Water Resources Research

RESEARCH ARTICLE
10.1029/2021WR031673

Controls on Flood Trends Across the United States

Matthias Kemter[1,2,3], Norbert Marwan[3], Gabriele Villarini[4], and Bruno Merz[1,2]

[1]Institute for Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany, [2]Section of Hydrology, Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany, [3]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany, [4]IIHR—Hydroscience & Engineering, University of Iowa, Iowa City, IA, USA

Abstract Trends in flood magnitudes vary across the conterminous USA (CONUS). There have been attempts to identify what controls these regionally varying trends, but these attempts were limited to certain—for example, climatic—variables or to smaller regions, using different methods and datasets each time. Here we attribute the trends in annual maximum streamflow for 4,390 gauging stations across the CONUS in the period 1960–2010, while using a novel combination of methods and an unprecedented variety of potential controlling variables to allow large-scale comparisons and minimize biases. Using process-based flood classification and complex networks, we find 10 distinct clusters of catchments with similar flood behavior. We compile a set of 31 hydro-climatological and land use variables as predictors for 10 separate Random Forest models, allowing us to find the main controls the flood magnitude trends for each cluster. By using Accumulated Local Effect plots, we can understand how these controls influence the trends in the flood magnitude. We show that hydro-climatologic changes and land use are of similar importance for flood magnitude trends across the CONUS. Static land use variables are more important than their trends, suggesting that land use is able to attenuate (forested areas) or amplify (urbanized areas) the effects of climatic changes on flood magnitudes. For some variables, we find opposing effects in different regions, showing that flood trend controls are highly dependent on regional characteristics and that our novel approach is necessary to attribute flood magnitude trends reliably at the continental scale while maintaining sensitivity to regional controls.

Key Points:

- A wide variety of controls are necessary to explain flood magnitude trends across the United States between 1960 and 2010
- Climatic changes and land cover conditions are of similar importance for flood magnitude trends at the regional scale
- Controls on flood trends can have highly nonlinear effects and can have opposing effects in different hydro-climatological subregions

Supporting Information:

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

Correspondence to:

M. Kemter,
kemter@uni-potsdam.de

Citation:

Kemter, M., Marwan, N., Villarini, G., & Merz, B. (2023). Controls on flood trends across the United States. *Water Resources Research*, *59*, e2021WR031673. https://doi.org/10.1029/2021WR031673

Received 23 NOV 2021
Accepted 10 JAN 2023

1. Introduction

Floods are among the most harmful natural hazards in the conterminous United States (CONUS) and worldwide (Munich Re, 2021). The damages they cause could grow, as increasing global temperatures have led to higher water content in the atmosphere and an intensification of precipitation extremes throughout the CONUS (Yin et al., 2018). While we would expect this intensification to lead to higher flood peaks (Sharma et al., 2018), in recent decades annual maximum streamflow (AMS)—which in this study will represent flood magnitudes—has both increased and decreased in different parts of the CONUS (Archfield et al., 2016; Hirsch & Ryberg, 2012; Sauer et al., 2021; Slater & Villarini, 2016; Villarini & Slater, 2017). This is partly caused by the superposition of different climate change effects. It has been suggested that increasing extreme rainfall can be offset by decreasing soil moisture due to higher evapotranspiration (Ivancic & Shaw, 2015), which is especially relevant for large catchments and floods of relatively low magnitude (Wasko & Nathan, 2019; Wasko & Sharma, 2017). Bertola et al. (2021) show that the effects of climatic drivers on flood trends vary with flood magnitude, for instance showing that in southern Europe soil moisture is of higher relevance for trends of small floods compared to floods with higher return periods. As our study deals with AMS, which contains mostly low magnitude floods, the potentially different trend controls for small floods are of particular importance here.

Further explanations are reduced meltwater availability from decreasing snowpacks (Yan et al., 2019) or reduced catchment wetness due to smaller rainstorm extents (Sharma et al., 2018). Additionally, the effects of climate change on flood generation can be modulated by direct human influences in the catchment through land use and land use changes (Blum et al., 2020; Hodgkins et al., 2019). Understanding this variety of controls on changes in flood magnitudes and their interactions represents critical information for the assessment of future changes to prevent growing damage and fatalities in regions of increasing AMS magnitudes and avoid unwarranted costs for flood protection in areas with a decreasing flood hazard.

Recent studies have attributed AMS magnitude trends in the CONUS to changes in selected weather variables (Do et al., 2020; Hirsch & Ryberg, 2012; Hodgkins et al., 2017; Slater & Villarini, 2016; Villarini &

Slater, 2018), mainly precipitation, soil moisture, snowmelt, and temperature. Others have focused on certain hydro-meteorologically homogeneous parts of the CONUS considering only the variables that seem most relevant in that region (Armstrong et al., 2014; Mallakpour & Villarini, 2015). Additionally, there has been work on the influence of river regulation and urbanization on flood magnitude trends (Blum et al., 2020; Hodgkins et al., 2019; Vogel et al., 2011). This focus on either certain controls on flood trends or a smaller subregion was often seen as necessary, given the diversity in climatology, flood generation processes, and land cover across the CONUS. However, the different catchment selection criteria and methodological approaches limit the comparability among regions and studies, making it impossible to compare the relative contributions from different variables. Nonetheless, it is necessary to consider regional differences, as the same outside forcings could have opposing effects on magnitude trends in hydro-climatologically different regions.

Here we aim to identify which factors influenced peak AMS magnitude trends in 4,390 catchments for the period 1960–2010 using the same consistent dataset and methods for the entire CONUS. We intentionally include anthropogenically influenced catchments—even those with reservoirs or other forms of streamflow regulation, which allows us to compare the impacts of direct human influences and a changing climate, as we know which of the catchments are regulated. This approach is similar to Ficklin et al. (2018), who compared the streamflow at pristine and human-modified watersheds in the United States and Canada, finding a dominant role of changes in the climate system over human alterations of the basins. Using complex networks, we find clusters of catchments with similar flood behavior. We train separate Random Forest models (Breiman, 2001) for each cluster to select the most important variables to explain the spatial variability in flood magnitude trends, drawing from the same set of 31 hydro-meteorological and land cover variables in all clusters. Accumulated Local Effect (ALE) plots are a recent advance in interpretable machine learning (Apley & Zhu, 2020), which we use to understand how each selected predictor contributes to AMS magnitude trends within the different regions. This combination of dataset and methodology allows us—for the first time—to detect and compare a wide range of controls on flood magnitude trends across CONUS regions and among processes (from hydro-metereology to land cover), while minimizing biases in variable selection. Through this approach, we aim to understand why flood magnitude trends differ regionally and how important climate change and land use are, respectively, for these trends and their differences. New or improved insights into this topic could reveal ways to better project or even mitigate future flood hazard changes.

2. Data

2.1. Streamflow Data

This work is based on annual maximum instantaneous streamflow (AMS) data provided by the United States Geological Survey (USGS). We start with data from 5,002 catchments that are located completely within the CONUS and have at least 30 years of record within the period 1960–2010. Data consist of magnitude and date of the AMS, and catchment characteristics (e.g., catchment area) for each gauging station. We selected the period 1960–2010 because: (a) the number of stations with sufficient data declines drastically for earlier and later years, and (b) the reanalysis dataset used to characterize potential controls on flood trends is only available until 2011. For each gauging station, the corresponding catchment boundaries are available from the USGS Streamgage NHDPlus Version 1 dataset.

2.2. Climate and Land Use Data

We use the Livneh daily reanalysis dataset with 0.0625° spatial resolution (Livneh et al., 2013), and calculate catchment-averaged daily time series for precipitation (mm), snowmelt (mm of water equivalent), total evapotranspiration (mm), and soil moisture (mm). Snowmelt is calculated from the daily difference in snow water equivalent. Soil moisture refers to the vertical soil level 2 of the underlying Variable Infiltration Capacity (VIC) hydrological model, which represents the soil layer that reflects daily to weekly changes in weather conditions. The upper boundary of this layer lies at 10 cm depth, while its thickness varies between 50 and 110 cm for each pixel as it was iteratively estimated during model calibration (Tang et al., 2012). To calculate rainfall from precipitation, we subtract any daily net gain of snow water equivalent (reflecting snowfall) in a cell from that day's precipitation value. From these hydro-climatological variables, we derive two types of data for different purposes: (a) daily time series for the classification of all floods in terms of flood generating processes; and (b)

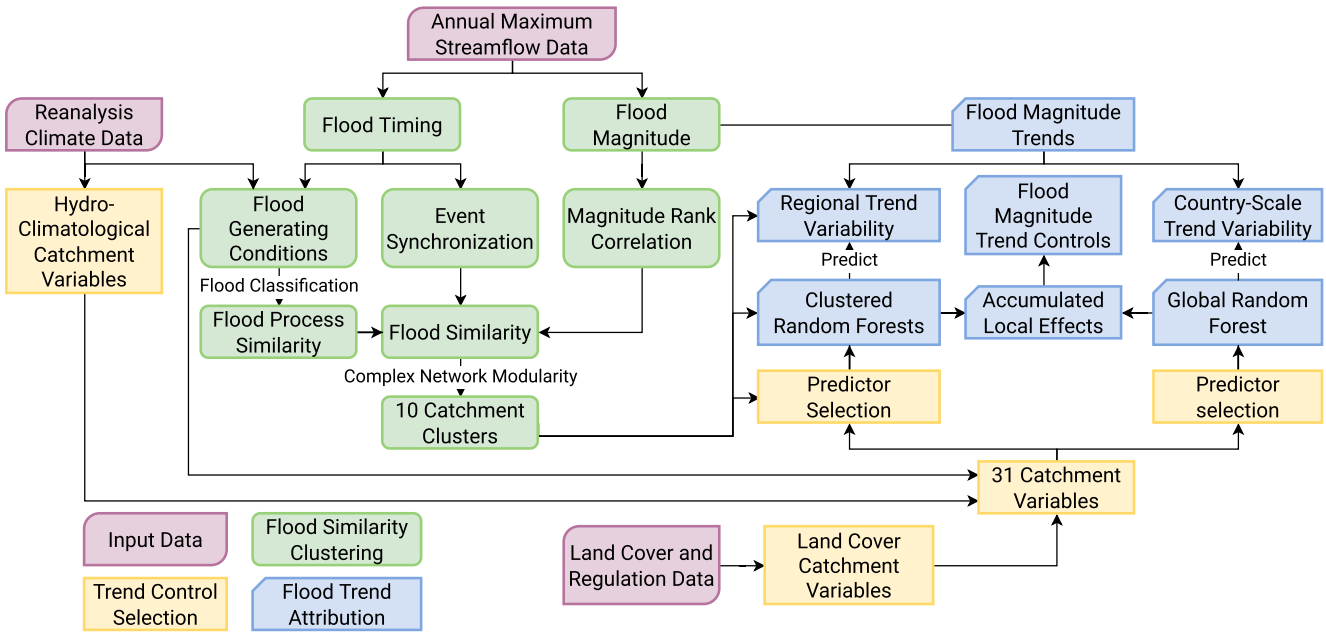

Figure 1. Workflow overview. Different parts of the workflow are separated by color and shape (see legend in the bottom left corner). We use flood and climate data to cluster catchments by flood similarity. For each cluster (regional variability) we train a Random Forest to find the most important controls on flood magnitude trends among 31 catchment variables. Additionally, we train one global Random Forest to find conterminous USA (CONUS)-scale controls on flood magnitude trends (country-scale variability). We investigate the effects of the controls on flood magnitude trends, using Accumulated Local Effects. See Section 2 and Table S1 in Supporting Information S1 for more details on data. See Section 3 for more details on methods.

long-term averages (e.g., mean annual rainfall, mean annual snowfall) and trends of the annual values, serving as potential controls on flood trends.

To study the effects of human activities on flood magnitude trends, we also consider averages and trends of land cover data from the GAGES-II dataset (Falcone, 2017; Falcone & LaMotte, 2016; Homer et al., 2015), which provides catchment averages for multiple land cover variables at different time steps (Table S1 in Supporting Information S1) between 1960 and 2010 and is available for 4,663 stations of the AMS dataset. The remaining 339 stations are not considered for further analysis.

3. Methods

In this section we explain the sequence of methods (Figure 1) we use to create the set of 31 predictors variables (Section 3.1), find clusters of similar flood behavior (Section 3.2), identify the most important predictors (Sections 3.3 and 3.4), understand their effects on flood magnitude trends (Section 3.5), and contextualize them at the CONUS scale (Section 3.6). Throughout these sections we highlight and discuss advantages of the methods we use and how they contribute to understand flood magnitude trends.

3.1. Predictor Variables

To find the controls on the flood magnitude trends we calculate 31 predictor variables. Details on their calculation can be found in Table S1 in Supporting Information S1, while we here explain their potential relevance for flood magnitude trends. We divide the predictors into five categories: flood characteristics, climate, flood type, land use, and topography. The flood characteristics category contains variables describing direct flood water contributions from the three main sources (rain, snowmelt, and soil moisture). To estimate their catchment averages prior to the floods, we calculate the concentration time t_c for every catchment:

$$t_c = 0.1 * A^{0.3}, \quad (1)$$

where A is the catchment area in km^2 (Corradini et al., 1995; Robinson & Sivapalan, 1997). For rain and snowmelt, we calculate the mean values during t_c prior to the flood day, which is the time period during which water

from these sources can affect the flood generation directly. For soil moisture, we use the value of the day prior to t_c to capture the antecedent catchment wetness state. Using the Theil-Sen estimator (Sen, 1968), we calculate trends in these three values, as they can be expected to have a direct impact on flood magnitude trends. Additionally, we consider AMS date trend as a variable in this category, as a shift in AMS timing can represent a change in flood generation that could explain flood magnitude trends. We calculate this based on the Theil-Sen estimator in a form that is adapted to circular values (Blöschl et al., 2017). Finally, the flood characteristics category contains the mean flood magnitude, which is an obvious choice, as we are aiming to understand its changes.

The climate category contains variables describing the long-term availability of the three flood generating water sources. We calculate the annual mean and the trends of these annual means for rain, snowmelt, and soil moisture. It was previously shown that flood magnitudes change differently depending on catchment wetness or snow availability (Bertola et al., 2021; Blöschl et al., 2019). Trends in rain, snowmelt, and soil moisture can directly affect flood magnitudes due to increased/decreased water availability. Furthermore, we consider precipitation seasonality (Walsh & Lawler, 1981) and its trend, as flood magnitudes in catchments that are mainly influenced by seasonal precipitation events [e.g., hurricanes, atmospheric rivers (ARs)] could change differently from those in catchments with less seasonal precipitation. The final variable in the climate category is aridity, as the ratio of total annual evapotranspiration and precipitation, which factors in differences in runoff generation in arid/humid catchments (Farquharson et al., 1992; Metzger et al., 2020). Aridity has been shown to have a strong effect on the flood frequency distribution especially in terms of flood magnitude variance (Guo et al., 2014).

The flood type category contains the relative importance of five flood generating processes, which are based on the flood process classification described in Section 3.2. Information about the processes controlling flood generation is crucial to understand flood magnitude trends, as catchments will react differently to changes in a flood cause that is often causing AMS, than to changes in a cause that is less relevant for flood generation in those catchments (Tarasova et al., 2019). Similarly, how the relevance of these processes has changed over time could be a crucial factor in flood magnitude trends (Blöschl et al., 2019; Kemter et al., 2020), which is why the trends in their occurrence comprise the remaining five variables in this category.

The land use category includes six variables representing natural and anthropogenic land cover, its changes, as well as direct human influences on streamflow. The relative area of the catchment covered by canopy can affect flood magnitudes, for example, due to higher water storage capacity or delayed snowmelt in forested areas (Blöschl et al., 2007; Marks et al., 1998). As no data on changes in canopy cover are available for most of our study period, we only consider a single value in 2011. Impervious surfaces greatly affect flood generation, by decreasing flood concentration times and increasing surface flow (Blum et al., 2020). Therefore, we use both mean catchment imperviousness and its trend as variables for our study. Comparably, cropland cover alters soil surface properties, affecting runoff generation, evapotranspiration, and thereby flood magnitudes (Schilling et al., 2014), which is why we consider cropland cover and its trend as variables. Finally, water regulation and diversion are targeted human alterations of streamflow, often directly aimed at reducing flood magnitudes (Magilligan & Nislow, 2005) and have been found to influence flood magnitude trends (Vorogushyn & Merz, 2013).

We do not include purely spatial variables (i.e., latitude, longitude, and elevation) as they do not have an effect on floods themselves, but would only be a proxy for other variables such as snowfall or aridity. Their inclusion could prevent the selection of the actual flood magnitude trend controls in the Random Forests. Therefore, the topography category only contains the catchment area, which has been shown to have an impact on flood magnitude trends (Bertola et al., 2020). Analogously to the purely spatial variables, we do not consider the mean flood date as a variable, which is often used to link floods to their generating processes and seasonal effects (e.g., Berghuijs et al., 2016), as these processes and their seasonality are already reflected in the variables of the flood type and flood characteristics category. Furthermore, the usefulness of these mean flood dates would differ regionally depending on the length or existence of a flood season. Nevertheless, we calculate mean AMS dates for all catchments using the circular mean, to contextualize some of our results in the discussion.

3.2. Flood Similarity and Clustering

As climate, topography, flood generation, land use, and the temporal variations of potential controls vary regionally throughout the CONUS, we assume that controls on flood magnitude trends vary regionally as well. Therefore, we divide the catchments into clusters with similar behavior in terms of flood generating processes, flood

magnitude fluctuations, and flood timing. This approach does not aim to find clusters of similar flood magnitude trends. First, we use catchment-average reanalysis data and a simple classification tree (Figure S1 in Supporting Information S1) to estimate the importance of the five different flood generating processes (i.e., long rain, short rain, snowmelt, rain-on-snow, and soil moisture excess) in each catchment. This classification is adapted from a previous study in Europe (Kemter et al., 2020). It is based on the catchment and weather conditions during the concentration time t_c prior to the day of the flood peak. We classify each AMS and estimate the flood generating process similarity $PS_{x,y}$ for each pair of catchments x and y as:

$$PS_{x,y} = \frac{1}{N}\sum_{n=1}^{N} p \quad \text{with } p = \begin{cases} 1 \; if \; P_{x,n} = P_{y,n} \\ 0 \; else \end{cases}, \tag{2}$$

where N is the total number of years for which data exist for the two stations, n is one of those years, and P is the classified generating process for a flood at a given station in a certain year. PS is one (zero) for stations that have all (no) processes in common.

Second, we test each pair of stations x and y for correlation in terms of flood magnitudes, which we call $RC_{x,y}$. We use the Spearman rank correlation, which compares the magnitudes solely based on relative rank, as absolute flood magnitudes for catchments with similar flood conditions can vary greatly, mainly because of catchment area. We normalize all $RC_{x,y}$ values to the range 0–1 by adding 1 to each value and dividing it by 2. Therefore, values of zero describe complete anti-correlation between the ranks of the flood magnitudes at two stations, while one denotes perfect correlation.

Third, we calculate the event synchronization (ES) of AMS events between all stations (Malik et al., 2010). We do not limit this analysis to commonly recorded years, as this would lead to high ES values at station pairs with low temporal overlap in the flood record. Let t_l^x and t_m^y be vectors of the flood dates at the stations x and y, where $l = 1, 2, 3, \ldots L_x$ and $m = 1, 2, 3, \ldots M_y$, with L_x and M_y as the total number of flood events at the respective stations. We allow for a time lag τ of up to 7 days between floods at two stations for them to be considered synchronous. We use a fixed time lag instead of the more commonly applied variable time lag (Agarwal et al., 2017), because our use of AMS values—as opposed to peak over threshold events—prevents multiple counting of synchronous events. We count the amount of events $C_{x,y}$ that occur at both stations within τ:

$$C_{x,y} = \sum_{l=1}^{L_x}\sum_{m=1}^{M_y} c \quad \text{with } c = \begin{cases} 1 \; if \; 0 \le |t_l^x - t_m^y| \le 7 \\ 0 \; else \end{cases}. \tag{3}$$

We then define the event synchronization $ES_{x,y}$ as:

$$ES_{x,y} = \frac{C_{x,y}}{\sqrt{L_x * M_y}}. \tag{4}$$

A value of 0 means that no events are synchronous between two stations, while a value of 1 indicates that all floods at one station occurred within a week of a flood at the other station and vice versa.

Finally, the three similarity measures are combined into the flood similarity measure $FS_{x,y}$ by multiplication:

$$FS_{x,y} = PS_{x,y} * RC_{x,y} * ES_{x,y}. \tag{5}$$

This measure combines information about similarity in terms of flood generating processes, flood magnitude distribution, and flood timing, weighing all three aspects equally. Our approach is a more comprehensive procedure compared to previous catchment clustering approaches (see Section 5.1 for further discussion). We use FS to find clusters of similar flood behavior using a complex network approach. A complex network consists of nodes, in this case the gauging stations, and links which connect certain nodes based on real-world connection or similarity (Boccaletti et al., 2006; Newman, 2003). Complex networks are therefore ideal for this clustering, as they represent the pairwise similarity of catchments in an intuitive way. We link only those nodes that are exceptionally similar to each other, without factoring in spatial proximity of catchments directly. We define exceptional similarity as the 99th percentile of all FS values. The links, for which FS exceeds this similarity

threshold, are denoted as ones in a binary adjacency matrix in which each row and column represents one catchment. To find clusters within this network, we compute modularity, a measure for how well the nodes within a cluster are connected to each other in comparison to their connections to nodes outside of the cluster (Clauset et al., 2004). Modularity has been successfully used to find clusters of similar flood and extreme rainfall behavior in previous studies (Conticello et al., 2020, 2018; Yang et al., 2019). We use the Louvian modularity algorithm (Blondel et al., 2008), implemented in the *Brain Connectivity Toolbox for MATLAB* (Rubinov & Sporns, 2010), to maximize the modularity of the clusters. This method is among the most commonly used network clustering approaches due to the low computational cost and accurate clustering performance, especially for relatively small networks like ours (Lancichinetti & Fortunato, 2009). To have a decent sample size to train the Random Forests, we merge small clusters (details on the threshold in Section 4.1) with the respective large cluster to which they have the most links. Very small clusters (<10 stations) and stations that are not connected to any other station within the network are excluded from further analysis.

3.3. Trend Detection and Attribution

We use the Theil-Sen slope estimator to estimate the values and quantify the significance ($p < 0.05$) of the trends in flood magnitudes for each station. To understand what controls the spatial variability in flood magnitude trends (regardless of their significance) across the CONUS we use Random Forest models (Breiman, 2001). This method can be used for the regression of a response variable based on several predictor variables. The prediction combines the results of a large number—in this Case 500—of decision trees, which can chose from a random subset of all predictors at each split of the tree. This kind of ensemble regression with a random element reduces overfitting, that is, improves the model performance on new data (Breiman, 2001). We use Random Forests, as we cannot a priori reject the possibility that some controls affect magnitude trends in a nonlinear way, which would not be captured in a linear regression model. Furthermore, we employ forests based on conditional inference trees, which have the advantage of resulting in less biased predictor selection for correlated predictors compared to commonly used partition trees (Hothorn et al., 2006). This allows us to reduce biases in initial variable selection. We use the *cforest_unbiased* function of the *party*-package in *R*, which set. all user parameters of the Random Forests to further minimize bias (Strobl et al., 2007).

Here we use our set of 31 variables as predictors for Random Forests to detect flood magnitude trend controls, that is, those variables that explain most of the spatial variability in flood magnitude trends. It is not necessary to focus only on significant trends here, as nonsignificant magnitude trends can be caused by significant opposing effects from different controls, which are averaged at the catchment scale (Kormann et al., 2015). While one could view these mixing effects at the seemingly nonsignificant center of the trend distribution within each cluster as random variability, they are important for fitting the Random Forest models to all flood magnitude trends and their inclusion is crucial for successful trend attribution (Kormann et al., 2015).

We then train one Random Forest for each cluster and interpret their results separately. To evaluate the performance of our clustered Random Forest approach we measure the predictive performance of the model using mean squared errors (MSE). MSE is computed based on out-of-bag errors, meaning that the calculated error for the prediction of a magnitude trend value is only based on the trees that did not include that value in their training data. This is similar to a training- and test-data split and reduces overfitting of the model. We compare the distribution of squared errors of the clustered Random Forests with those of a global Random Forest trained on all data without cluster information and use a Mann-Whitney test to determine significance of the difference in error distributions (Mann & Whitney, 1947).

3.4. Predictor Selection

To avoid overfitting by the Random Forests (i.e., good predictive capability for training data, but poor results for new data), it is advisable to reduce the amount of model predictors to as few as possible while retaining predictive accuracy. This process is called feature selection, where feature is synonymous with predictor. For this purpose, we use the Recursive Feature Elimination (RFE) algorithm of the *featureselection* function in the *R*-package *moreparty*. RFE is a method that is well suited to select the best set of predictors, even if some of them are correlated (Gregorutti et al., 2017). We use an iterative approach to reduce the number of predictors both at the cluster level and across all clusters. This is accomplished by first applying RFE to each cluster separately and removing predictors using the one standard deviation rule of *featureselection*. All predictors that have not been

selected for any of the clusters are then removed from the entire set of predictors. Afterward, we again apply RFE to all clusters using this reduced set of predictors and repeat this process until only those predictors are left that were selected in at least one cluster. For each cluster, this final set of selected predictors is used to train a Random Forest. The iterative use of relative total effect (RTE) reduces the number of considered predictors in each iteration, decreasing the uncertainty of the predictor selection and reducing the final number of selected predictors per cluster to an interpretable amount. For the global Random Forest, we only select the predictors once, skipping the iterative process.

3.5. Predictor Effects

To understand how the predictors in the Random Forest models affect the flood magnitude trends, we use ALE plots, which describe how a model predictor affects the model prediction throughout the predictor's value range (Apley & Zhu, 2020). In hydrology, this method has been recently used to study the effects of catchment variables on flood generating processes (Stein et al., 2021). The uncentered ALE $\hat{g}_{j,\mathrm{ALE}}(x)$ is calculated based on differences in the prediction within quantiles of the predictor:

$$\hat{g}_{j,\mathrm{ALE}}(x) = \sum_{k=1}^{K_j} \frac{1}{n_j(k)} \sum_{i:x_{i,j}\in N_j(k)} \left[g\left(z_{k,j}, x_{i,\setminus j}\right) - g\left(z_{k-1,j}, x_{i,\setminus j}\right) \right], \tag{6}$$

where x is one of the values of the predictor j for which the ALE plot is calculated, k is one of K_j quantiles into which the range of x is divided, $n_j(k)$ is the number of values of x that fall into this quantile $N_j(k)$ ranging from $i = 1, 2, \ldots n_j(k)$, and $z_{k,j}$ denotes the values of x at the boundary of that quantile. Furthermore, g is the output of the prediction model, and $x_{i,\setminus j}$ are the values of instance i for all other predictors except for j. This means that for each quantile we calculate the mean difference in the model response between the upper and the lower margin of that quantile. We use a variable number of quantiles depending on the number of stations within a cluster. We chose the number of quantiles so that each of them contains at least 15 stations, but limit the maximum number of quantiles at 15, to prevent the appearance of unrealistically high resolution and artifacts in the ALE plots. The uncentered ALE values are centered to retrieve $\hat{f}_{j,\mathrm{ALE}}(x)$ by subtracting its mean across all quantiles:

$$\hat{f}_{j,\mathrm{ALE}}(x) = \hat{g}_{j,\mathrm{ALE}}(x) - \frac{1}{K_j} \sum_{k=1}^{K_j} \hat{g}_{j,\mathrm{ALE}}(x_k) \tag{7}$$

Furthermore, we classify the predictor effect as "positive," "negative," or "non-monotonic." For this, we first test for the existence of a significant ($p < 0.05$) monotonic pattern in the ALE data using the Mann-Kendall test (Mann, 1945). If a significant relationship exists, we use the sign of its slope to assign "positive" or "negative" accordingly. If the Mann-Kendall test finds no significant relationship, we assign "non-monotonic." Note that this does not mean non-monotonic in the strict mathematic sense, but rather a distinction between a linear relationship and all other types of relationship. While the slope of the ALE plot gives us information about the directional effect of a predictor on the magnitude trends, we use the mean absolute ALE values to derive the absolute predictor importance, similar to Stein et al. (2021). For each cluster, we divide these values by the largest value within this cluster and call the resulting measure of predictor importance the RTE with a range between 0 and 1.

3.6. Country-Scale Variability

With our setup, the Random Forests explain the spatial variability of flood magnitude trends within each cluster, which we call regional variability. This approach however has a potential shortcoming for clusters in which the regional variability is not evenly spread around zero, but highly skewed toward the negative or positive. For example, if all flood magnitude trends in a cluster were positive (negative), the Random Forests could only explain why some catchments show more positive trends than others, but not why the overall trend of the cluster is positive (negative). We call the differences in the overall trends of different clusters the country-scale variability, reflecting how much the overall trends of the clusters differ from zero. Scenarios can be imagined in which the country-scale variability is controlled by different processes than the regional variability, which makes it important to study both types of spatial variability.

For clusters with a clear majority (at least two in three stations) of trends with a certain sign, we therefore additionally analyze the country-scale variability. For this, we use boxplots of the 31 catchment variables, grouped by clusters, to find the most prominent differences to neighboring clusters. Additionally, we study the selected

features and ALE plots of the global Random Forest, which contains more information about country-scale variability, as it includes all catchments from every cluster.

4. Results

4.1. Clustering Results

The clustering algorithm found 13 clusters. There were 273 catchments that were not assigned to any of the clusters and were excluded from further analysis, leaving a total of 4,390 catchments. Three of the 13 clusters consisted of less than 90 catchments each: 58 in a cluster on the Florida Peninsula, 64 along the eastern flank of the Rocky Mountains, and 88 on the Colorado Plateau, compared to 218 for the next smallest cluster. We considered these sample sizes as too small to guarantee reliable results, and the clusters were therefore merged with the cluster to which they had the most connections within the complex network, the Southeast, Rocky Mountains, and Pacific SW clusters, respectively. This produced a final set of 10 clusters consisting of 218–654 catchments (Figure 2a).

Flood magnitude trends (Figure 2b) range from −5.5%/yr to 4.6%/yr with a median of −0.0009%/yr. The 5th and 95th percentile of trends are −1.4%/yr and 1.1%/yr, respectively. There are 743 stations that show significant trends ($p < 0.05$, Theil Sen), but even for those without significant trends there are distinct regions of predominantly positive and negative trends. The clusters Midwest, Southeast, and New England consist mainly (i.e., at least 66.6%) of catchments with one direction of trends (Figure 3a). In the seven other clusters, both positive and negative flood magnitude trends are almost equally common.

The distribution of catchment variables shows that the clusters differ substantially from each other (Figure 3). All clusters are unique in at least one aspect and are clearly distinct from their neighboring clusters in several metrics. As the clustering is solely based on flood behavior, this highlights how a river is aggregating the different hydro-climatological and land cover processes of a catchment.

4.2. Variable Selection and Effects

During the iterative predictor selection process for the Random Forests, 16 predictors were chosen in at least one cluster, following the fourth iteration, which found no unselected predictors. We only show the three most important predictors here (Figure 4), while the detailed ALE plots and exact importance values can be found in Section 5.2 (Figure 5) and in Figures S3–S11 in Supporting Information S1. The 15 predictors that have not been selected in any cluster are mean flood magnitude, flood date trend, relevance of long rain, rain on snow and soil moisture floods, all trends in process relevance, annual snowfall trend, annual soil moisture trend, flood generating snowmelt trend, cropland trend, and imperviousness trend. If any of these variables is important for flood magnitude trends, it is not at the regional level or only for very few catchments within a cluster.

During the iterative predictor selection process, between two and five predictors were selected per cluster. Catchment area, annual rainfall trend, and flood soil moisture trend were only selected in one cluster each. The variables imperviousness (four times), mean annual rainfall, aridity, precipitation seasonality trend, flood rainfall trend, and canopy cover (three times each) were selected most often. The categories climate and land use are almost equally important, with predictors from the respective categories being selected 15 and 11 times. At least one climate predictor was selected in all but one cluster (Midwest).

The combined prediction performance of the clustered Random Forest models (MSE = 0.409) is significantly better ($p < 0.05$ based on the Mann-Whitney test) than that of the global Random Forest for the entire CONUS (MSE = 0.445). In the global Random Forest, only five predictors were selected: mean annual rainfall, annual rainfall trend, aridity, flood rainfall trend, and imperviousness. All of these are important in at least one of the clustered Random Forests as well, and four are among the most frequently selected predictors in the clustered models (all but annual rainfall trend).

4.3. Controls on Flood Magnitude Trends

Based on Figure 4 and the underlying ALE plots (Figure 5 and Figures S3–S11 in Supporting Information S1) we can draw conclusions on the processes that control regional variability in flood magnitude trends for the different regions. In this section, we present the controls of regional variability for the clusters in the context of previous

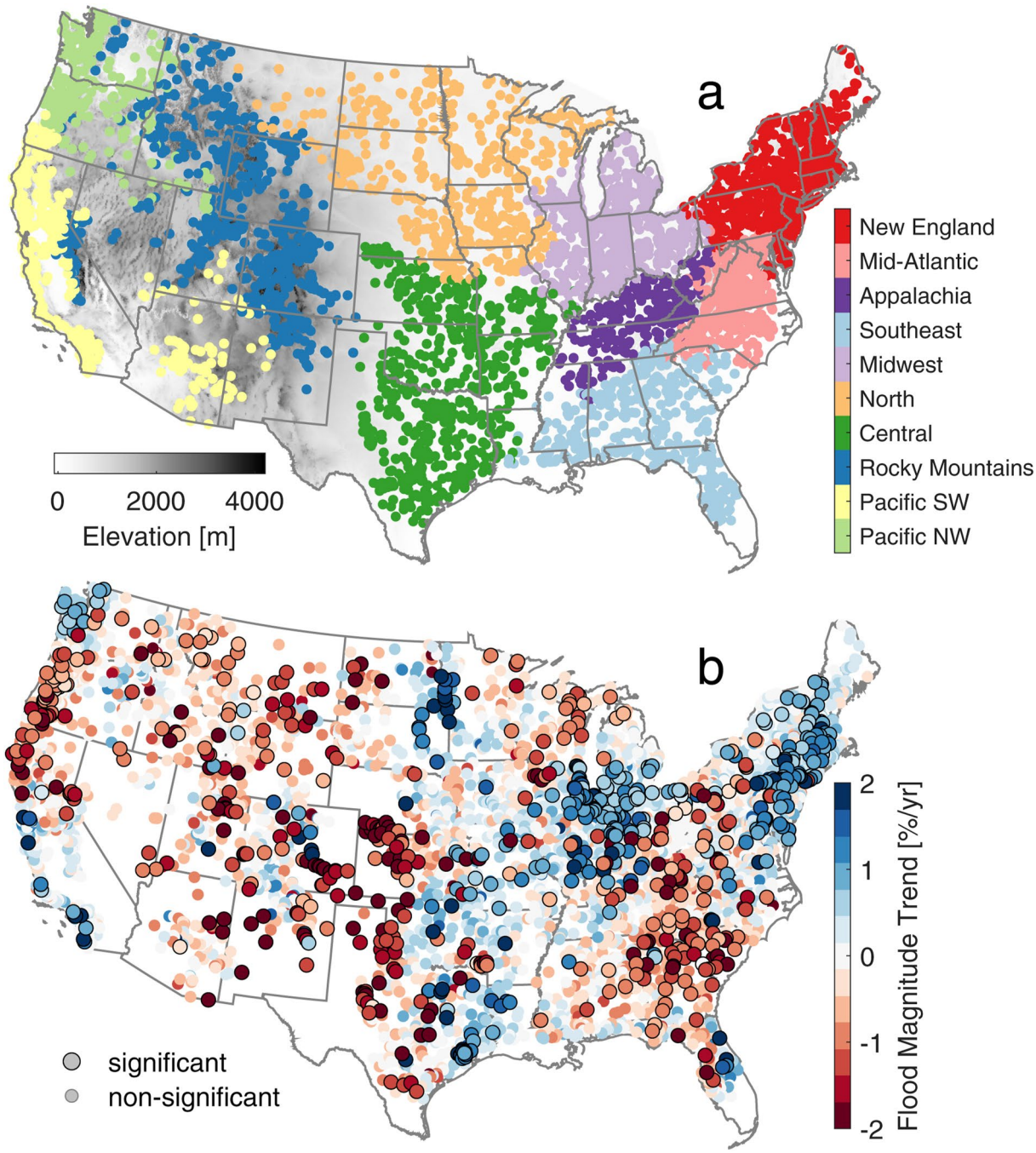

Figure 2. Flood behavior clustering and flood magnitude trends for the conterminous United States. (a) Based on flood generating processes, flood magnitude distribution, and flood timing, 10 distinct clusters of catchments have been found. Clusters are based on complex network modularity, meaning that floods in catchments within a cluster are similar to each other, but distinct from those in all other clusters. Note that no spatial information is part of the clustering process and that the clustering does not aim to group catchments of similar flood magnitude trends. (b) Trends in the magnitudes of annual maximum floods for 4,390 stations between 1960 and 2010 based on Sen's slope. Large dots mark significant trends ($p < 0.05$). Values are relative to the mean of the station time series. Color bar range is truncated to improve readability.

work. For the clusters with at least 66.6% of catchments showing a trend of the same sign, we additionally discuss country-scale variability based on Figure 3.

In the Pacific NW cluster, the most important predictor is aridity, which affects flood magnitude trends negatively (Figure 5). Simultaneously, flood generating rainfall trends have a positive effect. This means that flood

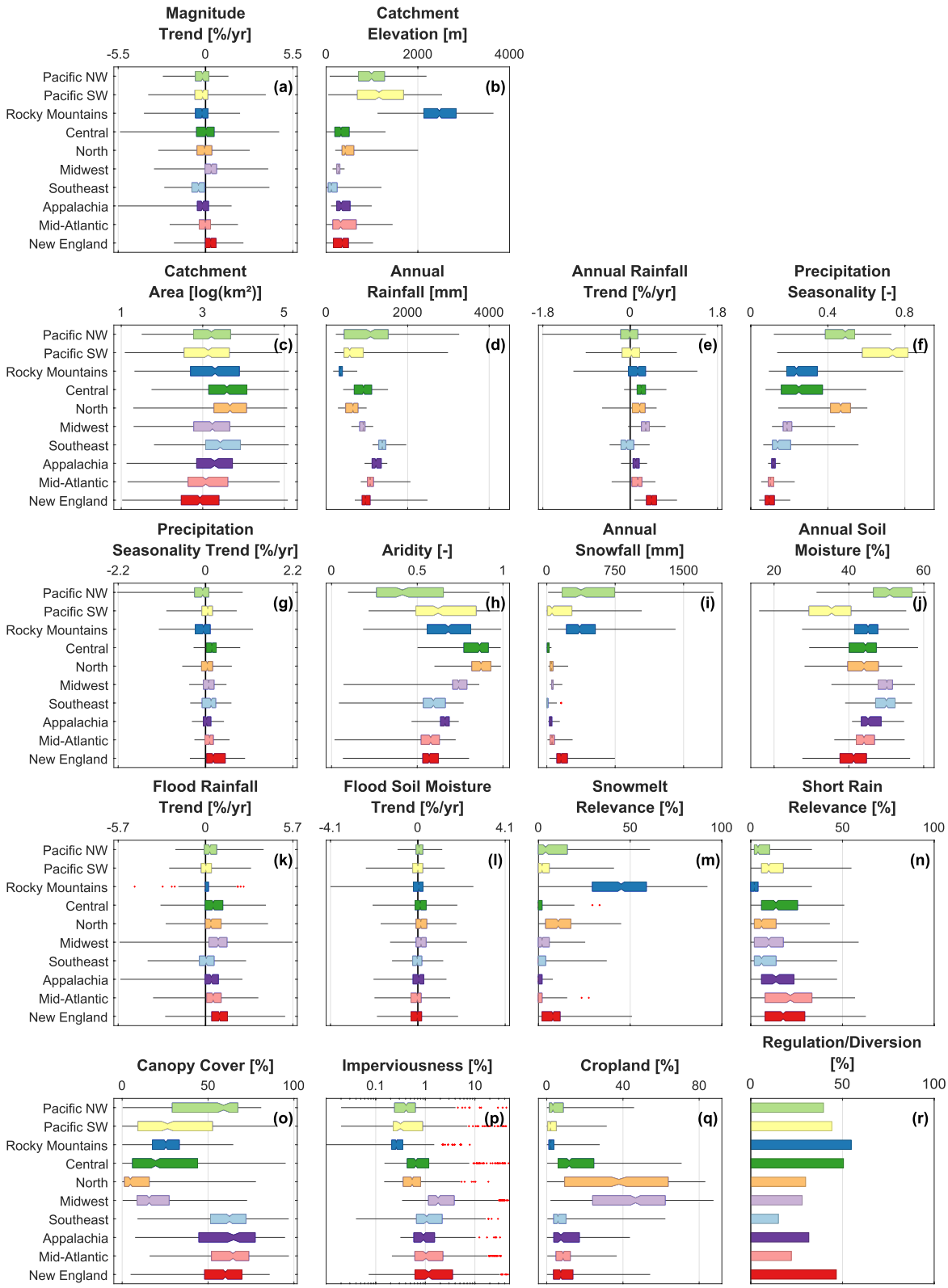

Figure 3. Cluster property comparison. Distribution of flood magnitude trends (a), catchment elevation (b), and the final set of 16 selected variables grouped by clusters (d–r). Panel titles indicate the values depicted on the *x*-axes. Boxes range from the 25th to the 75th percentile with a notch indicating the median. Whiskers reach to 10 times the interquartile range or to the furthest data point, if the latter is closer. Outliers marked in red. Imperviousness is shown on a logarithmic scale (p). The binary variable regulation/diversion is shown as a bar plot of the percentage of regulated catchments per cluster (r). All unselected variables are shown in Figure S2 in Supporting Information S1.

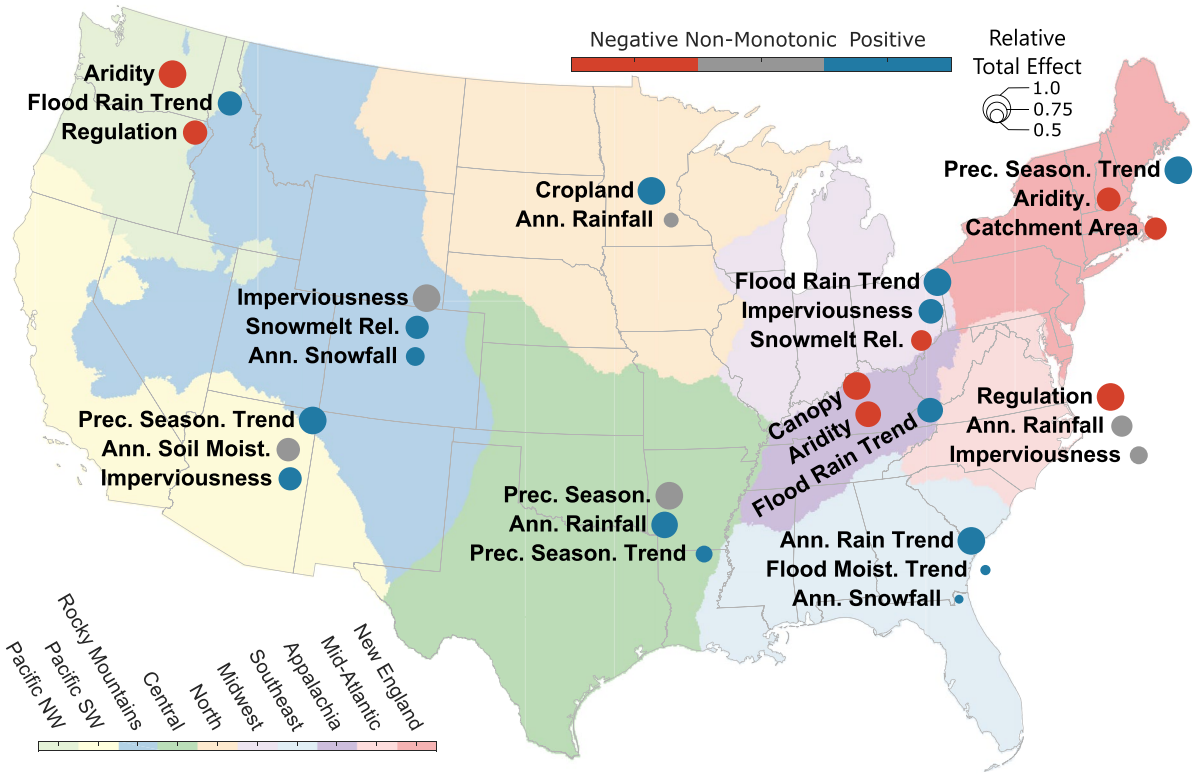

Figure 4. Catchment variable effects on flood magnitude trends. Overall predictor importance (dot size) of the three most important variables in each cluster (see Figure 5 and Figures S3-S11 in Supporting Information S1 for exact values and variables of lower importance). Only two variables were selected in the North cluster. Predictor importance is calculated as the relative total effect, that is, the mean absolute ALE divided by the maximum mean absolute ALE of the cluster. The directional effect (dot color) is estimated from the ALE slope. A positive (negative) effect means that the predicted flood magnitude trend increases (decreases) with increasing variable values. Interpolated cluster map in the background depicts the most common cluster among the 20 nearest stations at each 0.05° × 0.05° pixel.

magnitude trends are the highest in very wet catchments, in which flood generating rainfall has increased, and magnitude trends are lower in other catchments. The very dry catchments in which magnitude trends are high are situated in the rain shadow of the Cascade Range in the eastern part of the cluster, where precipitation patterns and AR influence differ greatly from the coastal part of the cluster (Hu et al., 2017). This could also be linked to an increasing amount of precipitation falling as rain instead of snow (Hatchett, 2018), mainly due to increasing temperatures (Yan et al., 2019) and partly due to warming ARs—which are essential for flood generation (Barth et al., 2017; Neiman et al., 2011) in the region. The positive flood magnitude trends are mostly (82%) limited to unregulated catchments, which shows in the negative effect of regulation/diversion.

In the Pacific SW cluster (Figure S3 in Supporting Information S1), the most important predictor is the trend in precipitation seasonality, meaning that relatively high (low) magnitude trends occur in catchments in which precipitation has become more (less) seasonal. The second highest importance is found for mean annual soil moisture, which shows a distinct, nonlinear change point in the ALE plots, with a positive effect for catchments with values below 33%. This means that AMS trends are high for the dryer catchments in southern California and the Colorado Plateau, where short rain events (positive ALE) are more important for flood generation. This indicates that the different types of flood generation in this cluster (i.e., short-lived spring rain in the south and east; long rain, ARs and a snowmelt component in the north) have changed differently, which can lead to drastic changes in flood magnitudes (Davenport et al., 2020). Additionally, there is a strong positive effect of imperviousness mainly for the highly urbanized areas of Los Angeles and the San Francisco Bay.

In the Rocky Mountains cluster (Figure S4 in Supporting Information S1), we find that imperviousness has the highest importance, but that its highly positive effect is limited to the few (2.6%) catchments with values above 2%. The mean annual snowfall and the relevance of snowmelt have a positive effect on flood magnitude trends. These two effects fit previous findings, suggesting that in low-to-medium elevation catchments in this region,

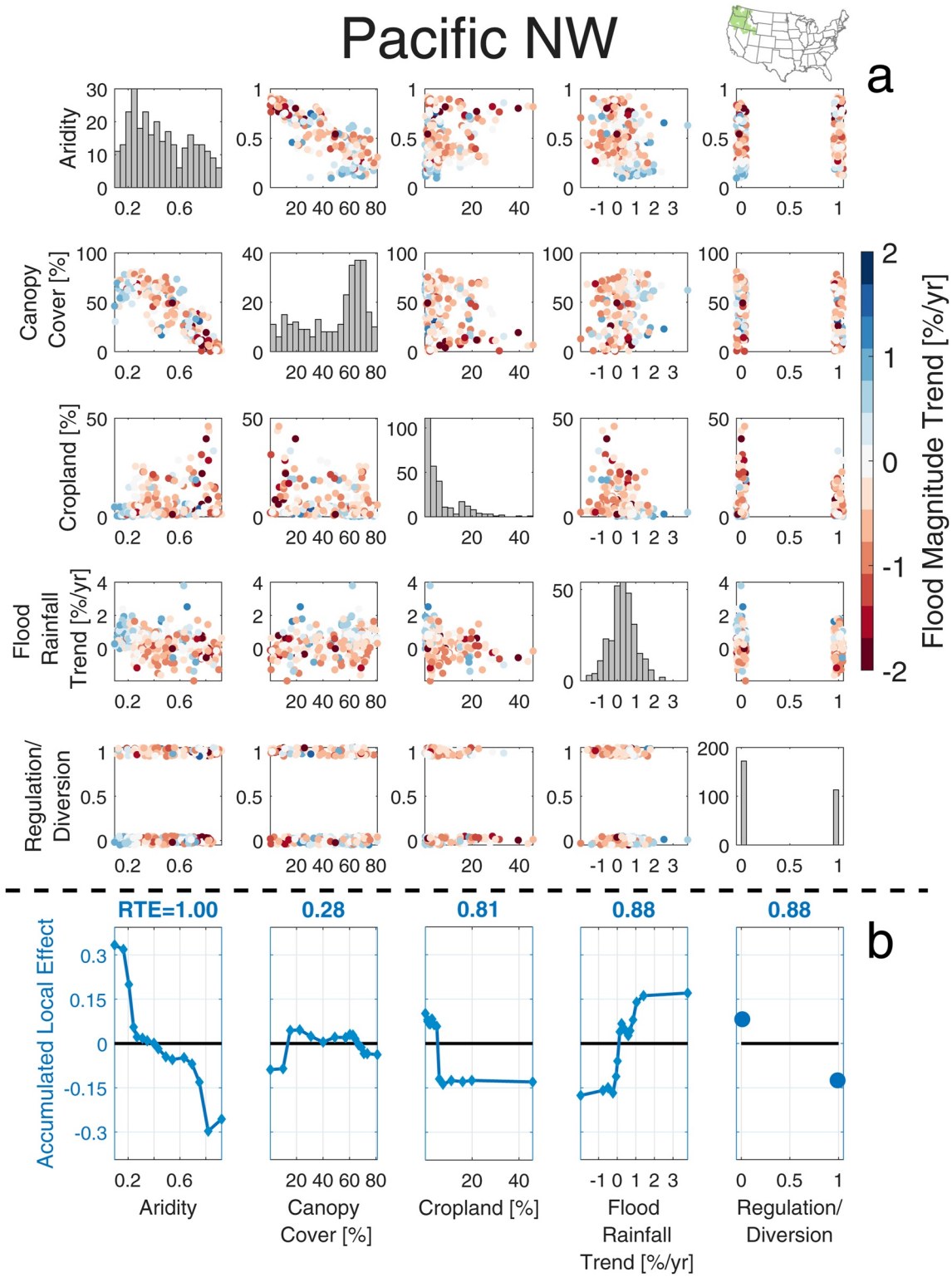
Pacific NW
a
Aridity
Canopy Cover [%]
Cropland [%]
Flood Rainfall Trend [%/yr]
Regulation/ Diversion
Flood Magnitude Trend [%/yr]
b
RTE=1.00
0.28
0.81
0.88
0.88
Accumulated Local Effect
Aridity
Canopy Cover [%]
Cropland [%]
Flood Rainfall Trend [%/yr]
Regulation/ Diversion

Figure 5.

increasing temperature has led to an earlier onset of snowmelt and reduced snowpacks, causing lower streamflow in spring (Rood et al., 2016), which is the region's main AMS season. At the higher elevations with more annual snowfall and a higher relevance of snowmelt for flood generation, snowpacks and thereby floods are not yet affected by increasing temperatures (Rood et al., 2016).

The Central and North clusters show similarities in the spatial distribution of flood magnitude trends (Figure 2b). Both have an average magnitude trend of about zero, but with a high variance in both directions and some of the highest and lowest trends in the entire CONUS (Figure 3a). There is a division roughly along the 100th western meridian separating negative AMS magnitude trends to the west and positive trends to the east, except for a patch of negative trends in the northeast of the North cluster. This east-west separation is common among climatological variables in this region, as it separates the arid highland prairie of the "Great Plains" in the west from the more humid, low-lying regions in the east. An influence on flood magnitudes that is not directly represented in our model is the lowering of the groundwater table due to human activity, which has been shown to have decreased peak streamflow in Kansas and Nebraska (Rasmussen & Perry, 2001), which are part of these two clusters. Slater and Villarini (2016) have shown that this water withdrawal has decreased basin wetness in this area.

Our model finds similar importance for the nonlinear negative effect of precipitation seasonality and the positive effect of mean annual rainfall on flood magnitude trends in the Central cluster (Figure S5b in Supporting Information S1). Both variables are highly anti-correlated here (Figure S5a in Supporting Information S1). Precipitation seasonality trend, with its almost linear, positive effect, is the third most important predictor. This suggests that, where precipitation is converging toward the flooding season, flood magnitudes increase, while catchments in which seasonality remains constant or diverges show no or negative flood magnitude trends. In the eastern half of the cluster, mean flood dates lie in late winter and early spring, when ARs cause most AMS (Lavers & Villarini, 2013; Nayak & Villarini, 2017), while supercell thunderstorm and hurricanes cause many floods in summer and fall in the west and south (Aryal et al., 2018; Smith et al., 2001). It is noteworthy that none of the land use variables we considered was selected in this cluster, suggesting that human activity is less important than hydro-climatological changes for flood magnitude trends, despite 48.2% of rivers being regulated/diverted. This high percentage does however not say anything about the degree of regulation at each of these rivers or whether water management has changed in the study period—aspects that we cannot study with this dataset. Furthermore, this seeming irrelevance of human activity does not include potential effects from water withdrawal of the Ogallala Aquifer, which is an important water source for rivers in the cluster and has been depleting between 1950 and 2015 (McGuire, 2017).

In the North cluster, cropland cover is the most important variable (Figure S6 in Supporting Information S1). It has a positive effect with a step change at a value of 30%, suggesting a direct human influence on floods through surface alteration and tile drainage, which only manifests itself in the flood magnitude trends if a large enough part of the catchment is affected. In contrast to the Central cluster, annual rainfall has a more complex effect on AMS magnitude trends here. The effect is negative for very dry catchments, positive for medium dry ones (400–600 mm/yr), neutral for medium wet ones (700–900 mm/yr), and positive again for wetter catchments. It has been previously shown that precipitation is the most reliable predictor for flood frequency trends in this region (Neri et al., 2019), but for magnitude trends the situation seems to be more complicated.

In the Midwest cluster flood magnitudes have increased in 74.0% of catchments and its mean flood magnitude trend of 0.34%/yr is the highest among all clusters (Figure 3a). Annual rainfall has increased in 98.2% of catchments (Figure 3e), which is most likely a key contributor to these AMS magnitude trends. This cluster shows a narrow window of mean AMS dates with 89% of them in March and April, suggesting that flood generation is very homogeneous. It is the cluster with the highest mean imperviousness and cropland cover (Figures 3p and 3q), as well as river regulation or diversion in 31.8% of catchments, indicating a strong anthropogenic influence on flood magnitudes. The most important variable in this cluster is the trend of flood generating rainfall, which has a

Figure 5. Predictor effects for the Pacific NW cluster. (a) Bivariate scatter plots for the five predictors that were selected for this cluster. Red (blue) dots indicate decreasing (increasing) flood magnitude trends of a given catchment. Random noise added to regulation/diversion *y*-values [it does not affect the variable it is plotted against or the Accumulated Local Effect (ALE) plots] to improve visibility of single dots. Histograms plotted on the diagonal. *Y*-axis labels on the left apply to the entire row, except for the histograms, which show counts on the *y*-axes. *X*-axis labels in the bottom of panel (b) apply to the entire column in panel (a). (b) Accumulated Local Effect plots for the five predictors in blue. RTE = relative total effect, that is, the mean absolute ALE divided by the maximum mean absolute ALE across selected variables in the cluster. Diamonds mark the quantile edges for which ALEs were calculated, highlighting the variable quantile ranges. The plot can be analyzed by comparing each ALE plot in panel (b) to the bivariate plots in the respective column of panel (a), noting the relationship between magnitude trends and different predictor values.

positive effect (Figure S7 in Supporting Information S1). Flood generating rainfall has increased in 82.2% of the catchments, likely driven by the generally increasing rainfall amounts mentioned above. Previous work suggests that about 50% of AMS in this cluster are related to ARs (Nayak & Villarini, 2017). Even though no trends in the frequency or duration of the ARs in this region could be found (Nayak & Villarini, 2017), an intensification of the associated rainfall could have driven the rainfall and flood magnitude trends. The second most important variable is imperviousness, which has a strong positive effect for the 94 catchments with imperviousness values above 6%, 44 of which are situated in the Chicago metropolitan area. Increasing imperviousness due to urbanization has been linked to increasing flood frequencies in the Chicago area (Villarini et al., 2013). Canopy cover mainly has a negative effect for catchments with more than 40% canopy cover. This means that flood magnitude trends are lower in the less urbanized or cultivated catchments of this cluster, hinting at a buffering effect of forested areas for increasing rainfall sums. Snowmelt relevance has a negative effect, with catchments in which snowmelt generates more than 12% of floods showing mainly negative flood magnitude trends.

Flood magnitudes have decreased in 78.3% of the catchments in the Southeast cluster (Figure 3a). The trend in annual rainfall has by far the highest importance and shows a positive effect (Figure S8 in Supporting Information S1). As 76.6% of stations in this cluster have their mean AMS date in February or March, we calculated rainfall trends for these 2 months. With a cluster-wide mean of −0.43%/yr these are more than seven times lower than the annual trends (−0.059%/yr). This suggests that especially the decreasing rainfall amounts in February and March are responsible for the generally negative flood magnitude trends in the cluster. This is likely related to the decreasing number and intensity of extra-tropical cyclones (Wang et al., 2013), which are important for spring precipitation in this region (Hawcroft et al., 2012). With a much lower importance, the trend in flood generating soil moisture (positive effect) was selected. As high antecedent soil moisture is responsible for the generation of 35.7% of AMS in this cluster—which is the highest value among all clusters (Figure S2e in Supporting Information S1)—a change in soil moisture can be expected to affect flood magnitudes substantially here. Annual rainfall sums have decreased the most in the catchments along the Atlantic Coast, east of the 86th Meridian. We find especially negative flood magnitude trends in the catchments in the north of the cluster, which are fed by precipitation in the Appalachians, explaining the negative effect of mean annual snowfall.

In the Appalachia cluster, canopy cover is the most important predictor, and has a negative effect in the Random Forest model (Figure S9 in Supporting Information S1). This effect stems from the highly forested regions along the western flank of the Appalachians. Aridity has the second highest effect, with relatively dry catchments showing the lowest flood magnitude trends. The trend in flood generating rainfall is of similar importance but shows a positive effect. It has been previously shown that ARs are responsible for 50%–80% of AMS in this region (Nayak & Villarini, 2017). Similar to the Midwest cluster, a change in the intensity of ARs could have had a significant impact on flood generating rainfall and thereby flood magnitudes in the cluster.

In the Mid-Atlantic cluster, regulation/diversion was found to be the most important predictor (Figure S10 in Supporting Information S1). The predictor effect is negative and only 12.3% of the stations with a positive flood magnitude trend are regulated. Annual rainfall, the second most important predictor, has a nonlinear effect. Annual rainfall trends are gradually increasing from the southwest to the northeast of the cluster. The effect of imperviousness is overall minor, but highly positive for the 13% of catchments with an imperviousness above 4%, which are mainly located in the Washington, D.C. and Richmond metropolitan areas.

Most catchments in the New England cluster show positive flood magnitude trends (73.5%) and the mean flood magnitude trend of 0.31%/yr is the second highest among all clusters (Figure 3a). These generally positive trends are most likely due to increasing annual rainfall in all the catchments (Figure 3e). There is also a high mean flood generating rainfall trend of 0.92%/yr (Figure 3k), which, together with the decreasing relevance of rain on snow or snowmelt for flood generation in 66.2% of catchments (Figures S2h and S2j in Supporting Information S1), suggests a general shift from snow-to rain-associated floods. The ALE plots found the precipitation seasonality trend (positive) as the most important predictor in the cluster (Figure S11 in Supporting Information S1). To reiterate, a low precipitation seasonality means that precipitation is spread out evenly throughout the year, meaning that flood magnitude trends are the highest in catchments in which precipitation has become more seasonal. Similar to the nearby Appalachia cluster, high aridity has a negative effect on flood magnitude trends here. The model found a negative, linear effect for catchment area, meaning that large catchments are experiencing smaller flood magnitude trends. Furthermore, short rain relevance has a minor positive effect. In this region, short rain floods are often caused by late spring or summer storms, which are highly sensitive to the local topography and land cover heterogeneity (Yeung et al., 2011).

5. Discussion

5.1. Clustering

The clustering process produced 10 spatially distinct (Figure 2a) and hydro-meteorologically meaningful clusters (Figure 3). The number of clusters is below those in previous studies (Brunner et al., 2020; Yang et al., 2019), who found 15 clusters; their analyses were based on different datasets, clustering approaches, and similarity measures. The difference can be explained by our relatively high minimum cluster size, which is necessary to guarantee a sufficient sample size for Random Forest training. Compared to the results of Brunner et al. (2020), our clusters are similar in extent and some of them are combinations of two or three clusters in their paper. A comparison to Yang et al. (2019) is less straightforward, as their limited dataset of 242 catchments leads to some very small clusters. Their six largest clusters, which contain 226 catchments, are similar in extent to our Pacific NW, Rocky Mountains, North, Midwest, and New England clusters as well as a combination of the Southeast, Mid-Atlantic and Appalachian clusters. In comparison to both results, our method generated less spatial overlap between clusters. These comparisons show that our clustering approach produced physically reasonable, distinct clusters, in spite of the larger size of the dataset and the ignorance of spatial information during the clustering process.

5.2. General Predictor Effects

We find that in all clusters in which regulation/diversion was selected, it has a negative effect on flood magnitude trends. While it can be expected due to the role of reservoirs in attenuating flood peaks to prevent damage, note that this predictor is a static value with no information about whether regulation has increased during the study period. As the GAGES data only provides a start and an end year for the influence codes, a more detailed analysis of the effects of trends in regulation/diversion is not possible in this study. The negative effects of canopy cover suggest that—similar to the effect of regulation/diversion—the flood attenuating impacts of forests could also mitigate the effects of a changing climate on flood magnitudes.

The ALE plots reveal that imperviousness generally has almost no effect at low values but a very high positive effect above a certain threshold (between 2% and 10% depending on the cluster). Both of these findings are in agreement with previous work (Hodgkins et al., 2019; Vogel et al., 2011). Like in the case of regulation/diversion, imperviousness is a static value for the entire study period. In this case, we additionally use imperviousness trends as a possible predictor, which was however not selected in any of the clusters nor in the global Random Forest model. This suggests that trends in imperviousness—all of which are positive in our dataset—have a smaller effect on flood magnitudes than the amplification of other trends in flood generating mechanisms through existing impervious surfaces. Note however, that the trends of imperviousness and cropland are only based on five and six time steps, respectively (Table S1 in Supporting Information S1). Therefore, the calculated trends might not be accurate. A dataset at a higher temporal resolution would be needed to confirm our finding of a smaller effect of imperviousness trends. To our knowledge, no such dataset exists for the study period.

Cropland cover, which was selected in two clusters, shows strong opposing effects in the Pacific NW and North cluster. While 98% of catchments in the Pacific NW cluster have less than 30% cropland cover, this number is exceeded in 59% of catchments in the North cluster (Figure 3q). The 30% threshold is relevant because the strong positive effect of cropland cover in the North cluster mainly occurs from this value upward (Figure S6b in Supporting Information S1). This discrepancy in cropland value ranges explains the opposing effects and hints at a nonlinear step change in the influence of cropland cover on flood generation which should be studied in more detail in future work to rule out the possibility that they could be artifacts due to insufficient land-cover data quality.

The generally positive effects of annual rainfall trends, flood rainfall trends, and flood soil moisture trends are unsurprising, as a higher (lower) amount of available water in a catchment will naturally lead to increasing (decreasing) flood magnitudes. More interestingly, high precipitation seasonality generally has a negative effect, while its trend has a positive effect. This means that in some clusters, catchments with a more seasonal precipitation distribution have lower magnitude trends, while in other clusters increasing seasonality is linked to higher flood magnitude trends. In the Central cluster, both effects exist simultaneously, which seems contradictory and should be studied further.

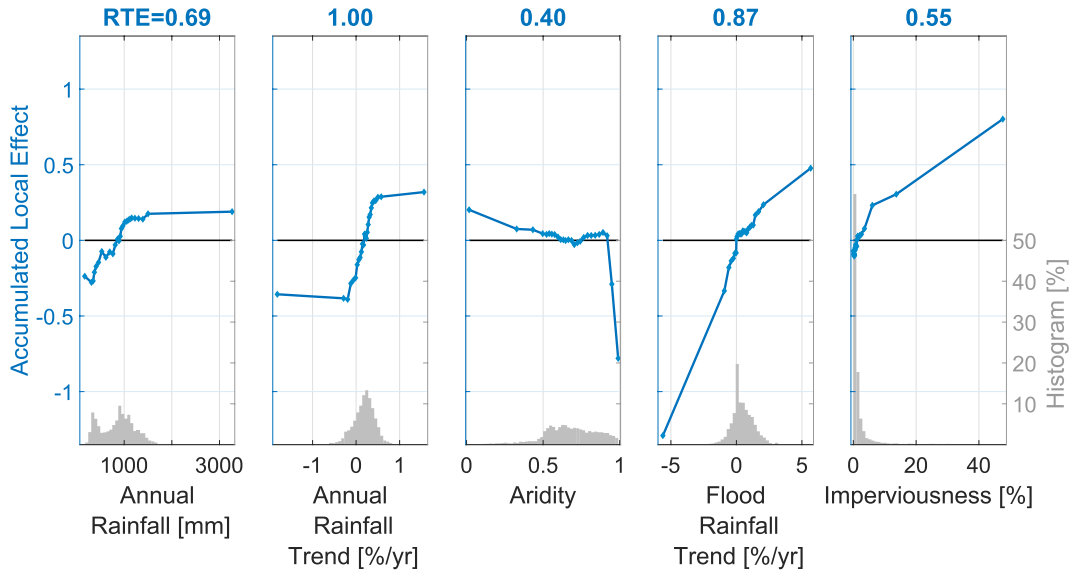

Figure 6. Accumulated local effects for the predictors of the global Random Forest. Blue lines show the ALE values for the five variables that were selected in the conterminous USA (CONUS)-wide Random Forest. Gray bars show a histogram of each variable. The trends in flood generating and annual rainfall are the most important predictors of flood magnitude trends at the CONUS level. RTE = relative total effect, that is, the mean absolute Accumulated Local Effect (ALE) divided by the maximum mean absolute ALE of the cluster. Diamonds mark the quantile edges for which ALEs were calculated, showing the variable quantile ranges.

In three clusters, aridity has a strong negative effect on flood magnitude trends, while mean annual rainfall sums have a positive or nonlinear effect in three others, which is consistent with previous findings that climate change is making wet seasons wetter in rainy regions, but not in dry ones (Chou et al., 2013), and that wet regions in the CONUS have become wetter in the past decades (Greve et al., 2014). The positive effect of short rain relevance in two clusters can be explained by the increase in daily extreme rainfall magnitude across the CONUS (Yin et al., 2018), which affects short rain floods the most, which we define here as floods that are triggered mainly by the rainfall from a single day. Mean annual snowfall and snowmelt relevance show opposing effects in the two respective clusters in which they were selected. For both variables, the effects are positive in the Rocky Mountains cluster, which stems—as mentioned in Section 4.3—from the effects of decreasing snowpacks and earlier snowmelt in the relatively low-lying regions (Rood et al., 2016), where less precipitation falls as snow than in the higher regions. In the two clusters with negative effects for these variables (Midwest and Southeast), snow is generally less relevant for flood generation but in the catchments where it is relevant, flood magnitude trends seem to be lower due to decreasing snow abundance.

5.3. Country-Scale Variability

While the predictive performance of the clustered Random Forests is higher than that of the global model (see Section 4.2), the characteristics of the latter can still yield additional insights into controls on flood magnitude trends at the CONUS scale. In the global Random Forests three climatic predictors, one flood process predictor, and only one land cover variable were selected (Figure 6), in contrast to the clustered Random Forests in which climate and land cover predictors are of almost similar importance. This suggests that climatic changes were of higher importance for past flood magnitude trends at the CONUS scale, while land cover conditions were almost equally important as climatic changes at the regional scale.

In the three clusters with predominantly negative (South Atlantic) or positive (Midwest and New England) flood magnitude trends, we identified a respective negative or positive trend in annual and flood generating rainfall sums as the most likely controls on these predominant trends. This result is reflected in the ALE plots for the global Random Forest, in which these two variables have the highest importance and a positive effect on flood magnitudes.

5.4. Uncertainty Assessment

All climatological variables are based on a single reanalysis dataset (Livneh et al., 2013). We note that this dataset was not initially calibrated to study flood events and that its daily resolution is too coarse for the quick flood generation in the smaller catchments of our dataset. Therefore, it might not represent all aspects of flood generation correctly. However, previous studies have successfully used the dataset to model floods and high flows in Texas (Zhao et al., 2016) and across the CONUS (Oudin et al., 2018). The fact that the controls we found in many clusters fit well to results from previous regional studies (see Section 4.3) further support the feasibility of the reanalysis data for this study. Nevertheless, all results have to be seen with this uncertainty in mind, meaning that variable importance should be viewed in relative, rather than absolute terms.

The flood classification approach we are using is a simplification of the real world in several ways. First, we only consider five different types of flood generation and treat them as distinct even though real world floods are always the result of the interplay of water from multiple sources. This degree of simplification is in line with most flood type classifications (Berghuijs et al., 2016; Tarasova et al., 2019). More nuanced classifications exist (e.g., Tarasova et al., 2020), but the aim of this study—to investigate a large, heterogeneous region with the same set of variables—required a simpler approach. Furthermore, the calculation of trends in flood types requires a certain number of events of a given type at each station within the study period. A larger number of flood types would therefore decrease the chance of gaining a sufficient number of events per type. Second, the thresholds in the classification scheme are to some extent arbitrary, even though they are based on previous studies (Kemter et al., 2020; Stein et al., 2020). A sensitivity analysis (not shown here) to small changes in these thresholds showed no substantial impact on our main results. Third, the classification is based on a simple representation of catchment concentration time t_c. While more realistic representations of t_c are available (Tarasova et al., 2019), which take into account the flood generating process, we decided to use the approach we present here. We did so because one aim of t_c in this study is to detect the flood generating process itself, which excludes it from the information that could be used to calculate t_c. Furthermore, we did not consider the variation of t_c between AMS events by analyzing event hydrographs. This would complicate the comparison and trend calculation of flood contributions from different water sources (rain, snowmelt, soil moisture), which we also calculate based on t_c. In addition, such an approach would not be possible for the smaller, fast-responding catchments in the dataset, as our analysis is based on daily streamflow data.

Sensitivity analysis (not shown here) of the clustering approach shows that it can be prone to shifting cluster assignment due to changes of connections in the complex network, leading to a shift of a group of catchments from one class to another or the merging or splitting of certain clusters. However, we found that, apart from these occasional shifts of small groups, the boundaries between clusters are stable. We furthermore found, that these changes in cluster assignment do not affect the main results of the Random Forests substantially: the ratio of predictor selections between the categories is stable, the most important predictors are always similar, and in cases when groups are transferred to other clusters the importance of a predictor with which this group is highly associated migrates with it to the new cluster.

Given the sparse data availability for land cover changes (only five to six time steps from 1960 to 2010) and no information about canopy cover or regulation/diversion changes, the studied effects of land cover changes carry high uncertainties. The fact that static land cover variables were selected in seven out of 10 clusters suggests that their trends—which were not selected at all—should also be of some relevance for flood magnitude trends. Trends in land use should be assessed in future studies focusing on more recent decades for which more data are available from remote sensing products.

6. Conclusions

Through flood behavior clustering and Random Forests, we found that hydro-climatological and anthropogenic factors have both substantially contributed to past changes in AMS magnitudes across the CONUS. Their similar importance, which could not be shown in previous studies drawing from a less diverse set of controls, highlights the necessity to represent land cover and water regulation as well as their potential changes in models of future flood risk, which so far focus mainly on climatic changes. The land use factors that have been found as important controls on flood magnitude trends are always static indicators (for imperviousness and cropland) or the available information does not allow to quantify whether the factor has changed within the study period (for

regulation/diversion and canopy cover). The result that static land use factors are more important than their trends carries some uncertainty given the sparse data availability in the study period, but suggests that land use is able to attenuate (in the case of forested areas) or amplify (in the case of urbanized areas) the effects of climatic changes on flood magnitudes. This suggests that land use change and land use planning can be crucial aspects of flood risk management at large scales, for instance through reforestation. The adequate consideration of both static and dynamic land cover variables in planning will require additional, dedicated studies, as flood magnitude trend controls can have nonlinear effects—like in the case of imperviousness, for which in four clusters catchments beyond a certain degree of urbanization have witnessed significantly higher flood magnitude trends than those below the threshold. Therefore, land cover changes and effects should be studied more thoroughly in hydrologic modeling research.

The novel combination of the flood behavior clusters with interpretable machine learning (ALE plots) allowed us to show how the different controls influence the flood magnitude trends in different regions of a large study area while minimizing biases in initial variable selection. We identified variables that generally had a negative or positive effect on flood magnitude trends independent of the cluster, while other variables showed opposing effects for different regions due to the diverse hydro-climatological conditions. Such regionally varying effects of flood trend controls can remain undetected in continental scale studies, potentially leading to incomplete interpretations. The ALE plots furthermore allowed us to uncover non-monotonic relationships between catchment variables and flood magnitude trends.

The clustered Random Forests have a higher predictive ability than a global Random Forest trained on the same data. As the clustered Random Forests focus on attributing flood magnitude trends within the clusters, the global Random Forest model is nonetheless useful to identify the controls on country-scale variability. Our results show that climatic changes control the overall flood magnitude trends across the CONUS, while land use modulates these affects within the regions. This highlights that a combination of a clustering approach and a global model is necessary to fully grasp controls on flood trends at a continental scale without neglecting regional differences.

Our findings highlight the importance of a holistic view on flood magnitude trends, as hydro-climatological and land cover conditions can regionally interact in complex ways, which can go unnoticed when entire continents are studied or only certain aspects, subregions, or catchment types are taken into account. Models of future changes in flood risk will have to consider these interactions to fully capture the interconnected nature of flood generation and more reliably project trends.

Data Availability Statement

USGS peak streamflow is available at https://nwis.waterdata.usgs.gov/usa/nwis/peak. The USGS Streamgage NHDPlus catchment shapes is available at https://water.usgs.gov/GIS/metadata/usgswrd/XML/streamgagebasins.xml. Livneh data provided by the NOAA/OAR/ESRL PSL at https://psl.noaa.gov. The GAGES-II Catchment attribute data is available at https://www.sciencebase.gov/catalog/item/59692a64e4b0d1f9f05fbd39.

Acknowledgments
This research was funded by the DFG Research Training Group "Natural Hazards and Risks in a Changing World" (NatRiskChange GRK 2043). NM acknowledges funding by the BMBF project climXtreme (01LP1902J). GV acknowledges funding by the USACE Water Institute. Open Access funding enabled and organized by Projekt DEAL.

References

Agarwal, A., Marwan, N., Rathinasamy, M., Merz, B., & Kurths, J. (2017). Multi-scale event synchronization analysis for unravelling climate processes: A wavelet-based approach. *Nonlinear Processes in Geophysics*, *24*(4), 599–611. https://doi.org/10.5194/npg-24-599-2017

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society – Series B: Statistical Methodology*, *82*(4), 1059–1086. https://doi.org/10.1111/rssb.12377

Archfield, S. A., Hirsch, R. M., Viglione, A., & Blöschl, G. (2016). Fragmented patterns of flood change across the United States. *Geophysical Research Letters*, *43*(19), 10232–10239. https://doi.org/10.1002/2016GL070590

Armstrong, W. H., Collins, M. J., & Snyder, N. P. (2014). Hydroclimatic flood trends in the Northeastern United States and linkages with large-scale atmospheric circulation patterns. *Hydrological Sciences Journal*, *59*(9), 1636–1655. https://doi.org/10.1080/02626667.2013.862339

Aryal, Y. N., Villarini, G., Zhang, W., & Vecchi, G. A. (2018). Long term changes in flooding and heavy rainfall associated with north Atlantic tropical cyclones: Roles of the North Atlantic Oscillation and El Niño-Southern Oscillation. *Journal of Hydrology*, *559*, 698–710. https://doi.org/10.1016/j.jhydrol.2018.02.072

Barth, N. A., Villarini, G., Nayak, M. A., & White, K. (2017). Mixed populations and annual flood frequency estimates in the Western United States: The role of atmospheric rivers. *Water Resources Research*, *53*(1), 257–269. https://doi.org/10.1002/2016WR019064

Berghuijs, W. R., Woods, R. A., Hutton, C. J., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. *Geophysical Research Letters*, *43*(9), 4382–4390. https://doi.org/10.1002/2016GL068070

Bertola, M., Viglione, A., Lun, D., Hall, J., & Blöschl, G. (2020). Flood trends in Europe: Are changes in small and big floods different? *Hydrology and Earth System Sciences*, *24*(4), 1805–1822. https://doi.org/10.5194/hess-24-1805-2020

Bertola, M., Viglione, A., Vorogushyn, S., Lun, D., Merz, B., & Blöschl, G. (2021). Do small and large floods have the same drivers of change? A regional attribution analysis in Europe. *Hydrology and Earth System Sciences*, *25*(3), 1347–1364. https://doi.org/10.5194/hess-25-1347-2021

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Blöschl, G., Ardoin-Bardin, S., Bonell, M., Dorninger, M., Goodrich, D., Gutknecht, D., et al. (2007). At what scales do climate variability and land cover change impact on flooding and low flows? *Hydrological Processes*, *21*(9), 1241–1247. https://doi.org/10.1002/hyp.6669

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., et al. (2017). Changing climate shifts timing of European floods. *Science*, *357*(6351), 588–590. https://doi.org/10.1126/science.aan2506

Blöschl, G., Hall, J., Viglione, A., Perdigão, R., Parajka, J., Merz, B., et al. (2019). Changing climate both increases and decreases European floods. *Nature*, *573*, 108–111. https://doi.org/10.1038/s41586-019-1495-6

Blum, A. G., Ferraro, P. J., Archfield, S. A., & Ryberg, K. R. (2020). Causal effect of impervious cover on annual flood magnitude for the United States. *Geophysical Research Letters*, *47*(5). https://doi.org/10.1029/2019GL086480

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*(4–5), 175–308. https://doi.org/10.1016/j.physrep.2005.10.009

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

Brunner, M. I., Gilleland, E., Wood, A., Swain, D. L., & Clark, M. (2020). Spatial dependence of floods shaped by spatiotemporal variations in meteorological and land-surface processes. *Geophysical Research Letters*, 2019, 1–13. https://doi.org/10.1029/2020gl088000

Chou, C., Chiang, J. C. H., Lan, C. W., Chung, C. H., Liao, Y. C., & Lee, C. J. (2013). Increase in the range between wet and dry season precipitation. *Nature Geoscience*, *6*(4), 263–267. https://doi.org/10.1038/ngeo1744

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, *70*(6), 066111. https://doi.org/10.1103/PhysRevE.70.066111

Conticello, F. R., Cioffi, F., Lall, U., & Merz, B. (2020). Synchronization and delay between circulation patterns and high streamflow events in Germany. *Water Resources Research*, *56*(4), 1–16. https://doi.org/10.1029/2019WR025598

Conticello, F. R., Cioffi, F., Merz, B., & Lall, U. (2018). An event synchronization method to link heavy rainfall events and large-scale atmospheric circulation features. *International Journal of Climatology*, *38*(3), 1421–1437. https://doi.org/10.1002/joc.5255

Corradini, C., Melone, F., & Singh, V. P. (1995). Some remarks on the use of GIUH in the hydrological practice. *Hydrology Research*, *26*(4–5), 297–312. https://doi.org/10.1016/S0140-6736(02)23618-1

Davenport, F. V., Herrera-Estrada, J. E., Burke, M., & Diffenbaugh, N. S. (2020). Flood size increases nonlinearly across the Western United States in response to lower snow-precipitation ratios. *Water Resources Research*, *56*(1), 1–19. https://doi.org/10.1029/2019WR025571

Do, H. X., Mei, Y., & Gronewold, A. D. (2020). To what extent are changes in flood magnitude related to changes in precipitation extremes? *Geophysical Research Letters*, *47*(18), 1–10. https://doi.org/10.1029/2020GL088684

Falcone, J. A. (2017). Coefficient-based consistent mapping of imperviousness in the conterminous U.S. at 60-m resolution for 1974, 1982, 1992, 2002, and 2012.

Falcone, J. A., & LaMotte, A. E. (2016). *National 1-kilometer rasters of selected Census of Agriculture statistics allocated to land use for the time period 1950 to 2012*. US Geological Survey Data Release.

Farquharson, F. A. K., Meigh, J. R., & Sutcliffe, J. V. (1992). Regional flood frequency analysis in arid and semi-arid areas. *Journal of Hydrology*, *138*(3), 487–501. https://doi.org/10.1016/0022-1694(92)90132-F

Ficklin, D. L., Abatzoglou, J. T., Robeson, S. M., Null, S. E., & Knouft, J. H. (2018). Natural and managed watersheds show similar responses to recent climate change. *Proceedings of the National Academy of Sciences*, *115*(34), 8553–8557. https://doi.org/10.1073/pnas.180102611

Gregorutti, B., Michel, B., & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Statistics and Computing*, *27*(3), 659–678. https://doi.org/10.1007/s11222-016-9646-1

Greve, P., Orlowsky, B., Mueller, B., Sheffield, J., Reichstein, M., & Seneviratne, S. I. (2014). Global assessment of trends in wetting and drying over land. *Nature Geoscience*, *7*(10), 716–721. https://doi.org/10.1038/NGEO2247

Guo, J., Li, H.-Y., Leung, L. R., Guo, S., Liu, P., Sivapalan, M., et al. (2014). Links between flood frequency and annual water balance behaviors: A basis for similarity and regionalization. *Water Resources Research*, *50*(2), 937–953. https://doi.org/10.1002/2013WR014374

Hatchett, B. J. (2018). Snow level characteristics and impacts of a spring typhoon-originating atmospheric river in the Sierra Nevada, USA. *Atmosphere*, *9*(6), 1–13. https://doi.org/10.3390/atmos9060233

Hawcroft, M. K., Shaffrey, L. C., Hodges, K. I., & Dacre, H. F. (2012). How much Northern Hemisphere precipitation is associated with extratropical cyclones? *Geophysical Research Letters*, *39*(24), 1–7. https://doi.org/10.1029/2012GL053866

Hirsch, R. M., & Ryberg, K. R. (2012). Has the magnitude of floods across the USA changed with global CO_2 levels? *Hydrological Sciences Journal*, *57*(1), 1–9. https://doi.org/10.1080/02626667.2011.621895

Hodgkins, G. A., Dudley, R. W., Archfield, S. A., & Renard, B. (2019). Effects of climate, regulation, and urbanization on historical flood trends in the United States. *Journal of Hydrology*, *573*, 697–709. https://doi.org/10.1016/j.jhydrol.2019.03.102

Hodgkins, G. A., Whitfield, P. H., Burn, D. H., Hannaford, J., Renard, B., Stahl, K., et al. (2017). Climate-driven variability in the occurrence of major floods across North America and Europe. *Journal of Hydrology*, *552*, 704–717. https://doi.org/10.1016/j.jhydrol.2017.07.027

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., et al. (2015). Completion of the 2011 national land cover database for the conterminous United States – Representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*, *81*(5), 345–354. https://doi.org/10.1016/S0099-1112(15)30100-2

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational & Graphical Statistics*, *15*(3), 651–674. https://doi.org/10.1198/106186006X133933

Hu, H., Dominguez, F., Wang, Z., Lavers, D. A., Zhang, G., & Ralph, F. M. (2017). Linking atmospheric river hydrological impacts on the U.S. West Coast to Rossby wave breaking. *Journal of Climate*, *30*(9), 3381–3399. https://doi.org/10.1175/JCLI-D-16-0386.1

Ivancic, T. J., & Shaw, S. B. (2015). Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge. *Climatic Change*, *133*(4), 681–693. https://doi.org/10.1007/s10584-015-1476-1

Kemter, M., Merz, B., Marwan, N., Vorogushyn, S., & Blöschl, G. (2020). Joint trends in flood magnitudes and spatial extents across Europe. *Geophysical Research Letters*, *47*(7), 1–8. https://doi.org/10.1029/2020GL087464

Kormann, C., Francke, T., Renner, M., & Bronstert, A. (2015). Attribution of high resolution streamflow trends in Western Austria – An approach based on climate and discharge station data. *Hydrology and Earth System Sciences*, *19*(3), 1225–1245. https://doi.org/10.5194/hess-19-1225-2015

Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical review E*, *80*(5), 056117. https://doi.org/10.1103/PhysRevE.80.056117

Lavers, D. A., & Villarini, G. (2013). Atmospheric rivers and flooding over the central United States. *Journal of Climate*, *26*(20), 7829–7836. https://doi.org/10.1175/JCLI-D-13-00212.1

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., et al. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *Journal of Climate*, *26*(23), 9384–9392. https://doi.org/10.1175/JCLI-D-12-00508.1

Magilligan, F. J., & Nislow, K. H. (2005). Changes in hydrologic regime by dams. *Geomorphology*, *71*(1–2), 61–78. https://doi.org/10.1016/j.geomorph.2004.08.017

Malik, N., Marwan, N., & Kurths, J. (2010). Spatial structures and directionalities in Monsoonal precipitation over South Asia. *Nonlinear Processes in Geophysics*, *17*(5), 371–381. https://doi.org/10.5194/npg-17-371-2010

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

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, *13*(3), 245–259.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. Retrieved from www.jstor.org/stable/2236101

Marks, D., Kimball, J., Tingey, D., & Link, T. (1998). 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. https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10/11%3C1569::AID-HYP682%3E3.0.CO;2-L

McGuire, V. L. (2017). *Water-level and recoverable water in storage changes, High Plains aquifer, predevelopment to 2015 and 2013–15*. U.S. Geological Survey Scientific Investigations Report 2017–5040, 14 p. https://doi.org/10.3133/sir20175040

Metzger, A., Marra, F., Smith, J. A., & Morin, E. (2020). Flood frequency estimation and uncertainty in arid/semi-arid regions. *Journal of Hydrology*, *590*, 125254. https://doi.org/10.1016/j.jhydrol.2020.125254

Munich, Re. (2021). NatCatService. Retrieved from www.munichre.com

Nayak, M. A., & Villarini, G. (2017). A long-term perspective of the hydroclimatological impacts of atmospheric rivers over the central United States. *Water Resources Research*, *53*(2), 1144–1166. https://doi.org/10.1002/2016WR019033

Neiman, P. J., Schick, L. J., Martin Ralph, F., Hughes, M., & Wick, G. A. (2011). Flooding in Western Washington: The connection to atmospheric rivers. *Journal of Hydrometeorology*, *12*(6), 1337–1358. https://doi.org/10.1175/2011JHM1358.1

Neri, A., Villarini, G., Slater, L. J., & Napolitano, F. (2019). On the statistical attribution of the frequency of flood events across the U.S. Midwest. *Advances in Water Resources*, *127*(February), 225–236. https://doi.org/10.1016/j.advwatres.2019.03.019

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167–256. https://doi.org/10.1137/S00361445 0342480

Oudin, L., Salavati, B., Furusho-Percot, C., Ribstein, P., & Saadi, M. (2018). Hydrological impacts of urbanization at the catchment scale. *Journal of Hydrology*, *559*, 774–786. https://doi.org/10.1016/j.jhydrol.2018.02.064

Rasmussen, T. J., & Perry, C. A. (2001). *Trends in peak flows of selected streams in Kansas. U.S. Geol. Survey Water-Resources Investigations Report 1-4203*. USGS.

Robinson, J. S., & Sivapalan, M. (1997). An investigation into the physical causes of scaling and heterogeneity of regional flood frequency. *Water Resources Research*, *33*(5), 1045–1059. https://doi.org/10.1029/97WR00044

Rood, S. B., Foster, S. G., Hillman, E. J., Luek, A., & Zanewich, K. P. (2016). Flood moderation: Declining peak flows along some Rocky Mountain rivers and the underlying mechanism. *Journal of Hydrology*, *536*, 174–182. https://doi.org/10.1016/j.jhydrol.2016.02.043

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, *52*(3), 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Sauer, I., Reese, R., Otto, C., Geiger, T., Willner, S., Guillod, B., et al. (2021). Climate signals in river flood damages emerge under sound regional disaggregation. *Nature Communications*, *12*, 2128. https://doi.org/10.21203/rs.3.rs-37259/v1

Schilling, K. E., Gassman, P. W., Kling, C. L., Campbell, T., Jha, M. K., Wolter, C. F., & Arnold, J. G. (2014). The potential for agricultural land use change to reduce flood risk in a large watershed. *Hydrological Processes*, *28*(8), 3314–3325. https://doi.org/10.1002/hyp.9865

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389. https://doi.org/10.1080/01621459.1968.10480934

Sharma, A., Wasko, C., & Lettenmaier, D. P. (2018). If precipitation extremes are increasing, why aren't floods? *Water Resources Research*, *54*(11), 8545–8551. https://doi.org/10.1029/2018WR023749

Slater, L. J., & Villarini, G. (2016). Recent trends in U.S. flood risk. *Geophysical Research Letters*, *43*(24), 12428–12436. https://doi.org/10.1002/2016GL071199

Smith, J. A., Baeck, M. L., Zhang, Y., & Doswell, C. A. (2001). Extreme rainfall and flooding from supercell thunderstorms. *Journal of Hydrometeorology*, *2*(5), 469–489. https://doi.org/10.1175/1525-7541(2001)002<0469:ERAFFS>2.0.CO;2

Stein, L., Clark, M. P., Knoben, W. J. M., Pianosi, F., & Woods, R. (2021). Process oriented insights from interpretable machine learning - what influences flood generating processes? *Authorea*. https://doi.org/10.1002/essoar.10503704.1

Stein, L., Pianosi, F., & Woods, R. (2020). Event-based classification for global study of river flood generating processes. *Hydrological Processes*, *34*(7), 1514–1529. https://doi.org/10.1002/hyp.13678

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*, 25. https://doi.org/10.1186/1471-2105-8-25

Tang, Q., Vivoni, E. R., Muñoz-Arriola, F., & Lettenmaier, D. P. (2012). Predictability of evapotranspiration patterns using remotely sensed vegetation dynamics during the North American Monsoon. *Journal of Hydrometeorology*, *13*(1), 103–121. https://doi.org/10.1175/JHM-D-11-032.1

Tarasova, L., Basso, S., Wendi, D., Viglione, A., Kumar, R., & Merz, R. (2020). A process-based framework to characterize and classify runoff events: The event typology of Germany. *Water Resources Research*, *56*(5). https://doi.org/10.1029/2019WR026951

Tarasova, L., Merz, R., Kiss, A., Basso, S., Blöschl, G., Merz, B., et al. (2019). Causative classification of river flood events. *Wiley Interdisciplinary Reviews: Water*, *6*(4), e1353. https://doi.org/10.1002/wat2.1353

Villarini, G., & Slater, L. J. (2017). Climatology of flooding in the United States. In S. L. Cutter (Ed.), *Oxford research encyclopedia of natural hazard science*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199389407.013.123

Villarini, G., & Slater, L. J. (2018). Examination of changes in annual maximum gauge height in the continental United States using quantile regression. *Journal of Hydrologic Engineering*, *23*(3), 06017010. https://doi.org/10.1061/(asce)he.1943-5584.0001620

Villarini, G., Smith, J. A., Baeck, M. L., Smith, B. K., & Sturdevant-Rees, P. (2013). Hydrologic analyses of the July 17–18, 1996, flood in Chicago and the role of urbanization. *Journal of Hydrologic Engineering*, *18*(2), 250–259. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000462

Vogel, R. M., Yaindl, C., & Walter, M. (2011). Nonstationarity: Flood magnification and recurrence reduction factors in the United States. *Journal of the American Water Resources Association*, *47*(3), 464–474. https://doi.org/10.1111/j.1752-1688.2011.00541.x

Vorogushyn, S., & Merz, B. (2013). Flood trends along the Rhine: The role of river training. *Hydrology and Earth System Sciences*, *17*(10), 3871–3884. https://doi.org/10.5194/hess-17-3871-2013

Walsh, R. P. D., & Lawler, D. M. (1981). Rainfall seasonality: Description, spatial patterns and change through time. *Weather*, *36*(7), 201–208. https://doi.org/10.1002/j.1477-8696.1981.tb05400.x

Wang, X. L., Feng, Y., Compo, G. P., Swail, V. R., Zwiers, F. W., Allan, R. J., & Sardeshmukh, P. D. (2013). Trends and low frequency variability of extra-tropical cyclone activity in the ensemble of twentieth century reanalysis. *Climate Dynamics*, *40*(11–12), 2775–2800. https://doi.org/10.1007/s00382-012-1450-9

Wasko, C., & Nathan, R. (2019). Influence of changes in rainfall and soil moisture on trends in flooding. *Journal of Hydrology*, *575*, 432–441. https://doi.org/10.1016/j.jhydrol.2019.05.054

Wasko, C., & Sharma, A. (2017). Global assessment of flood and storm extremes with increased temperatures. *Scientific Reports*, *7*(1), 1–8. https://doi.org/10.1038/s41598-017-08481-1

Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Leung, L. R., Coleman, A., & Hou, Z. (2019). Observed spatiotemporal changes in the mechanisms of extreme water available for runoff in the Western United States. *Geophysical Research Letters*, *46*(2), 767–775. https://doi.org/10.1029/2018GL080260

Yang, W., Yang, H., & Yang, D. (2019). Identification of homogeneous regions in terms of flood seasonality using a complex network approach. *Journal of Hydrology*, *576*, 726–735. https://doi.org/10.1016/j.jhydrol.2019.06.082

Yeung, J. K., Smith, J. A., Villarini, G., Ntelekos, A. A., Baeck, M. L., & Krajewski, W. F. (2011). Analyses of the warm season rainfall climatology of the northeastern US using regional climate model simulations and radar rainfall fields. *Advances in Water Resources*, *34*(2), 184–204. https://doi.org/10.1016/j.advwatres.2010.10.005

Yin, J., Gentine, P., Zhou, S., Sullivan, S. C., Wang, R., Zhang, Y., & Guo, S. (2018). Large increase in global storm runoff extremes driven by climate and anthropogenic changes. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-06765-2

Zhao, G., Gao, H., & Cuo, L. (2016). Effects of urbanization and climate change on peak flows over the San Antonio River basin, Texas. *Journal of Hydrometeorology*, *17*(9), 2371–2389. https://doi.org/10.1175/JHM-D-15-0216.1