

1 On the statistical attribution of the frequency of flood events across the U.S.
2 Midwest
3

4 ANDREA NERI¹, GABRIELE VILLARINI², LOUISE J. SLATER³, AND FRANCESCO
5 NAPOLITANO¹
6
7

8 ¹ Department of Civil, Constructional and Environmental Engineering, University of Rome “La
9 Sapienza”, Rome, Italy.

10 ² IIHR-Hydrosience & Engineering, The University of Iowa, Iowa City, Iowa, USA.

11 ³ School of Geography and the Environment, University of Oxford, Oxford, UK.
12

13
14 Manuscript submitted to
15 *Advances in Water Resources*
16 18 May 2018
17

18 Revised February 2019
19

20 *Corresponding author address:*

21 Andrea Neri, Department of Civil, Constructional and Environmental Engineering, University of
22 Rome “La Sapienza”, Via Eudossiana 18, 00184, Rome, Italy. E-mail: andrea.neri@uniroma1.it.
23 Tel.: +390644585063
24

ABSTRACT

The frequency of flood events has increased across most of the U.S. Midwest in the past 50-70 years; however, little is known about what is driving these changes. Using an observation-driven approach, we develop a statistical framework to attribute the changes in the frequency of flood peak events to changes in the climate system and to land use / land cover. We focus on 287 U.S. Geological Survey sites with at least 50 years of daily discharge measurements between the second half of the 20th century and the present. Our analyses are performed at the seasonal level and consider five predictors: precipitation, temperature, antecedent wetness conditions, agriculture, and population density. Even though we use simple models, we are able to reproduce well the interannual variability in the frequency of flood events as well as the overall long-term tendencies. Results indicate that precipitation and antecedent wetness conditions are the strongest predictors, with the role of the latter increasing as we lower the threshold for the event identification. Temperature is an important predictor only in the northern Great Plains during spring, where snow-related processes are most relevant. Population (as a proxy of urbanization) and agriculture are less important compared to the climate predictors.

Keywords: flood peaks; attribution; statistical modeling; peak-over-threshold

1 Introduction

Floods are often associated with a high toll in terms of economic impacts and fatalities, and are one of the most costly natural hazards in the United States (Centre for Research on the Epidemiology of Disasters, 2017). In the U.S. Midwest alone, floods have caused dozens of billions of dollars in economic losses over the most recent decades, with a loss of \$7.7 billion and 80 fatalities in just 2013-2017 (NOAA National Centers for Environmental Information (NCEI), 2018). Moreover, recent studies (Mallakpour and Villarini, 2015; Slater and Villarini, 2016) have found that the frequency of flood events has been generally increasing, likely exacerbating these negative impacts. Therefore, an improved understanding of the physical processes responsible for these events can provide basic information to help with our preparedness and response against flood hazards.

Most of the literature has focused on the detection of changes in the frequency of flood events (e.g., Mallakpour and Villarini 2015; Slater and Villarini 2016), showing statistically significant trends across large areas of the central United States. Most of these trends are increasing (Figure 1), regardless of the season or the event rarity. Albeit scientifically important, the knowledge and insight gained from the detection of trends do not advance our understanding of the drivers responsible for those changes. From that perspective, attribution studies related to flooding are essential (Hirsch and Archfield, 2015; Merz et al., 2012) and there has been a growing interest toward this line of research. For instance, (Slater and Villarini, 2017) focused their analysis on establishing which set of drivers allows a better prediction of streamflow, ranging from low to high flows. Slater and Villarini (2017; 2016) and Berghuijs et al. (2016) confirmed that antecedent wetness is an important predictor of streamflow magnitude. Mishra et al. (2010), Schottler et al. (2014), Villarini and Strong (2014) observed that land use / land cover

(LULC) can influence streamflow regime, potentially leading to increases in high flows. Berghuijs et al. (2016), Ivancic and Shaw (2015), Pradhanang et al. (2013) and Tang et al. (2012) showed that temperature can affect the magnitude of flood peaks through evapotranspiration and snow-related processes.

Only few recent studies have attempted to understand the causes of the changes in the frequency of flood events. Mallakpour and Villarini (2015) focused on daily discharge records at 774 stream gage stations across the central United States, and detected statistically significant trends in the frequency of flood counts over the 1962-2011 period; they preliminarily related these changes to more frequent extreme rainfall events. Slater and Villarini (2016) focused on stage height exceeding the National Weather Service's flood thresholds over the past 30 years and found that, while the precipitation preceding the flood events is important, the overall basin wetness is a critical driver of the frequency of flood events also.

Despite the few studies attempting to explain the drivers of year-to-year variations in the flood counts, a framework capable of explaining the observed changes in a comprehensive way is still lacking. In fact, while climatic processes are expected to play an important role (e.g., more frequent heavy rainfall events leading to more frequent flood events (Mallakpour and Villarini, 2015)), much less is known about the effect of changes in LULC on the frequency of flood counts. The role of LULC is expected to be especially important across the U.S. Midwest, which has been experiencing large changes in agricultural intensity and urbanization. For instance, in the Upper Mississippi River Basin, half of the land has been converted from grassland and forest to crops, especially corn and soybean (Frans et al., 2013). In the states of Iowa, Minnesota and Michigan, soybean expansion increased by 1000% at the expense of pastures or other sod-based crops (Zhang and Schilling, 2006). Across the entire Midwest, house density grew by 146%

between 1940 and 2000 (Radeloff et al., 2005), while population increased by 34.4% from 1958 to 2016 (United States Regional Economic Analysis Project, 2017).

Therefore, the main research questions we seek to address are:

- What are the major drivers, whether climate- or LULC-related, of the year-to-year variability and multidecadal changes in the frequency of flood events?
- How do these drivers representing different physical processes change across the different seasons?
- Do we identify the same drivers as we move from rarer to more frequent events?

The paper is organized as follows. We describe the data and the methodology in Section 2, which is followed by Section 3 where we present and discuss our results. Section 4 summarizes the main points of the study and concludes the paper.

2 Materials and Methods

2.1 Data

We focus on stream gages located in the U.S. Midwest (defined here to include Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota and Wisconsin) with daily average discharge measurements up to December 2016. We include only the sites that were classified as not affected by any significant regulation (i.e., the codes “5” or “6” are not present in their peak flow records), and that have at least 50 complete years (we consider a year complete if it has at least 330 daily observations). Figure 2 shows the location of the 287 stream gages.

To obtain the time series of the number of flood events that occurred during each season, we use a peak-over-threshold (POT) approach. We choose a threshold value that gives us one,

two, three and four events per year on average among the entire streamflow record length, allowing only one peak within a time window of 5 days plus the logarithm of the drainage area in square miles (Lang et al., 1999; see also Silva et al. (2012)). In this study, we will refer to the events exceeding the selected thresholds as “floods”, acknowledging that some of the events identified in this way will not lead to water outside the river banks. The examination of a different number of events allows us to evaluate the sensitivity of our results to different flood severity thresholds, to evaluate whether the drivers of those events tend to be the same or they change as we move from rarer to more frequent events.

In terms of predictors, we use precipitation, temperature, wetness conditions, agricultural land use, and population density (as a proxy for urbanization; e.g., DeWalle et al. 2000; Gluck and McCuen 1975). We hypothesize that these drivers explain the frequency of flood events, and that their relative importance is site-, threshold-, and/or season-specific. Precipitation (x_p) and temperature (x_T) are both derived from the PRISM Climate Group (PRISM Climate Group, 2017), which provides monthly climatic observations from 1895 to the present with a resolution of ~4km. For each basin we compute the basin-averaged seasonal temperature and accumulated precipitation, using the USGS Streamgauge NHDPlus Version 1 boundaries shapefiles (U.S. Geological Survey, 2006). We use the monthly precipitation and accumulate it also for the three months prior to the analyzed season (e.g., spring precipitation is used to represent the antecedent wetness conditions for the summer) as a simple way of capturing the water basin wetness conditions (x_M) (e.g., Kam and Sheffield 2016; Slater and Villarini 2016).

The fourth driver is the agricultural land cover (x_A), which reflects the effects of land cover on the rainfall-runoff relationship (Gupta et al., 2015; Schilling et al., 2008; Zhang and Schilling, 2006). We focus on the corn and soybean acreage combined because these two crops

represent the highest percentage of harvested acreage in the Midwest (U.S. Department of Agriculture, 2007). Here we use data from the National Agricultural Statistics Service website (U.S. Department of Agriculture, 2015) at a county level (Figure 2), and compute a basin-averaged value by measuring the area of the basin within each county and assuming that the corn and soybean acreage is uniformly distributed within each county (e.g., Slater and Villarini 2017; Villarini and Strong 2014). Similar to Slater and Villarini (2017), we include agricultural intensity as a predictor only for basins where at least 33% of the area is cultivated with corn and/or soybean during any year of the study period. The data have a time resolution of one year, so we used the same time series for all the four seasons.

The last covariate we consider is the population density (x_{pop}), which we use as a proxy of the urbanization of the water basin and assume to reflect municipal water use. Because there was no unique time series covering the entire period, we gathered population information from three different sources, similar to Slater and Villarini (2017): a 10-year time scale dataset was obtained from the “Population of Counties by Decennial Census: 1900 to 1990” (Forstall, 1995); an annual time series from 1970 to 2014 was obtained from the National Bureau of Economics Research (Roth, 2016); population data for 2015 and 2016 were obtained from the Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2016 (U.S. Census Bureau, 2016). We then merged these three datasets and fitted the time series with a locally weighted scatterplot smoothing (LOWESS) curve to obtain annual estimates. Because all three datasets provide population information at a county level (Figure 2), we computed basin-averaged population density in the same way as for agriculture (see also Slater and Villarini 2017). Given that the finest resolution was annual, we used the same value for each of the four seasons in a year.

Finally, we include population as a predictor only for the basins with a population density higher than 50 persons/km² in at least one year of the time series (Slater and Villarini, 2017).

2.2 Methods

Because of the discrete nature of the independent variable, we use Poisson regression to model the interannual variability in flood counts. Here we consider seven different regression models relating the response variable (i.e., flood counts) to the different predictors described in the previous section (see in Table 1). Recognizing that there may be sites with no flood events during a particular season (e.g., Northern Great Plains during the winter) and to be able to estimate the parameters of the Poisson model, we only model the seasonal records at the locations with at least five years where flood counts are greater than zero. Moreover, we exclude sites with negative β values (i.e., the parameter related to precipitation) as that would mean that we would expect fewer events for more precipitation.

As shown in Table 1, all the models we considered have seasonal accumulated precipitation as predictor, either in isolation (model P) or combined with other covariates. For instance, the $P.T$ and $P.M$ models also include the basin-averaged mean temperature (x_T) and wetness conditions (x_M), respectively. Models $P.PA$ and $P.Ppop$ consider agricultural land cover (x_A) and population density (x_{pop}) not as separate covariates (like for $P.T$ and $P.M$) but as an interaction factor, as we view them as important in modifying the relationship between precipitation and discharge (akin to the runoff coefficient) (Villarini and Strong, 2014). Model $P.PA.PM$ is similar to the previous ones, but it considers a total of three covariates: the total precipitation (x_P) as a separate predictor, and x_A and x_M as interaction terms. The last model is the *Mixed* model, which uses precipitation, antecedent wetness conditions and temperature as three separate covariates; this model uses two different temperature values according to the

season that is analyzed: during spring, the temperature covariate is the average temperature between March and April ($x_{T_{Mar-Apr}}$), as a simple way of representing snowmelt processes. In summer and fall, we include the average June-August temperature ($x_{T_{Summer}}$) as a way of accounting for evapotranspiration in the summer and drying soils in the fall. Overall, the selected sets of predictors and their rationale for being included here are consistent with the approach taken in Slater and Villarini (2017).

Rather than selecting a priori one of these seven models, we use the Bayesian Information Criterion (BIC) (Schwarz, 1978) and the Akaike Information Criterion (AIC) (Akaike, 1974) as a way of compromising goodness-of-fit and parsimony, with the former penalizing the inclusion of additional variables more than the latter for our sample sizes. For each station, threshold and season, we select the “best” model as the one with the lowest BIC or AIC value. In addition to a visual examination of all the time series to quantify how well our models perform (figures not shown), we compute the correlation coefficient R between the observations and the median of the fitted Poisson distribution. Furthermore, we evaluate the models’ performance via leave-one-out cross validation. At each station, and for each season and threshold value, the leave-one-out cross-validation is applied by estimating the parameters of the best model recursively by withholding one observation at the time, and predicting its value based on the value of the predictors for that season/year; we then compute the correlation coefficient between the observed time series and the predicted time series obtained this way.

3 Results and Discussion

Figure 3 (Figure 4) shows the location of the best-fitting models for each season and for each flood threshold (from one to four peaks per year on average) using BIC (AIC). As expected, less parsimonious models (e.g., *P.M* and *Mixed* models) are selected when using AIC rather

than BIC, even though the models selected are often the same regardless of the criterion (see Figure S2 of the supplemental material for more details). The two models that are most frequently selected are the *P* and *P.M* models, followed by the *Mixed* model. During the spring, the *P* model tends to be selected in areas where storm rainfall plays an important role, in particular in association with extratropical storms and atmospheric rivers (e.g., Nayak and Villarini 2017; Villarini 2016). Across the Northern Great Plains, on the other hand, the *P.M* and *Mixed* models are the most frequently selected, likely because of processes related to snowmelt and rain on frozen ground. During the summer, the model with only precipitation as covariate is the one most often selected (especially when BIC is used as selection criterion), while in the fall it is *P.M* model. Aside from seasonal differences, the choice of the threshold also plays a role in the identification of the best-fitting models. More specifically, as we decrease the flood threshold value from one to four events per year on average (moving from left to right in Figures 3-4), we notice a transition from the precipitation-only model as being the most frequently-selected model, to the *P.M* model. This transition suggests that the largest flood events tend to be driven by storm rainfall, and that the role of basin wetness increases as we deal with lower-magnitude but more frequent flood events (Slater and Villarini, 2017; Smith et al., 2013). Our results also indicate that the models that include population or agriculture are rarely selected, suggesting that the year-to-year changes in flood counts are more strongly related to climate than LULC drivers. This statement holds regardless of the threshold or season. Note that at some stations none of the seven models was selected either because the coefficient of the precipitation predictor was negative for all the seven models or because the observed time series had less than five years with flood counts greater than zero (Figure S1).

As an example of how well these models can reproduce the observational records, Figure 5 shows our modeling results for the USGS station 03230500 (Big Darby Creek at Darbyville, Ohio) for two different flood threshold. Albeit simple, our models can capture the interannual variability exhibited by the historical record very well, both at times when there is a lot of activity (e.g., summer and spring) as well as during quieter times (e.g., winter and, in particular, fall). While the results in Figure 5 refer to a single site, we produce the same plots for all the sites and for the 4 different flood threshold values (the figures are omitted to facilitate the reading of the paper): the good performance of our statistical models holds across the sites considered here, regardless of the season or threshold value.

To quantitatively evaluate the performance of our models across the study area, we show in Figure 6 the values of the correlation coefficient between observations and the median of the fitted Poisson distribution using the BIC to select the best models (Figure S3 of the supplemental material indicates the same results but referred to the best models selected according to the AIC). The average value of the correlation coefficient among all stations, seasons and threshold definitions is 0.58, and it increases when the threshold is lowered, ranging from an average value of 0.50 (one peak per year on average) to 0.63 (four peaks per year on average) among all seasons. Two main clusters with lower R values can be observed (especially for higher threshold values): one in the winter in the upper Mississippi and Missouri River Basins (across the states of Iowa, North Dakota and South Dakota), and a second one during spring and summer in the eastern part of the domain. Aside from these two clusters, the values of the correlation coefficient are high for the rest of the stations, as also highlighted by the visual examination of the time series. Figure 7 shows the differences in correlation coefficients when using BIC and AIC as selection criterion. Overall, the differences are very small, with a global average

difference equal to -0.02: this indicates the models based on AIC tend to exhibit a small improvement over those selected based on BIC, even though this limited improvement comes at the expenses of adding additional predictors, leading to more complex models. To further assess the performance of the models, we performed a leave-one-out cross-validation, which confirms the good prediction skills of these statistical models (Figure 8; the same results obtained using the AIC are presented in Figure S4 of the supplemental material). As shown by these results, the models exhibit good skills, especially during the flood active periods like spring and summer (Mallakpour and Villarini, 2017; Villarini, 2016), highlighting the potential applicability of these models for flood forecasting (see also Slater and Villarini (2018)).

Figure 9 stratifies the models' performance across the different sets of predictors when the BIC is used to select the best models (similar results are obtained when using the AIC to select the best models, as shown in Figure S5). Model P is very frequently selected, and the model fit improves as we move from rarer to more frequent events, with values of the correlation coefficient increasing from ~ 0.4 to ~ 0.6 . The $P.T$ model is selected only at a small number of stations, with a better fit during spring. For all seasons, the values of the coefficient related to the x_T predictor (γ_2 , Table 1) are negative, indicating that the number of flood events increases (decreases) with a decrease (increase) in temperature. One possible explanation for the improved model fit in spring is related to rain-on-snow events, where the colder temperatures can affect the duration of the snow cover and reduce infiltration, thus increasing the probability of higher discharge values during heavy rainfall events (Pradhanang et al., 2013). During summer and fall, instead, lower temperatures can result in lower evapotranspiration, maintaining high wetness conditions and thus facilitating the occurrence of higher discharges (Berghuijs et al., 2016).

Model $P.M$ is the second most widespread model, and is selected more frequently as we consider lower flood thresholds (i.e., higher number of annual exceedances). Compared to model P , the values of the correlation coefficient are slightly higher (Figure 9 and Figure S5). The high number of stations where $P.M$ was selected as the best-fitting model and the value of the average correlation coefficient across all stations, seasons and threshold values ($R = 0.63$ when BIC is used as the selection criterion) confirm the importance of the wetness conditions in predicting high flow discharge values (Ivancic and Shaw, 2015; Slater and Villarini, 2017; Villarini and Slater, 2017) and flood counts. Models $P.PA$, $P.Ppop$, and $P.PA.PM$ are rarely selected, suggesting that the covariates x_A and x_{pop} do not add much to the precipitation-related information, likely because LULC changes mostly affect low and medium discharges (Slater and Villarini, 2017).

The *Mixed* model performs quite well during spring in the northern part of the study region (Figures 3-4), where both antecedent wetness and temperature are important for snow-related processes. Also in this case, the relationship between temperature and number of events is negative, possibly due to rain-on-snow events and rain on frozen ground. Compared to the $P.T$ model, where only precipitation and temperature are included as predictors, the *Mixed* model performs better in the spring, indicating that accounting for the snow that fell during the winter can help improving the model performance, with an average correlation coefficient among all threshold values of 0.62 (against 0.59 for the $P.T$ model) when the BIC is used to select the best models (Figure 9). During fall (where x_T is the average summer temperature) the behavior is similar: the model is selected in more stations compared to the $P.T$ model, suggesting that the evapotranspiration and its impact in the water stored during the summer plays a relevant role (Slater and Villarini, 2016; Villarini and Slater, 2017). Overall, the values of the correlation

coefficients are large (Figure 9 and Figure S5). During summer, where the predictors are precipitation, wetness conditions and average summer temperature, the model was selected in fewer stations than model $P.T$, suggesting that the penalization of considering a third predictor (wetness conditions x_M) is higher than the improvement of the goodness-of-fit that it provides. Similar results are obtained when using AIC to select the best models.

To quantify the improvement in model fit due to the addition of a predictor to model P , we computed, for each season, flood threshold and for every station where P model was not selected as the best model, the difference between the R value of the best model and the R value of the P model, considering both BIC and AIC as selection criterion. As shown in Figure 10, the addition of other predictors improves the overall performance. The biggest improvements are found during spring and fall, where temperature and wetness conditions are important predictors capturing the processes responsible for the large discharge values and therefore flood events. During winter and summer, on the other hand, the improvement is overall more muted. Moreover, the average improvement is higher when considering AIC, given that it tends to select more complex models. These statements are valid regardless of the selected threshold.

At the beginning of this study, we showed the trends in the frequency of flood events across the U.S. Midwest (Figure 1), and then developed a methodology that allows the attribution of the interannual variability of flood events frequency in terms of simple predictors. To further evaluate the performance of our modeling framework, we circle back to the results in Figure 1, and see if they can be explained in terms of the drivers we have identified. Results indicate that the observed trends in Figure 1 are very well reproduced by our simple models (Figure 11). Using the results from the fitted distribution rather than the observations, we find the same trends for most of the stations in our domain, from Ohio to Missouri, to the Dakotas. The

areas where we observe the largest discrepancies are Kansas and Nebraska, where groundwater abstraction and the construction of ponds and terraces have led to declining water tables (e.g., Rasmussen and Perry 2001).

4 Conclusions and Future Directions

In this work we modeled the interannual variability in the frequency of flood events at 287 stream gage stations across the central United States. Analyses were performed at the seasonal scale and for different threshold values. The main findings of the study can be summarized as follows:

- We have developed seven simple Poisson regression models that used a combination of five predictors related to climatic conditions (i.e., precipitation, antecedent wetness, and temperature) and LULC properties (i.e., urbanization and agricultural intensity). Despite the simplicity of these statistical models, they were able to describe very well the interannual variability in the flood counts across different flood threshold values and seasons. The models' performance was evaluated through visual examination of the time series and by computing the correlation coefficient between the observational record and the median of the fitted Poisson distribution. The goodness-of-fit of our models was further substantiated through leave-one-out cross validation.
- The model with precipitation as the only predictor was the most frequently selected, especially during spring and summer and for rarer flood events (i.e., higher flood threshold values). The second most important covariate was the antecedent wetness, which influences the frequency of flood counts either in terms of snow-related processes during winter and spring, or in terms of evapotranspiration during summer

and fall; moreover, we found that this predictor played an increasing role as we considered more frequent events (i.e., lower threshold values). Temperature constitutes a good predictor only when combined with antecedent wetness conditions. LULC drivers (i.e., agricultural intensity and urbanization) are rarely selected, suggesting that much of the interannual variability in flood counts is driven by climate processes. Therefore, based on these modeling results, we have found that a very large component of the detected increasing trends across this area can be largely attributed to changes in precipitation and wetness conditions.

- The framework developed in this study could be improved by considering other predictors that would be directly related to the processes we are trying to capture through proxies (e.g., evapotranspiration or solid precipitation). One of the advantages of using only temperature and precipitation is that these quantities are readily available from global climate models, potentially extending the applicability of these models for flood predictions and projections.

Acknowledgements: This study was supported in part by the Broad Agency Announcement (BAA) Program and the Engineer Research and Development Center (ERDC)–Cold Regions Research and Engineering Laboratory (CRREL) under Contract No. W913E5-16-C-0002, and by the National Science Foundation under CAREER Grant AGS-1349827. The comments by two anonymous reviewers are gratefully acknowledged.

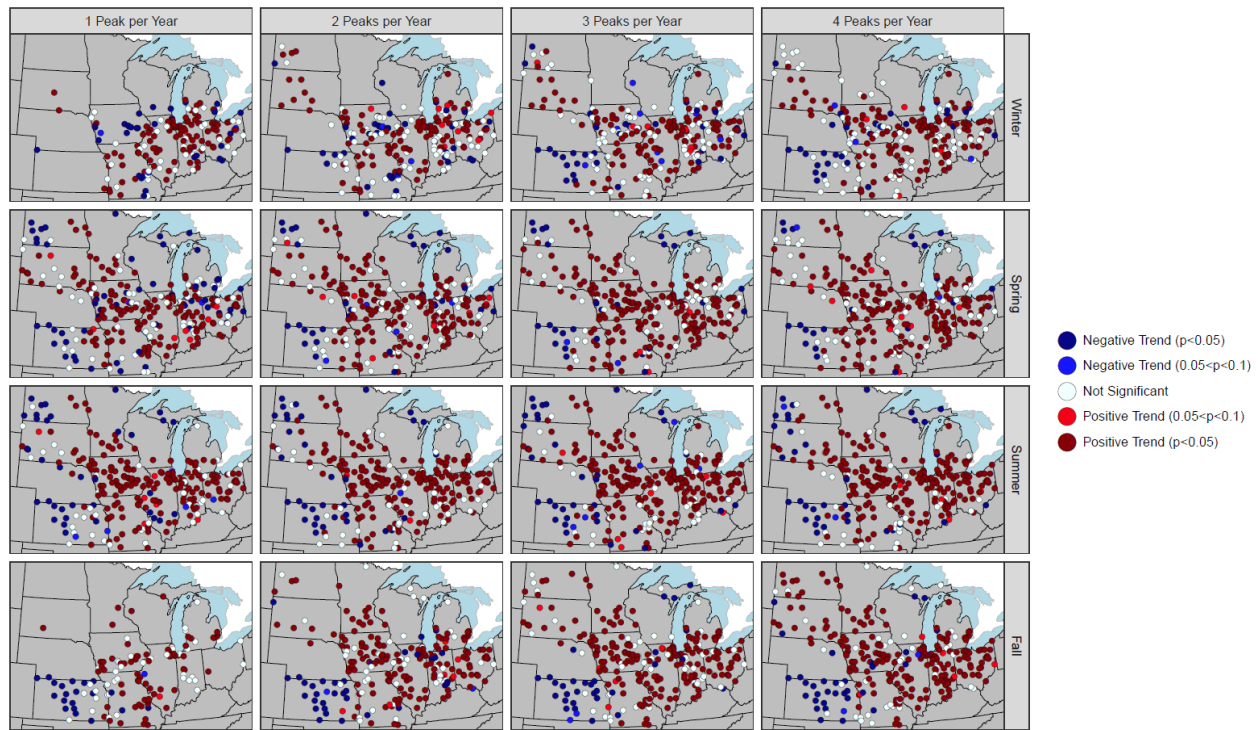


Figure 1 – Maps showing trends in flood counts (obtained using a peak-over-threshold approach; see Section 2) from 1940 to 2016 for 287 stream gage stations for each season (rows). Moving from the left to the right column, the threshold used to identify the flood events is lowered (see Section 2 for more details). The different color shades refer to the statistical significance of the results (dark blue/red: 5%; blue/red: 10%). The white circles indicate sites with trends not significant at the 10% level. The trends are computed using Poisson regression with time as the predictor. The trends are calculated only at those stations that have at least five years with at least one event.

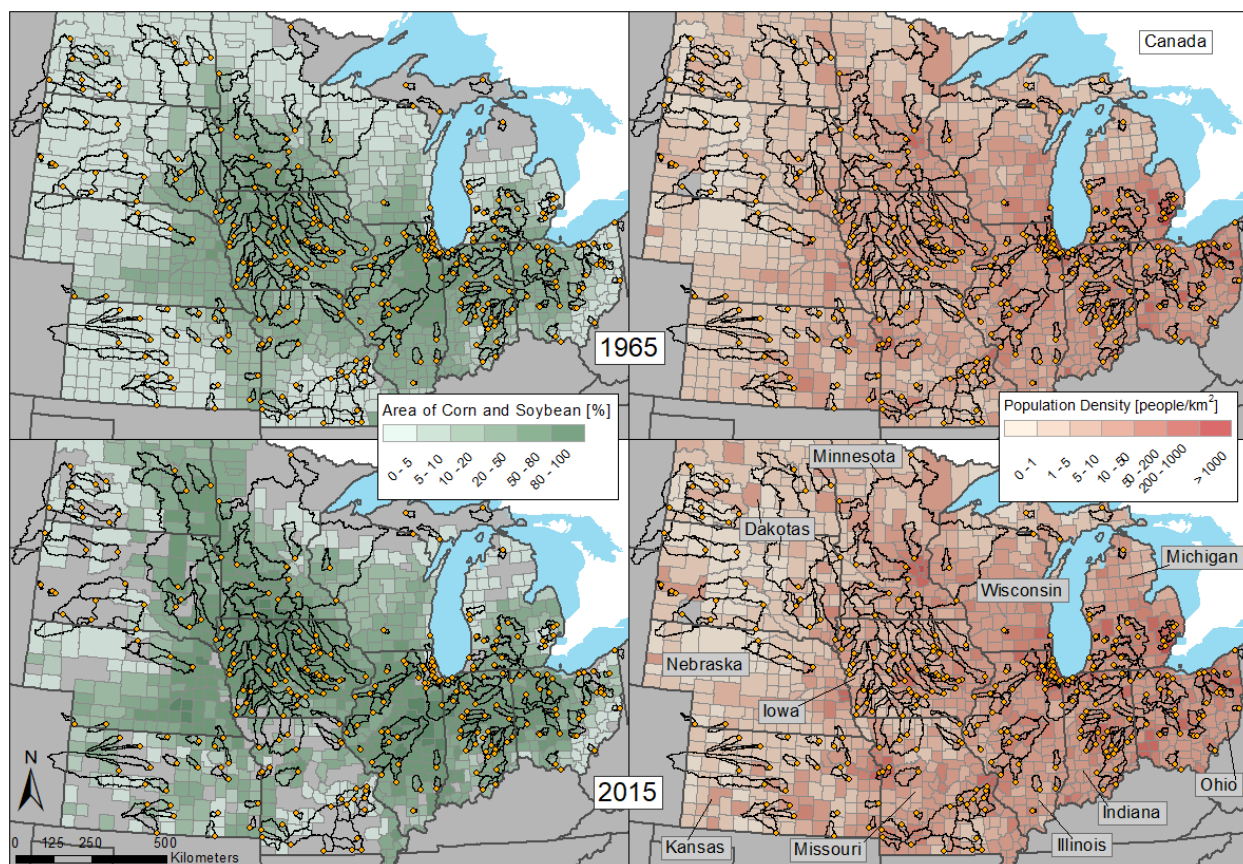


Figure 2 – Maps showing the location of the 287 stream gages considered in this study (orange circles; the black outlines represent the basin boundaries), the percentage of the counties used for corn and soybean production (left panels) and the population density in each county (right panels). The maps in the top row refer to the conditions in 1965, while those in the bottom panels to 2015.

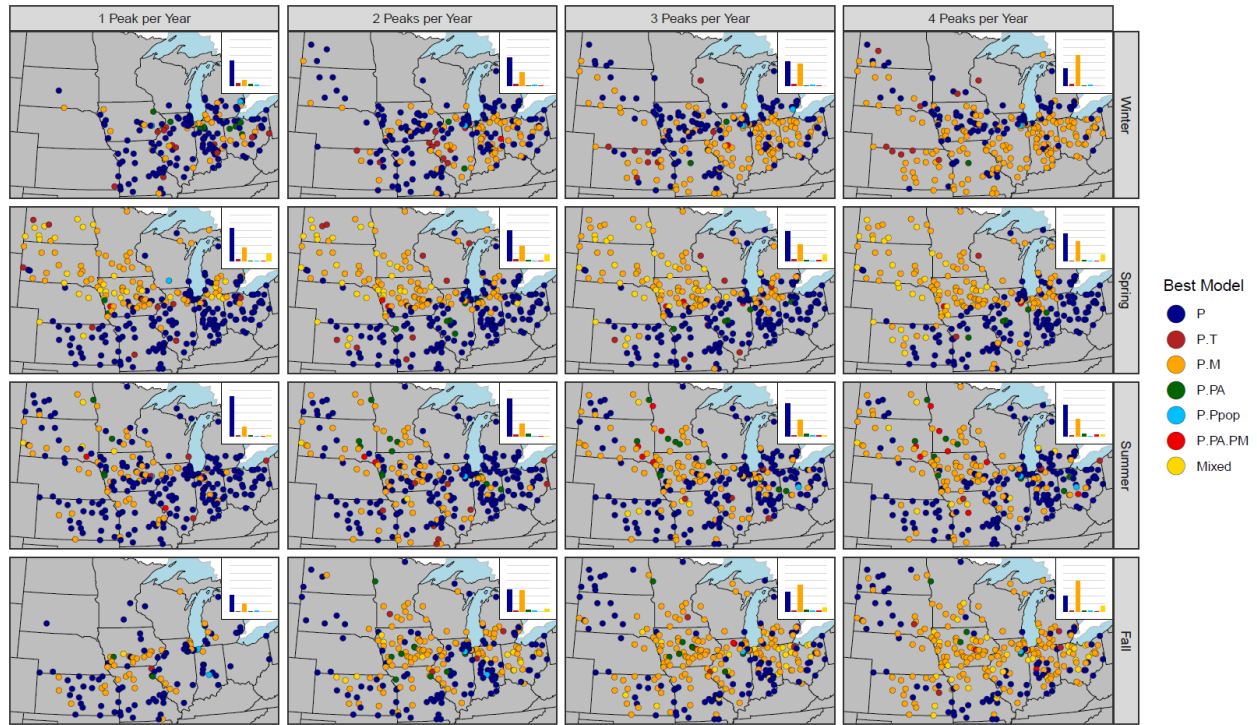


Figure 3 – Location of the best fitting models (dark blue: P , dark red: $P.T$, orange: $P.M$, dark green: $P.PA$, light blue: $P.Ppop$, red: $P.PA.PM$, yellow: $Mixed$) using BIC as selection criterion. Each row indicates a different season, while each column indicates different threshold values (from higher to lower moving from left to right). On the top right corner of each panel there is an indicative bar plot that shows the number of times each model has been chosen. Note that some stations do not have a best model, either because none of the seven candidate models have a positive coefficient for precipitation or because there are less than five years with non-zero events (see Figure S1 of the supplemental material for more details).

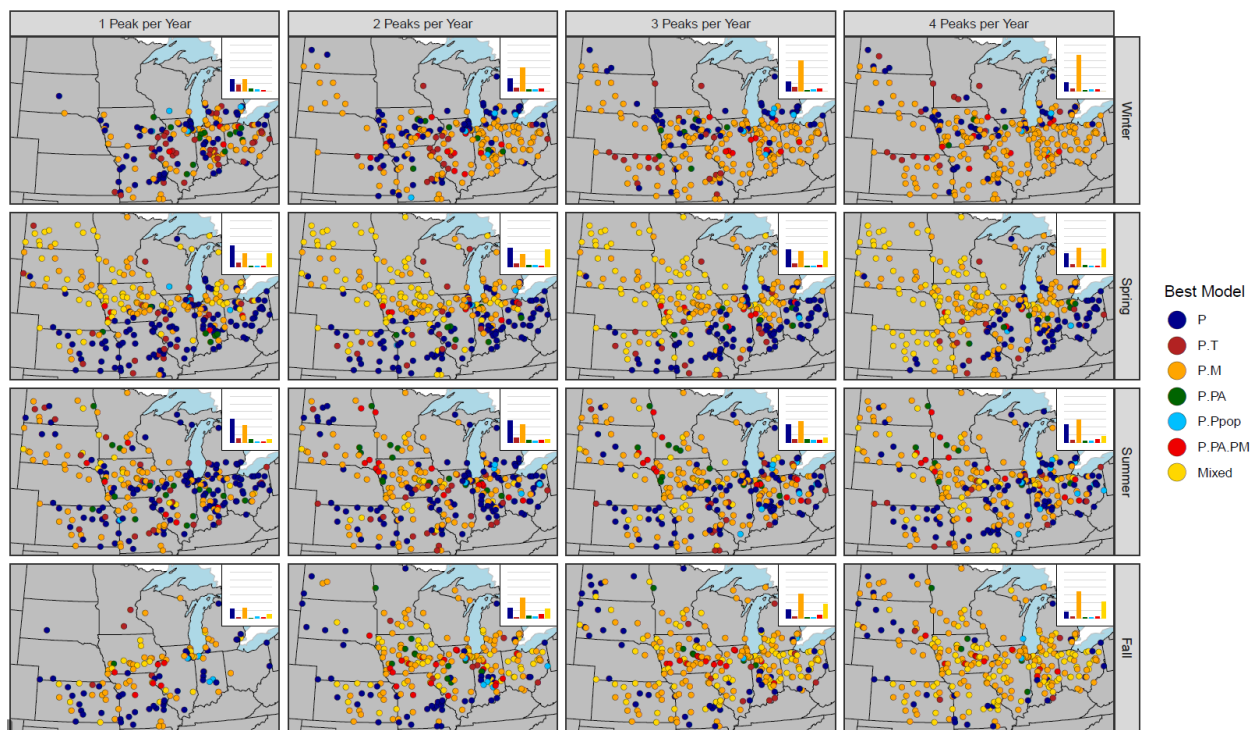


Figure 4 – Same as Figure 3, but using the AIC as selection criterion.

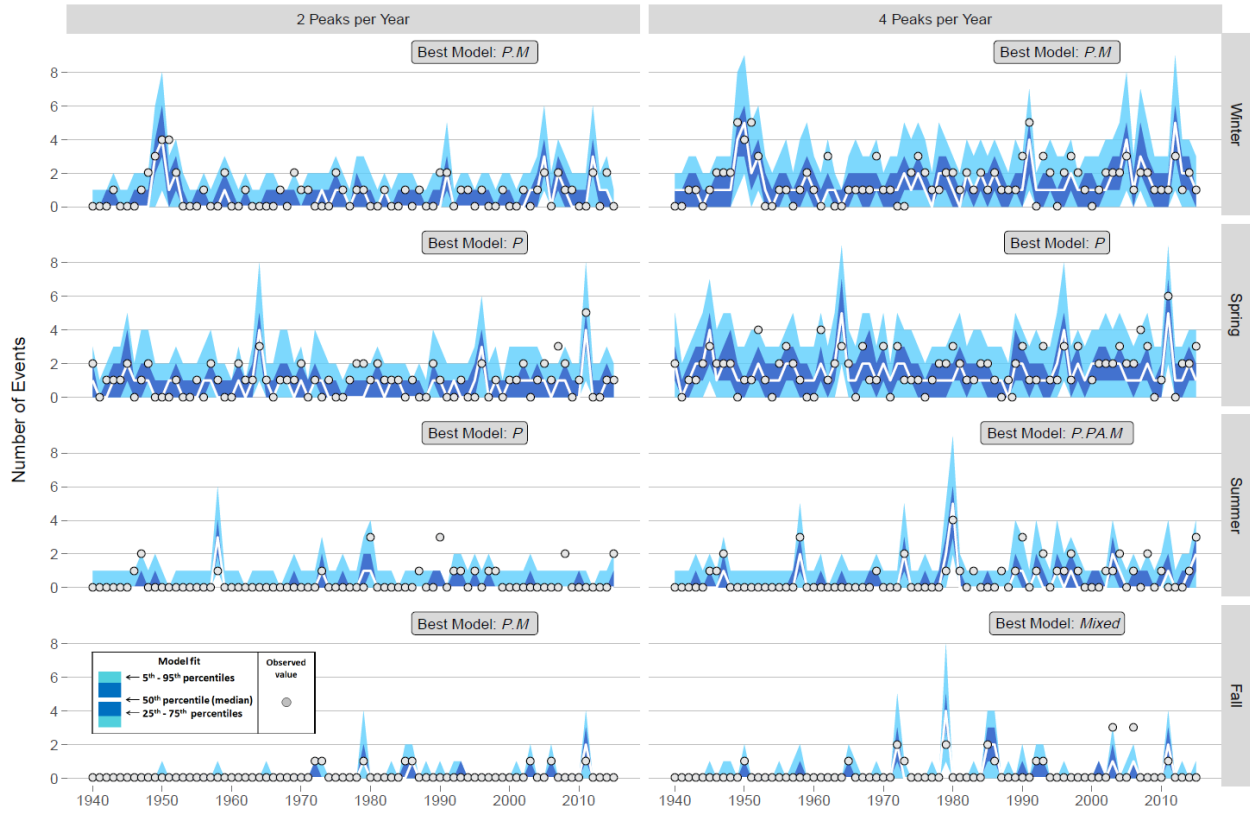


Figure 5 – Observed and modeled time series for the USGS station 03230500 (Big Darby Creek at Darbyville, Ohio) for each season and two threshold values (left and right columns indicate 2 or 4 peaks per year on average, respectively) and using BIC as criterion to select the best models (indicated in each time series in a gray text box). The gray circles represent the observations. The white line is the 50th percentile (median) of the fitted Poisson distribution; the blue and light blue ribbons represent the 25th – 75th and 5th – 95th percentile ranges, respectively.

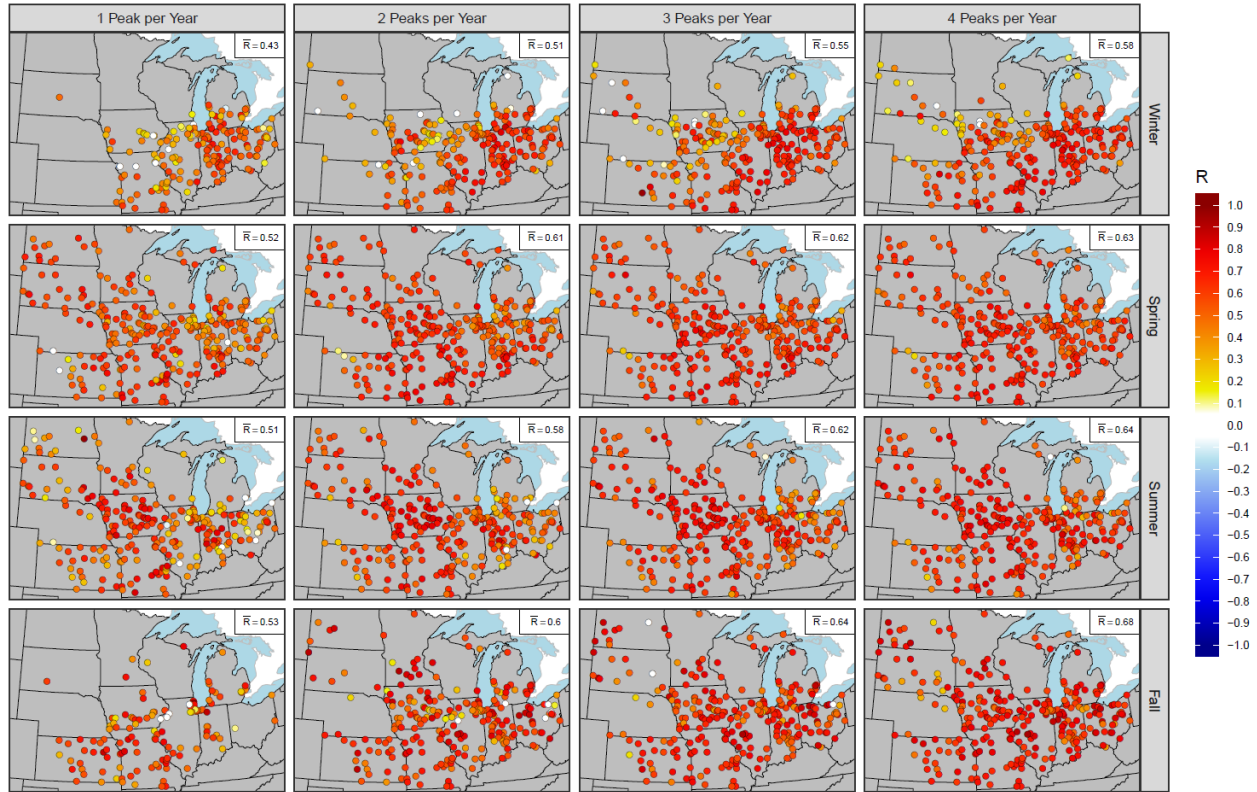


Figure 6 – Map of the Pearson correlation coefficient between the observations and the median of the fitted best models based on BIC (Figure S3 shows the same results but with respect to AIC), for every season and flood threshold value. On the top right corner of each panel, the average value of the correlation coefficient R is shown.

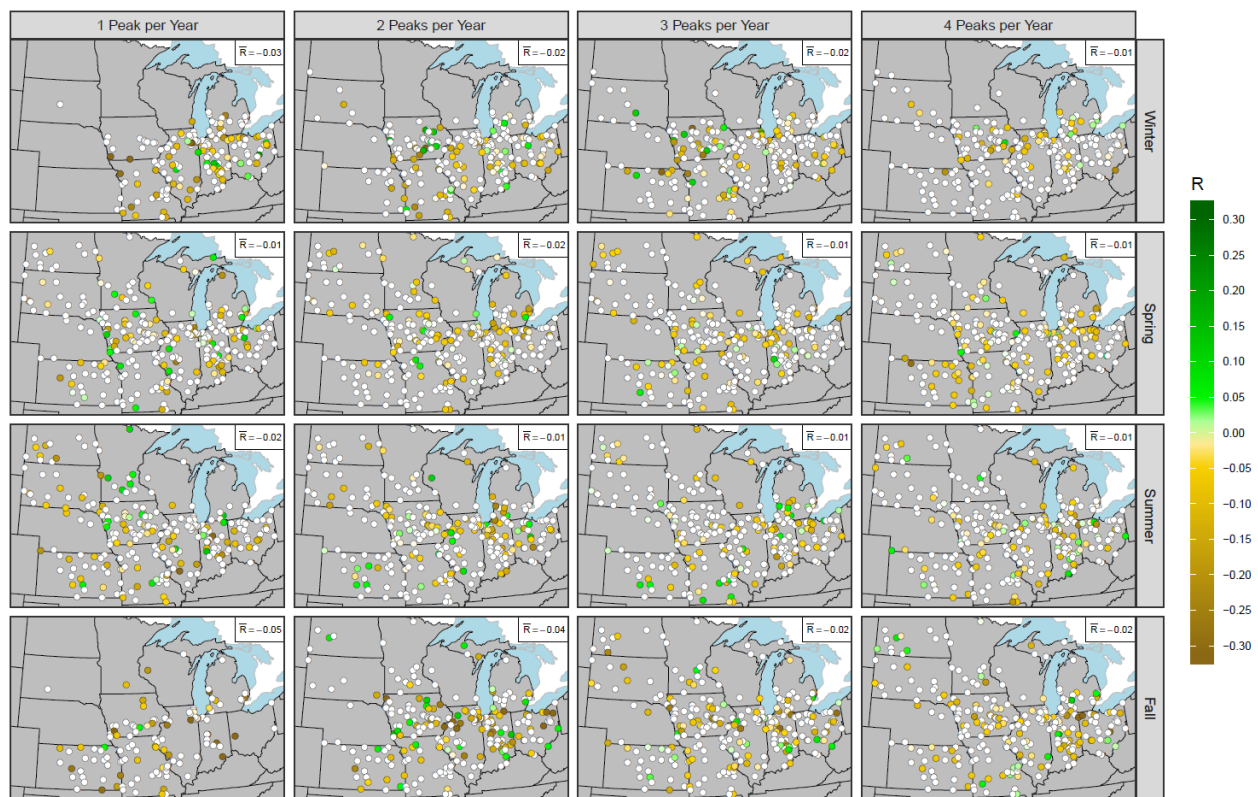


Figure 7 – Map showing the differences between the correlation coefficient of the best models selected using BIC and AIC, for each season and flood threshold value. The white circles represent a difference equal to zero (i.e., when the AIC and BIC select the same model). On the top right corner of each panel, the average value of the difference of correlation coefficients is shown.

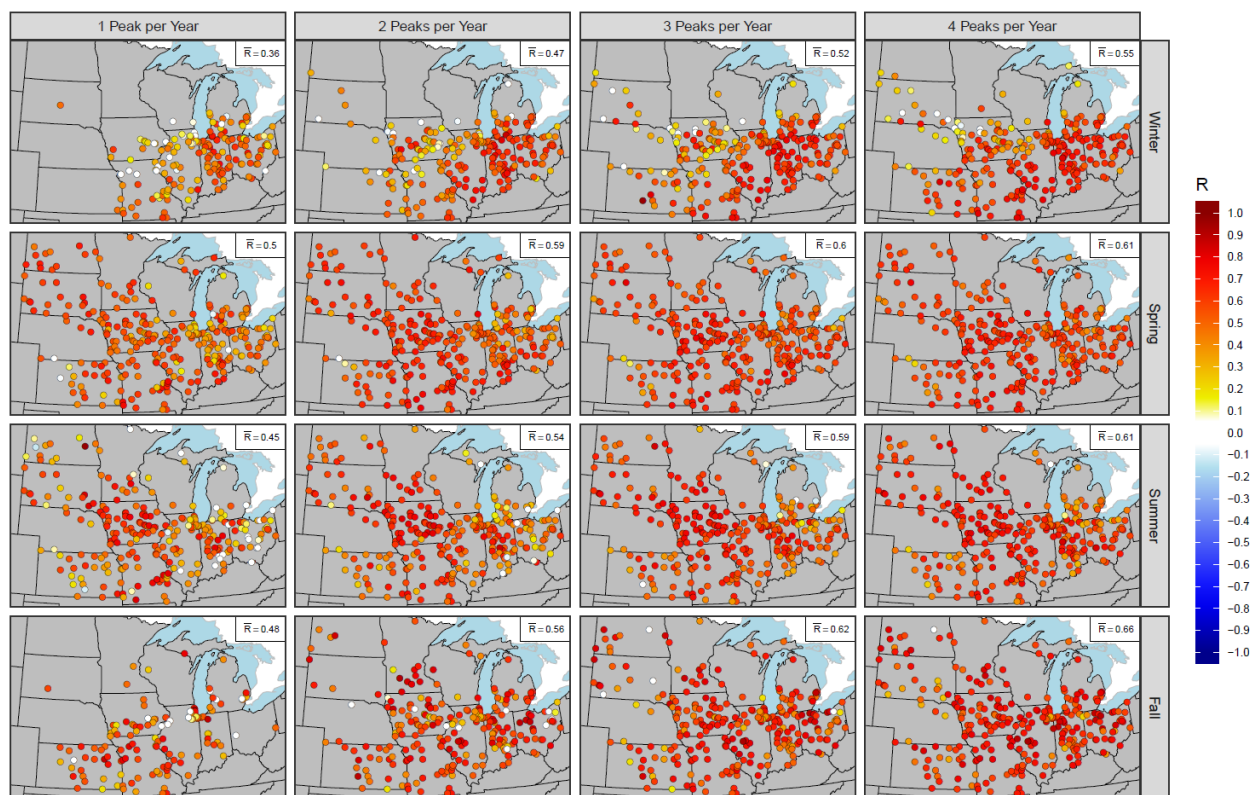


Figure 8 – Same as Figure 6 but for the leave-one-out cross validation. Figure S4 shows the same results but using AIC to select the best models.

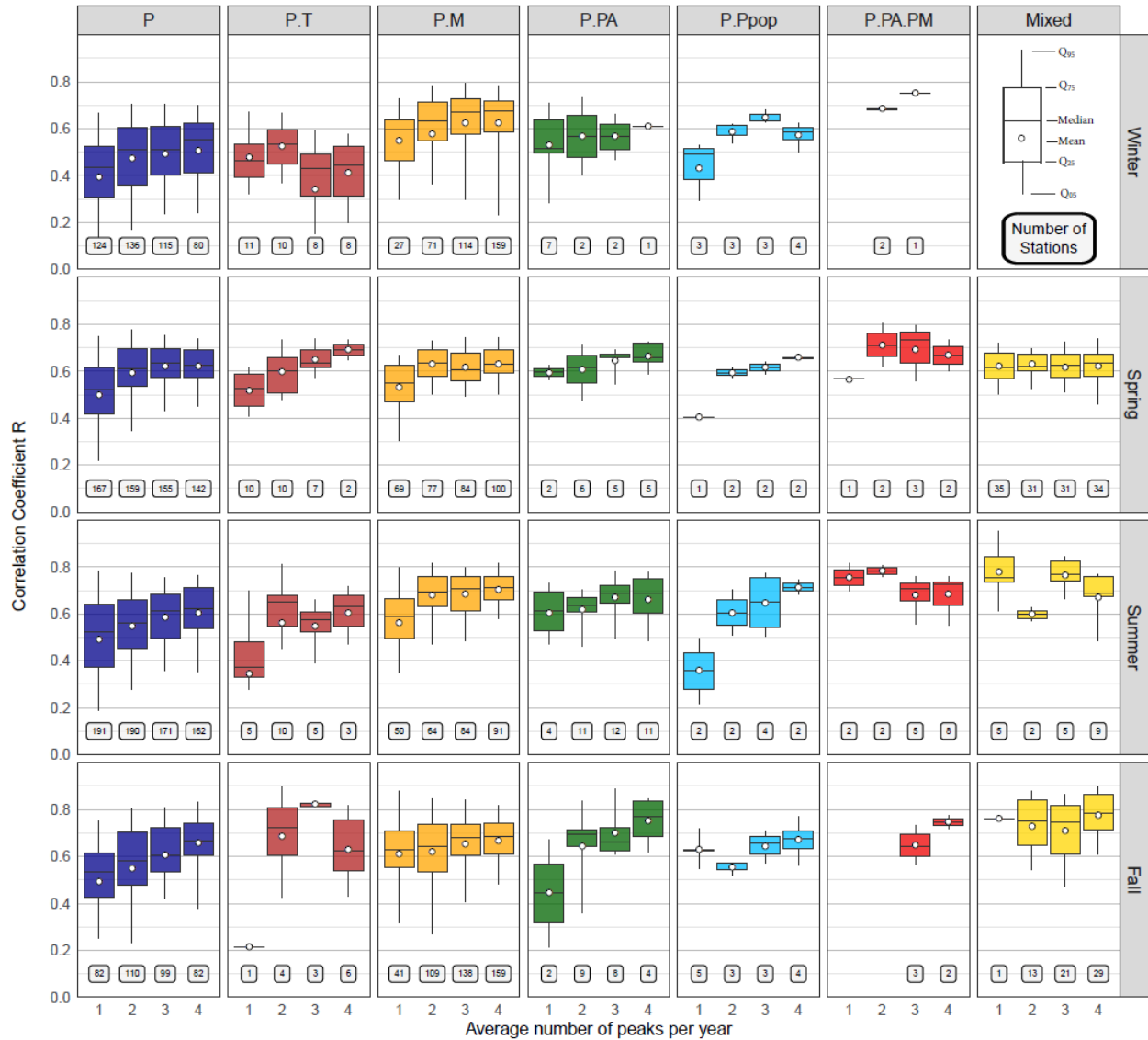


Figure 9 – Boxplots of the values of the Pearson correlation coefficient for the best models (selected according to BIC), for each season and for the four flood threshold values. The white circle represents the mean value of the R coefficients; the horizontal line is the median (50th percentile); the edges of the boxes are the 25th and 75th percentiles; the limits of the whiskers indicate the 5th and 95th percentiles. Each column represents a model, and each row a season. Within each panel, there are up to four boxplots, one for each flood threshold value. The numbers at the bottom of each boxplot represent the number of sites used to construct them. Note that the mixed model is not selected during winter. The same results obtained by selecting the best models using AIC are presented in Figure S5.

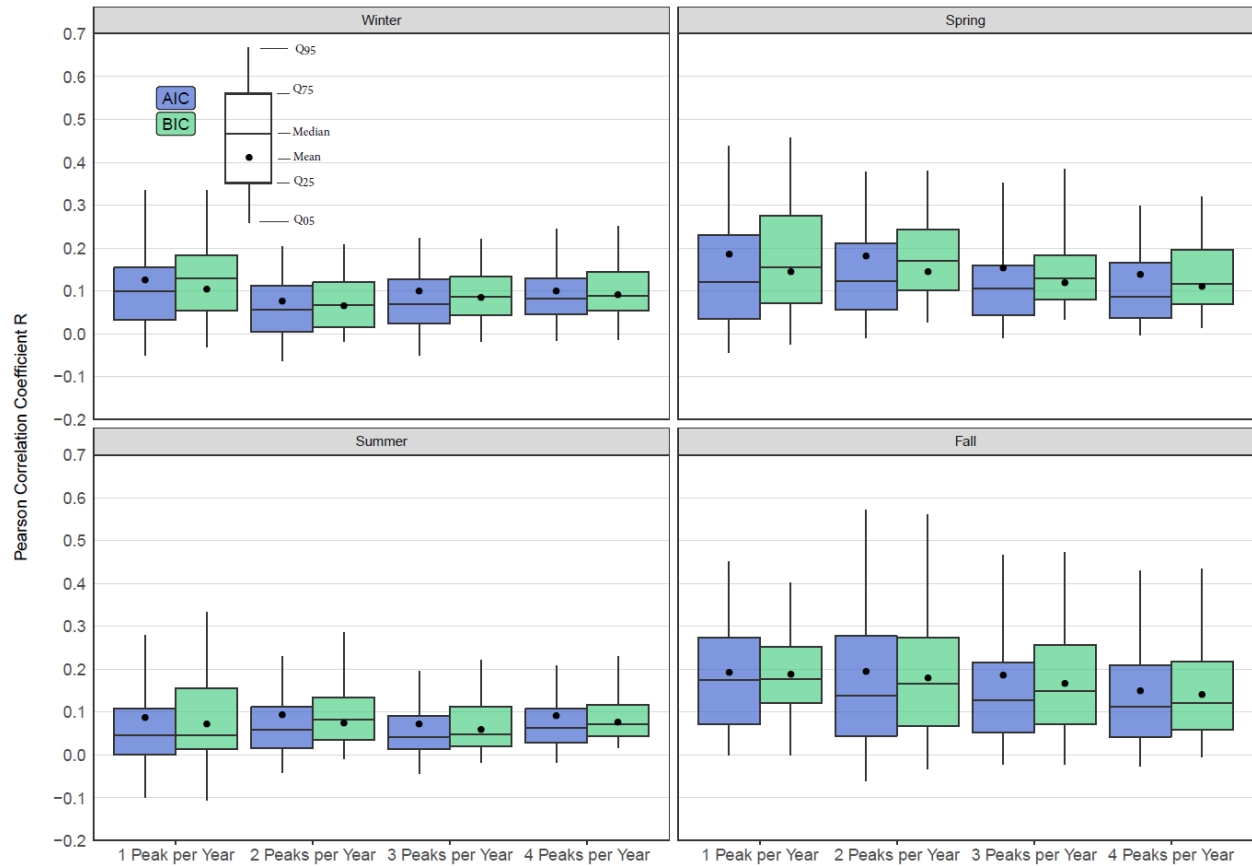


Figure 10 – Improved model fit associated with the addition of one or more predictors to the baseline model (model *P*) according to AIC (violet boxplots) and BIC (green boxplots). Each panel is for a given season, with four pairs of boxplots each (one per flood threshold value). The boxplot is constructed from the differences between the correlation coefficient of model *P* (for the stations where model *P* was not the best-fitting model), and the correlation coefficient of the selected best models for the same stations.

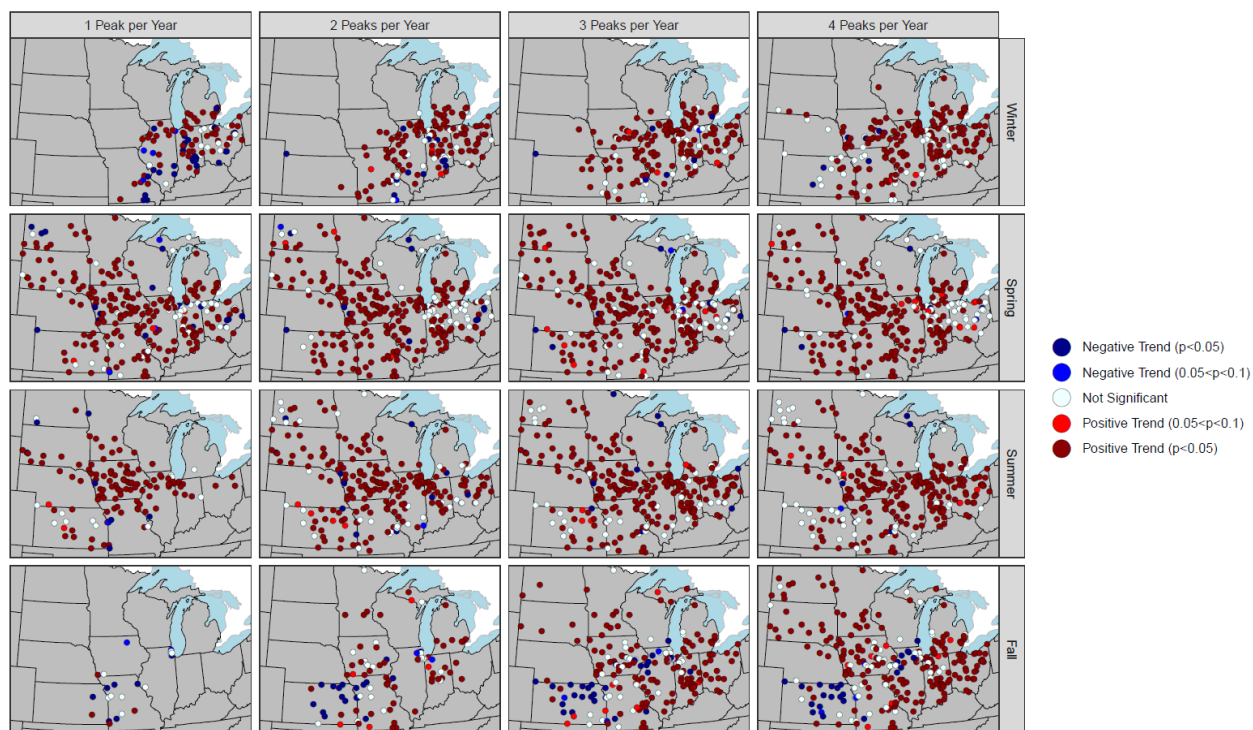


Figure 11 – Maps showing trends in flood counts from 1940 to 2016 for 287 stream gage stations for each season. In contrast to Figure 1, here the trends are computed with respect to the median of the fitted Poisson distribution, and not the observations, using BIC to select the best models (see Figure S6 for the same results using AIC). Moving from left to right, the threshold used to identify the flood events is lowered (see Section 2 for more details). The different color shades indicate the statistical significance of the results (dark blue/red: 5%; blue/red: 10%). The white circles are for sites with trends not significant at the 10% level. The trends are calculated only at those stations that have at least five years of non-zero events.

Model Name	Model Hypothesis	Model Predictors
<i>P</i>	$\log \lambda_1 = \alpha_1 + \beta_1 x_P$	x_P - Precipitation
<i>P.T</i>	$\log \lambda_2 = \alpha_2 + \beta_2 x_P + \gamma_2 x_T$	x_P - Precipitation x_T - Temperature
<i>P.M</i>	$\log \lambda_3 = \alpha_3 + \beta_3 x_P + \gamma_3 x_M$	x_P - Precipitation x_M - Wetness Conditions
<i>P.PA</i>	$\log \lambda_4 = \alpha_4 + \beta_4 x_P + \gamma_4 x_P x_A$	x_P - Precipitation x_A - Agricultural Cover
<i>P.Ppop</i>	$\log \lambda_5 = \alpha_5 + \beta_5 x_P + \gamma_5 x_P x_{pop}$	x_P - Precipitation x_{pop} - Population Density
<i>P.PA.PM</i>	$\log \lambda_6 = \alpha_6 + \beta_6 x_P + \gamma_6 x_P x_A + \delta_6 x_P x_M$	x_P - Precipitation x_A - Agricultural Cover x_M - Wetness Conditions
<i>Mixed</i>	$\log \lambda_7 = \alpha_7 + \beta_7 x_P + \gamma_7 x_M + \delta_7 x_{T_{Mar-Apr}}$ $\log \lambda_7 = \alpha_7 + \beta_7 x_P + \gamma_7 x_M + \delta_7 x_{T_{Summer}}$	x_P - Precipitation x_M - Wetness Conditions x_T - Temperature

Table 1 – List of the seven statistical models used to relate the seasonal occurrence of flood events to the five covariates. The first column refers to the acronym used throughout the paper to refer to a specific model. The second column refers to the relationship between the parameter λ of the Poisson distribution and the predictor(s). The right column summarizes the predictors used for each of the models.

Bibliography

- Akaike, H., 1974. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.* 19, 716–723. <https://doi.org/doi:10.1109/TAC.1974.1100705>, MR 0423716
- Berghuijs, W.R., Woods, R.A., Hutton, C.J., Sivapalan, M., 2016. Dominant flood generating mechanisms across the United States. *Geophys. Res. Lett.* 43, 4382–4390. <https://doi.org/10.1002/2016GL068070>
- Centre for Research on the Epidemiology of Disasters, 2017. Total economic damage caused by natural disasters in the U.S. 1900–2016 [WWW Document]. URL <https://www.statista.com/statistics/236530/economic-damage-caused-by-natural-disasters-in-the-us/> (accessed 1.20.18).
- DeWalle, D.R., Swistock, B.R., Johnson, T.E., McGuire, K.J., 2000. Potential effects of climate change and urbanization on mean annual streamflow in the United States. *Water Resour. Res.* 36, 2655–2664. <https://doi.org/10.1029/2000WR900134>
- Forstall, R.L., 1995. Population of Counties by Decennial Census: 1900 to 1990 [WWW Document]. URL <https://www.census.gov>
- Frans, C., Istanbuluoglu, E., Mishra, V., Munoz-Arriola, F., Lettenmaier, D.P., 2013. Are climatic or land cover changes the dominant cause of runoff trends in the Upper Mississippi River Basin? *Geophys. Res. Lett.* 40, 1104–1110. <https://doi.org/10.1002/grl.50262>
- Gluck, W.R., McCuen, R.H., 1975. Estimating land use characteristics for hydrologic models. *Water Resour. Res.* 11, 177–179. <https://doi.org/10.1029/WR011i001p00177>
- Gupta, S.C., Kessler, A.C., Brown, M.K., Zvomuya, F., 2015. Climate and agricultural land use change impacts on streamflow in the upper midwestern United States. *Water Resour. Res.* 51, 5301–5317. <https://doi.org/10.1002/2015WR017323>
- Hirsch, R.M., Archfield, S.A., 2015. Flood trends: Not higher but more often. *Nat. Clim. Chang.* 5, 198–199. <https://doi.org/10.1038/nclimate2551>
- 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. *Clim. Change* 133, 681–693. <https://doi.org/10.1007/s10584-015-1476-1>
- Kam, J., Sheffield, J., 2016. Changes in the low flow regime over the eastern United States (1962–2011): variability, trends, and attributions. *Clim. Change* 135, 639–653. <https://doi.org/10.1007/s10584-015-1574-0>
- Lang, M., Ouarda, T.B.M.J., Bobée, B., 1999. Towards operational guidelines for over-threshold modeling. *J. Hydrol.* 225, 103–117. [https://doi.org/10.1016/S0022-1694\(99\)00167-5](https://doi.org/10.1016/S0022-1694(99)00167-5)
- Mallakpour, I., Villarini, G., 2017. Analysis of changes in the magnitude, frequency, and seasonality of heavy precipitation over the contiguous USA. *Theor. Appl. Climatol.* 130, 345–363. <https://doi.org/10.1007/s00704-016-1881-z>

488 Mallakpour, I., Villarini, G., 2015. The changing nature of flooding across the central United
489 States. *Nat. Clim. Chang.* 5, 250–254. <https://doi.org/10.1038/nclimate2516>

490 Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., Hundedcha, Y., 2012. HESS Opinions:
491 “More efforts and scientific rigour are needed to attribute trends in flood time series.”
492 *Hydrol. Earth Syst. Sci.* 16, 1379–1387. <https://doi.org/10.5194/hess-16-1379-2012>

493 Mishra, V., Cherkauer, K.A., Niyogi, D., Lei, M., Pijanowski, B.C., Ray, D.K., Bowling, L.C.,
494 Yang, G., 2010. A regional scale assessment of land use/land cover and climatic changes on
495 water and energy cycle in the upper Midwest United States. *Int. J. Climatol.* 30, 2025–2044.
496 <https://doi.org/10.1002/joc.2095>

497 Nayak, M.A., Villarini, G., 2017. A long-term perspective of the hydroclimatological impacts of
498 atmospheric rivers over the central United States. *Water Resour. Res.* 53, 1144–1166.
499 <https://doi.org/doi:10.1002/2016WR019033>

500 NOAA National Centers for Environmental Information (NCEI), 2018. Billion-Dollar Weather
501 and Climate Disasters: Table of Events [WWW Document]. URL
502 <https://www.ncdc.noaa.gov/billions/events/US/1980-2018> (accessed 1.20.18).

503 Pradhanang, S.M., Frei, A., Zion, M., Schneiderman, E.M., Steenhuis, T.S., Pierson, D., 2013.
504 Rain-on-snow runoff events in New York. *Hydrol. Process.* 27, 3035–3049.
505 <https://doi.org/10.1002/hyp.9864>

506 PRISM Climate Group, 2017. PRISM Climate Data [WWW Document]. URL
507 <http://www.prism.oregonstate.edu/> (accessed 12.20.17).

508 Radeloff, V.C., Hammer, R.B., Stewart, S.I., 2005. Rural and Suburban Sprawl in the U.S.
509 Midwest from 1940 to 2000 and Its Relation to Forest Fragmentation. *Conserv. Biol.* 19,
510 793–805. <https://doi.org/10.1111/j.1523-1739.2005.00387.x>

511 Rasmussen, T.J., Perry, C.A., 2001. Trends in peak flows of selected streams in Kansas, Water-
512 Resources Investigations Report. <https://doi.org/https://doi.org/10.3133/wri014203>

513 Roth, J., 2016. Census U.S. Intercensal County Population Data, 1970-2014 [WWW Document].
514 URL http://www.nber.org/data/census_popest.html (accessed 12.15.17).

515 Schilling, K.E., Jha, M.K., Zhang, Y.-K., Gassman, P.W., Wolter, C.F., 2008. Impact of land use
516 and land cover change on the water balance of a large agricultural watershed: Historical
517 effects and future directions. *Water Resour. Res.* 44.
518 <https://doi.org/10.1029/2007WR006644>

519 Schottler, S.P., Ulrich, J., Belmont, P., Moore, R., Lauer, J.W., Engstrom, D.R., Almendinger,
520 J.E., 2014. Twentieth century agricultural drainage creates more erosive rivers. *Hydrol.*
521 *Process.* 28, 1951–1961. <https://doi.org/10.1002/hyp.9738>

522 Schwarz, G., 1978. Estimating the Dimension of a Model. *Ann. Stat.* 6, 461–464.
523 <https://doi.org/10.1214/09-AOS712>

524 Silva, A.T., Portela, M.M., Naghettini, M., 2012. Nonstationarities in the occurrence rates of
525 flood events in Portuguese watersheds. *Hydrol. Earth Syst. Sci.* 16, 241–254.
526 <https://doi.org/10.5194/hess-16-241-2012>

Slater, L., Villarini, G., 2017. Evaluating the Drivers of Seasonal Streamflow in the U.S. Midwest. *Water* 9, 695. <https://doi.org/10.3390/w9090695>

Slater, L.J., Villarini, G., 2018. Enhancing the Predictability of Seasonal Stream flow With a Statistical-Dynamical Approach. *Geophys. Res. Lett.* 45, 6504–6513. <https://doi.org/10.1029/2018GL077945>

Slater, L.J., Villarini, G., 2016. Recent trends in U.S. flood risk. *Geophys. Res. Lett.* 43, 12,428–12,436. <https://doi.org/10.1002/2016GL071199>

Smith, J.A., Baeck, M.L., Villarini, G., Wright, D.B., Krajewski, W., 2013. Extreme Flood Response: The June 2008 Flooding in Iowa. *J. Hydrometeorol.* 14, 1810–1825. <https://doi.org/10.1175/JHM-D-12-0191.1>

Tang, C., Crosby, B.T., Wheaton, J.M., Piechota, T.C., 2012. Assessing streamflow sensitivity to temperature increases in the Salmon River Basin, Idaho. *Glob. Planet. Change* 88–89, 32–44. <https://doi.org/10.1016/J.GLOPLACHA.2012.03.002>

U.S. Census Bureau, 2016. Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2016 [WWW Document]. URL <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk> (accessed 12.20.17).

U.S. Department of Agriculture, 2015. Spatial Analysis Research Section [WWW Document]. URL <ftp://ftp.nass.usda.gov/quickstats/> (accessed 12.20.17).

U.S. Department of Agriculture, 2007. Ag Atlas Maps—Crops and Plants [WWW Document]. URL <https://www.nass.usda.gov/Publications/AgCensus/2007/> (accessed 12.20.17).

U.S. Geological Survey, 2006. Vector Digital Data - NHDPlus Version 1 [WWW Document]. URL https://www.usgs.gov/core-science-systems/ngp/national-hydrography/watershed-boundary-dataset?qt-science_support_page_related_con=4#qt-science_support_page_related_con (accessed 12.20.17).

United States Regional Economic Analysis Project, 2017. Midwest vs. United States Comparative Trends Analysis: Population Growth and Change, 1958-2016 [WWW Document]. URL <https://united-states.reaproject.org/analysis/comparative-trends-analysis/population/tools/10030000/0/> (accessed 1.25.18).

Villarini, G., 2016. On the seasonality of flooding across the continental United States. *Adv. Water Resour.* 87, 80–91. <https://doi.org/10.1016/J.ADVWATRES.2015.11.009>

Villarini, G., Slater, L.J., 2017. Examination of Changes in Annual Maximum Gauge Height in the Continental United States Using Quantile Regression. *J. Hydrol. Eng.* 23, 6017010. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001620](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001620)

Villarini, G., Strong, A., 2014. Roles of climate and agricultural practices in discharge changes in an agricultural watershed in Iowa. *Agric. Ecosyst. Environ.* 188, 204–211. <https://doi.org/10.1016/J.AGEE.2014.02.036>

Zhang, Y.-K., Schilling, K.E., 2006. Increasing streamflow and baseflow in Mississippi River since the 1940 s: Effect of land use change. *J. Hydrol.* 324, 412–422.

566 <https://doi.org/10.1016/J.JHYDROL.2005.09.033>

567

568