Global and Planetary Change 155 (2017) 98–108

Contents lists available at ScienceDirect

Global and Planetary Change

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

# On the use of Cox regression to examine the temporal clustering of flooding and heavy precipitation across the central United States

Iman Mallakpour[a,1], Gabriele Villarini[a,*], Michael P. Jones[b], James A. Smith[c]

[a] IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, IA, USA
[b] Department of Biostatistics, The University of Iowa, Iowa City, IA, USA
[c] Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA

## ABSTRACT

The central United States is plagued by frequent catastrophic flooding, such as the flood events of 1993, 2008, 2011, 2013, 2014 and 2016. The goal of this study is to examine whether it is possible to describe the occurrence of flood and heavy precipitation events at the sub-seasonal scale in terms of variations in the climate system. Daily streamflow and precipitation time series over the central United States (defined here to include North Dakota, South Dakota, Nebraska, Kansas, Missouri, Iowa, Minnesota, Wisconsin, Illinois, West Virginia, Kentucky, Ohio, Indiana, and Michigan) are used in this study. We model the occurrence/non-occurrence of a flood and heavy precipitation event over time using regression models based on Cox processes, which can be viewed as a generalization of Poisson processes. Rather than assuming that an event (i.e., flooding or precipitation) occurs independently of the occurrence of the previous one (as in Poisson processes), Cox processes allow us to account for the potential presence of temporal clustering, which manifests itself in an alternation of quiet and active periods. Here we model the occurrence/non-occurrence of flood and heavy precipitation events using two climate indices as time-varying covariates: the Arctic Oscillation (AO) and the Pacific-North American pattern (PNA). We find that AO and/or PNA are important predictors in explaining the temporal clustering in flood occurrences in over 78% of the stream gages we considered. Similar results are obtained when working with heavy precipitation events. Analyses of the sensitivity of the results to different thresholds used to identify events lead to the same conclusions. The findings of this work highlight that variations in the climate system play a critical role in explaining the occurrence of flood and heavy precipitation events at the sub-seasonal scale over the central United States.

## 1. Introduction

Flooding is one of the most common natural hazards in the United States. The National Weather Service (NWS) estimated that flooding was responsible for billions of dollars in damage and about 100 flood-related fatalities on average every year in the United States over the 1914–2013 period (NWS, 2014). In recent years, flood-related fatalities have ranked second among weather-related hazards in the United States (e.g., Ashley and Ashley, 2008). The central United States, in particular, is plagued by large flood events, especially over the most recent decades (e.g., 1993, 2008, 2011, 2013, 2014, and 2016), with flood damage in excess of several billion dollars and numerous fatalities (e.g., Hicks and Burton, 2008; Otto, 2009; Smith and Katz, 2013; NCDC, 2015).

Because of these significant societal and economic impacts, different studies have examined whether it is possible to identify changes in flood magnitude and/or frequency (e.g., Changnon and Demissie, 1996; Lins and Slack, 1999, 2005; Villarini et al., 2009, 2011a; Hirsch and Ryberg, 2012; Slater et al., 2015; Mallakpour and Villarini, 2015; Slater and Villarini, 2016). For instance, Mallakpour and Villarini (2015) focused on the changes in the frequency and magnitude of flood events over the central United States since the second half of the 20th century, finding stronger evidence for changes in the frequency rather than in the magnitude of flooding. Moreover, the strongest signal of change was for increasing trends, meaning that we have been experiencing a larger number of flood events over the past 50+ years. Mallakpour and Villarini (2015) tied their findings for flooding to changes in the frequency of heavy precipitation events, with other drivers tied to human modifications of the catchments playing a more limited impact. Indeed, studies have shown that large flood events affecting the central United

* Corresponding author at: IIHR-Hydroscience & Engineering, The University of Iowa, C. Maxwell Stanley Hydraulics Laboratory, Iowa City, 52242, IA, USA.
*E-mail address:* gabriele-villarini@uiowa.edu (G. Villarini).
[1] Current affiliation: Center for Hydrometeorology and Remote Sensing (CHRS), Department of Civil & Environmental Engineering, University of California, Irvine, California, USA.

http://dx.doi.org/10.1016/j.gloplacha.2017.07.001
Received 5 December 2016; Received in revised form 30 June 2017; Accepted 3 July 2017
Available online 04 July 2017
0921-8181/ 

States (e.g., 1993, 2008) are the product of heavy precipitation over a prolonged time period (e.g., Coleman and Budikova, 2010; Budikova et al., 2010; Smith et al., 2013; Villarini et al., 2013a; Smith and Baeck, 2015).

While the detection of changes in flood and heavy precipitation characteristics (i.e., magnitude and frequency) is important, it is also critical to investigate the driving forces that are responsible for the observed changes. By improving our understanding of the factors responsible for these changes, we will be in a potentially better position to improve our capabilities of predicting and projecting flood events. Most of the studies in the literature have linked the detected changes in extreme precipitation and flooding to either climatic changes that have been caused by human-induced global warming or to natural variability in the climate system (e.g., Jain and Lall, 2001; Milly et al., 2002; Gershunov and Cayan, 2003; Kunkel, 2003; Sankarasubramanian and Lall, 2003; Higgins et al., 2007; Min et al., 2011; Deser et al., 2012; Zhang et al., 2013; Fischer and Knutti, 2015; Yu et al., 2016; Mallakpour and Villarini, 2016a, 2016b). For instance, Fischer et al. (2013) noted that one of the main sources of uncertainty in the prediction of extreme events is the role of climate variability, especially on timescales from sub-seasonal to multi-decadal. Redmond et al. (2002) indicated that fluctuations in the spatial and temporal distribution of extreme hydrologic events are due to climate fluctuations. Yu et al. (2016) investigated the relationship between extreme precipitation events and climate variability over the contiguous United States for the 1979–2013 period. They found that climate variability was the main driver of the observed increase in extreme precipitation over the last three decades.

Furthermore, studies have shown that climate variability can affect the jet streams and storm tracks, which can cause changes in the atmospheric moisture content and transport patterns (e.g., Jain and Lall, 2001; Andersen and Shepherd, 2013; Villarini et al., 2011b). For instance, in both the 1993 and 2008 flood events over the central United States, atmospheric circulation and climate condition were favorable for bringing higher than average moisture from the Gulf of Mexico and the Atlantic Ocean to the central United States (e.g., Dirmeyer and Kinter, 2009; Coleman and Budikova, 2010; Budikova et al., 2010; Smith et al., 2013; Lavers and Villarini, 2013). Smith and Baeck (2015) analyzed the 1927 Mississippi River flood and linked this catastrophic event to enhanced moisture transport across the Gulf of Mexico, with anomalous conditions in the Pacific that controlled the development of baroclinic disturbances over the central United States. Mallakpour and Villarini (2016a) investigated the relationship between climate variability and the frequency of flooding and heavy precipitation over this area at the seasonal scale; they found that variability in the climate system driven by both the Atlantic and Pacific Oceans can affect the frequency of precipitation, and consequently flooding, over the central United States. However, all the analyses in their study focused on the modeling of the total number of events within a given season and they did not consider how events were distributed within the season.

Indeed, even though there has been some progress in connecting the occurrence of extreme meteorological events to the climate modes in recent years, much of the effort has been towards analyses at the annual or seasonal scales. However, when we focus on flood and heavy precipitation events at the sub-seasonal level, do these events occur at regular intervals, randomly or in clusters? Depending on the answer to this question, different modeling frameworks would be considered; moreover, given the same number of events within the season, the impacts of how they are distributed within the season will be different. In general, the fact that extreme events cluster in time can have a great impact on a number of fields as well as important societal and economic repercussions. For instance, Mumby et al. (2011) examined the temporal clustering of North Atlantic tropical cyclones and its effect on coral reefs. They found that the observed clustered behavior in tropical cyclones has a less detrimental impact on reef health than by assuming independent events. By accounting for the temporal clustering of hurricane occurrences in the vicinity of Florida, Jagger and Elsner (2012) were able to improve the retrospective forecasts of these storms. Vitolo et al. (2009) discussed the potential effects of clustering of extreme events for the insurance/reinsurance industry and explained that, given the same total losses from an individual event or from a cluster of events (before reinsurance), the cost to an insurance company after reinsurance from a cluster of events can be higher than from a single event.

The goals of this study are to examine whether hydrometeorological extreme events tend to occur independently of each other or in clusters, and whether it is possible to identify the climatic drivers responsible for the observed changes in the frequency of these events over the central United States at the sub-seasonal (daily) scale. Our hypothesis is that heavy rainfall and flood events tend to cluster in time, with the clustering that is driven by changes in the climate system. To be able to test this hypothesis, we cannot use a modeling framework that is based on Poisson processes, which are "memoryless," meaning that the occurrence of one event does not bear any information related to the occurrence of subsequent ones. Therefore, here we resort to Cox processes (Cox, 1972) to describe the occurrence/non-occurrence of flood and heavy precipitation events over the central United States (e.g., Smith and Karr, 1983, 1986; Villarini et al., 2013b; Slater and Villarini, 2016). Cox processes can be viewed as a generalization of Poisson processes with a randomly varying rate of occurrence. In the hydrometeorological community, there have been few studies that used Cox processes to model the occurrence of flood (e.g., Smith and Karr, 1986; Futter et al., 1991; Villarini et al., 2013b) and rainfall events (e.g., Smith and Karr, 1983, 1985). Smith and Karr (1983) introduced a point process framework based on Cox processes to investigate rainfall occurrences in the Potomac River basin, and found that clustering was an important feature of rainfall occurrences over their basin. Smith and Karr (1986) developed a flood frequency approach based on Cox regression to extract information related to the occurrence/non-occurrence of flood events depending on time-varying predictors; they found that snow pack and soil moisture can control the occurrence of flood events over the North Branch of the Potomac River. More recently, Villarini et al. (2013b) used Cox regression to investigate temporal clustering of flood events over Iowa, and found that the occurrence of flood events can be described in terms of the Pacific-North American pattern (PNA) and North Atlantic Oscillation (NAO) at several stations.

Rather than assuming that flood and heavy precipitation events occur independently of the occurrence of previous events as in Poisson processes, Cox processes allow accounting for the potential presence of temporal clustering. Under serial clustering, there is an alternation of quiet and active periods, in which the occurrence of an event has information about possible future events (Smith and Karr, 1985, 1986). We refer to a clustered process as one that is not Poisson (Karr, 1991; Smith and Karr, 1983). In particular, flooding and heavy precipitation can exhibit a concentration of events in a particular period of the year (e.g., seasonal, monthly), which can be described by a Poisson process with a seasonally varying rate of occurrence (e.g., Mallakpour and Villarini, 2016a). This inhomogeneous Poisson process does not represent what we mean by clustered process in this study, as it can be represented by a Poisson process under an appropriate transformation of the time axis, with the arrival time of the events that can be described by an exponential distribution. On the other hand, the time clock that characterizes the occurrence of events in a Cox model is not one with constant speed, but one that sometimes moves faster and sometimes slower (Zhou, 2001).

In this study, we use Cox regression to examine the relationship between observed climate variability and frequency of heavy precipitation and flood events at the daily time scale across the central United States; our results will increase our understanding of what controls the occurrence of these extremes over this area, and have the potential to provide critical information for short-term forecasting of heavy precipitation and flood events.