hess-18-2695-2014

Friedman, J. H. (2001, 10). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. doi: 10.2307/2699986

Gaál, L., Szolgay, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A., & Blöschl, G. (2012, 6). Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology. *Water Resources Research*, *48*(4). Retrieved from http://doi.wiley.com/10.1029/2011WR011509 doi: 10.1029/2011WR011509

Gudmundsson, L., Leonard, M., Do, H. X., Westra, S., & Seneviratne, S. I. (2019, 1). Observed Trends in Global Indicators of Mean and Extreme Streamflow. *Geophysical Research Letters*, *46*(2), 756–766. Retrieved from http://doi.wiley.com/10.1029/2018GL079725 doi: 10.1029/2018GL079725

Institute of Hydrology (IoH). (1999). *Flood Estimation Handbook.* IoH, Wallingford.

Jehn, F. U., Bestian, K., Breuer, L., Kraft, P., & Houska, T. (2020, 3). Using hydrological and climatic catchment clusters to explore drivers of catchment behavior. *Hydrology and Earth System Sciences*, *24*(3), 1081–1100. Retrieved from https://www.hydrol-earth-syst-sci.net/24/1081/2020/ doi: 10.5194/hess-24-1081-2020

Jennings, K. S., Winchell, T. S., Livneh, B., & Molotch, N. P. (2018, 12). Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere. *Nature Communications*, *9*(1), 1148. Retrieved from http://www.nature.com/articles/s41467-018-03629-7 doi: 10.1038/s41467-018-03629-7

Keller, L., Rössler, O., Martius, O., & Weingartner, R. (2018, 1). Delineation of flood generating processes and their hydrological response. *Hydrological Processes*, *32*(2), 228–240. Retrieved from http://doi.wiley.com/10.1002/hyp.11407 doi: 10.1002/hyp.11407

Kirchner, J. W. (2006, 3). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, *42*(3). Retrieved from http://doi.wiley.com/10.1029/2005WR004362 doi: 10.1029/2005WR004362

Knoben, W. J. M., Woods, R. A., & Freer, J. E. (2018, 7). A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data. *Water Resources Research*, *54*(7), 5088–5109. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1029/2018WR022913 doi: 10.1029/2018WR022913

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22. Retrieved from https://cran.r-project.org/doc/Rnews/

Lull, H., & Reinhart, K. (1972). *Forests and floods in the eastern United States (Vol. 226).* US Northeastern Forest Experiment Station.

Mallakpour, I., & Villarini, G. (2015, 3). The changing nature of flooding across the central United States. *Nature Climate Change*, *5*(3), 250–254. Retrieved from http://www.nature.com/articles/nclimate2516 doi: 10.1038/nclimate2516

Marks, D., Kimball, J., Tingey, D., & Link, T. (1998, 8). The sensitivity of snowmelt processes to climate conditions and forest cover during rain-on-snow: a case study of the 1996 Pacific Northwest flood. *Hydrological Processes*, *12*(10-11), 1569–1587. doi: 10.1002/(SICI)1099-1085(199808/09)12:10/11⟨1569::AID-HYP682⟩3.0.CO;2-L

Marks, D., Link, T., Winstral, A., & Garen, D. (2001). Simulating snowmelt processes during rain-on-snow over a semi-arid mountain basin. *Annals of Glaciol-*

ogy, *32*, 195–202. doi: 10.3189/172756401781819751

McCartney, M. P., Neal, C., & Neal, M. (1998). Use of deuterium to understand runoff generation in a headwater catchment containing a dambo. *Hydrology and Earth System Sciences*, *2*(1), 65–76. Retrieved from http://www.hydrol-earth-syst-sci.net/2/65/1998/ doi: 10.5194/hess-2-65-1998

Merz, R., & Blöschl, G. (2003, 5). A process typology of regional floods. *Water Resources Research*, *39*(12). Retrieved from http://doi.wiley.com/10.1029/2002WR001952 doi: 10.1029/2002WR001952

Merz, R., & Blöschl, G. (2005, 5). Flood frequency regionalisation—spatial proximity vs. catchment attributes. *Journal of Hydrology*, *302*(1), 283–306. Retrieved from http://www.sciencedirect.com/science/article/pii/S002216940400383X

Merz, R., & Blöschl, G. (2009, 5). A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. *Water Resources Research*, *45*(1). Retrieved from http://doi.wiley.com/10.1029/2008WR007163 doi: 10.1029/2008WR007163

Merz, R., Blöschl, G., & Parajka, J. (2006, 12). Spatio-temporal variability of event runoff coefficients. *Journal of Hydrology*, *331*(3-4), 591–604. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022169406003192 doi: 10.1016/J.JHYDROL.2006.06.008

Miller, D. (1964). *Interception processes during snow storms* (Tech. Rep.). Res. Paper PSW-RP-18. Berkeley, CA: Pacific Southwest Forest & Range Experiment.

Miller, D. A., White, R. A., Miller, D. A., & White, R. A. (1998, 1). A Conterminous United States Multilayer Soil Characteristics Dataset for Regional Climate and Hydrology Modeling. *Earth Interactions*, *2*(2), 1–26. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/1087-3562%281998%29002%3C0001%3AACUSMS%3E2.3.CO%3B2 doi: 10.1175/1087-3562(1998)002⟨0001:ACUSMS⟩2.3.CO;2

Molnar, C. (2019). *Interpretable machine learning.* Retrieved from https://books.google.com/books?hl=en&lr=&id=jBm3DwAAQBAJ&oi=fnd&pg=PP1&dq=interpretable+machine+learning&ots=EfzQ_rKIT-&sig=HqoVfuKmuG8EwFf_vqEL-ZXZB5I

Molnar, C., Bischl, B., & Casalicchio, G. (2018). iml: An R package for Interpretable Machine Learning. *JOSS*, *3*(26), 786. Retrieved from http://joss.theoj.org/papers/10.21105/joss.00786 doi: 10.21105/joss.00786

Musselman, K. N., Lehner, F., Ikeda, K., Clark, M. P., Prein, A. F., Liu, C., ... Rasmussen, R. (2018, 9). *Projected increases and shifts in rain-on-snow flood risk over western North America* (Vol. 8) (No. 9). Nature Publishing Group. doi: 10.1038/s41558-018-0236-4

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., ... Duan, Q. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrol. Earth Syst. Sci*, *19*, 209–223. Retrieved from www.hydrol-earth-syst-sci.net/19/209/2015/ doi: 10.5194/hess-19-209-2015

O'Connor, J. E., & Costa, J. E. (2004, 1). Spatial distribution of the largest rainfall-runoff floods from basins between 2.6 and 26,000 km2 in the United States and Puerto Rico. *Water Resources Research*, *40*(1). Retrieved from http://doi.wiley.com/10.1029/2003WR002247 doi: 10.1029/2003WR002247

Osterkamp, W. R., & Friedman, J. M. (2000, 11). The disparity between extreme rainfall events and rare floods - with emphasis on the semi-arid American West. *Hydrological Processes*, *14*(16-17), 2817–2829. Retrieved from http://doi.wiley.com/10.1002/1099-1085%28200011/12%2914%3A16/17%3C2817%3A%3AAID-HYP121%3E3.0.CO%3B2-B doi:

10.1002/1099-1085(200011/12)14:16/17⟨2817::AID-HYP121⟩3.0.CO;2-B

Pariente, S. (2002). Spatial patterns of soil moisture as affected by shrubs, in different climatic conditions. *Environmental Monitoring and Assessment*, *73*(3), 237–251. doi: 10.1023/A:1013119405441

Pianosi, F., & Wagener, T. (2015, 5). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, *67*, 1–11. Retrieved from https://www.sciencedirect.com/science/article/pii/S1364815215000237 doi: 10.1016/J.ENVSOFT.2015.01.004

Pitlick, J. (1994). Relation between peak flows, precipitation, and physiography for five mountainous regions in the western USA. *Journal of Hydrology*. doi: 10.1016/0022-1694(94)90055-8

Rogger, M., Pirkl, H., Viglione, A., Komma, J., Kohl, B., Kirnbauer, R., ... Blöschl, G. (2012, 5). Step changes in the flood frequency curve: Process controls. *Water Resources Research*, *48*(5). Retrieved from http://doi.wiley.com/10.1029/2011WR011187 doi: 10.1029/2011WR011187

Rosbjerg, D., Blöschl, G., Burn, D. H., Castellarin, A., Croke, B., Di Baldassarre, G., ... Viglione, A. (2013). Prediction of floods in ungauged basins. In G. Blöschl (Ed.), *Runoff prediction in ungauged basins: synthesis across processes, places and scales.* Cambridge University Press.

Ryan, J. A., & Ulrich, J. M. (2019). *quantmod: Quantitative Financial Modelling Framework.* Retrieved from https://cran.r-project.org/package=quantmod

Schumm, S. (1956). Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. *Geological society of America bulletin*. Retrieved from http://gsabulletin.gsapubs.org/content/67/5/597.short

Shafer, D. S., Young, M. H., Zitzer, S. F., Caldwell, T. G., & McDonald, E. V. (2007, 6). Impacts of interrelated biotic and abiotic processes during the past 125 000 years of landscape evolution in the Northern Mojave Desert, Nevada, USA. *Journal of Arid Environments*, *69*(4), 633–657. doi: 10.1016/j.jaridenv.2006.11.011

Sharma, A., Wasko, C., & Lettenmaier, D. P. (2018, 11). If precipitation extremes are increasing, why aren't floods? *Water Resources Research*. Retrieved from http://doi.wiley.com/10.1029/2018WR023749 doi: 10.1029/2018WR023749

Sikorska, A. E., Viviroli, D., & Seibert, J. (2015, 10). Flood-type classification in mountainous catchments using crisp and fuzzy decision trees. *Water Resources Research*, *51*(10), 7959–7976. Retrieved from http://doi.wiley.com/10.1002/2015WR017326 doi: 10.1002/2015WR017326

Singh, P., Spitzbart, G., Hübl, H., & Weinmeister, H. (1998). The role of snowpack in producing floods under heavy rainfall. In K. Kovar (Ed.), *Hydrology, water resources and ecology in headwaters* (pp. 89–95). IAHS Publ. 248.

Smith, A., Sampson, C., & Bates, P. D. (2015, 3). Regional flood frequency analysis at the global scale. *Water Resources Research*, *51*(1), 539–553. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/2014WR015814/full

Soil Survey Staff. (2019). *Gridded National Soil Survey Geographic (gNATSGO) Database for the Conterminous United States.* United States Department of Agriculture, Natural Resources Conservation Service.

Stein, L., Pianosi, F., & Woods, R. A. (2019, 12). Event-based classification for global study of river flood generating processes. *Hydrological Processes*, hyp.13678. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.13678 doi: 10.1002/hyp.13678

Storck, P., Lettenmaier, D. P., & Bolton, S. M. (2002, 11). Measurement of snow interception and canopy effects on snow accumulation and melt in a mountainous maritime climate, Oregon, United States. *Water Resources Research*, *38*(11), 5–1. doi: 10.1029/2002wr001281

Sui, J., & Koehler, G. (2001, 10). Rain-on-snow induced flood events in Southern Germany. *Journal of Hydrology*, *252*(1-4), 205–220. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0022169401004607` doi: 10.1016/S0022-1694(01)00460-7

Tarasova, L., Basso, S., Wendi, D., Viglione, A., Kumar, R., & Merz, R. (2020, 5). A Process-Based Framework to Characterize and Classify Runoff Events: The Event Typology of Germany. *Water Resources Research*, *56*(5). doi: 10.1029/2019wr026951

Tarasova, L., Merz, R., Kiss, A., Basso, S., Blöschl, G., Merz, B., ... Wietzke, L. (2019, 5). Causative classification of river flood events. *Wiley Interdisciplinary Reviews: Water*, e1353. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1353` doi: 10.1002/wat2.1353

Tetzlaff, D., Seibert, J., McGuire, K. J., Laudon, H., Burns, D. A., Dunn, S. M., & Soulsby, C. (2009, 3). How does landscape structure influence catchment transit time across different geomorphic provinces? *Hydrological Processes*, *23*(6), 945–953. Retrieved from `http://doi.wiley.com/10.1002/hyp.7240` doi: 10.1002/hyp.7240

Toloşi, L., & Lengauer, T. (2011, 7). Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics*, *27*(14), 1986–1994. Retrieved from `https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr300` doi: 10.1093/bioinformatics/btr300

Tooth, S. (2000, 8). Process, form and change in dryland rivers: a review of recent research. *Earth-Science Reviews*, *51*(1-4), 67–107. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0012825200000143#BIB248` doi: 10.1016/S0012-8252(00)00014-3

van Dijk, A. I., van Noordwijk, M., Calder, I. R., Bruijnzeel, S. L., Schellekens, J. A., & Chappell, N. A. (2009, 1). *Forest-flood relation still tenuous - Comment on 'Global evidence that deforestation amplifies flood risk and severity in the developing world'by C. J. A. Bradshaw, N.S. Sodi, K. S.-H. Peh and B.W. Brook* (Vol. 15) (No. 1). John Wiley & Sons, Ltd (10.1111). Retrieved from `http://doi.wiley.com/10.1111/j.1365-2486.2008.01708.x` doi: 10.1111/j.1365-2486.2008.01708.x

Viglione, A., & Blöschl, G. (2009, 2). On the role of storm duration in the mapping of rainfall to flood return periods. *Hydrology and Earth System Sciences*, *13*(2), 205–216. Retrieved from `http://www.hydrol-earth-syst-sci.net/13/205/2009/` doi: 10.5194/hess-13-205-2009

Ward, R. C. (1978). *Floods- a geographical perspective.*

Wasko, C., & Nathan, R. (2019, 8). Influence of changes in rainfall and soil moisture on trends in flooding. *Journal of Hydrology*, *575*, 432–441. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0022169419304998` doi: 10.1016/J.JHYDROL.2019.05.054

Weingartner, R., Barbena, M., & Spreafico, M. (2003, 11). Floods in mountain areas - An overview based on examples from Switzerland. *Journal of Hydrology*, *282*(1-4), 10–24. doi: 10.1016/S0022-1694(03)00249-X

Wohl, E. (2013). *Mountain Rivers Revisited* (Vol. 19). Washington, D. C.: American Geophysical Union. Retrieved from `http://doi.wiley.com/10.1029/WM019` doi: 10.1029/WM019

Wood, E. F., Sivapalan, M., & Beven, K. J. (1990, 2). Similarity and scale in catchment storm response. *Reviews of Geophysics*, *28*(1), 1. Retrieved from `http://doi.wiley.com/10.1029/RG028i001p00001` doi: 10.1029/RG028i001p00001

Woods, R. A. (2009, 10). Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks. *Advances in Water Resources*, *32*(10), 1465–1481. doi: 10.1016/j.advwatres.2009.06.011

Zhang, W., An, S., Xu, Z., Cui, J., & Xu, Q. (2011, 11). The impact of vegetation and soil on runoff regulation in headwater streams on the east Qinghai-Tibet Plateau, China. *Catena*, *87*(2), 182–189. doi: 10.1016/j.catena.2011.05.020