FINAL REPORT

Climate-informed Estimation of Hydrologic Extremes for Robust Adaptation to Non-stationary Climate

SERDP Project RC-2516

OCTOBER 2019

Casey Brown
University of Massachusetts, Amherst

Linda Mearns
NCAR

Upmanu Lall
Columbia University

Paul Moody
U.S. Military Academy

Jim Hall
Oxford University

SERDP
DOD ▪ EPA ▪ DOE

Page Intentionally Left Blank

Page Intentionally Left Blank

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
10/10/2019

2. REPORT TYPE
SERDP Final Report

3. DATES COVERED *(From - To)*
9/21/2015 - 9/20/2019

4. TITLE AND SUBTITLE
Climate-informed Estimation of Hydrologic Extremes for Robust Adaptation to Non-stationary Climate

5a. CONTRACT NUMBER
15-C-0052

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Casey Brown

5d. PROJECT NUMBER
RC-2516

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Massachusetts, Amherst
130 Natural Resources Road, 12B Marston Hall
Amherst, MA 01002

8. PERFORMING ORGANIZATION REPORT NUMBER
RC-2516

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Strategic Environmental Research and Development Program
4800 Mark Center Drive, Suite 17D03
Alexandria, VA 22350-3605

10. SPONSOR/MONITOR'S ACRONYM(S)
SERDP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
RC-2516

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This research develops and evaluates methods to produce the next generation of intensity-duration frequency (IDF) curves and hydrologic design events relevant for engineering design at DoD installations. The research demonstrates the utility of the methods that link non-stationary statistical analyses of observed hydrometeorological extremes to climate information produced through Earth system modeling. The effort is premised on the hypothesis that the biases and other failings of GCM projections may be overcome with innovative data science-based approaches to extracting meaningful and credible signals from the same. An assessment of climate modeling methods is made in terms of their ability to inform the key climate information needs that emerge from an analysis of historical non-stationarity already realized in the observed record. The researchers evaluated the relative advantage of various climate information tailoring methods, including different dynamical downscaling techniques, in terms of their ability to provide credible climate information relevant to hydrologic extremes.

15. SUBJECT TERMS
Climate-informed, Hydrologic Extremes, Robust Adaptation, Non-stationary Climate

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Casey Brown
				232	19b. TELEPHONE NUMBER *(Include area code)*
UNCLASS	UNCLASS	UNCLASS	UNCLASS		413-577-2337

Page Intentionally Left Blank

1. EXECUTIVE SUMMARY

This research develops and evaluates methods to produce the next generation of intensity-duration-frequency (IDF) curves and hydrologic design events relevant for engineering design at DoD installations. The research demonstrates the utility of the methods that link non-stationary statistical analyses of observed hydrometeorological extremes to climate information produced through Earth system modeling. The effort is premised on the hypothesis that the biases and other failings of GCM projections may be overcome with innovative data science-based approaches to extracting meaningful and credible signals from the same. An assessment of climate modeling methods is made in terms of their ability to inform the key climate information needs that emerge from an analysis of historical non-stationarity already realized in the observed record. We evaluate the relative advantage of various climate information tailoring methods, including different dynamical downscaling techniques, in terms of their ability to provide credible climate information relevant to hydrologic extremes. Through this research, we develop a robust method for estimating future changes in hydrologic extremes based on merged historical observations and credible climate projections, and highlight the implications for engineering practice by providing infrastructure design guidance. The innovative research methods developed in this effort are applied to several climate conditions in continental United-States with a larger focus on the Ohio River Basin, where precipitation extremes drive riverine flooding, and the upper Missouri River Basin, where snowpack is the primary source of moisture driving riverine flooding. Future research should continue to explore how these concepts can be incorporated in standard practice for design of hydrologic infrastructure on DoD installations by incorporating methods for recognizing and managing future climate trends.

Organization of the chapters

Chapter 2 provides a review of the challenges and advances in design of infrastructure for floods under non-stationarity. Based on more than 300 references, this review covers i) the potential sources of non-stationarity in time series of floods, ii) the methods for estimating design floods that rely on the stationary assumption, iii) the methods for estimating design floods that assume non-stationarity resulting from climate change and i) discussion on the current design methodologies in view of the pervasive uncertainties and strategies to manage the consequences of those uncertainties.

Chapters 3 and 4 explore the use of climate-informed parameter models for assessing future precipitation (Chapter 3) and streamflow (Chapter 4) extremes. Rather than using projections of local weather variables, such as precipitation variable, climate-informed approaches use large-scale climate variables (e.g., ENSO) to condition the parameters of the models. Chapter 3 analyses the performance of Global Circulation Models (GCMs) in reproducing large-scale climate variables leading to precipitation extremes. Chapter 4 proposes a general methodology to set up climate-informed regional models for streamflow extremes and drive climate projections projected large-scale climate variables. Both chapters focus on the Ohio River basin.

Chapters 5 and 6 focus on the role snow variable plays in flooding in the Upper Missouri basin. More specifically, Chapters 5 assesses the quality of the available datasets for temperature, precipitation and snow water equivalent variables across the Upper Missouri Basin. Climate projections for the region are then discussed across the basin. Chapter 6 investigates the consequences of changing snowpack across the Upper Missouri Basin on floods. A comparison between hydrological simulations and data-driven models (i.e., artificial neural networks) is conducted in order to better understand the role of snowpack in flooding in the region. Results show that uncertainty in snow water equivalent may affect significantly detection of change in streamflow extremes.

Chapter 7 examines the drivers of precipitation in the East-South-Central U.S during the cool season to advance understanding of the conditions that could lead to major flooding in the area. The study especially focuses on the different sources of bias in precipitation that are introduced by the regional climate models (RCMs). The main sources of bias in this region are linked to moisture flux into the region, transient, synoptic-scale low-pressure systems, Gulf of Mexico and Caribbean seas surface temperatures (SSTs), and ENSO-related teleconnections. Following these results, caution should be taken when using RCM projections to analyze hydro meteorological extremes.

Chapters 8 and 9 are extending the analysis of precipitation and streamflow extremes to the scale of the continental U.S. The two Chapters use the same set of nine catchments that covers rather well the range of climate conditions in continental U.S. Chapter 8 proposes a straightforward approach to define climate factors that account for climate change uncertainties. The suggested approach combines stationary statistical model for precipitation extremes with a sensitivity analyses regarding the model parameters of the precipitation distribution. Climate projections from GCMs and RCMs are used ex-post to infer the potential changes in precipitation distribution parameters. Climate factors can be chosen to provide satisficing levels of protection for a chosen range of uncertainty from climate projections. Chapter 9 is a comparison study across different climate areas and different models to predict change in streamflow extremes. The test-bed of models includes stationary approaches, climate-informed and trend-informed statistical models and hydrological simulations. Results show that no method performs better than others perform and thus suggests using several approaches for flood design.

Chapter 10 investigates the predictability of short- and long-term horizons for climate extremes in Ohio and Mississippi River Basins. Results show that both scales provide two distinct pieces of information with crucial implications in the management of water and crucial infrastructure systems in the region.

Chapters 11 and 12 focuses on decision-analysis and robust adaptation regarding risk from hydrological extremes. Chapter 11 is an application of the Decision Scaling approach for which the climate stress test is conducted using large-scale variable such as soil moisture across the Ohio River Basin and the seas surface temperature in the Pacific (e.g., ENSO). The considered case study for the application is Louisville, Kentucky. The study discusses some of the benefits and limitations of climate-informed stress test highlights areas of future research. Chapter 12 presents a set of stylized experiments to assess the uncertainties and biases involved in estimating future climate risk over a finite future period, given a limited observational record. Results suggest that shorter design lives are preferred for situations where inter-annual to decadal variability can be successfully identified and predicted, suggesting the importance of sequential investment strategies for adaptation.

2. INTRODUCTION: DESIGN CONSIDERATIONS FOR HYDROLOGIC EXTREMES IN A CHANGING CLIMATE

This chapter is published in Journal of Hydrology:

François, B., Schlef, K.E., Wi, S., Brown, C.M., 2019. Design considerations for riverine floods in a changing climate – A review. J. Hydrol. 574, 557–573. https://doi.org/10.1016/j.jhydrol.2019.04.068

2.1. EXECUTIVE SUMMARY

Conventional methods for designing infrastructure that is subject to flood risk, such as dams and levees, assume a stationary design flood. However, observed and potential non-stationarity in floods can result in costly over-design or dangerous under-design. Despite substantial attention, evidence from the literature makes clear there is no consensus methodology for estimating design variables under climate change. Practical guidance remains elusive. This paper presents a review of the challenges and advances in design of infrastructure for floods under non-stationarity. First, potential sources of non-stationarity in time series of floods are described to provide context and motivation. Second, methods for estimating design floods that rely on the stationary assumption are presented and their limitations are discussed. Third, methods for estimating design floods that assume non-stationarity resulting from climate change are summarized. Finally, the inadequacies of current design methodologies in view of the pervasive uncertainties are assessed and strategies to manage the consequences of those uncertainties are presented.

Key words: Flood, Design, Climate change, Uncertainty

2.2. INTRODUCTION

Infrastructure such as dams and levees are built to reduce risk from hydrological extremes such as riverine floods. Conventionally, such infrastructure is designed to protect against flood events up to a chosen magnitude. This so-called design flood is commonly defined as the flow quantile with a selected frequency of occurrence or return period. For example, the 100-year return flood is typical for designing levees and other protection structures (e.g., Olsen, 2006). More recently, the design value is sometimes chosen based on maximizing net economic benefits (i.e., avoided damages by implementing flood protection less the cost of the protection) or minimizing the expected damages of flooding (e.g., Lund, 2002). For infrastructure where failure would lead to tremendous damages, such as very large dams, some national and regional agencies use the concepts of Probable Maximum Precipitation or Flood (PMP or PMF) rather than statistical approaches for design (WMO, 2009). All of these approaches to design flood estimation make use of the assumption of stationarity.

Notably however, stationarity lacks a well-established definition within the hydrologic and water system analysis literature. For instance, Milly et al. (2008) described stationarity as “the idea that natural systems fluctuate within an unchanging envelope of variability”. Montanari and Koutsoyannis (2014), referencing Kolmogorov (1931, 1938) and Khintchine (1934), defined as stationary a process that “undergoes change, but its statistics are conserved”. Other studies, such as that by Salas et al. (2018), maintain that stationary means “the marginal distribution remains invariant in time”. In this review, we adopt the definition by Salas et al. (2018). In the context of design for riverine floods, this definition of stationary implies that the design flood is time-invariant and hence the probability of infrastructure failure remains constant for the duration of the design life.

There is now a significant body of literature claiming that it is no longer valid to assume the design flood is stationary. This claim is based on theoretical considerations and observational evidence from some locations that flood risk is changing due to climate change and socio-economic development (e.g., Dankers et al. 2014; Winsemius et al., 2015; Arnell and Gosling, 2016; Berghuijs et al., 2017 and references within). While flood risk combines hydrological hazard and vulnerability (Merz et al. 2010), here the focus is hydrological hazard, which is the probability of occurrence of a given flood magnitude. The concern is that if the design flood is indeed nonstationary, then an assumption of stationarity could result in over- or under-design (Jain and Lall, 2001; Rosner et al., 2014). Over-design occurs when infrastructure becomes too large as peak flows decrease, leading to sunk capital and operating costs. Conversely, under-design occurs when infrastructure becomes too small as peak flows increase, leading to societal impacts such as human casualties and economic damages, which are generally of higher concern than sunk costs (Rosner et al., 2014). For instance, the American Society of Civil Engineers estimates that about $150 billion in flood damages resulting from Hurricane Katrina were due to under-design of levees and other flood protection structures (ASCE, 2007).

To avoid over- or under-design, Hirsch (2011) recommends that “once we recognize that we have nonstationarity for a variety of reasons, […], we really have to rethink our approach to planning”. This quote summarizes the increasingly common motivation to improve current and develop new design approaches that account for non-stationarity in hydro-meteorological processes, and thus improve the robustness and resilience of flood infrastructure (Beven, 2011; Kundzewicz et al., 2017). For example, in 2010, a workshop in Boulder (Colorado, U.S.) gathered hydrologists, climatologists, engineers and scientists around the question: “if stationarity is dead, what do we do?“ (Galloway, 2011). One topic was whether the U.S. federal guidelines for Flood Frequency Analysis, as described in Bulletin 17B (IACWD, 1982), should be updated to account for non-stationarity. Eight years later, the update, Bulletin 17C, acknowledges evidence of non-stationarity but provides no concrete guidance for its incorporation into analysis and design, instead suggesting that in situations of sufficient evidence for climate-induced non-stationarity, water practitioners should “employ time-varying parameters or other appropriate techniques” (England Jr. et al., 2018).

However, water practitioners do not generally know the range of available methods for flood frequency analysis assuming non-stationarity (Serinaldi and Kilsby, 2015) even though the use of such approaches and models have gained popularity in academia (e.g., Strupczewski et al., 2001; Delgado et al., 2010; Katz, 2013; Prosodocimi et al., 2014; Yu et al., 2015; Spence and Brown, 2016; Šraj et al., 2016; see Salas et al., 2018 for a review). This lack of knowledge likely stems from the proliferation of approaches and the many challenges associated with assuming non-stationarity. Challenges include but are not limited to: comprehensively diagnosing the drivers of change (e.g., anthropogenic climate change or natural climate variability) (Vogel et al., 2011; Deser et al., 2012a; Elmer et al., 2012; Merz et al., 2012; Harrigan et al., 2014), determining how change affects the flood time series (e.g., when a trend is detected, a statistically significant mean shift can also be detected) or distribution moments (i.e., mean, variance, skewness and kurtosis; although, in practice, observational records are too short to detect changes in the latter two) (Coles, 2001; Xiong and Guo, 2004; Villarini et al., 2009a; Katz, 2013), and identifying appropriate mathematical models to represent change (e.g., multiple mathematical models have been used to represent change in annual maximum flows at the Little Sugar Creek (Charlotte, North Carolina), leading to widely different estimates of the 100-year flood) (Villarini et al., 2009b; Salas and Obeysekera, 2014; Serinaldi and Kilsby, 2015).

This paper addresses the need for guidance on design for riverine floods under non-stationarity, particularly as associated with climate change, by providing a comprehensive review of available design methods and guidelines. Section 2.3 reviews the main drivers that can lead to changes in peak flow magnitude and frequency and then discusses methods commonly used for detection of change and attribution to specific drivers. Section 2.4 reviews the available stationary approaches to design flood estimation and provides an example application with record of flood events showing non-stationarity. Section 2.5 describes current methods for estimating the design flood assuming non-stationarity. Section 2.6 discusses the available options for addressing uncertainty. Section 6 concludes with an outlook on future flood design and further research.

2.3. WHAT ASSUMPTION SHOULD BE MADE?

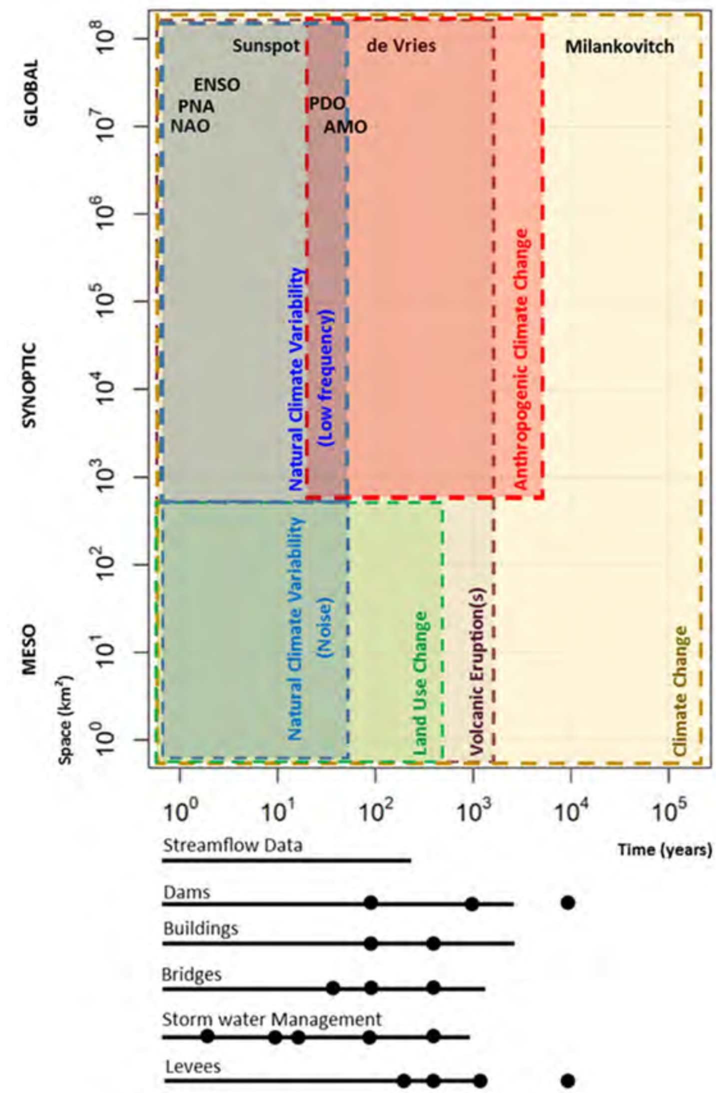


Figure 2-1 Generalized spatiotemporal scales for selected flood drivers (boundaries and placements are approximate). The examples of low-frequency oscillatory ocean-atmospheric phenomena are meant to be illustrative, not exhaustive (NAO is the North Atlantic Oscillation, PNA is the Pacific/North American pattern, ENSO is the El Nino-Southern Oscillation, PDO is the Pacific Decadal Oscillation, and AMO is the Atlantic Multidecadal Oscillation). For each infrastructure type, the solid line shows the possible lifespan (the upper bound was set to be approximately the age of the oldest existing and in-use element to-date) while the dots indicate common flood return periods used in design (the PMF, commonly used for dams, is approximated by the 10,000-year return period); the time scale of streamflow data records is shown for comparison. Climate change is defined (following the glossary of

IPCC, 2014) as the combined influence of natural climate variability, anthropogenic climate change, changes in land use, volcanic eruptions, and solar cycles.

The question of whether to assume a stationary or non-stationary design flood is closely tied to considerations of the spatiotemporal scales associated with the natural and anthropogenic processes that drive flood events and the intended infrastructure design (Figure 2-1). Clearly, the processes acting on the timescale of the design event for a storm water management system (e.g., 10-year flood) are vastly different from those acting on the timescale of a PMF used for dam design. In this context, this section explores: (1) potential non-stationarity induced by processes driving flood events, (2) methods for detecting change in peak flow series, and (3) methods for attributing observed change to specific drivers. The discussion is supported further by examples to illustrate the issues associated with the range of spatiotemporal scales that influence the magnitude and occurrence of flood events as well as their detection and attribution to particular drivers.

2.3.1. Evidence and drivers of non-stationarity in riverine floods

Riverine floods are primarily driven by (1) precipitation, which is driven by natural climate variability (e.g., Deser et al. 2012a) and anthropogenic climate change (e.g., Zhang and Delworth, 2018), and (2) land surface response, which is driven by land-use change, river regulation, and natural catastrophes. Each of these drivers are potential sources of non-stationarity. The Red River of the North at Fargo (North Dakota, U.S.) is a particularly salient example of nonstationary stream flow (Figure 2-2) (Mueller and Foley, 2014). Although flagged as regulated by the U.S. Geological Survey, the U.S. Army Corps of Engineers (USACE) demonstrated that changes in streamflow beginning in the early 1940's (Villarini et al., 2009a) cannot be explained by flow regulation (see discussion in Serinaldi and Kilsby, 2015); instead, tree ring analysis has shown that the river experiences "high and low flood modes […], which extend from several decades to nearly a century" (George and Nielsen, 2003).

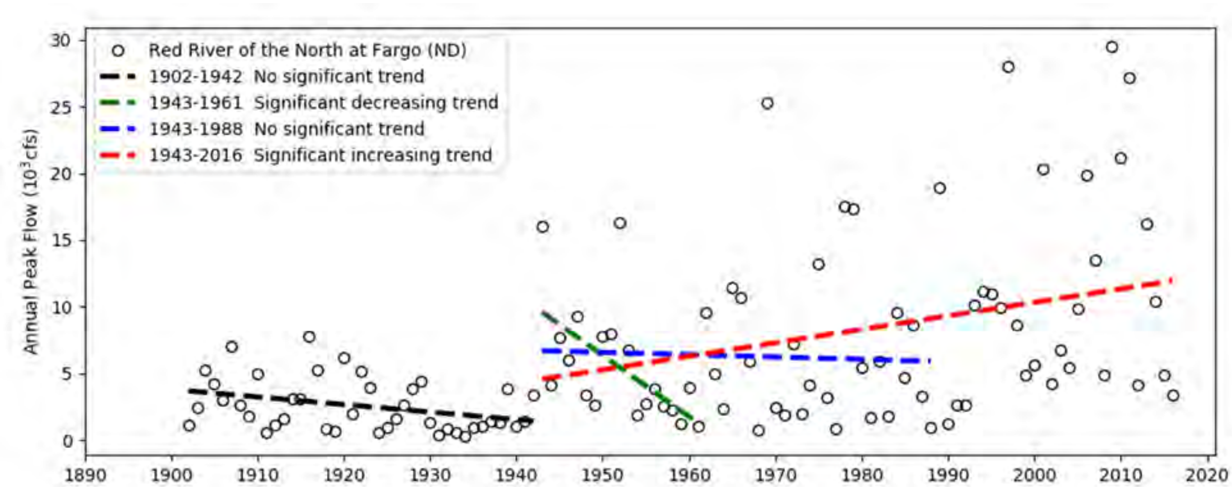


Figure 2-2 Annual peak flow for the Red River of the North at Fargo (North Dakota, USA) (USGS id: 05054000) Dashed curves show trends for different periods. The significance of the trends has been tested against the null-hypothesis of an i.i.d. process with the Mann-Kendall test at the 5% significance level. Flow values are in cubic feet per second (cfs). Note that Villarini et al. (2009a) detected a significant shift in mean peak flow in 1942.

Natural climate variability, sometimes called internal climate variability, is "variability of the climate system that occurs in the absence of external forcing, and includes processes intrinsic to the atmosphere, the ocean, and the coupled ocean-atmosphere system" (Deser et al., 2012b). Natural climate variability, generally considered to occur on timescales of 30 years or less (with exceptions), is comprised of two components acting at different spatiotemporal scales; (1) noise at local and regional scales and (2) low-frequency oscillations at regional to sub-global scales that have a global influence on hydrology (Deser et al., 2012a; Hulme et al., 1999). For example, the El Nino-Southern Oscillation (ENSO) occurs in the tropical Pacific but influences hydrology around the globe (Trenberth 1997; Ward et al., 2014; Lee et al. 2018). Natural climate variability can affect flood magnitude (e.g., Hannaford and Marsh, 2008; Schlef et al., 2018a; Zhang et al., 2015), frequency of occurrence (e.g., Andrews et al., 2004; Delgado et al., 2012; Hodgkins et al., 2017; Kiem et al., 2003; Mallakpour and Villarini, 2016a), and timing (e.g., Blöschl et al., 2017) (for other examples of the influence of natural climate variability of floods see also Petrow et al., 2009; Prudhomme and Genevier, 2011; Villarini et al., 2013; Wilby and Quinn, 2013; Armstrong et al., 2014; Li and Tan, 2015; Bracken et al., 2018). Natural climate variability is generally thought to cause temporal clustering or semi-cyclical patterns in floods (e.g., Jain and Lall, 2001).

Anthropogenic climate change is forcing of the atmosphere by anthropogenic greenhouse gas emission, leading to global-scale changes in the coupled ocean-atmospheric system (IPCC, 2014). Notably, in model simulations, natural climate variability dominates the signal in precipitation and temperature due to anthropogenic climate change until around 2050 (Hawkins and Sutton, 2011; Deser et al. 2012b; Hingray and Saïd, 2014; Whateley & Brown, 2016; Schlef et al. 2018b; Martel et al. 2018). At the horizon 2050, natural climate variability may still be difficult to separate from anthropogenic climate change effects on integrated variables such as streamflow and crop yields (e.g., Hulme et al. 1999). There is physically-based evidence that ongoing anthropogenic climate change may lead to an intensification of and changes in the water cycle worldwide (IPCC, 2014), leading to changes in floods (e.g., Milly et al., 2002; Hirabayashi et al., 2013; Kundzewicz et al., 2013; Alfieri et al., 2015; Arnell and Gosling, 2016). Warmer air temperatures will increase atmospheric moisture holding capacity, leading to an increase in the intensity and variability of extreme precipitation (Lenderink and Meijgaard, 2008; Trenberth, 2011; Berg and Haerter, 2013; Fischer and Knutti, 2016; Yin et al., 2018), likely affecting flash and short-rain floods (for flood definitions here and subsequently, see Merz and Blöschl, 2003). Warmer temperatures will also cause snowpack accumulation to shrink in mountainous regions and at high latitudes (Barnett et al., 2005) and change the timing and rate of snowmelt (e.g., McCabe and Clark, 2005; Blöschl et al., 2017; Musselman et al., 2017). In general, these modifications are likely to decrease rain-on-snow and snowmelt floods, but are dependent on local trends. Additionally, there will be changes in storm tracks and individual storm intensities, linked to changes in the underlying large-scale circulation patterns (Bengtsson et al., 2006; Boé and Terray, 2008; Knutson et al., 2010; Santos et al., 2016; Shaw et al., 2016), likely affecting short- and long-rain floods. Apart from the possibility of a tipping point, anthropogenic climate change is generally thought to produce slow and long-term trend-like changes in floods.

Land use changes include urbanization, forest management, and agricultural practices, which can affect floods for up to several centuries but can only be verified at spatial scales up to several tens of square kilometers (Blöschl et al., 2007; Rogger et al., 2017). In particular, the expansion of impervious areas due to urbanization decreases infiltration, which weakens the buffering effect of the natural ecosystem, causes changes hydrological extremes (e.g., Rose and Peters, 2001; Smith et al., 2002; Konrad, 2003; Moglen and Shivers, 2006; Saghafian et al., 2008; Villarini et al., 2009a; Villarini et al., 2009b; Vogel et al., 2011). Land use changes are thought to cause either change points or trend-like changes in floods.

River regulation (or training) can range from construction of dams and weirs to straightening meandering or braided rivers. The impacts of river regulation on floods varies (for one example of the magnitude of impact see Vorogushyn and Merz, 2013); for large rivers, the effects of river regulation on floods may be as much or greater than land use changes (Lammersen et al., 2002; Bronstert et al., 2007). In the case of channel modification, the impact is largest when the flood remains in the riverbed (Hall et al., 2014). River regulation is generally thought to introduce change points in floods.

Natural catastrophes, in particular volcanos, cause immediate and massive impacts at local and regional scales, cause global cooling and changes in hydrology in the first several years after the event, and may contribute to long-lasting and global climate change under certain conditions (Hofmann, 1987; Major & Mark, 2006; Schneider et al., 2009; Trenberth & Dai, 2007). Natural catastrophes are generally thought to introduce change points in floods but are not generally considered in non-stationary analyses.

2.3.2. Detection of change in peak flow records

Tracking or anticipating change in hydrologic extremes is the typical starting point when considering the adequacy of the current flood risk management strategy at a given location. There are a variety of methods available for detecting trends and shifts in hydrological time series (see the reviews by Kundzewicz and Robson, 2004; Khaliq et al., 2009; Madsen et al., 2014; Bayazit, 2015). The World Meteorological Organization (WMO, 2009) recommends the non-parametric Mann-Kendall test for trend analysis (MK; Kendall, 1975); it does not require an assumed probability distribution, has been widely applied to hydrologic extremes (e.g., Petrow and Mertz, 2009; Petrow et al., 2009; Villarini et al., 2009b; Mediero et al., 2014; Archfield et al., 2016), and has been adapted to account for long-term persistence (Hamed, 2008). For detecting shifts, the non-parametric Pettitt test (1979), which is easily implementable, has been widely used (e.g., Villarini et al., 2009a; Rougé et al., 2013; Prosdocimi et al., 2014; Li and Tan, 2015; Mallakpour and Villarini, 2016b). However, it only detects one shift within a time series; unlike the Rodionov test (2004).

These and other methods do not always correctly detect trends and shifts, as illustrated by simulation experiments where the trend is stochastically generated (Spence and Brown, 2016) and is highly dependent on spatiotemporal scale (Figure 2-1). Specifically, the outcome of statistical tests for trend detection depends on the hydrological record length (Blöschl and Montanari, 2010; Barros et al., 2014). Using the Red River of the North at Fargo as an example (Figure 2-2), we applied the MK test to three periods: the first (1943-1961) shows a significant downward trend, the second (1943-1988) shows no significant trend, and the third (1943-2016) shows a significant upward trend. Although these periods were obviously chosen specifically to illustrate the point, this analysis indicates that the overall trend was dominated by low-frequency variability for several decades. This example highlights that natural climate variability has its own structure, which can easily confound trend analysis. Although temporal and spatial structures of certain climate patterns are relatively well characterized (e.g., ENSO), this emphasizes the importance of diagnosing the structure of natural variability and adapting the trend analysis. For instance, if temporal clustering or semi-cyclical patterns are identified in peak streamflow, one solution is to apply trend tests over periods that include one or more cycles (Ishak et al., 2013). At longer timescales, the Hurst phenomenon (Hurst 1951), in which a trend may not be separable from persistence, applies (Koutsoyiannis, 2003, 2006). The outcome of the statistical tests also depends on spatial scale and may vary from one catchment to another (Archfield et al., 2016) in part due to land-use change or regulation (Vogel et al., 2011). Aggregation to large spatial scales tends to reduce the noise of natural climate variability and improves trend detection (Fischer and Knutti, 2014).

2.3.3. Attribution of change in peak flow records

Attribution, which consists of quantifying the contribution of various drivers to change in peak flows, is crucial for projection (Hall et al., 2014) and can raise awareness of the need for mitigation (e.g., Thompson and Otto, 2015; Schwab et al., 2017). Attribution studies of changes in hydro-climatic extremes in general and of the causes of individual events (e.g., Allen, 2003; Pall et al. 2011; 2017; Easterling et al. 2016; Viglione et al. 2016) are becoming more common. For flood change attribution, Merz et al. (2012) suggested a three-part framework that uses observed peak flows and simulation results. The first part, evidence of consistency, relates to whether the observed change in peak flows is consistent with change in the assumed driver. The second part, evidence of inconsistency, consists of demonstrating that change in peak flow is not the outcome of a different driver from the one originally assumed; the aim is to avoid wrong attribution when several drivers of changes are acting at the considered location. The third part is the provision of confidence level of the causality attribution. This framework is often reduced to comparing observed peak discharges with simulated discharges that are obtained with and without changes in the drivers (e.g., Wolski et al., 2014; Aich et al., 2015; Prosdocimi et al., 2015) (see Section 2.5.1.2 for a description of possible statistical approaches). The drawback of this causal framework and other sensitivity-type analyses are that inference is limited to the relative or qualitative weights for various drivers and is limited when several drivers are jointly responsible for changes in peak flows (Harrigan et al., 2014; Vogel et al., 2011).

2.3.4. IMPLICATION FOR DESIGN

The above sections highlight that clear physical reasoning supports non-stationarity in floods. However, physical evidence is often inconclusive. Natural climate variability and land use change make detection and attribution of change difficult, which could at least partly explain the lack of physical evidence. Thus, engineers and decision-makers could find a trend but not be sure if it is due to anthropogenic climate change. Alternatively, they may find no trend but since the climate change signal is yet to be detected, flood hazard may nevertheless increase or decrease in the future. In either case, these possibilities pose a clear design challenge. The sections below discuss the limitations of current approaches given this context.

2.4. STATUS QUO: DESIGN FLOOD ASSUMING STATIONARITY

The stationarity assumption has been used for decades for designing flood infrastructure and has been demonstrated to be the most pertinent assumption at many locations (Serinaldi and Kilsby, 2015; Luke et al., 2017). Assuming stationarity, there are two primary approaches to determining the design flood (Figure 2-3). The first is to perform Flood Frequency Analysis (FFA) and then use pre-determined return periods or risk-based approaches. The second is to use the concept of PMP and PMF. This section briefly reviews these methods.

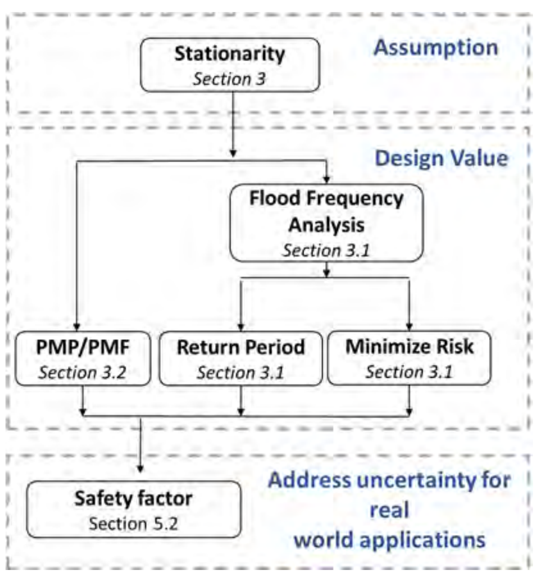


Figure 2-3 Pathway of choices for flood design under stationary assumption.

2.4.1. Design based on Flood Frequency Analysis

In stationary FFA, observed peak flow values are represented by a probability distribution with time-invariant parameters:

$$\boldsymbol{Y} \sim f(\theta_1, .. \theta_k, .., \theta_K), \tag{2.1}$$

where $Y=(y_1,y_2,..,y_N)$ is the peak flow time series, N is the number of observations, '~' means 'distributed as', and f is a chosen distribution function. The U.S. recommends the log Pearson type 3 distribution (England Jr. et al., 2018) and the National Environment Research Council in the United Kingdom recommends the Generalized Extreme Value distribution (NREC, 1975); other extreme value distributions may also be appropriate (Merz and Theiken, 2009). $\theta=(\theta_1,..,\theta_K)$ is a vector of K parameters (usually K=2 or 3). For more details on conducting FFA assuming stationarity, including parameter estimation, application to ungauged catchments or at regional scales, and handling missing data, the reader can refer to the aforementioned guidelines and the reviews by Coles (2001), Katz et al. (2002), and Khaliq et al. (2006). Uncertainty in FFA is introduced in streamflow measurements and in fitting extreme value distributions to limited records.

Once FFA is complete, there are two approaches to determining a design value. The first approach is to use a pre-determined return period. The 100-year return flood, in particular, has become commonly used in the U.S. as "a reasonable compromise between the need for building restrictions to minimize potential loss of life and property and the economic benefits to be derived from floodplain development" (FEMA, 2011), partially due to the National Flood Insurance Program, which mandates flood insurance for structures at risk of the 100-year flood. However, the choice of flood protection standard may be more political than based on scientific or economic justification. For example in the U.S., the 500-year standard mandated after Hurricane Sandy by the Obama administration was subsequently reversed by the Trump administration (Koerth-Baker, 2017). The degree of conservatism varies by country. In the Netherlands, standards for levee and dike design, which range anywhere from the 300- to 10,000-year flood depending on location, were developed from FFA of historic data and the analysis of costs of construction compared to damages and risk of death associated with past floods (Voorendt, 2015). The second is a risk-based approach that minimizes total expected cost (National Research Council, 2000; Lund, 2002; Jonkman et al., 2004; Tung, 2005). In practice, the objective is to balance the costs of protection with the expected costs associated with the hydrological hazard (i.e., flood damage costs plus the costs of any emergency actions taken during the flood event). Thus, the assumption of stationarity applies to not only the flood probability distribution, but also the socio-economic estimates that determine infrastructure and damage costs, which add another level of uncertainty.

2.4.2. Probable Maximum Precipitation and Flood

A significant part of the flood design literature relies on the concept of PMP and PMF. PMP is "the greatest depth of precipitation for a given duration meteorologically […] with no allowance made for long-term climatic trends" (WMO, 1986). The PMF is the flood that would result from the combination of the PMP and the most severe hydrologic conditions (i.e., initial soil moisture and/or snowpack) considered physically possible in the region. The PMF is used for designing infrastructure whose failure would cause tremendous damages to the downstream population and economy (e.g., large dams). Its return period has sometimes been associated with values ranging from 104 to 107 years (Fernandes et al. 2010; Nathan et al. 2016). It is primarily used in North America and Australia; most European countries use conventional FFA (Boes, 2011), because such high return periods are considered beyond the credible limit of extrapolation (Nathan and Weinmann, 2001). In the U.S., state-level regulations sometimes allow design within a range; the lower bound may be set by either a return period (e.g., the 500-year flood) or a percent of the PMF, while the upper bound is a percent of the PMF (FEMA, 2012).

The World Meteorological Organization provides guidelines for estimating PMP (WMO, 2009). According to those guidelines, there are two distinct categories of methods; those based on the statistical analysis of observed extreme rainfall and those that rely on physically-based storm modelling. The WMO recommends the statistical analysis of extreme rainfall (Hershfield, 1961) as an approximation for PMP for small catchments with surface area less than 1,000 km2. The most common physically-based approach, called storm maximization and storm transposition (e.g., Rakhecha et al. 1999), consists of boosting the atmospheric conditions (most often the atmospheric moisture) to the physical limits. More specifically, the PMP_T value for a storm of duration T at a given location can be obtained by (Stratz and Hossain, 2014):

$$PMP_T = Pobs_T\left(\frac{W_p(max)_T}{W_p(obs)}\right), \qquad 2.2$$

where $Pobs_T$ is the maximum observed depth of precipitation for the duration T at the location of the observed storm, $W_p(obs)$ is the precipitable water in the air column of the actual storm being maximized, and $W_p(max)_T$ is the maximum probable precipitable water in the moisture column in the transposed location where PMP is estimated. Several methods exist for estimating the latter. For instance, the USACE estimates $W_p(max)_T$ based on the maximum 12-hour persisting temperature dew point (Schreiner and Riedel, 1978). Another method is to define $W_p(max)_T$ as the 100-yr return precipitable water estimated from numerical simulations (e.g., a regional climate model) (Beauchamp et al., 2013). In Equation 2.2, PMP_T is defined for the same duration and for the same area as the observed storm. This value can be interpolated to another spatial area and another duration by using depth-area-duration curves (U.S. Department of Commerce, 1999).

The PMP is converted into the PMF using either a hydrological model or the unit hydrograph method. These methods require as input the Maximum Probable Storm (PMS), which is the hypothetical storm for a particular drainage area and duration that results from the PMP. The PMS temporal structure is commonly obtained by assuming a hyetograph specific to the climatic region (for the continental U.S., see Water Resources Program, 2009). The PMS spatial pattern is usually based on a standard isohyetal pattern that has a shape and orientation corresponding to that commonly observed at the location (National Weather Service, 1982). The PMF estimate is sensitive to both initial soil conditions and to the spatial and temporal patterns of the PMS, leading to uncertainty bounds that may be as large as 50% (Jakob, 2013). In locations where PMF is likely to occur during periods when snowmelt may significantly contribute to runoff, it is recommended to assume a 100-year return period snowpack (Debs et al. 1999; Beauchamp et al. 2013); additional uncertainty is introduced in estimating the snowpack.

2.4.3. Concerns in the use of traditional methods with non-stationarity

If the drivers discussed in Section 2.3.1 cause non-stationarity in hydrological extremes, the use of a stationary FFA or PMP/PMF may lead to poor design estimates (e.g., Jain and Lall, 2001; Sarewitz et al., 2003; Kunkel et al., 2013; Rosner et al., 2014). Poor design estimates can occur for reasons such as (1) the available historic data may only partially sample the range of peak flow variability resulting from low-frequency oscillations associated with natural climate variability or (2) a trend detection test may provide a false positive or false negative. To illustrate how non-stationarity can affect the design flood estimate, we use the example of the Red River of the North at Fargo. We fit the Generalized Extreme Value distribution (GEV; Jenkinson, 1955) for two sub-periods of the annual maximum streamflow records (1902-1942 and

1943-2016). Figure 2-4 displays the flood frequency curves for these two periods, together with their confidence intervals as obtained via bootstrap (Obeysekera and Salas, 2014). The peak flow quantiles obtained from the period 1902-1942 are significantly different from those from 1943-2017. Let us assume that, in 1942, local engineers designed a levee based on the 100-year flood. According to the available data (1902-1942), the design flood equals 17*103 cfs. However, this value was surpassed ten times during 1943-2017 (Figure 2-2), which would have potentially caused significant societal consequences. On the flood frequency curve obtained with the 1943-2016 data, the 17*103 cfs corresponds approximately to an 8-year flood. Using data from 1942-2016, the 100-year flood equals 57*103 cfs and corresponds to a 1,500-year return period according to the 1902-1942 data. This example illustrates how the use of FFA assuming stationarity may lead to under- or over-design should the peak flow distribution change.

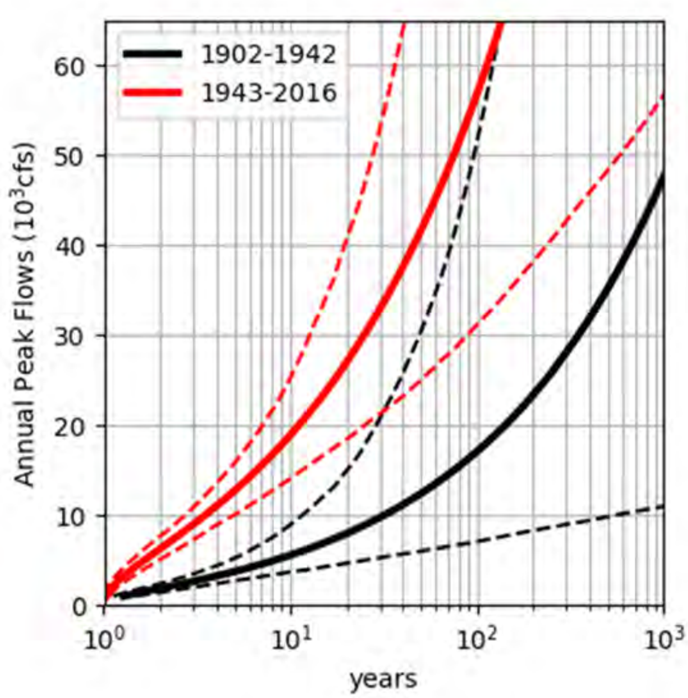


Figure 2-4 Flood frequency curves for the Red River of the North at Fargo (USGS id: 05054000) as obtained prior and after the year 1942 GEV parameters are calibrated regarding the observed peak flows by optimizing the maximum likelihood function (Martins and Stedinger, 2000). 90% confidence intervals (dashed lines) are obtained via bootstrap method (Obeysekera and Salas, 2014)

2.5. NONSTATIONARY DESIGN FLOOD ESTIMATION

This section describes the primary approaches to flood projection and subsequent design flood estimation under non-stationarity (Figure 2-5). These approaches are often referred to as ‘Predict-Then-Act’ approaches in the sense that predicted change in flood quantiles are used for infrastructure design (Lempert et al., 2004). This section focuses on anthropogenic climate change and natural climate variability because they are major sources of non-stationarity in the peak flow distribution or the PMP/PMF, although the overarching concepts are also relevant for changes in land-use and river regulation.

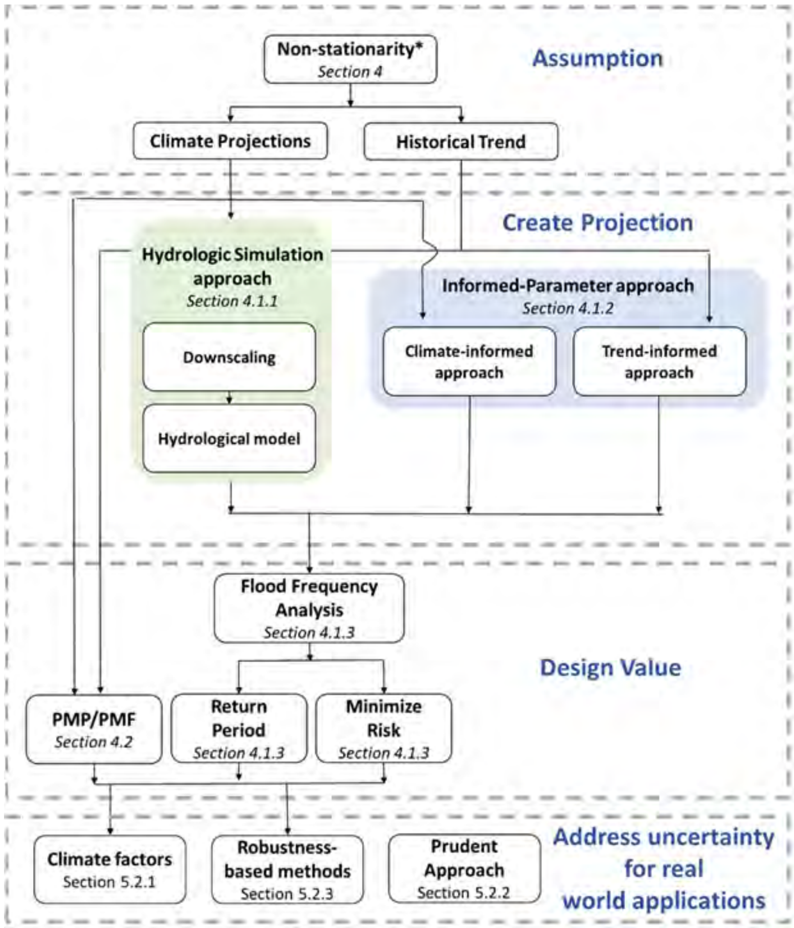


*Figure 2-5 Pathway of choices for flood design under nonstationary assumption (*only accounting for climate variability and change).*

2.5.1. DESIGN BASED ON FLOOD FREQUENCY ANALYSIS UNDER NON-STATIONARITY FROM CLIMATE

Assuming non-stationarity, the first step in design flood estimation via FFA is projection of flood risk; here the focus is on projection of the flood distribution, although projection of future infrastructure cost and flood damages is also necessary for risk assessment. There are two primary approaches to projection of the flood distribution; (1) hydrologic simulation approaches combine simulation output from General Circulation Models (GCMs) with downscaling and bias correction methods and hydrological modeling (green box in Figure 2-5) and (2) informed parameter approaches create statistical models with time-varying parameters (blue box in Figure 2-5). The latter include trend-informed models in which time is the only covariate and climate-informed models in which climate variables are used as covariates.

2.5.1.1. FLOOD PROJECTION VIA HYDROLOGIC SIMULATION APPROACHES

Climate projections from GCMs are often used for assessing future changes in river flows through a chain of models. GCMs are physically-based models that attempt to consistently simulate the behavior of the atmosphere, land-surface and ocean, including interactions among these components, at a global scale. Scenarios that attempt to span the range of plausible future greenhouse gas emissions (CMIP3 experiment, Nakicenovic et al., 20001) or concentrations (CMIP5 experiment, Taylor et al., 2012) drive GCM projections of future climate variables. Because GCM outputs are subject to biases (e.g., Dai, 2006;

Kundzewicz and Stakhiv, 2010; Sillmann et al., 2013; Crétat et al., 2014) and are simulated at spatial resolutions that are coarser than the scale required by the hydrologic models to represent the paramount processes generating runoff (Fowler et al., 2007), the outputs are commonly bias corrected and downscaled to create realistic inputs for hydrologic models (Maraun et al., 2010). The downscaled and bias corrected GCM projections are then used to force physical or conceptual hydrologic models to simulate extreme streamflow.

Use of this approach is wide-spread and ranges from the catchment scale (e.g., Camici et al., 2014; McMillan et al., 2010; Ngongondo et al., 2013; Prudhomme et al., 2003) to national, continental and global scales (e.g., Alfieri et al., 2015; Arnell and Gosling, 2016; Dankers and Feyen, 2009; Dankers et al., 2014; Hirabayashi et al., 2008, 2013; Leng et al., 2016; Rojas et al., 2011, 2012; Roudier et al., 2016). The major advantage of using the chain of models is that the evolution of flood-inducing weather variables under future atmospheric conditions is physically modeled. However, this framework suffers from several shortcomings discussed below (e.g., Kundzewicz and Stakhiv, 2010), and overall is subject to a myriad of sources of uncertainty that arise from each step of the modeling framework (Stainforth et al., 2007a; Wilby and Dessai, 2010).

The use of GCM projections for assessment of hydrological extremes is questionable because GCM estimates of precipitation are biased, particularly in extremes (e.g., Mehran et al., 2014) which affects simulation of hydrological processes (e.g., Leander and Buishand, 2007; Sperna Weiland et al., 2010; Rocheta et al., 2013). In particular, GCMs do not simulate the physical processes that generate extreme precipitation that trigger most flood events because those processes occur at spatial scales that are finer than model resolutions (e.g., Boberg et al., 2007; Leander and Buishand, 2007; Wuebbles et al., 2014; Crétat et al., 2014). In comparison to observations, historical simulations usually underestimate precipitation extremes, which suggests an underestimation of projected precipitation extremes (Kundzewicz et al., 2017), likely resulting in an underestimation of future peak flows.

Bias correction and downscaling, which sometimes involves implicit bias correction, are common approaches to mitigating the bias and coarse resolution of GCM outputs, with the goal of producing more hydrologically-relevant projections of climate. Bias correction is the "correction of model output towards observations in a post-processing step" (Ehret et al., 2012). Bias correction methods have multiple disadvantages: (1) bias is corrected without consideration of the forcing or structural errors which may be causing the bias, (2) bias in statistics other than the mean, such as the variance, is often not corrected (Teutschbein and Seibert; 2012), and (3) bias is assumed time-invariant (i.e., bias identified in the current period can be used to correct bias in future periods), which can cause misleading results when assessing hydrological extremes (e.g., Ehret et al., 2012; Maurer et al., 2013; Velázquez et al., 2015). Thus, bias correction alone is not recommended for future flood design.

Downscaling methods can be classified into three main categories: (1) delta change factor methods (or perturbation methods), (2) statistical methods, and (3) dynamical methods (Ekström et al., 2015; for comprehensive reviews see also Fowler et al., 2007; Madsen et al., 2014). Delta change factor methods perturb observed time series of precipitation and temperature with changes identified by GCM projections (e.g., Prudhomme et al., 2002, 2003; Lehner et al., 2006; Kay et al., 2006), more sophisticated variations of this method can account for seasonal effects and changes in variability (e.g., Hingray et al., 2007; Willems and Vrac, 2011). Statistical downscaling methods include (1) regression methods, which rely on multivariate statistics representing the link between large-scale climate predictors and local predictands (e.g., Bürger and Chen, 2005; Hessami et al., 2008), (2) resampling methods such as weather typing and climate analogs, which rely on the relationship between synoptic meteorological patterns and local predictands (e.g., Boé et al., 2006; Lafaysse et al., 2014; Pierce et al., 2014; Raynaud et al., 2017), and (3) stochastic weather generators, which simulate time series of chosen weather variables reproducing observed or pre-defined statistics or patterns conditioned on large-scale or synoptic circulations (e.g., Qian et al. 2002; Chen et al., 2018). Finally, dynamical downscaling methods consist of using Regional Climate Models (RCMs) which are similar to GCMs but have higher resolution and smaller modeling domains (see Rummukainen et al., 2010 and Xue et al., 2014 for reviews). Because RCMs produce their own bias in addition to inheriting bias from GCM boundary conditions, RCM projections are commonly further bias corrected and downscaled (Sunyer et al., 2012).

For analysis of possible future design floods, delta change factors are not recommended because they do not account for local differences that exist within the same GCM grid cell and for any feedback processes (Ekström et al., 2015). The computational cost associated with the use of RCMs is still sufficiently high to limit the temporal span of projections, reducing confidence in subsequent FFA. Also, RCM projections are usually available for only a few GCM/RCM realizations, and thus do not span the range of GCM internal variability. Rather, use of an ensemble of statistical methods is recommended, for two reasons. First, no single method out-performs others in intercomparison studies of statistical downscaling methods (Bronstert et al., 2007; Gutíerrez et al., 2018); thus, it is impossible to definitively choose one method. Second, downscaling contributes as much, if not more, uncertainty as the forcing scenario and GCM model (Bürger et al., 2013; Hingray and Saïd, 2014; Lafaysse et al., 2014); thus, it is necessary to estimate the uncertainty through the use of an ensemble. Regardless of what downscaling methods are chosen, it is crucial to evaluate the resulting climate scenarios for their climatic credibility and plausibility (i.e., are current climate conditions well-represented and is the realization of possible future climate physically sound) as well as their relevance and quality for hydrologic impact analysis (Bronstert et al., 2007).

The bias corrected and downscaled climate projections are then used to drive hydrologic models to simulate streamflow at a given location (e.g., Hingray et al. 2014). Although the comparison of different hydrologic models has a long-standing history in hydrology (e.g., Refsgaard and Knudsen, 1996), the implications of hydrologic modeling choices for assessing changing environmental conditions has recently emerged as a concern (e.g., Merz et al., 2011; Brigode et al., 2013; Mendoza et al., 2015). Hydrologic modeling choices, as discussed by Mendoza et al. (2016), include assuming time-invariant model parameters, model structure (e.g., Chen et al., 2013; Wi et al., 2015), the spatial resolution of input data (e.g., Lobligeois et al., 2014; Essou et al., 2016), the existence of multiple optimal parameter sets (e.g., Wi et al., 2015), calibration and validation periods (e.g., Coron et al., 2012, 2014) and the calibration objective function. Studies that have examined the effect of these choices on flood projection have found that while the importance of different sources of uncertainty varies by hydrologic regime and return period (Kay et al., 2009), the uncertainty from future climate can be as much or more than the uncertainty from the model structure (Booij, 2005; Steinschneider et al., 2015a) and that estimates of the 100-year can vary widely depending on the combination of modeling choices (Brigode et al., 2015; Wi et al., 2015). Due to the uncertainty stemming from hydrologic models, an ensemble of model structures and parameter sets is recommended, the latter could be inferred using various calibration and validation periods. If computational constraints make an ensemble undesirable, a simple sensitivity analysis could indicate the magnitude of uncertainty stemming from hydrologic models relative to GCMs and downscaling methods; if the hydrologic model uncertainty is of comparable or greater magnitude, then the ensemble should be reconsidered.

2.5.1.2. FLOOD PROJECTION VIA INFORMED PARAMETER APPROACHES

Informed parameter approaches use statistical models with time varying parameters to allow the flood distribution to be nonstationary (e.g., Serago and Vogel, 2018); in other words, Equation 2.1 becomes $Y(t) \sim f(\theta_1(t),..,\theta_K(t))$ and provides a peak flow distribution for each time step (usually annual) out to the planning horizon; for the most recent and exhaustive review of nonstationary statistical methods, see Salas et al. (2018). There are two main approaches. The first, termed the trend-informed approach, makes use of the historical trend to model non-stationarity in the peak flow distribution (e.g., El Adlouni et al., 2007;

Rootzén and Katz, 2013; Salas and Obeysekera, 2014; Luke et al., 2017). The second, termed the climate-informed approach (e.g., Sankarasubramanian and Lall, 2003; Kwon et al., 2008), is to model the temporal evolution of the flood distribution parameters using large-scale climate variables as covariates. A common practice is to try a variety of models (e.g., either trends or covariates for various distribution parameters or moments) and retain the best model, often assessed using a criterion such as the Akaike or Bayesian information criterion. However, assessment of the best model should also consider whether the modeled changes can be linked to physical processes and whether those links will remain valid under future climate conditions (e.g., Jiang et al., 2016). For example, a time-varying location parameter could correspond to changes in moisture content of storms and their intensity (Stedinger and Griffis, 2011) while a time-varying scale parameter could correspond to changes in storm features (e.g., winter storm vs. snowmelt).

In the trend-informed approach, model parameters are:

$$\theta_k(t) = \sum_{n=0}^{N} a_{k,n} t^n, \qquad 2.3$$

where $\theta_k(t)$ is the value of the kth parameter of the flood distribution for time t, $a_{k,n}$ are regression coefficients, and N is the regression order. Usually N equals 1 (i.e., linear relation) or 2 (i.e., quadratic relation). The trend-informed approach has been primarily used to model non-stationarity in the historical period for characterizing historical changes (e.g., Strupczewski et al., 2001; Delgado et al., 2010; Vogel et al., 2011; Prosdocimi et al., 2014; Šraj et al., 2016; Hu et al., 2017).

The use of the trend-informed approach for assessing future design values relies on the questionable assumption that historical trends will remain the same over the entire planning period (Jain and Lall, 2001; He et al., 2006; Sivapalan and Samuel, 2009; Blöschl and Montanari, 2010; Serinaldi and Kilsby, 2015; Luke et al., 2017). In an analysis of 1,250 stream gauges across the U.S., trend-informed models were rarely preferred even when a trend was detected in the first half of the record because it rarely persisted into the second half; however, for gauges where the physical processes underlying the trend did continue, an updated stationary model, which used the parameters of the trend-informed model at the end of the first half of the record, was generally preferred (Luke et al., 2017). Additionally, trend-informed models are often limited by sampling uncertainty and model structure uncertainty and robustness; for example, two models with comparable goodness-of-fit for the historical period can lead to significantly different flood quantiles for future periods (Serinaldi and Kilsby, 2015). Given these limitations, a high level of caution is needed when using trend-informed models to assess future flood design values.

Climate-informed model parameters are:

$$\theta_k(t) = \sum_{m=1}^{M_k} \sum_{n=0}^{N_{k,m}} a_{k,m,n} x_{k,m}^n(t), \qquad 2.4$$

where x_(k,m) is the value of the mth climate covariates at time t for the parameter $\theta_k(t)$,and $a_{k,m,n}$ are the regression coefficients. Compared to equation 2.3, the time variable has been replaced by covariates $x_{k,m}$, which are usually large-scale climate variables, but can also be non-climatic, such as a reservoir index (López and Francés, 2013) or carbon dioxide concentrations (Hirsch and Ryberg; 2012). The number of climate covariates will depend on the parameter and the regression order will depend on both the parameter and the number of covariates.

Implementing climate-informed models requires identifying the relevant climate patterns and/or variables that drive flood occurrence and magnitude over the study region. The choice of the relevant covariates usually follows from analysis of the hydro-climatology, scientific literature, and historic reports for the considered area (e.g., Delgado et al., 2012; Schlef et al., 2018a). In the case where identification of covariates is not straightforward, there are a variety of available methods: time series correlation (e.g., Kwon et al., 2008), correlation maps (e.g., Schlef et al., 2018a), composite analysis (e.g., Jain and Lall, 2001), weather typing (e.g., Robertson et al., 2015), simulation experiments (e.g., Cook, 1999), and use of a Bayesian frameworks to identify spatial patterns in gridded data (Renard and Lall, 2014). Projections of the covariates to drive changes in flood distribution parameters are usually obtained from GCMs (blue box on Figure 2-5). This requires the covariates to be well reproduced by GCMs, thus favoring covariates defined over large spatial domains and at monthly to seasonal temporal scales rather than daily localized precipitation used to force hydrologic models. For instance, maximum annual precipitation over a region should not be used as covariate because of low GCM skill (e.g., Boberg et al., 2007; Leander and Buishand, 2007; Wuebbles et al., 2014; Crétat et al., 2014). Similar to downscaling approaches (e.g., Greene et al., 2011), the use of both thermodynamic and dynamic covariates is expected to increase model robustness under climate change.

Climate-informed models have been used for analyzing or reconstructing historical variability of hydrological extremes (e.g., Griffis and Stedinger, 2007b; Kwon et al., 2008; Li and Tan, 2015; Bracken et al., 2018) and for projecting future peak flow (Delgado et al., 2014; Tramblay et al., 2014; Condon et al., 2015; Schlef et al., 2018a). Similar to the hydrologic simulation approach, the credibility of flood projections derived from climate-informed models is subject to limitations and uncertainties stemming from the use of GCMs, but unlike the simulation approach, avoids the uncertainty stemming from downscaling methods and hydrological models. Although GCM performance regarding highly-studied climate patterns such as ENSO and PDO is generally acceptable and well-characterized (e.g., Tashetto et al., 2014; Yim et al., 2015), this may not be the case for other climate variables relevant to a specific gauge location, such as soil moisture (Yuan and Quiring, 2017). Furthermore, GCMs often do not correctly reproduce correlations between climate patterns (e.g., Kim et al., 2017) and calculation of some large-scale climate indices from GCM outputs can be prohibitively difficult. Projections from climate-informed models are also highly influenced by the choice of covariates, which depends on the modeler's experience and knowledge of the study area and even the best combination of covariates is unlikely to fully explain peak flow variability (e.g., Delgado et al., 2012). Finally, uncertainty stemming from sampling, model structure, and robustness of covariates under future climate remains poorly characterized (e.g., Merz et al., 2014). Despite these limitations, climate-informed models are a relatively new and appealing alternative to flood projection via hydrologic simulation.

2.5.1.3. NONSTATIONARY RETURN PERIODS AND RISK ASSESSMENT FOR DESIGN

The definition of nonstationary return periods is still an area of on-going research and discussion. Flood projections derived from trend-informed models can use a revised concept of return period, defined as the mean expected time from the current date to the first occurrence of a flood event that exceeds the chosen design value (Cooley, 2013; Salas and Obeysekera, 2014) or equivalently, the time length for which the probability of exceedance is one (Parey et al., 2007). Calculation of this revised return period requires that time-varying probabilities of exceedance be summed to infinity (equation 4.3. in Cooley, 2013) and that the temporal evolution of flood events be correctly represented. This revised return period cannot be applied to projections derived from either the hydrologic simulation approach or climate-informed models because they do not meet those two conditions (i.e., most of the GCMs project climate for the next century only and cannot predict the temporality of events). Consequently, return periods for flood projections derived not using a trend-informed model are often calculated assuming stationarity within a chosen time window (e.g., 1950-2000 for a historic period and 2050-2100 for a future period) (Schlef et al. 2018a), which re-assumes stationarity and imposes short record lengths (e.g., 50 years) from which to calculate extreme values (e.g., the 100- or 500-year flood). Alternatives to either the revised return period or re-assuming stationarity include the design life level, which is the probability of the event occurring within the planning horizon (Rootzén and Katz, 2013), and reliability, which is the probability that an exceedance event will not occur over the planning period (i.e., expected performance, rather than expected failure) (Read and Vogel, 2015). Once nonstationary flood probabilities have been determined from the flood projections, as in the stationary case, design can be based on either pre-determined return periods or a risk-based approach (e.g., Schlef, 2018). Frequently, nonstationary risk-based approaches are embedded into sequential decision frameworks, such as a decision tree (cf. Loucks et al., 2005), to infer sequential adaptation measures (e.g., Manning et al., 2009; Gersonius et al., 2013; Rosner et al., 2014).

2.5.2. Probable Maximum Precipitation and Flood under non-stationarity from climate

Whether the PMP and subsequent PMF are likely to change at a given location is a subject of controversy. Jakob (2013) for instance highlighted a subtle change in the PMP definition in the most recent WMO guidelines (i.e., 2009 compared to 1986). The new definition states that "PMP is the theoretical maximum precipitation for a given duration under modern meteorological conditions", which emphasizes that the PMP may have changed in the past and could be modified in the future should the climate change. To date, there are no federal guidelines for dam design that account for potential changes in PMF, perhaps due to the perspective that "there is no compelling evidence that would support climate-related changes in PMFs" and that current design recommendations should be changed only when substantial research quantitatively demonstrates a link (USACE, 2016). However; recent academic research using both climate projections and historical trends, although still nascent, indicates there likely is a link. Significant increases in PMP have been shown for the whole contiguous U.S. (Kunkel et al., 2013), specific locations within the U.S. (Stratz and Hossain, 2014; Gangrade et al., 2018), and specific locations in Canada (Rousseau et al., 2014; Clavet-Gaumont et al., 2017). Regardless, to-date no method has become widely acknowledged for projection of future PMP and PMF under climate change, and thus we are unable to offer guidance on specific methods.

2.6. ADDRESSING DESIGN FLOOD UNCERTAINTY

At the most basic level, uncertainty in design flood values can be broadly conceptualized as being comprised of two components: aleatory uncertainty is essential, unavoidable unpredictability or chance (i.e., the residual unpredictability of events given models and parameters), while epistemic uncertainty is lack of knowledge or ignorance (i.e., roughly, lack of confidence about parameters and models) (Spiegelhalter & Riesch, 2011). However, the complexity of climate change has led to the concept of deep or severe uncertainty, which "results from myriad factors both scientific and social, and consequently is difficult to accurately define and quantify" (Kandlikar et al., 2005; for more discussion see also Lempert et al., 2004; Hall et al. 2007; Hall et al., 2012; Ray and Brown, 2015). Both aleatory and epistemic uncertainty are present under the assumption of stationarity; deep uncertainty is added under the assumption of non-stationarity. This section first describes uncertainty quantification and attribution, and then describes methods for incorporating uncertainty into design.

2.6.1. Uncertainty quantification and attribution

A common approach to quantifying aleatory and epistemic uncertainty is through confidence intervals (e.g., dashed curves in Figure 2-4), which have long been a topic of active research in hydrologic science (e.g., IACWD, 1982; Stedinger 1983; Ashkar et al., 1987; Hu, 1987; Schendel et al., 2015; England Jr. 2018). Under stationarity, confidence intervals are generally obtained through the variance of the design quantiles (e.g., Lu and Stedinger, 1992). Under non-stationarity, assessment of confidence intervals is less straightforward. Obeysekera and Salas (2014) described three methods: (1) a delta method, which is "based the large sample properties of maximum likelihood estimators", (2) a bootstrap method, which is "based on the bootstrap of standardized data, which are then used to fit an ensemble of nonstationary models" and (3) a profile likelihood method from Coles (2001), which "uses the log-likelihood function"; the delta and bootstrap methods are more computationally efficient but less accurate compared to the profile likelihood method.

For lack of a better approach, quantification of deep uncertainty has been generally based on the use of ensembles of forcing scenarios, initial conditions, and models or methods (e.g., the ensembles of GCM projections in IPCC, 2014). In this case, the spread of the ensemble is considered to provide some, albeit limited and usually underestimated, indication of the level of associated uncertainty (Stainforth et al., 2007b). There is significant research effort aimed at reducing ensemble spread by improving models; however, significant reductions are only possible with hundreds if not thousands of additional years of observations (Leach 2007). Alternatively, other efforts aim to achieve a more informative ensemble spread by selecting GCMs which demonstrate good performance or are independent (Knutti et al., 2013; Steinschneider et al., 2015b). However, since non-independent GCMs are more likely to project changes toward the same direction and/or with comparable magnitude, the use of an increasing number of non-independent models into an ensemble could lead to an apparent reduction of the uncertainty, although this reduction is likely to be meaningless (Knutti, 2010). The use of ensembles has led to many attribution studies (e.g., Hawkins and Sutton, 2009; Hingray and Saïd, 2014), which use methods of sensitivity analysis, such as ANOVA, to partition to the total uncertainty as represented by the ensemble spread, into portions attributable to specific sources, which may include forcing scenario, initial conditions, model structure, model parameters, etc. (Stainforth et al., 2007a).

2.6.2. Incorporating uncertainty into design

Under the assumption of stationarity, the common engineering approach to incorporating uncertainty in design is through safety factors; or freeboard in the context of levee or dam design, which in some cases is even applied to the already conservative PMF (e.g., FEMA, 2012; Shaw, 2009; NYC rules, 2013). Notably, such safety factors are usually independent of the confidence interval estimation of uncertainty, which is not generally used in design; as stated in Bulletin 17C, "application of confidence intervals in reaching water-resource planning decisions depends upon the needs of the user" (England Jr. et al., 2018) and proposed approaches are generally limited to research applications. If, for example, the risk of under-design must be reduced to 5%, a possible approach, although naïve, could be to use the value of the upper bound of 90% confidence interval could be used. While this use is straightforward and can easily be applied in practice, it overinterprets the mathematic definition of the intervals (e.g., Klemeš, 2002; Serinaldi and Kilsby, 2015). Another proposed use entails a cost-benefit analysis to determine a design value which accounts for epistemic uncertainty arising from parameter estimation (Botto et al., 2014, 2017; Gaume, 2018).

Under the assumption of non-stationarity, particularly as arises from climate change, there are three primary means of incorporating uncertainty in design: climate factors (i.e., the application of an additional safety factor specifically addressing the uncertainty from climate change), the prudent approach (i.e., use of information that is known with relatively high confidence to qualitatively inform whether additional protection should be considered in design) and robustness-based decision methods (i.e., finding a design which is satisficing over a wide array of plausible futures). As a side note, taking the (weighted) mean of the ensemble, which is an approach commonly used to summarize the oftentimes overwhelming amount of information from future climate projections, serves only to conceal the uncertainty and negatively impact characterization of extremes, rather than actively incorporate that uncertainty into design.

2.6.2.1. CLIMATE FACTORS

In the context of uncertainty associated with climate change, additional safety factors are called climate allowances or climate factors. Through a review of existing guidelines in Europe, Madsen et al. (2014) found that the use of climate factors with the explicit purpose of protecting against climate change is rare; exceptions include Germany, Norway, and the United Kingdom (Defra, 2006; Hennegriff et al., 2006; Lawrence and Hisdal, 2011; Environmental Agency, 2016). Although climate factors are easy to apply, such a "simplistic adjustment to peak flow estimates" is a result of "poorly understood impacts of future climate change" (e.g., Kuklicke and Demeritt, 2016). Climate factors lack flexibility because they are generally prescribed for a single time-horizon (one exception is the recent United Kingdom guidelines) and incorrectly estimate small-scale variability because they are usually defined by basin or political boundaries (Madsen et al., 2014). Calculation of climate factors is generally described in technical reports, which often lack clear descriptions of calculation methods and are based on the difference between historical and projected simulations of streamflow from GCMs. Furthermore, studies that assess climate factor performance are limited but needed, especially given that the resulting climate factors can be highly sensitive to the modeling choices (e.g., the first-generation of United Kingdom climate factors, see Reynard et al., 2005; Defra, 2006; Prudhomme et al., 2010).

2.6.2.2. PRUDENT APPROACH

Even under deep uncertainty, the direction of change of some key variables can be known with relatively high confidence; for example, temperature has and will continue to increase under climate change. This information can be used to qualitatively infer the likely direction of change in hydrological extremes (e.g., decreased snowpack would lead to smaller and earlier peak flows, or intensified extreme precipitation will increase peak flows), assuming sufficiently negligible feedback effects. Given this information, decision-makers might decide to opt for a prudent approach, based on the precautionary principle (Gollier and Treich, 2003). In the context of climate change mitigation, Kirkwood (2011) noted that "the prudent path lies somewhere between doing absolutely nothing about climate change and doing everything possible". In the context of decision-making for hydrologic design, the prudent approach consists of making design decisions based on expected changes in peak flows or their drivers that are known with relatively high confidence. The prudent approach is especially applicable for projects involving discrete choices, such as levee height or the size of a pump station. For example, given a discrete set of return periods used for design, if the 50-year flood was initially selected, but there was high confidence that future peak flows would increase, the prudent approach would choose the next discrete level, such as the 100-year flood. This is often justifiable when its marginal cost is small compared to the total infrastructure cost (Hallegate, 2009). An even more precautionary approach to flood risk, adapted by the Scottish Environment Protection Agency, is to prohibit development of flood sensitive projects in medium to high risk areas (SEPA, 2017).

2.6.2.3. ROBUSTNESS-BASED DECISION METHODS

To deal with deep uncertainty associated with climate change, Lempert (2002) introduced the idea of robustness. Unlike a risk-based approach, which finds an optimal design for one assumed future state of the world, a robustness-based approach finds a design that will be satisficing (i.e., perform well, Simon, 1956) for a large range of plausible futures. Multiple approaches using the concept of robustness have been developed: Robust Decision Making (Lempert, 2003), Info-Gap analysis (Ben-Haim, 2006), Scenario-Neutral approach (Prudhomme et al., 2010), and Decision Scaling (Brown et al., 2011, 2012). Stakhiv (2011) provides an extended discussion on the use of robust-decision making methods for water resource management under climate change. Both Info-Gap analysis and Decision Scaling have been specifically applied to design of flood infrastructure (Hine and Hall, 2010; Steinschneider et al., 2015a; Spence and Brown, 2016, 2018; Knighton et al., 2017) and the Scenario-Neutral approach has been applied to determine system vulnerability to hydrological extremes without prescribing a design value (Prudhomme et al., 2010). Despite the value of robustness-based decision methods, quantitative approaches to their implementation are relatively new and require increased application and improvement. For example, Hall et al. (2012) showed that Info-Gap and Robust Decision Making provide similar but not identical solutions for the same case study, yet note that the comparison improved understanding of the system and the proposed management options.

2.7. DISCUSSION AND CONCLUSION

This paper addresses the need for guidance on design for riverine floods under non-stationarity. In particular, Section 2.3 addresses evidence and drivers of non-stationarity as well as methods for detection and attribution. The key points are:

- Theoretical reasons/historical evidence indicate that flood hazard is possibly changing at some locations due to climate change
- For design of short-life infrastructure (< 30 years), the natural climate variability signal dominates the anthropogenic climate change signal; consequently, assuming stationarity may be adequate, but long historical records are recommended for inferring model parameters
- For design of long-life infrastructure (> 30 years), the assumption of stationarity may not be valid due to anthropogenic climate change impacting the magnitude, occurrence and typology of floods
- The structure inherent in natural climate variability can confound trend analysis and should be diagnosed and addressed, when possible.
- Methods for detecting change are imperfect; however, this effect can be reduced through regionalization and using long historical records

Section 2.4 addresses status quo flood design, which assumes stationarity. The key points are:

- Design of long-life infrastructure is usually based on either (1) FFA with pre-specified return periods or risk minimization or (2) the concepts of PMP and PMF
- If flood hazard is indeed nonstationary, the status quo can lead to costly over-design or dangerous under-design

Section 2.5 addresses methods for nonstationary flood design under climate change. The key points are:

- For nonstationary flood design, FFA must use projections from either hydrologic simulation or informed-parameter approaches; such projections rely on GCM projections of future climate which often have biased precipitation estimates and unknown future skill
- For the hydrologic simulation approach:
    - Bias correction alone and delta change downscaling methods are not recommended
    - Use of the projections currently available from RCMs are not recommended because the simulation periods are often too short to estimate flood frequency. The small number of GCM/RCM realizations does not allow a correct representation of the range of GCM internal variability.
    - The use of GCM projections can be improved through statistical downscaling but use of an ensemble of downscaling methods and evaluation of method credibility is recommended because no one method outperforms others and because downscaling adds considerable uncertainty

- Ensembles of hydrologic models are recommended in cases where the ensuing uncertainty is comparable to or greater than that from GCMs and downscaling methods. A sensitivity analysis could be used to infer the relative contribution of hydrological models to the total uncertainty
- For informed-parameter approaches:
  - Trend-informed models are not recommended for projection because historical trends will not necessarily persist during the entire planning period
  - Climate-informed models are an appealing alternative to hydrologic simulation, but are still a relatively new and untested methodology, and should be used as a complementary approach
- Nonstationary return periods are difficult to define and provide minimal design guidance; instead, the concepts of design life level and reliability, which transparently communicate probability of failure, are recommended for guiding design
- Since studies of changing PMP and PMF are nascent and no clear and well-established methods for assessing change exist, no method is recommended, despite the possibility that PMP and PMF are nonstationary under climate change

Section 2.6 addresses quantification and attribution of design flood uncertainty, and approaches to incorporating uncertainty associated with climate change into design. The key points are:

- Aleatory and epistemic uncertainty are present under the assumption of stationarity; deep uncertainty is added in the context of non-stationarity from climate change
- Confidence intervals are recommended for communicating aleatory and epistemic uncertainty, although their use for design is often unclear; ensembles are recommended for characterizing deep uncertainty, despite indicating only the lower bounds of the full uncertainty range
- For design, robustness-based decision methods are recommended to account for climate change uncertainty; if a robustness-based approach is not possible, the use of either climate factors or prudent approach should be considered

Some very clear, and relatively unsurprising, avenues of needed future research arise from this review. First, future research should aim to cultivate non-traditional sources of data that can be used to extend observed records, such as historical information (e.g., written records) and paleo-data (e.g., tree rings and sediment cores), and generate improved future records (e.g., satellite data and crowd-sourced streamflow measurement). Second, given the strengths and weaknesses of methods for flood hazard projection for FFA, future research should focus on improved understanding of the climate processes driving floods, more applications of the climate-informed approach, and methods for assessing the potential credibility of projections under climate change. For PMP and PMF, future research should address uncertainty reduction for PMP and PMF and development of methods for projection that can be widely accepted. Finally, more research is needed on application and evaluation of methods for incorporating (deep) uncertainty into design.

To conclude this discussion about design considerations under non-stationarity, specifically in the context of climate change, it is worth quoting Jakob (2013): “Design with change in mind. Not just climate. Think across disciplines”. Whether stationarity is “dead” (Milly et al., 2008), “alive” (Lins and Cohn, 2011), “immortal” (Montanari and Koutsoyiannis, 2014) or “undead” (Serinaldi and Kilsby, 2015) is not really important provided that the chosen approach to design can sufficiently represent the physical system and its evolution. The ultimate goal is robust (or even resilient) design that avoids both sunk capital costs and massive flood damages. Thus, when inference from climate projections lack significant change in flood drivers over the design horizon, the stationarity assumption may be preferred. However, in the presence of observed or theoretical evidence that flood hazard is changing or is likely to change in the future, then the non-stationarity assumption may be preferred. While this review focused on climate change, non-climatic factors also can cause changes in flood hazard, possibly on shorter time scales.

The impact of both climate and non-climate factors on flood hazard signals the "end of the static design paradigm" (Brown, 2010). This implies that the chosen design may fail more quickly than expected; hence the growing focus on resilient design across many disciplines, including flood design (e.g., Park et al., 2011; Sayers et al., 2012). Additionally, this implies that design should be able to tolerate various levels of failure, rather than existing solely in binary failure or non-failure states, meaning that non-structural flood solutions should be designed conjunctively with structural solutions. Finally, while a static infrastructure design may be optimal for the most likely future in a risk-based approach or satisficing for most plausible futures in a robustness-based approach, the end of static design promotes adaptive flood risk management which benefits from increasing knowledge through time (e.g. Hui et al. 2018). Adaptive flood risk management can reduce or delay initial and often large investment costs associated with infrastructure construction which may subsequently become unnecessary under future states of the world. There are two adaptive approaches which have been applied to flood risk management: (1) Dynamic Adaptive Policy Pathways (Haasnoot et al. 2013) and (2) real options analysis (e.g., Hino and Hall, 2017). These adaptive approaches should be combined with robustness-based approaches to promote design under non-stationarity from climate change.

2.8. REFERENCES

Aich, V., Liersch, S., Vetter, T., Andersson, J., Müller, E., and Hattermann, F. (2015) Climate or Land Use?—Attribution of Changes in River Flooding in the Sahel Zone. Water, 7(12), 2796–2820.

Allen, M. (2003) Liability for climate change. Nature, 421(6926), 891.

Alfieri, L., Burek, P., Feyen, L., and Forzieri, G. (2015) Global warming increases the frequency of river floods in Europe. Hydrol. Earth Syst. Sci., 19(5), 2247–2260.

Andrews, E. D., Antweiler, R. C., Neiman, P. J., and Ralph, F. M. (2004) Influence of ENSO on flood frequency along the California coast. Journal of Climate, 17(2), 337–348.

Archfield, S. A., Hirsch, R. M., Viglione, A., and Blöschl, G. (2016) Fragmented patterns of flood change across the United States. Geophysical Research Letters, 43(19), 10,232-10,239.

Armstrong, W. H., Collins, M. J., and 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.

Arnell, N. W. and Gosling, S. N. (2016) The impacts of climate change on river flood risk at the global scale. Climatic Change, 134(3), 387–401.

(ASCE) American Society of Civil Engineers (ed.) (2007) The New Orleans hurricane protection system: what went wrong and why: a report, Reston, Va, ASCE.

Ashkar, F., Bobée, B., and Fortier, L. (1987) "Confidence Intervals for Design Flood Events under Different Statistical Flood Models" in V. P. Singh (ed.), Application of Frequency and Risk in Water Resources. Dordrecht, Springer Netherlands, 139–151. [online] http://www.springerlink.com/index/10.1007/978-94-009-3955-4_11 (Accessed September 14, 2017).

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P. (2005) Potential impacts of a warming climate on water availability in snow-dominated regions. Nature, 438, 303–309.

Barros, A. P., Duan, Y., Brun, J., and Medina, M. A. (2014) Flood Nonstationarity in the Southeast and Mid-Atlantic Regions of the United States. Journal of Hydrologic Engineering, 19(10), 5014014.

Bayazit, M. (2015) Nonstationarity of Hydrological Records and Recent Trends in Trend Analysis: A State-of-the-art Review. Environmental Processes, 2(3), 527–542.

Beauchamp, J., Leconte, R., Trudel, M., and Brissette, F. (2013) Estimation of the summer-fall PMP and PMF of a northern watershed under a changed climate: PMP and PMF Estimation Under Climate Change. Water Resources Research, 49(6), 3852–3862.

Bengtsson, L., Hodges, K. I., and Roeckner, E. (2006) Storm Tracks and Climate Change. Journal of Climate, 19(15), 3518–3543.

Berg, P. and Haerter, J. O. (2013) Unexpected increase in precipitation intensity with temperature — A result of mixing of precipitation types? Atmospheric Research, 119, 56–61.

Berghuijs, W. R., Aalbers, E. E., Larsen, J. R., Trancoso, R., and Woods, R. A. (2017) Recent changes in extreme floods across multiple continents. Environmental Research Letters, 12(11), 114035.

Ben-Haim, Y. (2006). Info-gap decision theory: Decisions under severe uncertainty (2nd ed.). Cambridge, MA: Academic Press.

Beven, K. (2011) I believe in climate change but how precautionary do we need to be in planning for the future? Hydrological Processes, 25(9), 1517–1520.

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.

Blöschl, G. and Montanari, A. (2010) Climate change impacts—throwing the dice? Hydrological Processes, 24, 374–381.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A., Merz, B., Arheimer, B., et al. (2017) Changing climate shifts timing of European

floods. Science, 357(6351), 588–590.Boes, R. (ed.) (2011) “Inflow Design Flood and Dam Safety” in Dams and Reservoirs under Changing Challenges. CRC Press, 677–684.
Boberg F, Berg P, Thejll P, Christensen JH (2007) Analysis of temporal changes in precipitation intensities using PRUDENCE data. Danish Climate Centre Report 07-03, Copenhagen
Boes, R. (ed.) (2011) “Inflow Design Flood and Dam Safety” in Dams and Reservoirs under Changing Challenges. CRC Press, 677–684.
Boé, J., Terray, L., Habets, F., and Martin, E. (2006) A simple statistical-dynamical downscaling scheme based on weather types and conditional resampling. Journal of Geophysical Research, 111(D23), D23106.
Boé, J. and Terray, L. (2008) A Weather-Type Approach to Analyzing Winter Precipitation in France: Twentieth-Century Trends and the Role of Anthropogenic Forcing. Journal of Climate, 21(13), 3118–3133.
Booij, M. J. (2005) Impact of climate change on river flooding assessed with different spatial model resolutions. Journal of Hydrology, 303(1–4), 176–198.
Botto, A., Ganora, D., Claps, P., and Laio, F. (2017) Technical Note: Design flood under hydrological uncertainty. Hydrology and Earth System Sciences Discussions, 1–10.
Botto, A., Ganora, D., Laio, F., and Claps, P. (2014) Uncertainty compliant design flood estimation. Water Resources Research, 50(5), 4242–4253.
Bracken, C., Holman, K. D., Rajagopalan, B., and Moradkhani, H. (2018) A Bayesian Hierarchical Approach to Multivariate Nonstationary Hydrologic Frequency Analysis. Water Resources Research, 54(1), 243–255.
Brigode, P., Oudin, L., and Perrin, C. (2013) Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change? Journal of Hydrology, 476, 410–425.
Brigode, P., Paquet, E., Bernardara, P., Gailhard, J., Garavaglia, F., Ribstein, P., et al. (2015) Dependence of model-based extreme flood estimation on the calibration period: case study of the Kamp River (Austria). Hydrological Sciences Journal, 60(7–8), 1424–1437.
Bronstert, A., Kolokotronis, V., Schwandt, D., and Straub, H. (2007) Comparison and evaluation of regional climate scenarios for hydrological impact analysis: General scheme and application example. International Journal of Climatology, 27(12), 1579–1594.
Brown, C. (2010) The end of reliability. Journal of water resources planning and management, 136(2), 143–145.
Brown, C., Werick, W., Leger, W., and Fay, D. (2011) A Decision-Analytic Approach to Managing Climate Risks: Application to the Upper Great Lakes. JAWRA Journal of the American Water Resources Association, 47(3), 524–534.
Brown, C., Ghile, Y., Laverty, M., and Li, K. (2012) Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector. Water Resour. Res, 48(W09537).
Bürger, G. and Chen, Y. (2005) Regression-based downscaling of spatial variability for hydrologic applications. Journal of Hydrology, 311(1–4), 299–317.
Bürger, G., Sobie, S. R., Cannon, A. J., Werner, A. T., and Murdock, T. Q. (2013) Downscaling Extremes: An Intercomparison of Multiple Methods for Future Climate. Journal of Climate, 26(10), 3429–3449.
Camici, S., Brocca, L., Melone, F., and Moramarco, T. (2014) Impact of Climate Change on Flood Frequency Using Different Climate Models and Downscaling Approaches. Journal of Hydrologic Engineering, 19(8), 4014002.
Chen, X., Yang, T., Wang, X., Xu, C.-Y., and Yu, Z. (2013) Uncertainty Intercomparison of Different Hydrological Models in Simulating Extreme Flows. Water Resources Management, 27(5), 1393–1409.
Chen, J., Chen, H., and Guo, S. (2018) Multi-site precipitation downscaling using a stochastic weather generator. Climate Dynamics, 50(5–6), 1975–1992.
Clavet-Gaumont, J., Huard, D., Frigon, A., Koenig, K., Slota, P., Rousseau, A., Klein, I., Thiémonge, N., Houdré, F., Perdikaris, J., Turcotte, R., Lafleur, J., Larouche, B., 2017. Probable maximum flood in a changing climate: An overview for Canadian basins. J. Hydrol. Reg. Stud. 13, 11–25. https://doi.org/10.1016/j.ejrh.2017.07.003
Coles, G.S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, New York, p. 208.
Condon, L. E., Gangopadhyay, S., and Pruitt, T. (2015) Climate change and non-stationary flood risk for the upper Truckee River basin. Hydrology and Earth System Sciences, 19(1), 159–175.
Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., et al. (2012) Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. Water Resources Research, 48(5), W05552.
Coron, L., Andréassian, V., Perrin, C., Bourqui, M., and Hendrickx, F. (2014) On the lack of robustness of hydrologic models regarding water balance simulation: a diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments. Hydrology and Earth System Sciences, 18(2), 727–746.
Cook, K. H. (1999) Generation of the African Easterly Jet and Its Role in Determining West African Precipitation. Journal of Climate, 12(5), 1165–1184.
Cooley, D. (2013). “Return periods and return levels under climate change.”Chapter 4, Extremes in a changing climate: Detection, analysis and uncertainty, A. AghaKouchak, D. Easterling, and K. Hsu, eds., Vol. 65,Springer, New York.
Crétat, J., Vizy, E. K., and Cook, K. H. (2014) How well are daily intense rainfall events captured by current climate models over Africa? Climate Dynamics, 42(9–10), 2691–2711.
Dai, A. (2006) Precipitation Characteristics in Eighteen Coupled Climate Models. Journal of Climate, 19(18), 4605–4630.
Dankers, R. and Feyen, L. (2009) Flood hazard in Europe in an ensemble of regional climate scenarios. Journal of Geophysical Research, 114(D16)..

Dankers, R., Arnell, N. W., Clark, D. B., Falloon, P. D., Fekete, B. M., Gosling, S. N., et al. (2014) First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. Proceedings of the National Academy of Sciences, 111(9), 3257–3261.

Debs, I., Sparks, D., Rousselle, J., and Birikundavyi, S. (1999) Évaluation des méthodes utilisées pour l'estimation de la crue maximale probable en régions nordiques. Canadian Journal of Civil Engineering, 26(3), 355–367.

Defra (Department for Environment Food and Rural Affairs), 2006. Flood and Coastal Defence Appraisal Guidance (FCDPAG3), Economic Appraisal Supplementary Note to Operating Authorities – Climate Change Impacts. Department for Environment, Food and Rural Affairs, London, 9 pp. http://webarchive.nationalarchives.gov.uk/20090703211502/http://www.defra.gov.uk/environ/fcd/pubs/pagn/climatechangeupdate.pdf (Last access 2017-9-13).

Delgado, J. M., Apel, H., and Merz, B. (2010) Flood trends and variability in the Mekong river. Hydrology and Earth System Sciences, 14(3), 407–418.

Delgado, J. M., Merz, B., and Apel, H. (2012) A climate-flood link for the lower Mekong River. Hydrology and Earth System Sciences, 16(5), 1533–1541.

Delgado, J. M., Merz, B., and Apel, H. (2014) Projecting flood hazard under climate change: an alternative approach to model chains. Natural Hazards and Earth System Science, 14(6), 1579–1589.

Deser, C., Knutti, R., Solomon, S., and Phillips, A. S. (2012a) Communication of the role of natural variability in future North American climate. Nature Climate Change, 2(11), 775–779.

Deser, C., Phillips, A., Bourdette, V., and Teng, H. (2012b) Uncertainty in climate change projections: the role of internal variability. Climate Dynamics, 38(3–4), 527–546.

Easterling, D. R., Kunkel, K. E., Wehner, M. F., and Sun, L. (2016) Detection and attribution of climate extremes in the observed record. Weather and Climate Extremes, 11, 17–27.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J. (2012) Should we apply bias correction to global and regional climate model data? Hydrology and Earth System Sciences, 16(9), 3391–3404.

Elmer, F., Hoymann, J., Düthmann, D., Vorogushyn, S., and Kreibich, H. (2012) Drivers of flood risk change in residential areas. Natural Hazards and Earth System Science, 12(5), 1641–1657.

Environmental Agency (2016) Guidance: Flood risk assessments: climate change allowances, in Flooding and coastal change. https://www.gov.uk/guidance/flood-risk-assessments-climate-change-allowances#vulnerability (Last access 2017-09-13)

England, J.F., Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas, W.O., Jr., Veilleux, A.G., Kiang, J.E., and Mason, R.R., Jr., 2018, Guidelines for determining flood flow frequency—Bulletin 17C: U.S. Geological Survey Techniques and Methods, book 4, chap. B5, 148 p., https://doi.org/10.3133/tm4B5.

Ekström, M., Grose, M. R., and Whetton, P. H. (2015) An appraisal of downscaling methods used in climate change research. Wiley Interdisciplinary Reviews: Climate Change, 6(3), 301–319.

El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R., and Bobée, B. (2007) Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resources Research, 43(3), W03410, doi:10.1029/2005WR004545.

Essou, G. R. C., Sabarly, F., Lucas-Picher, P., Brissette, F., and Poulin, A. (2016) Can Precipitation and Temperature from Meteorological Reanalyses Be Used for Hydrological Modeling? Journal of Hydrometeorology, 17(7), 1929–1950.

FEMA. (2011). Answers to Questions About the NFIP, (March), 70. (Available at: https://www.fema.gov/media-library-data/20130726-1438-20490-1905/f084_atq_11aug11.pdf, last access: 18-1-2019)

FEMA (2012) Summary of Current State Hydrologic Design Guidelines, Federal Emergency Management Agency. (available online: https://www.fema.gov/media-library-data/20130726-1849-25045-9688/04_hydrosafetydam_ch_9_11.pdf (Accessed August 22, 2018).

Fernandes, W., Naghettini, M., and Loschi, R. (2010) A Bayesian approach for estimating extreme flood probabilities with upper-bounded distribution functions. Stochastic Environmental Research and Risk Assessment, 24(8), 1127–1143.

Fischer, E. M. and Knutti, R. (2014) Detection of spatially aggregated changes in temperature and precipitation extremes: Fischer and Knutti: Detection of changes in extremes. Geophysical Research Letters, 41(2), 547–554.

Fischer, E. M. and Knutti, R. (2016) Observed heavy precipitation increase confirms theory and early models. Nature Climate Change, 6(11), 986–991.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C. (2007) Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology, 27(12), 1547–1578.

Galloway, G. E. (2011) If Stationarity is Dead, What Do We Do Now? JAWRA Journal of the American Water Resources Association, 47(3), 563–570. doi: 10.1111/j.1752-1688.2011.00550.x

Gangrade, S., Kao, S.-C., Naz, B.S., Rastogi, D., Ashfaq, M., Singh, N., Preston, B.L., 2018. Sensitivity of Probable Maximum Flood in a Changing Environment. Water Resour. Res. https://doi.org/10.1029/2017WR021987

Gaume, E. (2018) Flood frequency analysis: The Bayesian choice. Wiley Interdisciplinary Reviews: Water, 5(4), e1290.

George, S. S. and Nielsen, E. (2003) Palaeoflood records for the Red River, Manitoba, Canada, derived from anatomical tree-ring signatures. The Holocene, 13(4), 547–555.

Gersonius, B., Ashley, R., Pathirana, A., and Zevenbergen, C. (2013) Climate change uncertainty: building flexibility into water and flood risk infrastructure. Climatic Change, 116(2), 411–423.

Gollier, C., Treich, N., 2003. Decision-making under scientific uncertainty: the economics of the precautionary principle. J. Risk Uncertain. 27, 77–103.

Greene, A. M., Robertson, A. W., Smyth, P., and Triglia, S. (2011) Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model. Quarterly Journal of the Royal Meteorological Society, 137(655), 347–359.
Griffis, V. W. and Stedinger, J. R. (2007) “Incorporating Climate Change and Variability into Bulletin 17B LP3 Model” in World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat, 1–8. American Society of Civil Engineers. Available online : http://ascelibrary.org/doi/10.1061/40927%28243%2969 (Accessed February 14, 2017).
Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., et al. (2018) An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. International Journal of Climatology. [online] http://doi.wiley.com/10.1002/joc.5462 (Accessed January 29, 2019).
Haasnoot, M., Kwakkel, J. H., Walker, W. E., and ter Maat, J. (2013) Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. Global Environmental Change, 23(2), 485–498.
Hall, J., Fu, G., and Lawry, J. (2007) Imprecise probabilities of climate change: aggregation of fuzzy scenarios and model uncertainties. Climatic Change, 81(3–4), 265–281.
Hall, J. W., Lempert, R. J., Keller, K., Hackbarth, A., Mijere, C., and McInerney, D. J. (2012) Robust Climate Policies Under Uncertainty: A Comparison of Robust Decision Making and Info-Gap Methods. Risk Analysis, 32(10), 1657–1672.
Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., et al. (2014) Understanding flood regime changes in Europe: a state-of-the-art assessment. Hydrology and Earth System Sciences, 18(7), 2735–2772.
Hallegatte, S. (2009) Strategies to adapt to an uncertain climate change. Global Environmental Change, 19(2), 240–247.
Hamed, K. H. (2008) Tred detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. Journal of Hydrology, 349(3–4), 350–363.
Hannaford, J. and Marsh, T. J. (2008) High-flow and flood trends in a network of undisturbed catchments in the UK. International Journal of Climatology, 28(10), 1325–1338.
Harrigan, S., Murphy, C., Hall, J., Wilby, R. L., and Sweeney, J. (2014) Attribution of detected changes in streamflow using multiple working hypotheses. Hydrology and Earth System Sciences, 18(5), 1935–1952.
Hawkins, E. and Sutton, R. (2011) The potential to narrow uncertainty in projections of regional precipitation change. Climate Dynamics, 37(1–2), 407–418.
He, Y., Bardossy, A., and Juergen, B. (2006) “Non-stationarity flood frequency analysis in southern Germany” in Philadelphia, United States, Drexel University, College of Engineering.
Hennegriff, W., Kolokotronis, V., Weber, H., and Bartels, H. (2006) Climate Change and Floods – Findings and Adaptation Strategies for Flood Protection. KA - Abwasser, Abfall, 53(8). [online] http://citeseerx.ist.psu.edu/viewdoc/citations?doi=10.1.1.529.8978.
Hershfield, D. . (1961) Estimating the probable maximum precipitation. Proc. Am. Soc. Civil Eng., J. Hydraulics Division, 87, 99–106.
Hessami, M., Gachon, P., Ouarda, T. B. M. J., and St-Hilaire, A. (2008) Automated regression-based statistical downscaling tool. Environmental Modelling & Software, 23(6), 813–834.
Hine, D. and Hall, J. W. (2010) Information gap analysis of flood model uncertainties and regional frequency analysis. Water Resources Research, 46(1). doi: 10.1029/2008WR007620
Hingray, B., Mouhous, N., Mezghani, A., Bogner, K., Schaefli, B., and Musy, A. (2007) Accounting for global-mean warming and scaling uncertainties in climate change impact studies: application to a regulated lake system. Hydrology and Earth System Sciences, 11(3), 1207–1226.
Hingray, B., Picouet, C., and Musy, A. (2014) Hydrology: a science for engineers, Boca Raton, FL, CRC Press, Taylor & Francis Group.
Hingray, B. and Saïd, M. (2014) Partitioning Internal Variability and Model Uncertainty Components in a Multimember Multimodel Ensemble of Climate Projections. Journal of Climate, 27(17), 6779–6798.
Hino, M. and Hall, J. W. (2017) Real Options Analysis of Adaptation to Changing Flood Risk: Structural and Nonstructural Measures. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 3(3), 4017005.
Hirabayashi, Y., Kanae, S., Emori, S., Oki, T., and Kimoto, M. (2008) Global projections of changing risks of floods and droughts in a changing climate. Hydrological Sciences Journal, 53(4), 754–772.
Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., et al. (2013) Global flood risk under climate change. Nature Climate Change, 3(9), 816–821.
Hirsch, R. M. (2011) A Perspective on Nonstationarity and Water Management1: A Perspective on Nonstationarity and Water Management. JAWRA Journal of the American Water Resources Association, 47(3), 436–446.
Hirsch, R. M. and 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.
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.
Hofmann, D. J. (1987) Perturbations to the global atmosphere associated with the El Chichon volcanic eruption of 1982. Reviews of Geophysics, 25(4), 743.
Hu, S. (1987) Determination of confidence intervals for design floods. Journal of Hydrology, 96(1–4), 201–213.
Hu, Y., Liang, Z., Chen, X., Liu, Y., Wang, H., Yang, J., et al. (2017) Estimation of design flood using EWT and ENE metrics and uncertainty analysis under non-stationary conditions. Stochastic Environmental Research and Risk Assessment, 31(10), 2617–2626.

Hui, R., Herman, J., Lund, J., and Madani, K. (2018) Adaptive water infrastructure planning for nonstationary hydrology. Advances in Water Resources, 118, 83–94.
Hulme, M., Barrow, E. M., Arnell, N. W., Harrison, P. A., Johns, T. C., and Downing, T. E. (1999) Relative impacts of human-induced climate change and natural climate variability. Nature, 397(6721), 688–691.
Huntington, T. G. (2006) Evidence for intensification of the global water cycle: Review and synthesis. Journal of Hydrology, 319(1–4), 83–95.
Hurst, H. E. (1951) Long term storage capacities of reservoirs. Trans. Am. Soc. Civil Engrs 116, 776–808.
(IACWD) Interagency Advisory Committee on Water Data, 1982. Guidelines for Determining Flood Flow Frequency, Bulletin 17B of the Hydrology Committee, USGS, Office of Water Data Coordination, Reston, VA, March 1982.
IPCC (2014) Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the IPCC Fifth Assessment Report., Cambridge, Cambridge University Press.
Ishak, E. H., Rahman, A., Westra, S., Sharma, A., and Kuczera, G. (2013) Evaluating the non-stationarity of Australian annual maximum flood. Journal of Hydrology, 494, 134–145.
Jakob, D. (2013) “Nonstationarity in Extremes and Engineering Design” in A. AghaKouchak, D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian (eds.), Extremes in a Changing Climate. Dordrecht, Springer Netherlands, 363–417.
Jain, S. and Lall, U. (2001) Floods in a changing climate: Does the past represent the future? Water Resources Research, 37(12), 3193–3205.
Jenkinson, A. F. (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society, 81(348), 158–171.
Jiang, P., Yu, Z., Gautam, M. R., Yuan, F., and Acharya, K. (2016) Changes of storm properties in the United States: Observations and multimodel ensemble projections. Global and Planetary Change, 142, 41–52.
Jonkman, S. N., Brinkhuis-Jak, M., and Kok, M. (2004) Cost benefit analysis and flood damage mitigation in the Netherlands. Heron, 49(1). [online] https://repository.tudelft.nl/islandora/object/uuid:8507e21d-8427-4ba5-9a15-329237ada846/datastream/OBJ/view (Accessed September 13, 2017).
Kandlikar, M., Risbey, J., and Dessai, S. (2005) Representing and communicating deep uncertainty in climate-change assessments. Comptes Rendus Geoscience, 337(4), 443–455.
Katz, R. W., Parlange, M. B., and Naveau, P. (2002) Statistics of extremes in hydrology. Advances in Water Resources, 25(8–12), 1287–1304.
Katz, R. W. (2013) “Statistical Methods for Nonstationrity Extremes” in A. AghaKouchak, D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian (eds.), Extremes in a Changing Climate. Water Science and Technology Library. Dordrecht, Springer Netherlands, 23.
Kay, A. L., Jones, R. G., and Reynard, N. S. (2006) RCM rainfall for UK flood frequency estimation. II. Climate change results. Journal of Hydrology, 318(1–4), 163–172.
Kay, A. L., Davies, H. N., Bell, V. A., and Jones, R. G. (2009) Comparison of uncertainty sources for climate change impacts: flood frequency in England. Climatic Change, 92(1–2), 41–63.
Kendall, M.G. (1975) Rank Correlation Methods. 4th Edition, Charles Griffin, London.
Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J.-C., Gachon, P., and Bobée, B. (2006) Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. Journal of Hydrology, 329(3–4), 534–552.
Khaliq, M. N., Ouarda, T. B. M. J., Gachon, P., Sushama, L., and St-Hilaire, A. (2009) Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers. Journal of Hydrology, 368(1–4), 117–130.
Khintchine, A. (1934), Korrelationstheorie der station€aren stochastischen Prozesse, Math. Ann., 109, 604–615.
Kiem, A. S., Franks, S. W., and Kuczera, G. (2003) Multi-decadal variability of flood risk. Geophysical Research Letters, 30(2).
Kim, S., Son, H.-Y., and Kug, J.-S. (2017) How well do climate models simulate atmospheric teleconnections over the North Pacific and East Asia associated with ENSO? Climate Dynamics, 48(3–4), 971–985.
Kirkwood, John B. (2011) A Prudent Approach to Climate Change. Seattle Journal of Environmental Law: Vol. 1 : Iss. 1 , Article 8.
Klemeš V. Risk analysis: the unbearable cleverness of bluffing. In: Bogardi JJ, Kundzewicz ZW, editors. Risk, reliability, uncertainty, and robustness of water resource systems. International hydrology series. Cambridge University Press; 2002. p. 22–9. http://dx.doi.org/10.1017/ CBO9780511546006.003.
Knighton, J., Steinschneider, S., and Walter, M. T. (2017) A Vulnerability-Based, Bottom-up Assessment of Future Riverine Flood Risk Using a Modified Peaks-Over-Threshold Approach and a Physically Based Hydrologic Model. Water Resources Research, 53(12), 10043–10064.
Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K., Holland, G., Landsea, C., et al. (2010) Tropical cyclones and climate change. Nature Geoscience, 3(3), 157–163.
Knutti, R. (2010) The end of model democracy?: An editorial comment. Climatic Change, 102(3–4), 395–404.
Knutti, R., Masson, D., and Gettelman, A. (2013) Climate model genealogy: Generation CMIP5 and how we got there. Geophysical Research Letters, 40(6), 1194–1199.
Koerth-Baker, M. (2017) It’s Time To Ditch The Concept Of “100-Year Floods.” fivethirtyeight. [online] https://fivethirtyeight.com/features/its-time-to-ditch-the-concept-of-100-year-floods/.
Konrad, C. P. (2003) Effects of urban development on floods, US Geological Survey Fact Sheet FS-076-03. Available at:

https://pubs.usgs.gov/fs/fs07603/pdf/fs07603.pdf
Kolmogorov, A. N. (1931), Uber die analytischen Methoden in der Wahrscheinlichkcitsrechnung, Math. Ann., 104, 415–458. (English translation, pp. 62–108, Kluwer Acad., Dordrecht, Netherlands, 1992).
Kolmogorov, A. N. (1938), A simplified proof of the Birkhoff-Khinchin ergodic theorem, Usp. Mat. Nauk., 5, 52–56. (English edition, pp. 271–276, Kluwer Acad., Dordrecht, Netherlands, 1991).
Koutsoyiannis, D. (2003) Climate change, the Hurst phenomenon, and hydrological statistics. Hydrological Sciences Journal, 48(1), 3–24.
Koutsoyiannis, D. (2006) Nonstationarity versus scaling in hydrology. Journal of Hydrology, 324(1–4), 239–254.
Kuklicke, C. and Demeritt, D. (2016) Adaptive and risk-based approaches to climate change and the management of uncertainty and institutional risk: The case of future flooding in England. Global Environmental Change, 37, 56–68.
Kundzewicz, Z. W. and Robson, A. J. (2004) Change detection in hydrological records—a review of the methodology. Hydrological Sciences Journal, 49(1), 7–19.
Kundzewicz, Z. W. & Stakhiv, E. Z. (2010) Are climate models “ready for prime time” in water resources management applications, or is more research needed? Editorial. Hydrol. Sci. J. 55(7), 1085–1089.
Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., et al. (2013) Flood risk and climate change: global and regional perspectives. Hydrological Sciences Journal, 59(1), 1–28.
Kundzewicz, Z. W., Krysanova, V., Dankers, R., Hirabayashi, Y., Kanae, S., Hattermann, F. F., et al. (2017) Differences in flood hazard projections in Europe - their causes and consequences for decision making. Hydrological Sciences Journal, 62:1, 1-14 Doi: 10.1080/02626667.2016.1241398
Kunkel, K.E., Karl, T.R., Easterling, D.R., Redmond, K., Young, J., Yin, X., Hennon, P., 2013. Probable maximum precipitation and climate change. Geophys. Res. Lett. 40, 1402–1408. https://doi.org/10.1002/grl.50334
Kwon, H.-H., Brown, C., and Lall, U. (2008) Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling. Geophysical Research Letters, 35(5).
Lafaysse, M., Hingray, B., Mezghani, A., Gailhard, J., and Terray, L. (2014) Internal variability and model uncertainty components in future hydrometeorological projections: The Alpine Durance basin. Water Resources Research, 50(4), 3317–3341.
Lammersen, R., Engel, H., van de Langemheen, W., and Buiteveld, H. (2002) Impact of river training and retention measures on flood peaks along the Rhine. Journal of Hydrology, 267(1–2), 115–124.
Lawrence, D., Hisdal, H., (2011) Hydrological Projections for Flooding in Norway under a Future Climate. NVE Report 5-2011. Norwegian Water Resources and Energy Directorate, Oslo, 47 pp. ISBN: 978-82-410-0753-8.
Leach, A. J. (2007) The climate change learning curve. Journal of Economic Dynamics and Control, 31(5), 1728–1752.
Leander, R. and Buishand, T. A. (2007) Resampling of regional climate model output for the simulation of extreme river flows. Journal of Hydrology, 332(3–4), 487–496.
Lee, D., Ward, P., and Block, P. (2018) Attribution of Large-Scale Climate Patterns to Seasonal Peak-Flow and Prospects for Prediction Globally. Water Resources Research, 54(2), 916–938.
Lempert, R. J. (2002) A new decision sciences for complex systems. Proceedings of the National Academy of Sciences, 99(suppl 3), 7309–7313.
Lempert, R. J., Popper, S. W., and Bankes, S. C. (2003) Shaping the next one hundred years: new methods for quantitative, long-term policy analysis, Santa Monica, CA, RAND.
Lempert, R., Nakicenovic, N., Sarewitz, D., and Schlesinger, M. (2004) Characterizing climate-change uncertainties for decision-makers. An editorial essay. Climatic Change, 65(1), 1–9.
Lenderink, G. and van Meijgaard, E. (2008) Increase in hourly precipitation extremes beyond expectations from temperature changes. Nature Geoscience, 1(8), 511–514.
Leng, G., Tang, Q., Huang, S., Zhang, X., and Cao, J. (2016) Assessments of joint hydrological extreme risks in a warming climate in China. International Journal of Climatology, 36(4), 1632–1642.
Lehner, B., Döll, P., Alcamo, J., Henrichs, T., and Kaspar, F. (2006) Estimating the Impact of Global Change on Flood and Drought Risks in Europe: A Continental, Integrated Analysis. Climatic Change, 75(3), 273–299.
Li, J. and Tan, S. (2015) Nonstationary Flood Frequency Analysis for Annual Flood Peak Series, Adopting Climate Indices and Check Dam Index as Covariates. Water Resources Management, 29(15), 5533–5550.
Lins, H. F. and Cohn, T. A. (2011) Stationarity: Wanted Dead or Alive? JAWRA Journal of the American Water Resources Association, 47(3), 475–480.
Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., and Loumagne, C. (2014) When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. Hydrology and Earth System Sciences, 18(2), 575–594.
López, J. and Francés, F. (2013) Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. Hydrology and Earth System Sciences, 17(8), 3189–3203.
Loucks, D. P., van Beek, E., Stedinger, J. R., Dijkman, J. P. M., and Villars, M. T. (2005) Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications, Paris : UNESCO.
Lu, L.-H. and Stedinger, J. R. (1992) Variance of two- and three-parameter GEV/PWM quantile estimators: formulae, confidence intervals, and a comparison. Journal of Hydrology, 138(1–2), 247–267.
Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R., and Sanders, B. F. (2017) Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. Water Resources Research, 53 doi:

10.1002/2016WR019676.
Lund, J. R. (2002) Floodplain planning with risk-based optimization. Journal of water resources planning and management, 128(3), 202–207.
Madsen, H., Lawrence, D., Lang, M., Martinkova, M., and Kjeldsen, T. R. (2014) Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. Journal of Hydrology, 519, 3634–3650.
Major, J. and Mark, L. (2006) Peak flow responses to landscape disturbances caused by the cataclysmic 1980 eruption of Mount St. Helens, Washington. Geological Society of America Bulletin, 118(7/8), 938–958.
Mallakpour, I. and Villarini, G. (2016a) Investigating the relationship between the frequency of flooding over the central United States and large-scale climate. Advances in Water Resources, 92, 159–171.
Mallakpour, I. and Villarini, G. (2016b) A simulation study to examine the sensitivity of the Pettitt test to detect abrupt changes in mean. Hydrological Sciences Journal, 61(2), 245–254.
Manning, L. J., Hall, J. W., Fowler, H. J., Kilsby, C. G., and Tebaldi, C. (2009) Using probabilistic climate change information from a multimodel ensemble for water resources assessment. Water Resources Research, 45(11).
Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., et al. (2010) Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Reviews of Geophysics, 48(3). [online] http://eprints.uni-kiel.de/11662/ (Accessed March 19, 2013).
Martins, E. S. and Stedinger, J. R. (2000) Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resources Research, 36(3), 737–744.
Martel, J.-L., Mailhot, A., Brissette, F., and Caya, D. (2018) Role of Natural Climate Variability in the Detection of Anthropogenic Climate Change Signal for Mean and Extreme Precipitation at Local and Regional Scales. Journal of Climate, 31(11), 4241–4263.
Maurer, E. P., Das, T., and Cayan, D. R. (2013) Errors in climate model daily precipitation and temperature output: time invariance and implications for bias correction. Hydrology and Earth System Sciences, 17(6), 2147–2159.
McCabe, G. J. and Clark, M. P. (2005) Trends and variability in snowmelt runoff in the western United States. Journal of Hydrometeorology, 6(4), 476–482.
McMillan, H., Jackson, B., and Poyck, S. (2010) Flood Risk Under Climate Change: A framework for assessing the impacts of climate change on river flow and floods, using dynamically-downscalled climate scenarios - A case study for the Uawa (East cape) and Waihou (Northland) catchments., Ministry of Agriculture and Forestry. [online] https://www.niwa.co.nz/sites/niwa.co.nz/files/import/attachments/chc2010_033_Flood_Risk_CC.pdf (Accessed September 11, 2017).
Mediero, L., Santillán, D., Garrote, L., and Granados, A. (2014) Detection and attribution of trends in magnitude, frequency and timing of floods in Spain. Journal of Hydrology, 517, 1072–1088.
Mehran, A., AghaKouchak, A., and Phillips, T. J. (2014) Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations. Journal of Geophysical Research: Atmospheres, 119(4), 1695–1707.
Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., et al. (2015) Effects of Hydrologic Model Choice and Calibration on the Portrayal of Climate Change Impacts. Journal of Hydrometeorology, 16(2), 762–780.
Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D., et al. (2016) How do hydrologic modeling decisions affect the portrayal of climate change impacts?: Subjective Hydrologic Modelling Decisions in Climate Change Impacts. Hydrological Processes, 30(7), 1071–1095.
Merz, R. and Blöschl, G. (2003) A process typology of regional floods. Water Resources Research, 39(12).
Merz, B. and Thieken, A. H. (2009) Flood risk curves and uncertainty bounds. Natural Hazards, 51(3), 437–458.
Merz, B., Hall, J., Disse, M., and Schumann, A. (2010) Fluvial flood risk management in a changing world. Natural Hazards and Earth System Sciences, 10(3), 509–527.
Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., and Hundecha, Y. (2012) HESS Opinions “More efforts and scientific rigour are needed to attribute trends in flood time series.” Hydrology and Earth System Sciences, 16(5), 1379–1387.
Merz, B., Aerts, J., Arnbjerg-Nielsen, K., Baldi, M., Becker, A., Bichet, A., et al. (2014) Floods and climate: emerging perspectives for flood risk assessment and management. Natural Hazards and Earth System Science, 14(7), 1921–1942.
Merz, R., Parajka J., and Blöschl, G. (2011), Time stability of catchment model parameters: Implications for climate impact analyses, Water Resour. Res. 47, W02531, doi: 10.1029/2010WR009505.
Milly, P. C. D., Wetherald, R. T., Dunne, K. A., and Delworth, T. L. (2002) Increasing risk of great floods in a changing climate. Nature, 415(6871), 514–517.
Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., et al. (2008) Stationarity Is Dead: Whither Water Management? Science, 319(5863), 573–574.
Moglen, G. E. and Shivers, D. E. (2006) Methods for adjusting US Geological Survey rural regression peak discharges in an urban setting. [online] http://drum.lib.umd.edu/handle/1903/7135 (Accessed September 19, 2017).
Montanari, A. and Koutsoyiannis, D. (2014) Modeling and mitigating natural hazards: Stationarity is immortal! Water Resources Research, 50(12), 9748–9756.
Mueller, C., and Foley, P. Discussion of ‘’Uncertainty of the Assumptions Required for Estimating the Regulatory Flood: Red River of the North’’ by Paul E. Todhuner. J. Hydrol Eng, 2014; 19(1):275-6
Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., and Rasmussen, R. (2017) Slower snowmelt in a warmer world. Nature Climate Change, 7(3), 214–219.

Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S., et al. (2001) Special Report on Emissions Scenarios : a special report of Working Group III of the Intergovernmental Panel on Climate Change, [online] http://www.osti.gov/energycitations/servlets/purl/15009867-Kv00FB/native/ (Accessed April 20, 2012).

(NRC) National Research Council (2000) Risk Analysis and Uncertainty in Flood Damage Reduction Studies, Washington, D.C., National Academies Press. https://doi.org/10.17226/9971.

National Weather Service, 1982. Hydrometorological Report No. 52, Application of Probable Maximum Precipitation Estimates – United States East of the 105th Meridian, National Oceanic and Atmospheric Administration, Silver Spring, Maryland.

Natural Environment Research Council (NERC) (1975), Flood Studies Report, vol. 1, London, 1975.

Nathan RJ, Weinmann PE (2001) Estimation of large and extreme floods for medium and large catchments: book VI. Australian rainfall and runoff—a guide to flood estimation, 4th edn. Engineers Australia, National Committee for Water Engineering, Australia

Nathan, R., Jordan, P., Scorah, M., Lang, S., Kuczera, G., Schaefer, M., et al. (2016) Estimating the exceedance probability of extreme rainfalls up to the probable maximum precipitation. Journal of Hydrology, 543, 706–720.

Ngongondo, C., Li, L., Gong, L., Xu, C.-Y., and Alemaw, B. F. (2013) Flood frequency under changing climate in the upper Kafue River basin, southern Africa: a large scale hydrological model application. Stochastic Environmental Research and Risk Assessment, 27(8), 1883–1898.

NYC rules (2013) Flood Elevation: Adopted Rules. http://rules.cityofnewyork.us/tags/flood-elevation (Last Access 2017-9-13).

Obeysekera, J. and Salas, J. D. (2014) Quantifying the Uncertainty of Design Floods under Nonstationary Conditions. Journal of Hydrologic Engineering, 19(7), 1438–1446.

Olsen, J. R. (2006) Climate Change and Floodplain Management in the United States. Climatic Change, 76(3–4), 407–426.

Pall, P., Aina, T., Stone, D. A., Stott, P. A., Nozawa, T., Hilberts, A. G. J., et al. (2011) Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. Nature, 470(7334), 382–385.

Pall, P., Patricola, C. M., Wehner, M. F., Stone, D. A., Paciorek, C. J., and Collins, W. D. (2017) Diagnosing conditional anthropogenic contributions to heavy Colorado rainfall in September 2013. Weather and Climate Extremes, 17, 1–6.

Parey, S., Malek, F., Laurent, C., and Dacunha-Castelle, D. (2007) Trends and climate evolution: Statistical approach for very high temperatures in France. Climatic Change, 81(3–4), 331–352.

Park, J., Seager, T. P., and Rao, P. S. C. (2011) Lessons in risk- versus resilience-based design and management. Integrated Environmental Assessment and Management, 7(3), 396–399.

Petrow, T. and Merz, B. (2009) Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002. Journal of Hydrology, 371(1–4), 129–141.

Petrow, T., Zimmer, J., and Merz, B. (2009) Changes in the flood hazard in Germany through changing frequency and persistence of circulation patterns. Natural Hazards and Earth System Sciences, 9(4), 1409.

Pettitt, A. N. (1979) A Non-Parametric Approach to the Change-Point Problem. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(2), 126–135.

Pierce, D. W., Cayan, D. R., and Thrasher, B. L. (2014) Statistical downscaling using localized constructed analogs (LOCA). Journal of Hydrometeorology, 15(6), 2558–2585.

Prosdocimi, I., Kjeldsen, T. R., and Svensson, C. (2014) Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK. Natural Hazards and Earth System Science, 14(5), 1125–1144.

Prosdocimi, I., Kjeldsen, T. R., and Miller, J. D. (2015) Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models. Water Resources Research, 51(6), 4244–4262.

Prudhomme, C., Reynard, N., and Crooks, S. (2002) Downscaling of global climate models for flood frequency analysis: where are we now? Hydrological Processes, 16(6), 1137–1150.

Prudhomme, C., Jakob, D., and Svensson, C. (2003) Uncertainty and climate change impact on the flood regime of small UK catchments. Journal of Hydrology, 277(1–2), 1–23.

Prudhomme, C., Wilby, R.L., Crooks, S., Kay, A.L., Reynard, N.S. (2010). Scenario neutral approach to climate change impact studies: application to flood risk. J. Hydrol. 390 (3–4), 198–209.

Prudhomme, C. and Genevier, M. (2011) Can atmospheric circulation be linked to flooding in Europe? Hydrological Processes, 25(7), 1180–1190.

Qian, B., Corte-Real, J., and Xu, H. (2002) Multisite stochastic weather models for impact studies. International Journal of Climatology, 22(11), 1377–1397.

Rakhecha, P. R. and Clark, C. (1999) Revised estimates of one-day probable maximum precipitation (PMP) for India. Meteorological Applications, 6(4), 343–350.

Ray, P. A. and Brown, C. (2015) Confronting climate uncertainty in water resources planning and project design: the decision tree framework, Washington, DC, World Bank Group.

Raynaud, D., Hingray, B., Zin, I., Anquetin, S., Debionne, S., and Vautard, R. (2017) Atmospheric analogues for physically consistent scenarios of surface weather in Europe and Maghreb. International Journal of Climatology, 37(4), 2160–2176.

Read, L. K. and Vogel, R. M. (2015) Reliability, return periods, and risk under nonstationarity. Water Resources Research, 51(8), 6381–6398.

Refsgaard, J. C. and Knudsen, J. (1996) Operational validation and intercomparison of different types of hydrological models. Water Resour. Res, 32(7), 2189–2202.

Renard, B., & Lall, U. (2014). Regional frequency analysis conditioned on large-scale atmospheric or oceanic fields. Water

Resources Research, 50, 9536–9554. https://doi.org/10.1002/ 2014WR016277
Reynard, N. S., Crooks, S. M., Kay, A. L., DEFRA/Environment Agency Flood and Coastal Defence R & D Programme (Great Britain), Great Britain, and Environment Agency (2005) Impact of climate change on flood flows in river catchments: final report, Bristol, Environment Agency.
Robertson, A. W., Kushnir, Y., Lall, U., and Nakamura, J. (2015) "Weather and Climatic Drivers of Extreme Flooding Events over the Midwest of the United States" in M. Chavez, M. Ghil, and J. Urrutia-Fucugauchi (eds.), Geophysical Monograph Series. Hoboken, NJ, John Wiley & Sons, Inc, 113–124. [online] http://doi.wiley.com/10.1002/9781119157052.ch9 (Accessed June 26, 2018).
Rocheta, E., Sugiyanto, M., Johnson, F., Evans, J., and Sharma, A. (2014) How well do general circulation models represent low-frequency rainfall variability? Water Resources Research, 50(3), 2108–2123.
Rodionov, S. N. (2004) A sequential algorithm for testing climate regime shifts. Geophysical Research Letters, 31(9), L09204.
Rogger, M., Agnoletti, M., Alaoui, A., Bathurst, J. C., Bodner, G., Borga, M., et al. (2017) Land-use change impacts on floods at the catchment scale - Challenges and opportunities for future research. Water Resources Research. [online] http://doi.wiley.com/10.1002/2017WR020723 (Accessed June 21, 2017).
Rojas, R., Feyen, L., Dosio, A., and Bavera, D. (2011) Improving pan-European hydrological simulation of extreme events through statistical bias correction of RCM-driven climate simulations. Hydrology and Earth System Sciences, 15(8), 2599–2620.
Rojas, R., Feyen, L., Bianchi, A., and Dosio, A. (2012) Assessment of future flood hazard in Europe using a large ensemble of bias-corrected regional climate simulations. Journal of Geophysical Research: Atmospheres, 117(D17109).
Rootzén, H. and Katz, R. W. (2013) Design Life Level: Quantifying risk in a changing climate: Design Life Level. Water Resources Research, 49(9), 5964–5972.
Rose, S. and Peters, N. E. (2001) Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): a comparative hydrological approach. Hydrological Processes, 15(8), 1441–1457.
Rosner, A., Vogel, R. M., and Kirshen, P. H. (2014) A risk-based approach to flood management decisions in a nonstationary world. Water Resources Research, 50(3), 1928–1942.
Roudier, P., Andersson, J. C. M., Donnelly, C., Feyen, L., Greuell, W., and Ludwig, F. (2016) Projections of future floods and hydrological droughts in Europe under a +2°C global warming. Climatic Change, 135(2), 341–355.
Rougé, C., Ge, Y., and Cai, X. (2013) Detecting gradual and abrupt changes in hydrological records. Advances in Water Resources, 53, 33–44.
Rousseau, A. N., Klein, I. M., Freudiger, D., Gagnon, P., Frigon, A., and Ratté-Fortin, C. (2014) Development of a methodology to evaluate probable maximum precipitation (PMP) under changing climate conditions: Application to southern Quebec, Canada. Journal of Hydrology, 519, 3094–3109.
Rummukainen, M. (2010) State-of-the-art with regional climate models. Wiley Interdisciplinary Reviews: Climate Change, 1(1), 82–96.
Saghafian, B., Farazjoo, H., Bozorgy, B., and Yazdandoost, F. (2008) Flood Intensification due to Changes in Land Use. Water Resources Management, 22(8), 1051–1067.
Salas, J. D. and Obeysekera, J. (2014) Revisiting the Concepts of Return Period and Risk for Nonstationary Hydrologic Extreme Events. Journal of Hydrologic Engineering, 19(3), 554–568.
Salas, J. D., Obeysekera, J., and Vogel, R. M. (2018) Techniques for assessing water infrastructure for nonstationary extreme events: a review. Hydrological Sciences Journal, 63(3), 325–352.
Santos, J. A., Belo-Pereira, M., Fraga, H., and Pinto, J. G. (2016) Understanding climate change projections for precipitation over western Europe with a weather typing approach: Precipitation Projections for Europe. Journal of Geophysical Research: Atmospheres, 121(3), 1170–1189.
Sankarasubramanian, A., Lall, U., 2003. Flood quantiles in a changing climate: Seasonal forecasts and causal relations. Water Resour. Res. 39.
Sarewitz, D., Pielke Jr., R., and Keykhah, M. (2003) Vulnerability and Risk: Some Thoughts from a Political and Policy Perspective. Risk Analysis, 23(4), 805–810.
Sayers, P. B., Galloway, G. E., and Hall, J. (2012) "Robust decision making under uncertainty - Towards adaptive and resilient flood risk management infrastructure" in P. B. Sayers (ed.), Flood Risk: Planning, Design and Management of Flood Defence Infrastructure. ICE Publishing.
Schendel, T. and Thongwichian, R. (2015) Flood frequency analysis: Confidence interval estimation by test inversion bootstrapping. Advances in Water Resources, 83, 1–9.
Schlef, K. E., François, B., Robertson, A. W., and Brown, C. (2018a) A General Methodology for Climate-Informed Approaches to Long-Term Flood Projection-Illustrated With the Ohio River Basin. Water Resources Research, 54(11), 9321–9341.
Schlef, K. E., Steinschneider, S., and Brown, C. M. (2018b) Spatiotemporal Impacts of Climate and Demand on Water Supply in the Apalachicola-Chattahoochee-Flint Basin. Journal of Water Resources Planning and Management, 144(2), 5017020.
Schlef, K. (2018) Flood Risk Assessment, Management, and Perceptions in a Changing World. PhD Thesis, University of Massachusetts, Amherst, United States. Available at: https://scholarworks.umass.edu/dissertations_2/1273/ (last access August 2018)
Schneider, D. P., Ammann, C. M., Otto-Bliesner, B. L., and Kaufman, D. S. (2009) Climate response to large, high-latitude and low-latitude volcanic eruptions in the Community Climate System Model. Journal of Geophysical Research, 114(D15). doi:/10.1029/2008JD011222.

Schreiner, L. C. and Riedel, J. T. (1978) Probable maximum precipitation estimates, United States east of the 105th meridian, Department of Commerce, National Oceanic and Atmospheric Administration Silver Spring, MD.

Schwab, M., Meinke, I., Vanderlinden, J.-P., and von Storch, H. (2017) Regional decision-makers as potential users of Extreme Weather Event Attribution - Case studies from the German Baltic Sea coast and the Greater Paris area. Weather and Climate Extremes, 18, 1–7.

(SEPA) Scottish Environment Protection Agency, 2017. SEPA Planning Background Paper: Flood Risk. Scottish Environment Protection Agency guidance on land use and flood risk. Available online: https://www.sepa.org.uk/media/162837/lups-bp-gu2a-land-use-planning-background-paper-on-flood-risk.pdf. Last access: 09-2018.

Serago, J. M. and Vogel, R. M. (2018) Parsimonious nonstationary flood frequency analysis. Advances in Water Resources, 112, 1–16.

Serinaldi, F. and Kilsby, C. G. (2015) Stationarity is undead: Uncertainty dominates the distribution of extremes. Advances in Water Resources, 77, 17–36.

Shaw, W. (2009) Raise your Home, Lower Your Monthly Payments. Protect buildings and reduce monthly expenses with freeboard. Boston, MA, Massachusetts Office of a Coastal Zone Management. [online] http://www.mass.gov/eea/docs/czm/stormsmart/ssc/ssc5-freeboard.pdf (Accessed January 10, 2018).

Shaw, T. A., Baldwin, M., Barnes, E. A., Caballero, R., Garfinkel, C. I., Hwang, Y.-T., et al. (2016) Storm track processes and the opposing influences of climate change. Nature Geoscience, 9(9), 656–664.

Sperna Weiland, F. C., van Beek, L. P. H., Kwadijk, J. C. J., and Bierkens, M. F. P. (2010) The ability of a GCM-forced hydrological model to reproduce global discharge variability. Hydrology and Earth System Sciences, 14(8), 1595–1621.

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., and Bronaugh, D. (2013) Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate: CLIMATE EXTREMES INDICES IN CMIP5. Journal of Geophysical Research: Atmospheres, 118(4), 1716–1733.

Simon, H. (1956) Rational choice and the structure of the environment. Psychological Review 63: 81-97

Sivapalan, M. and Samuel, J. M. (2009) Transcending limitations of stationarity and the return period: process-based approach to flood estimation and risk assessment. Hydrological Processes, 23(11), 1671–1675.

Smith, J. A., Baeck, M. L., Morrison, J. E., Sturdevant-Rees, P., Turner-Gillespie, D. F., and Bates, P. D. (2002) The regional hydrology of extreme floods in an urbanizing drainage basin. Journal of Hydrometeorology, 3(3), 267–282.

Spence, C. M. and Brown, C. M. (2016) Nonstationary decision model for flood risk decision scaling. Water Resources Research, 52(11), 8650–8667.

Spence, C. M. and Brown, C. M. (2018) Decision Analytic Approach to Resolving Divergent Climate Assumptions in Water Resources Planning. Journal of Water Resources Planning and Management, 144(9), 4018054.

Spiegelhalter, D. J. and Riesch, H. (2011) Don’t know, can’t know: embracing deeper uncertainties when analysing risks. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 369(1956), 4730–4750.

Šraj, M., Viglione, A., Parajka, J., and Blöschl, G. (2016) The influence of non-stationarity in extreme hydrological events on flood frequency estimation. Journal of Hydrology and Hydromechanics, 64(4).

Stainforth, D. ., Allen, M. ., Tredger, E. ., and Smith, L. . (2007a) Confidence, uncertainty and decision-support relevance in climate predictions. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1857), 2145–2161.

Stainforth, D. A., Downing, T. E., Washington, R., Lopez, A., and New, M. (2007b) Issues in the interpretation of climate model ensembles to inform decisions. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1857), 2163–2177.

Stakhiv, E. Z. (2011) Pragmatic Approaches for Water Management Under Climate Change Uncertainty. JAWRA Journal of the American Water Resources Association, 47(6), 1183–1196.

Stephens, G. L., L’Ecuyer, T., Forbes, R., Gettelmen, A., Golaz, J.-C., Bodas-Salcedo, A., et al. (2010) Dreary state of precipitation in global models. Journal of Geophysical Research: 115(D24211). doi:10.1029/2010JD014532.

Stratz, S.A., Hossain, F., 2014. Probable maximum precipitation in a changing climate: Implications for dam design. J. Hydrol. Eng. 19, 6014006.

Stedinger, J. R. (1983) Confidence Intervals for Design Events. Journal of Hydraulic Engineering, 109(1), 13–27.

Stedinger, J. R. and Griffis, V. W. (2011) Getting From Here to Where? Flood Frequency Analysis and Climate1: Getting From Here to Where? Flood Frequency Analysis and Climate. JAWRA Journal of the American Water Resources Association, 47(3), 506–513.

Steinschneider, S., Wi, S., and Brown, C. (2015a) The integrated effects of climate and hydrologic uncertainty on future flood risk assessments. Hydrological Processes, 29(12), 2823–2839.

Steinschneider, S., McCrary, R., Mearns, L. O., and Brown, C. (2015b) The effects of climate model similarity on probabilistic climate projections and the implications for local, risk-based adaptation planning. Geophysical Research Letters, 42(12), 5014–5044.

Strupczewski, W. G., Singh, V. P., and Mitosek, H. T. (2001) Non-stationary approach to at-site flood frequency modelling. III. Flood analysis of Polish rivers. Journal of Hydrology, 248(1), 152–167.

Sunyer, M. A., Madsen, H., and Ang, P. H. (2012) A comparison of different regional climate models and statistical downscaling methods for extreme rainfall estimation under climate change. Atmospheric Research, 103, 119–128.

Taschetto, A. S., Gupta, A. S., Jourdain, N. C., Santoso, A., Ummenhofer, C. C., and England, M. H. (2014) Cold Tongue and Warm Pool ENSO Events in CMIP5: Mean State and Future Projections. Journal of Climate, 27(8), 2861–2885.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012) An Overview of CMIP5 and the Experiment Design. Bulletin of the American Meteorological Society, 93(4), 485–498.

Teutschbein, C. and Seibert, J. (2012) Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. Journal of Hydrology, 456–457, 12–29.

Thompson, A. and Otto, F. E. L. (2015) Ethical and normative implications of weather event attribution for policy discussions concerning loss and damage. Climatic Change, 133(3), 439–451.

Tramblay, Y., Amoussou, E., Dorigo, W., and Mahé, G. (2014) Flood risk under future climate in data sparse regions: Linking extreme value models and flood generating processes. Journal of Hydrology, 519, 549–558.

Trenberth, K. E. (1997) The definition of el nino. Bulletin of the American Meteorological Society, 78(12), 2771–2778.

Trenberth, K. E. and Dai, A. (2007) Effects of Mount Pinatubo volcanic eruption on the hydrological cycle as an analog of geoengineering. Geophysical Research Letters, 34(15). doi:/10.1029/2007GL030524.

Trenberth, K. (2011) Changes in precipitation with climate change. Climate Research, 47(1), 123–138.

Tung, Y.-K. (2005) Flood Defense Systems Design by Risk-Based Approaches. Water International, 30(1), 50–57.

USACE (2016) Guidance for Incorporating Climate Change Impacts to Inland Hydrology in Civil Work Studies, Designs, and Projects, U.S. Army Corps of Engineers. [online] https://www.wbdg.org/FFC/ARMYCOE/COEECB/ecb_2016_25.pdf (Accessed August 22, 2018).

U.S. Department of Commerce. (1999). Hydrometeorological Rep. No. 59, Probable Maximum Precipitation for California, National Oceanic and Atmospheric Administration, U.S. Dept. of the Army, Corps of Engineers, Washington, DC.

Velázquez, J. A., Troin, M., Caya, D., and Brissette, F. (2015) Evaluating the Time-Invariance Hypothesis of Climate Model Bias Correction: Implications for Hydrological Impact Studies. Journal of Hydrometeorology, 16(5), 2013–2026.

Viglione, A., Merz, B., Viet Dung, N., Parajka, J., Nester, T., and Blöschl, G. (2016) Attribution of regional flood changes based on scaling fingerprints. Water Resources Research, 52(7), 5322–5340.

Villarini, G., Serinaldi, F., Smith, J. A., and Krajewski, W. F. (2009a) On the stationarity of annual flood peaks in the continental United States during the 20th century: stationarity of annual flood peaks. Water Resources Research, 45(8), W08417.

Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D., and Krajewski, W. F. (2009b) Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. Advances in Water Resources, 32(8), 1255–1266.

Villarini, G., Smith, J. A., Vitolo, R., and Stephenson, D. B. (2013) On the temporal clustering of US floods and its relationship to climate teleconnection patterns. International Journal of Climatology, 33(3), 629–640.

Vogel, R. M., Yaindl, C., and Walter, M. (2011) Nonstationarity: Flood Magnification and Recurrence Reduction Factors in the United States. JAWRA Journal of the American Water Resources Association, 47(3), 464–474.

Voorendt, M. (2015). The development of the Dutch Flood safety strategy, Communications on Hydraulic and Geotechnical Engineering 2015-01 Delft University of Technology

Vorogushyn, S. and Merz, B. (2013) Flood trends along the Rhine: the role of river training. Hydrology and Earth System Sciences, 17(10), 3871–3884.

Wang, Q. J. (1991) The Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-Runoff Models. Water Resources Research, 27(9), 2467–2471.

Ward, P. J., Eisner, S., Flörke, M., Dettinger, M. D., and Kummu, M. (2014) Annual flood sensitivities to El Niño–Southern Oscillation at the global scale. Hydrology and Earth System Sciences, 18(1), 47–66.

Water Resources Program (2009) Design Storm Construction, Washington State, Water Resources Program - Dam Safety Office. Report Number: 92-55G, 67p (Available online: https://fortress.wa.gov/ecy/publications/documents/9255g.pdf (Accessed August 24, 2018).

Whateley, S. and Brown, C. (2016) Assessing the relative effects of emissions, climate means, and variability on large water supply systems. Geophysical Research Letters, 43(21), 11,329-11,338.

Wi, S., Yang, Y. C. E., Steinschneider, S., Khalil, A., and Brown, C. M. (2015) Calibration approaches for distributed hydrologic models in poorly gaged basins: implication for streamflow projections under climate change. Hydrology and Earth System Sciences, 19(2), 857–876.

Wilby, R. L. and Dessai, S. (2010) Robust adaptation to climate change. Weather, 65(7), 180–185.

Wilby, R. L. and Quinn, N. W. (2013) Reconstructing multi-decadal variations in fluvial flood risk using atmospheric circulation patterns. Journal of Hydrology, 487, 109–121.

Willems, P. and Vrac, M. (2011) Statistical precipitation downscaling for small-scale hydrological impact investigations of climate change. Journal of Hydrology, 402(3–4), 193–205.

Winsemius, H. C., Aerts, J. C. J. H., van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., et al. (2015) Global drivers of future river flood risk. Nature Climate Change, 6(4), 381–385.

Wolski, P., Stone, D., Tadross, M., Wehner, M., and Hewitson, B. (2014) Attribution of floods in the Okavango basin, Southern Africa. Journal of Hydrology, 511, 350–358.

(WMO) World Meteorological Organization (1986) Manual for estimation of Probable Maximum Precipitation. Second ed. Operational hydrology report no. 1, WMO – No. 332, Geneva

(WMO) World Meteorological Organization (2009) Extreme Values Analysis in: The Guide to Hydrological Practices. Volume II: Management of Water Resources and Application of Hydrological Practices, [online]

http://www.whycos.org/hwrp/guide/index.php (Accessed June 2, 2017).
Wuebbles, D., Meehl, G., Hayhoe, K., Karl, T. R., Kunkel, K., Santer, B., et al. (2014) CMIP5 Climate Model Analyses: Climate Extremes in the United States. Bulletin of the American Meteorological Society, 95(4), 571–583.
Xiong, L. and Guo, S. (2004) Trend test and change-point detection for the annual discharge series of the Yangtze River at the Yichang hydrological station. Hydrological Sciences Journal, 49(1), 99–112.
Xue, Y., Janjic, Z., Dudhia, J., Vasic, R., and De Sales, F. (2014) A review on regional dynamical downscaling in intraseasonal to seasonal simulation/prediction and major factors that affect downscaling ability. Atmospheric Research, 147–148, 68–85.
Yim, B. Y., Kwon, M., Min, H. S., and Kug, J.-S. (2015) Pacific Decadal Oscillation and its relation to the extratropical atmospheric variation in CMIP5. Climate Dynamics, 44(5–6), 1521–1540.
Yin, J., Gentine, P., Zhou, S., Sullivan, S. C., Wang, R., Zhang, Y., et al. (2018) Large increase in global storm runoff extremes driven by climate and anthropogenic changes. Nature Communications, 9(1).
Yu, X., Cohn, T. A., and Stedinger, J. R. (2015) "Flood Frequency Analysis in the Context of Climate Change" in American Society of Civil Engineers, 2376–2385. [online] http://ascelibrary.org/doi/10.1061/9780784479162.233 (Accessed July 11, 2017).
Yuan, S. and Quiring, S. M. (2017) Evaluation of soil moisture in CMIP5 simulations over the contiguous United States using in situ and satellite observations. Hydrology and Earth System Sciences, 21(4), 2203–2218.
Zhang, Q., Gu, X., Singh, V. P., Xiao, M., and Chen, X. (2015) Evaluation of flood frequency under non-stationarity resulting from climate indices and reservoir indices in the East River basin, China. Journal of Hydrology, 527, 565–575.
Zhang, H. and Delworth, T. L. (2018) Robustness of anthropogenically forced decadal precipitation changes projected for the 21st century. Nature Communications, 9(1). [online] http://www.nature.com/articles/s41467-018-03611-3 (Accessed January 17, 2019)

3. REGIONAL EXTREME PRECIPITATION EVENTS: ROBUST INFERENCE FROM CREDIBLY SIMULATED GCM VARIABLES

This Chapter is published in Water Resource Research:

Farnham, D. J., Doss-Gollin, J., & Lall, U. (2018). Regional Extreme Precipitation Events: Robust Inference From Credibly Simulated GCM Variables. Water Resources Research. *https://doi.org/10.1002/2017wr021318*

Supplementary materials are available at:

https://agupubs.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2F2017WR021318&file=wrcr23208-sup-0001-2017WR021318-s01.pdf

3.1. EXECUTIVE SUMMARY

General circulation models (GCMs) have been demonstrated to produce estimates of precipitation, including the frequency of extreme precipitation, with substantial bias and uncertainty relative to their representation of other fields. Thus, while theory predicts changes in the hydrologic cycle under anthropogenic warming, there is generally low confidence in future projections of extreme precipitation frequency for specific river basins. In this paper, we explore whether a GCM simulates large-scale atmospheric circulation indices that are associated with Regional Extreme Precipitation (REP) days more accurately than it simulates REP days themselves, and thus whether conditional simulation of the precipitation events based on the circulation indices may improve the simulation of REP events. We show that a coupled Geophysical Fluid Dynamics Laboratory GCM simulates too many springtime REP days in the Ohio River Basin in historical (1950–2005) simulations. The GCM, however, does credibly simulate the distributional and persistence properties of several indices (which represent the large-scale atmospheric pressure features, local atmospheric moisture content, and local vertical velocity) that are shown to modulate the likelihood of REP occurrence in the reanalysis/observational record. We show that simulation of REP events based on the GCM-based atmospheric indices greatly reduces the bias of GCM REP frequency relative to the observed record. The simulation is conducted via a Bayesian regression model by imposing the empirical relationship between observed REP occurrence and the reanalysis-based atmospheric indices. Application of this model to future (2006–2100) representative concentration pathway 8.5 scenario suggests an increasing trend in springtime REP incidence in the study region. The proposed approach of simulating precipitation events of interest, particularly those poorly represented in GCMs, with a statistical model based on climate indices that are reasonably simulated by GCMs could be applied to subseasonal to seasonal forecasts as well as future projections.

3.2. TECHNICICAL APPROACH

Floods are responsible for significant loss of life and economic damages both within the United States (U.S.) and worldwide. Flood impacts in the U.S. are estimated at $USD 8 billion (in 2014 dollars) and 82 fatalities per year from 1984 to 2013 (NWS Internet Services Team, 2015), while worldwide flood losses were estimated to be about $USD 85 billion (in 2012 U.S. dollars) in 1993 alone (Kundzewicz et al., 2013). Furthermore, trends in population and urbanization are expected to increase exposure to hydroclimate extremes (including floods) into the future (Jongman et al., 2012). Given that projections of extreme precipitation changes remain highly uncertain (IPCC, 2012), particularly in the midlatitudes, improved estimation of future hydroclimate extremes is a key ingredient for the mitigation of future flood impacts.

The poor representation of precipitation fields (particularly extreme precipitation) in general circulation model (GCM) simulations (Dai, 2006; Kendon et al., 2012; Stephens et al., 2010) complicate the projections of future hydroclimate extremes. Simulated precipitation fields are often used as inputs to hydrologic models (e.g., Hirabayashi et al., 2013; Kundzewicz et al., 2010; Lehner et al., 2006; Winsemius et al., 2015) after some form of bias correction (e.g., quantile-quantile mapping; Gudmundsson et al., 2012) or downscaling is applied. It is often difficult to justify a bias-correction approach, especially for extrapolation into the future, since there is no accompanying insight as to the underlying cause for the bias, or whether the bias correction used would be applicable in the future. In this paper, we explore whether some atmospheric variables that are closely related to the occurrence of regional extreme precipitation (REP) are well simulated by GCMs, such that their use for conditional prediction of REPs under seasonal forecasts or for climate change projections can be an effective strategy.

An important question is whether a GCM reproduces REP events well in the historical record. Since GCMs represent the coupled dynamics of the ocean-atmosphere-land systems, answering such a question is highly dependent on the physical parameterizations of each individual GCM. One possibility is that the GCMs credibly simulate large-scale climate circulations but that grid-scale (and subgrid-scale) precipitation mechanisms are not well represented. In this case, it may be possible to use credibly simulated state variables from GCM simulations to derive or simulate credible sequences of REP events associated with major floods. We explore this possibility by focusing on a single GCM and a set of atmospheric circulation indices relevant to floods in the Ohio River Basin. The following set of questions provide the framework for our overall goal of identifying the causal structure associated with REP events and developing an empirical model that allows the causal structure to be tested and used in a predictive context.

Q1 For the Ohio River Basin, are the extreme springtime precipitation events that are relevant for floods well simulated by the GCM?

Q2 Can atmospheric indices that are associated with the onset of REP events be identified from reanalysis?

Q3 Are suitably derived atmospheric indices associated with REP events in atmospheric reanalysis credibly simulated by the GCM?

Q4 If GCMs represent the large-scale atmospheric indices more credibly than they do the REP events, can we use the GCM derived atmospheric indices to directly simulate extreme precipitation events in the current and future climate?

3.3. DATA AND METHODS

3.3.1. CASE STUDY

We use the Ohio River Basin, which has a long history of regional flooding, to examine the questions presented in section 3.2. Major events in 1933, 1937, 1945, 1997, and 2011 are among the numerous floods that have had high financial and human life costs. The springtime flood of 1913 caused over 450 deaths (Perry, 2000), while the springtime flood of 2011 is estimated to have cost over $3 billion in damages (Smith et al., 2016). Although floods are influenced by water management strategies, land use, and soil characteristics, the floods in the Ohio River Basin are generally associated with heavy and/or persistent precipitation events and/or snowmelt (Nakamura et al., 2012). The dominance of the precipitation signal is also supported by Mallakpour and Villarini (2015), who primarily attribute changes in flood frequency in the central U.S. to changes in heavy rainfall frequency and temperatures while noting that land surface changes play a secondary role.

In the study region, and in the midlatitudes more generally, intense rainfall over a large area typically requires large-scale advection of moisture from the tropics (Knippertz & Wernli, 2010; Lu et al., 2013; Steinschneider & Lall, 2016). Tropical moisture export-related precipitation over the central and eastern United States is dominated by the Great Plains activity center, which sources moisture primarily from the Gulf of Mexico and Caribbean Sea (Gimeno et al., 2010; Lavers & Villarini, 2013; Steinschneider & Lall, 2016). Dirmeyer and Kinter (2010) showed that large-scale flooding across the U.S. Midwest is often associated with moisture sources extending through Texas, Eastern Mexico, the western Gulf of Mexico, and the Caribbean Sea (termed the “Maya Express”). Nakamura et al. (2012) showed that springtime extreme streamflow in the Ohio River Basin is driven by a unique, recurrent, persistent, and strong atmospheric anticyclonic circulation anomaly located to the east of the U.S. Atlantic coast (i.e., the Bermuda High), which forces anomalous northward moisture transport from the Gulf of Mexico and tropical Atlantic.

3.3.2. Methodological Overview

We build on the diagnostic literature discussed in section 3.2 in this paper and focus directly on predicting whether or not a REP process is likely to occur on a given day based on atmospheric conditions as summarized by a set of indices. The REP event is defined here as a day when at least 4 of the 15 subregions in the region of interest experiences a daily rainfall that exceeds the 99th percentile of daily rainfall at that location. Subregions are defined by the blue grid in Figure 3-1 and are based on the GCM's spatial gridding. Thus, a spatiotemporal extreme precipitation process is implicitly considered conditional on variables that are derived from a climate model. Notably, we do not explicitely address issues related to the ability of GCMs to simulate extreme precipitation as a function of spatial resolution (such as in Wehner et al., 2010).

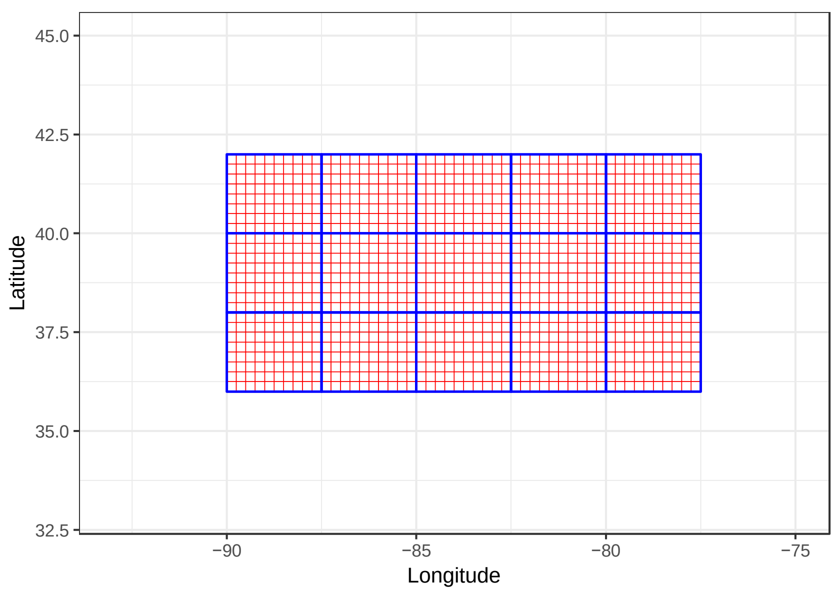


Figure 3-1 Map of study area. Blue grid shows resolution of Geophysical Fluid Dynamics Laboratory CM3 coupled model cells. Red grid shows native resolution of CPC precipitation data cells. The shaded area indicates the Ohio River Basin (~530,000 km2) as defined by the United States Geological Survey.

We focus on flood-relevant extreme precipitation events and fit and simulate from a Bayesian model that propagates the parameter estimation uncertainties to the future simulations. This latter point is vital for decision making since understanding the range of possible future outcomes, via various prediction intervals, is helpful for determining our level of confidence in the projections and thus whether the projections represent actionable information or not.

Our approach is conceptually similar to a Nonhomogeneous Hidden Markov Model (NHMM; Cioffi et al., 2016, 2017; Holsclaw et al., 2015; Hughes et al., 1999; Kwon et al., 2009; Robertson & Smyth, 2003) for precipitation downscaling. In the NHMM approach, a stochastic model is considered for the daily rainfall process, where rainfall occurrence is modeled conditionally on a latent (unobserved) state, and the probability of being in a particular hidden state is informed by a set of appropriate atmospheric circulation variables. This approach is useful in the context of flood modeling, since it preserves the sequence of rainfall occurrence and hence of antecedent conditions and event rainfall, both of which are important for determining flood potential. A challenge with this approach is that rainfall extremes may or may not be well represented, since often they are not explicitly conditioned on changing climate state. The end result of simulating a credible precipitation index time series from dynamical model outputs is common to both our proposed method and many bias-correction and statistical downscaling techniques (e.g., Gutmann et al., 2014; Maraun et al., 2010; Wilby et al., 2002). Our method, however, places a central focus on identifying and representing the underlying dynamics of the process. We discuss bias-correction and downscaling approaches common to the literature in section 3.5.2.

Lastly, we focus on the spring (March–April–May, MAM) season in the Ohio River Basin (Figure 3-1), following the observation in Nakamura et al. (2012) that this is the dominant season for major regional floods. Our historical study period is from 1 March 1950 through 30 May 2005, and our future study period is from 1 March 2006 through 30 May 2100. All anomalies are estimated relative to the historical monthly mean unless otherwise noted.

3.3.3. Regional Extreme Precipitation Days and Extreme Streamflow

We use the Climate Prediction Center (CPC) U.S. unified gauge-based surface precipitation (P) data at horizontal resolution of 0.25° by 0.25° (Xie et al., 2010). The data are defined as the precipitation accumulated in the prior 24 h at 12 UTC and are available online from the International Research Institute's Data library at https://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.UNIFIED_PRCP/.GAUGE_BASED/.GLOBAL/.v1p0/. We upscale the CPC precipitation data by taking the spatial average of the daily precipitation over the coarser horizontal gridding of the dynamical climate model introduced below (2.5° longitude by 2.0° latitude). We refer to this upscaled CPC precipitation data as observed precipitation throughout the manuscript.

The 99th percentile precipitation exceedances, used to define the REP days, are defined from the full-year daily record for each individual grid cell within the region of interest. In this case, the region refers to all of the area covered by the blue and red grids in Figure 3-1. The 99th percentile thresholds used to derive the REP days are estimated separately for the observed and GCM records from the historic record (1950–2005) unless noted otherwise. This means that our REP record is insensitive to bias in the 99th percentile precipitation in the GCM, which in turn means that this work does not address GCM bias in precipitation intensity (such as in Maraun et al., 2010). Using the available data shown in Figure 3-1, a REP day means that 4 or more of the region's 15 grid cells experience a 99th percentile exceedance of daily rainfall. We use the Hydro-Climatic Data Network streamflow data from the United States Geological Survey data downloaded with the dataRetrieval package of the R statistical programming language, and retain only sites with drainage areas larger than 15,000 km2 and with fewer than 25 missing days over the historical study period. Six streamflow stations in the Ohio River Basin meet these criteria and are shown in Figure 3-2.

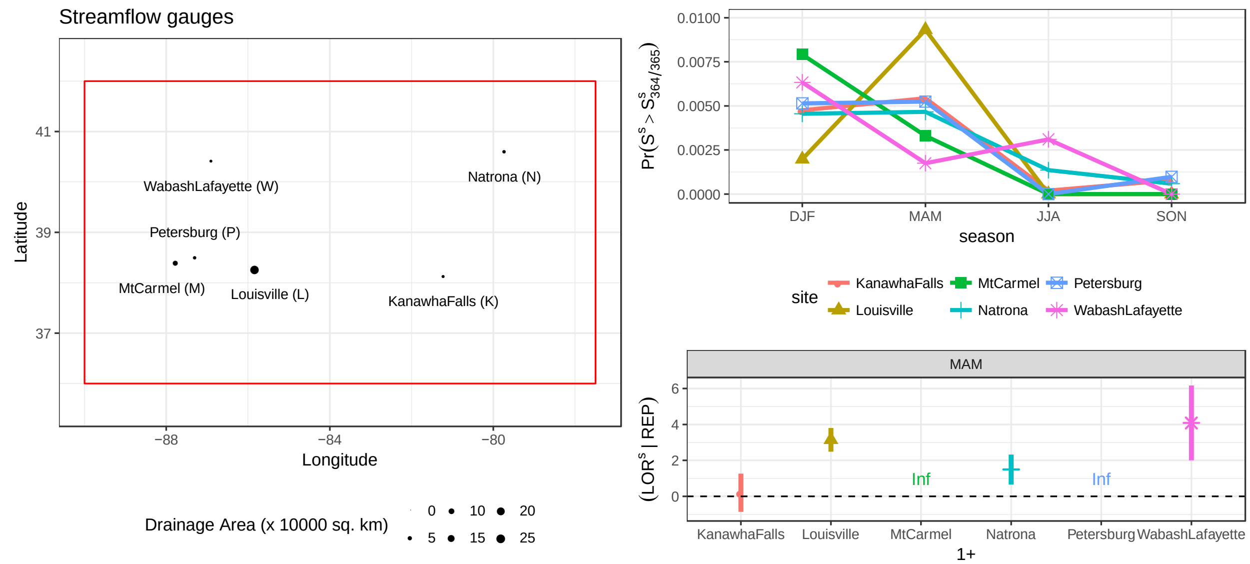


Figure 3-2 (left) Locations and drainage areas of the six long record streamflow stations. (top, right) The seasonality of extreme streamflow (>≈99.7th percentile) for each site in colors as expressed through the probability of extreme streamflow occurrence during each season. (bottom, right) The log odds ratio (equation 3.1) and confidence interval associated with MAM days when one of more REP days have occurred in the previous 15 days versus those when no REP days have occurred in the previous 15 days and streamflow being above or below the ≈99.7th percentile. The odds ratio confidence interval was calculated via the unconditional maximum likelihood estimation (or the Wald method) via the epitools package of the R statistical programing language.

Our first goal is to investigate the relationship between the REP days and extreme streamflow days, the latter of which we define as streamflow greater than the 1 in 365 day streamflow (≈99.7th percentile), defined from each site's full record. We use the log odds ratio of equation 3.1 to assess the extent to which REP day occurrence in the previous 15 days corresponds to enhanced probabilities of extreme streamflow at the six long record streamflow gauges.

$$(\mathrm{LOR}^s|\mathrm{REP}) = \ln\left[\frac{\mathrm{P}\left(S_t^s > S_{364/365}^s \middle| \sum_{t'=(t-15)}^{t} \mathrm{REP}_{t'} \geq 1\right) \Big/ \mathrm{P}\left(S_t^s \leq S_{364/365}^s \middle| \sum_{t'=(t-15)}^{t} \mathrm{REP}_{t'} \geq 1\right)}{\mathrm{P}\left(S_t^s > S_{364/365}^s \middle| \sum_{t'=(t-15)}^{t} \mathrm{REP}_{t'} = 0\right) \Big/ \mathrm{P}\left(S_t^s \leq S_{364/365}^s \middle| \sum_{t'=(t-15)}^{t} \mathrm{REP}_{t'} = 0\right)}\right] \qquad 3.1$$

where S_t^s is the streamflow at time step *t* and streamflow station *s*, $S_{364/365}^t$ is the 1 in 365 day streamflow at site *s*, and t' is a dummy variable to loop from *t-15* to *t*.

3.3.4. Atmospheric Reanalysis for Event Diagnostics

We use atmospheric specific humidity (*Q*), geopotential height (*Z*), upward velocity (*ω*), and zonal wind (*U*) fields from the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis 1 data set (Kalnay et al., 1996). The NCEP/NCAR reanalysis data set has a horizontal resolution of 2.5° by 2.5° and 17 pressure levels. We download six hourly data and define each day as the average value between 12 UTC and 12 UTC to ensure that the atmospheric reanalysis data is on the same temporal grid as the CPC precipitation. The NCEP/NCAR Reanalysis 1 data is available from NOAA/OAR/ESRL PSD, Boulder, CO, USA, online at http://www.esrl.noaa.gov/psd/.

3.3.5. General Circulation Model

We use the P, Q, Z, ω, and U fields from the Geophysical Fluid Dynamics Laboratory (GFDL) global coupled model (Donner et al., 2011), called CM3. The surface and atmosphere in CM3 has a resolution of 2.5° longitude by 2.0° latitude (Figure 3-1). CM3 outputs are available online at https://www.gfdl.noaa.gov/coupled-physical-model-cm3/. Based on the atmospheric variables and daily resolution that we required for this work, we could only acquire two historic ensemble member simulations and one future simulation.

3.4. RESULTS AND DISCUSSION

3.4.1. Regional Extreme Precipitation Days and Streamflow

Figure 3-2 highlights the positive relationship between REP incidence and subsequent extreme streamflows during MAM in the study basin. Boreal winter (DJF) and spring (MAM) dominate the record of extreme streamflows (≈99.7th percentile) and the station with the largest drainage area (Louisville) shows a clear maximum in MAM. The estimated log odds ratio defined in equation 3.1 is positive for all stations during MAM (Figure 3-2), a clear indication that the occurrence of REP days is strongly associated with the occurrence of extreme streamflows during MAM in the Ohio River Basin. The extreme streamflow seasonality and enhanced odds of occurrence following REP days are similar when extreme streamflow is defined using the 99th and 99.9th percentiles, indicating that the relationship between high streamflows and antecedent REP events is not sensitive to the definition of extreme streamflow.

3.4.2. Regional Extreme Precipitation in a GCM Versus Observations

We next turn our attention to **Q1** by comparing REP day frequency and persistence in the observed and GCM records.

The CM3 model simulates too many MAM REP events in the study region and too few back-to-back MAM REP days when compared to the observed record (Figure 3-3). This is supported quantitatively by highly significant Wilcoxon rank sum tests in supporting information Table S1. The MAM REP frequency bias stems from a seasonality bias in the GCM that results in too many (few) local extreme precipitation days in the spring (summer) and higher spatial coherence of local extreme precipitation days in the GCM. The origin of the persistence bias in the GCM appears to be related to faster storm propagation speeds due to bias in the climatological jet stream.

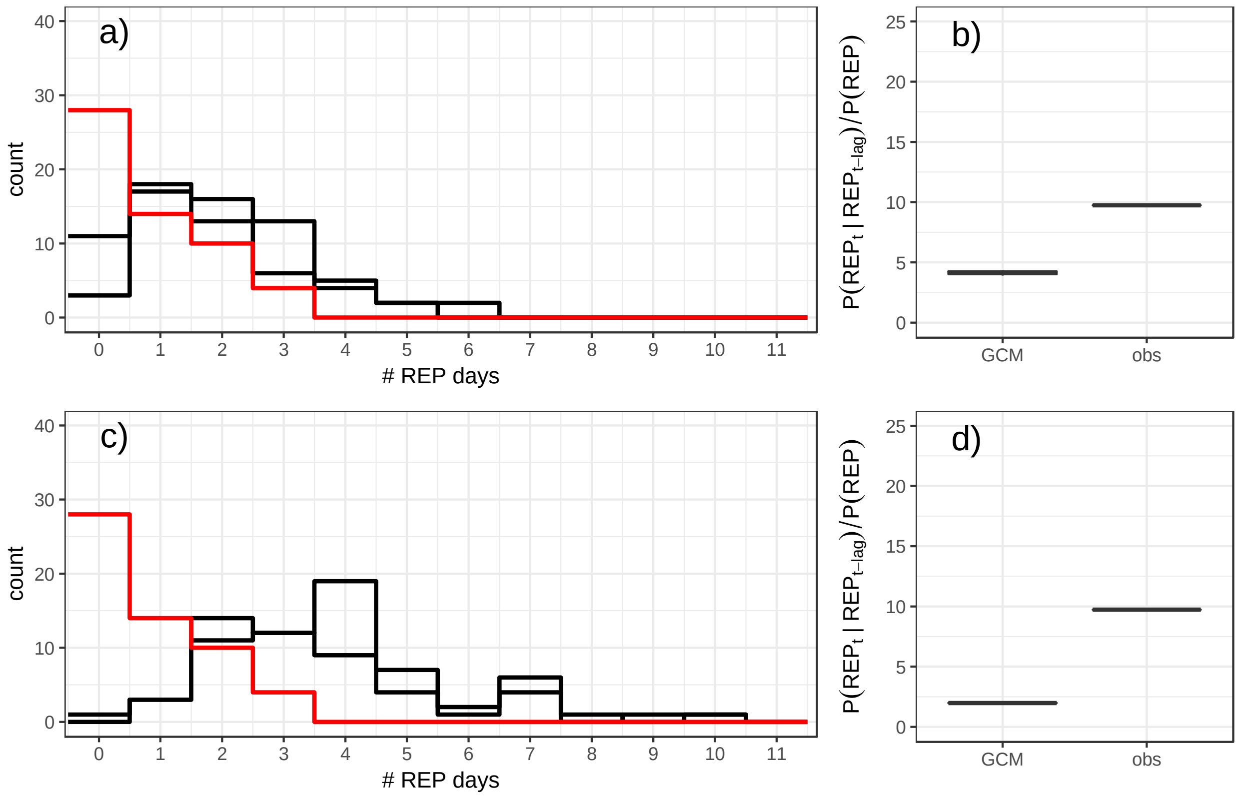


Figure 3-3 (top left; a) The frequency distribution of the number of MAM REP days by year for the observed record (red solid line) and the two GFDL CM3 ensemble members (black solid lines). (top right; b) The probability of a REP event on a day given that a REP event occurred the day prior divided by the marginal probability of a REP event for the MAM season for the observed record and the two ensemble members. (bottom; c, d) Same as (top) but with the observed 99th percentile precipitation thresholds used to derive the model REP records. The bottom panels show that the discrepancy between the GCM runs and the observed REP records is even more stark when the observed precipitation data is used to calculate the 99th percentile thresholds for the model and REP records, an indication of a significant positive bias with respect to the GCM's 99th percentile precipitation. In fact, the median of the study region's 99th percentiles is 31 mm/d in the GFDL CM3 model, and only 25 mm/d in the CPC data.

While the CM3 model exhibits a wet bias in the 99th percentile precipitation, the approach used to define REP events means that this does not explain the inflated MAM REP counts in the GCM. Since the total number of local (one cell) extreme precipitation days (i.e., >99th percentile) is the same for both the observed and GCM records, the REP frequency bias can stem from a bias in the seasonal distribution of the local extreme precipitation days or a bias in the spatial correlation across the study region.

There is clearly a bias in the seasonality of the extreme precipitation days, which contributes to the over-simulation of MAM REP days. The CM3 model ensemble members show 37 and 38% of their local (single-grid) extreme precipitation days occurring during MAM, while the observed record shows only 27% (see supporting information Figure S1). Conversely, the CM3 members simulate between 10 and 11% of local extreme precipitation days during JJA, less than the observed value of 26%. This seasonality bias is manifest in the REP climatology itself (supporting information Figure S2) with the GCM simulating relatively few REP days during the summer (JJA) and relatively more during MAM. Deficiencies in simulating extreme precipitation during boreal midlatitude summer has been observed and discussed for other models (e.g., Durman et al., 2001) and may be attributable to parameterizations of subgrid-scale convective processes (Liang et al., 2006).

The second reason for the inflated MAM REP counts is a tendency of the CM3 model to produce too many cooccurring local extreme precipitation days. More precisely, REP days occur during 22 and 24% of all MAM days when there is at least one local extreme precipitation event in the two CM3 ensemble members, respectively, compared to just 11% in the observed records (see supporting information Figure S1). This indicates that when the model produces extreme precipitation in any part of the study region, it has a tendency to simultaneously produce extreme precipitation in several grid cells. This high regional covariance, or smearing, of the extreme precipitation can be seen in supporting information Figure S3. This high spatial covariance is not surprising given that the effective resolution of numerical models is known to be significantly greater than the grid spacing (e.g., Grasso, 2000). This point is noteworthy for any regional flood hazard assessment that uses GCMs.

In addition to the frequency bias, the CM3 model undersimulates the occurrence of back-to-back REP days (Figure 3-3, right). The probability of a REP day following the occurrence of a REP day is about 4 times more than the marginal probability of REP occurrence in the GCM, compared to about 10 times in the observed record. This appears to be related to representation of the storm tracks, which in CM3 propagate primarily from west to east, underrepresenting observed south to north movement. This causes the precipitation (particularly along cold fronts) to exit the study region more quickly (supporting information Figure S4). We conclude that the relevant precipitation events are not well simulated by the CM3 model (i.e., no to Q1) and turn our attention to Q2 by investigating the atmospheric circulations associated with REP days.

3.4.3. Circulation Patterns Associated With Regional Extreme Precipitation

The atmospheric circulation during the REP days is similar in the reanalysis record and the CM3 historical runs, aside from a modest southward shift in the composite storm location in the GCM that appears to be a manifestation of latitudinal bias in the jet.

Figure 3-4 shows the composite time-lagged geopotential height and specific humidity anomalies at 700 hPa (Z_{700} and Q_{700}) preceding and during the MAM REP days for the observed record. The dominant features of the atmospheric development of the REP are similar to those found in Nakamura et al. (2012) for the top 20 floods in the Ohio River Basin and include the following:

1. A zonal dipole pattern in the anomalous Z_{700} field at latitudes between about 35°N and 45°N preceding and accompanying the REP events.
2. The dipole pattern migrates eastward beginning approximately 3 days prior to the REP events, accompanied by an intensification of the dipole and significant northward low-level wind anomalies (not shown).
3. A well-organized positive anomaly in the Q_{700} field over the Ohio River Basin along the interface of low and high Z_{700} anomalies that peaks during the day of the event.
4. An anomalous warm surface and low-level temperature anomaly that stretches from the Gulf of Mexico up to the Northeast U.S. (not shown), indicating that the REP events are often associated with frontal systems which in turn are often coupled with extratropical cyclones (not shown).
5. An anomalous high-pressure ridge in the northwest Pacific Ocean south of the Gulf of Alaska that starts to intensify at least 4 days prior to the REP day and persists through the day after the REP day. This north Pacific ridge appears to be a lower frequency pattern that together with the pressure dipole (noted above) forms a tripole structure spanning from the eastern Pacific to the western Atlantic during REP days that is reminiscent of the wavenumber 6 pattern.

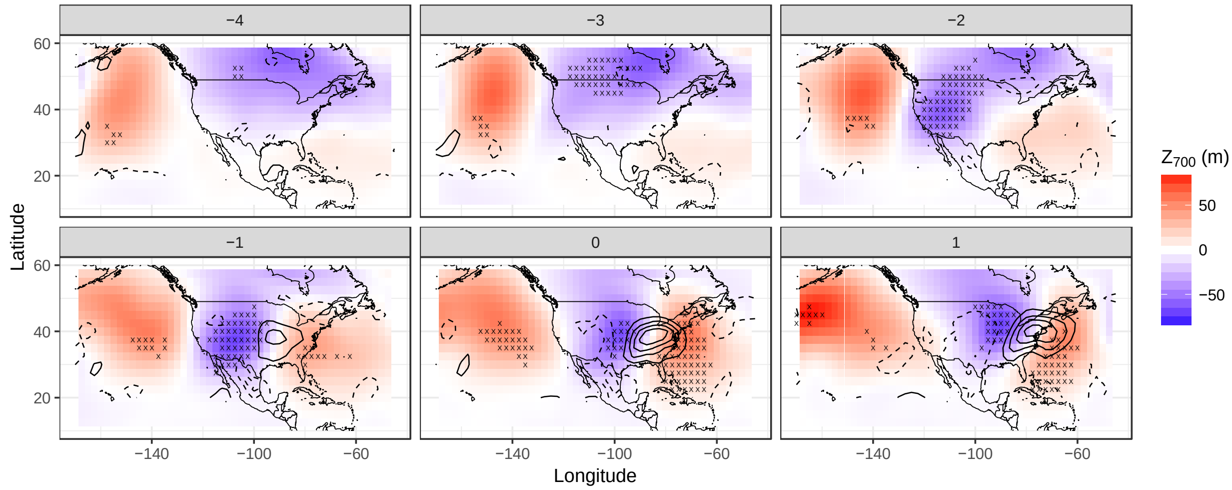


Figure 3-4 Daily composites of Z700 anomalies (shades) and Q700 (contours at 4 × 10−4 kg kg−1) from 4 days before each MAM REP event to 1 day following the event for the observed-reanalysis record. Solid contours represent positive anomalies and dashed contours represent negative anomalies. A cross indicates that at least 80% of composite members (i.e., at least 37 of the 46 REP events) had Z700 anomalies of the same sign in that location.

The most consistent of the atmospheric features associated with the REP days is a high-pressure system (Western Atlantic ridging) which is for some events related to an intensified and westward-extended subtropical high. Another consistent feature is the presence of a low-pressure system in the western U.S. that forms about 2–3 days prior to the REP days.

Despite the bias in the rainfall field, the CM3 ensemble member composites of Z_{700} (Figure 3-5 and supporting information Figure S5) during MAM REP events show a similar pattern of troughing west of the basin and ridging east of the basin, compared to the reanalysis record. There are, however, a few subtle differences. The ridging patterns associated with REP days in the CM3 model have a tendency to extend to the north-east of the study area, while in the reanalysis record the ridging tends to extend over locations to the south-east of the study area. The CM3 model also shows a southward displacement of the low-pressure center relative to the reanalysis record, evident in the extent and location of precipitation during study region REP days (stronger/weaker southeast/northwest precipitation during GCM REP events can be seen in the difference between the GCM and observation percentile precipitation during REP events in supporting information Figure S3). This is likely related to a southward displacement of the storm tracks in the CM3 model, which can be seen in the enhanced (suppressed) standard deviation of MAM 700 hPa geopotential height to the south (north) of 30°N (35°N) in the GCM ensemble members compared to reanalysis (supporting information Figure S6) and the clear southward displacement of the springtime jet (supporting information Figure S7). We also note the absence of the REP-associated ridging in the north Pacific in the GCM, which along with the higher frequency wave train associated with REPs in the GCM, suggests that the GCM can produce REP days in the Ohio River Basin without the presence of hemispherically organized flow compared to the observed-reanalysis record. Despite the modest latitudinal bias, and the lack of a clear tripole pattern, we highlight that the Z_{700} patterns associated with MAM REP events are largely similar between the GCM and reanalysis.

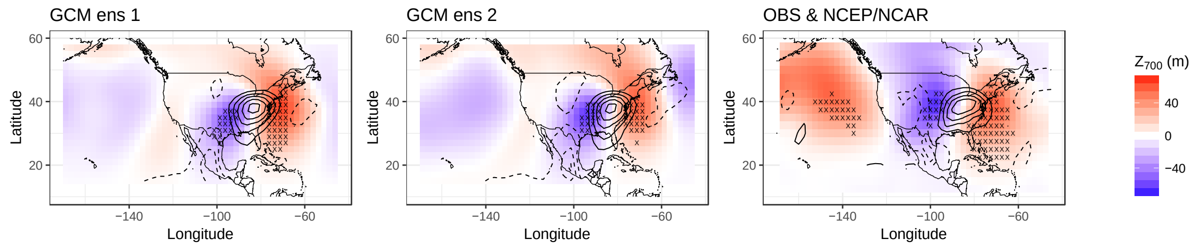


Figure 3-5 Same as Figure 3-4 but for the day of the REP event (lag = 0) and each of the GFDL CM3 GCM ensemble members and the observed-reanalysis record (plots). As in Figure 3-4, a cross indicates that at least 80% of composite members had Z_{700} anomalies of the same sign in that location. This 80% criteria translates to at least 83 out of 103 REP events, 92 out of 115 REP events, and 37 out of 46 REP events, for the two CM3 ensemble members and the observed-reanalysis record, respectively.

3.4.4. Atmospheric Indices

In this section, we show that the GCM appears to reasonably simulate the distributional and persistence features of five atmospheric indices that modulate the likelihood of REP events. This is critical to the conditional simulation strategy proposed in section 3.4.5.

Given that the CM3 model credibly represents the pressure dipole associated with MAM REP events, we define two indices by geopotential heights in boxes to the east and west of the Ohio River Basin. We call these indices the and (for the low and high pressure systems associated with the REP days) and define them as the mean of Z_{700} in the western and eastern boxes, respectively, shown in Figure 3-6. The boxes have a large meridional extent to capture both the center of the low-pressure storms in the GCM REP days and the observed REP days (Figure 3-5). We also define an index by the mean Z_{700} in the large box in the northwest Pacific during the 3 days prior to the current day. We call this index and include it to represent the impact of a strong wavetrain with a center of high pressure in the North Pacific on the probability of REP event (Figure 3-4 and Figure 3-6). We also define two indices to capture the atmospheric conditions over the Ohio River Basin. The first of these indices is defined as the mean of Q_{700} over the basin and is called HUM; we assume that higher values of moisture over the basin increase the probability of a REP day. The next of these indices is the mean of ω_{700} over the basin and is called OMG. This index is important since it represents the existent or absence of local convergence and uplift that is important for the occurrence of precipitation.

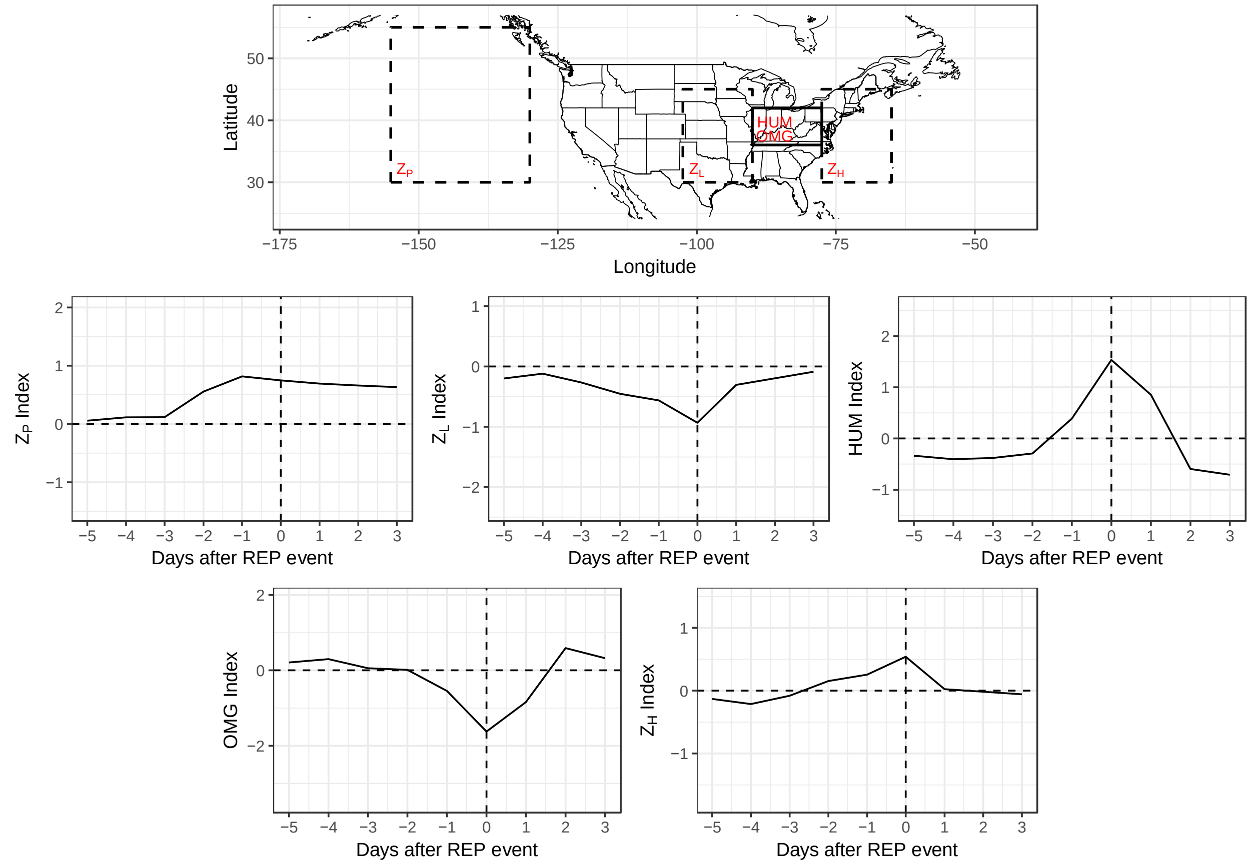


Figure 3-6 (top) The regions that define each of the atmospheric indices. The index names are shown in red. The Ohio River Basin, shown in more detail in Figure 3-1 is shaded in dark gray. The index is defined by the average Z_{700} within the area between 130°W and 155°W and 30°N and 55°N (leftmost dashed box), the index is defined by the average Z_{700} within the area between 87.5°W and 102.5°W and 30°N and 45°N (middle dashed box), and the index is defined by the average Z_{700} within the area between 62.5°W and 77.5°W and 30°N and 45°N (rightmost dashed box). The OMG and HUM indices are defined using the average atmospheric vertical velocity and specific humidity within the area between 77.5°W and 90°W and 36°N and 42°N (solid box). (middle and bottom) The index values prior to and after the REP events. The black line shows the median index value. The dark shaded area shows the range capturing the middle 50% of days, while the light shaded area shows the range capturing the middle 90% of days. All figures use the observed REP record and the corresponding reanalysis-based atmospheric indices.

All five of these indices are defined as their standardized quantities (subtracting their seasonal mean and dividing by their seasonal standard deviation) following Karl et al. (1990). Most importantly, all five of these indices modulate the probability of REP occurrence (Figure 3-6). It should be noted, however, that the daily reanalysis-based indices have been defined by the 12 UTC to 12 UTC values to match the temporal grid of the CPC data while the CM3 indices have been defined on a standard daily grid that begins and ends with 0 UTC to match the daily temporal grid of the CM3 precipitation. We assume that the relationship between the indices and REP occurrence is insensitive to this temporal grid difference. Based on (Figure 3-6), we conclude that indices that are associated with the onset of REP events can be identified from reanalysis (i.e., yes to **Q2**) and turn our attention to **Q3** by investigating the simulation of the atmospheric indices in the CM3 GCM.

Figure 3-7 illustrates that the distributional and persistence properties of REP occurrence of the indices are reasonably well simulated by the GCM (i.e., *yes* to **Q3**). Supporting information Table S1 quantitatively illustrates (based on Wilcoxon rank sum tests) that the distributions of the atmospheric indices based on the GCM and reanalysis are more similar than the distributions of REP days per year based on the GCM precipitation and the observed precipitation. There are, however, a few differences between the GCM and reanalysis indices. These differences include slightly lower HUM index autocorrelations, slightly higher autocorrelation, and slightly higher persistence of index values in its lowest tenth percentile for the GCM (Figure 3-7; bottom second). It seems likely that the persistence bias of the HUM index partially explains the reduced persistence in the GCM-based REP days compared to the observed. On the other hand, the other notable persistence biases of the and indices should increase the probability of back-to-back REP days in the GCM compared to the observed record. Based on (Figure 3-7), we conclude that the atmospheric indices associated with REP events are credibly simulated by the GCM (i.e., yes to **Q3**). We now turn our attention to the problem of directly using these indices to simulate the REP events (i.e., **Q4**).

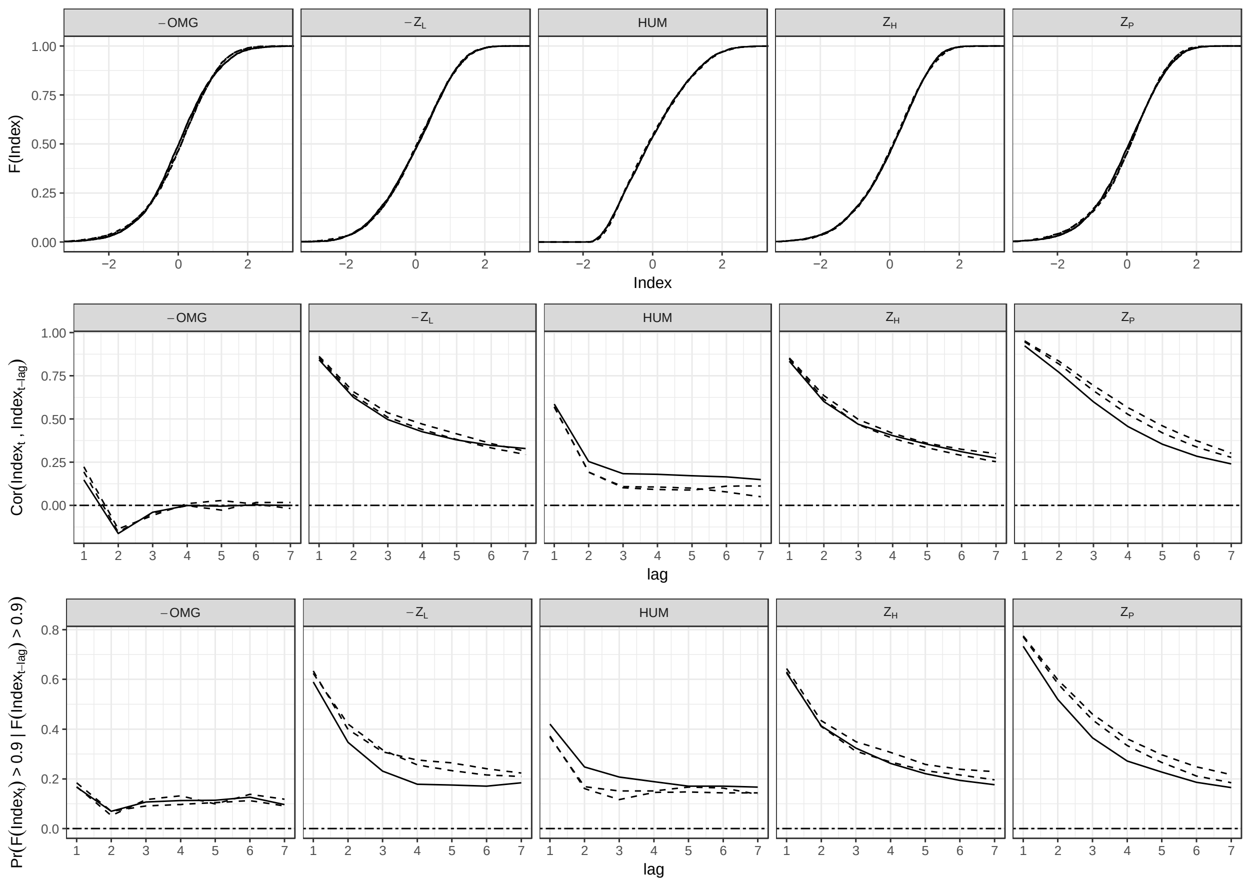


Figure 3-7 (top) Cumulative distribution function for the MAM indices. (middle) The serial correlation function for the MAM indices. (bottom) The serial tail persistence of the MAM indices when in high states as shown by the probability of the index being above the 90th percentile on day t, given that the index was above that percentile on day t-lag, where lag values of 1–10 are shown along the x axis. In all figures, the solid line is the reanalysis-based indices and the dashed lines are the GCM ensemble member-based indices. Negative OMG and are shown for easier interpretation since low values of these two indices are associated with REP days.

3.4.5. Conditional Simulation

In this section, we turn our attention to **Q4** and demonstrate that

1. the conditional simulation of REP days based on a regression on the atmospheric indices addresses the bias in the historic record;
2. a future upward trend in REP day frequency is projected both when using the raw GCM precipitation fields and when using the conditional simulation model based on GCM-derived atmospheric indices;
3. this positive trend appears to be driven both by a trend in the moisture index (which is in turn at least partially the result of increasing temperatures), and by trends in the other indices.

To set up the logistic regression-based simulation model, with a binary response variable (REP or no REP), we assume that the Z_H, Z_L, Z_P, OMG, and HUM indices on day *t* linearly modulate the probability of REP occurrence on day *t*. Based on this assumption, we define a logistic regression model to estimate the probability of a REP day given the five indices (equation 3.2). We estimate from the observation-derived REPs and reanalysis-derived indices (equation 3.2). We refer to these parameter estimates as b_{Z_L}, b_{Z_H}, b_{Z_p}, $b_{Z_{HUM}}$ and $b_{Z_{OMG}}$. We use a fully Bayesian model implemented in Stan (Carpenter et al., 2017) in **R**. We use diffuse normal prior distributions with means of 0 and standard deviations of 25 and 5 for the α and β parameters, respectively.

$$\mathrm{P}(\mathrm{REP}_t^{\mathrm{obs}}=1)=\frac{\exp\left[\alpha+\beta_{Z_L}\left(\mathrm{Z_L}_{t+1}^{\mathrm{reanal}}\right)+\beta_{Z_H}\left(\mathrm{Z_H}_t^{\mathrm{reanal}}\right)+\beta_{Z_P}\left(\mathrm{Z_P}_t^{\mathrm{reanal}}\right)+\beta_{\mathrm{HUM}}\left(\mathrm{HUM}_t^{\mathrm{reanal}}\right)+\beta_{\mathrm{OMG}}\left(\mathrm{OMG}_t^{\mathrm{reanal}}\right)\right]}{1+\exp\left[\alpha+\beta_{Z_L}\left(\mathrm{Z_L}_{t+1}^{\mathrm{reanal}}\right)+\beta_{Z_H}\left(\mathrm{Z_H}_t^{\mathrm{reanal}}\right)+\beta_{Z_P}\left(\mathrm{Z_P}_t^{\mathrm{reanal}}\right)+\beta_{\mathrm{HUM}}\left(\mathrm{HUM}_t^{\mathrm{reanal}}\right)+\beta_{\mathrm{OMG}}\left(\mathrm{OMG}_t^{\mathrm{reanal}}\right)\right]} \tag{3.2}$$

where *t* is a time index and REP is the regional extreme precipitation indicator (either 0 or 1).

After fitting this model on the observed/reanalysis record, we are able to simulate REP days from the GCM-derived indices using equation 3.3. Specifically, we sample from a Bernoulli distribution for each day with probability of a REP as computed from equation 3.3. We retain 1,000 samples for each day.

$$\mathrm{P}(\widehat{\mathrm{REP}}_t^{\mathrm{mod}}=1)=\frac{\exp\left[a+b_{Z_L}\left(\mathrm{Z_L}_{t+1}^{\mathrm{mod}}\right)+b_{Z_H}\left(\mathrm{Z_H}_t^{\mathrm{mod}}\right)+b_{Z_P}\left(\mathrm{Z_P}_t^{\mathrm{mod}}\right)+b_{\mathrm{HUM}}\left(\mathrm{HUM}_t^{\mathrm{mod}}\right)+b_{\mathrm{OMG}}\left(\mathrm{OMG}_t^{\mathrm{mod}}\right)\right]}{1+\exp\left[a+b_{Z_L}\left(\mathrm{Z_L}_{t+1}^{\mathrm{mod}}\right)+b_{Z_H}\left(\mathrm{Z_H}_t^{\mathrm{mod}}\right)+b_{Z_P}\left(\mathrm{Z_P}_t^{\mathrm{mod}}\right)+b_{\mathrm{HUM}}\left(\mathrm{HUM}_t^{\mathrm{mod}}\right)+b_{\mathrm{OMG}}\left(\mathrm{OMG}_t^{\mathrm{mod}}\right)\right]} \tag{3.3}$$

3.4.5.1. MODEL CHECKING

To verify that our model captures a substantial portion of the variance in the record, we first evaluate the ability of our model to reproduce the observed record by fitting the model on the first 42 years (1950–1991; about three quarters of the data) and predicting the last 14 years. We use these time intervals so that the calibration sample contains at least several years of the relatively data rich period after the introduction of satellite observations systems in the late 1970s and early 1980s. The model is only able to capture a small portion of the day-to-day variation with daily hit rates of 12% and 11% for the calibration and testing samples, respectively, and false alarm rates of 88% for both the calibration and testing samples. If we allow the simulation to be off by 1 day in either direction, however, then we have hit rates of 22% and 14% and false alarm rates of 0%. Supporting information Figures S8 and S9 show that the model captures a portion of the interannual variation and has a negative bias with respect to representing the persistence of REP days. Lastly, the proposed model explains 33% of the deviance in the data and partially reproduces the spectral peaks at 3–4 years and 7–8 years when fit on the full historical data (supporting information Figure S10). In summary the physical variables that we have identified only explain a portion of the variance in the REP record and can therefore be seen as necessary but not sufficient to predict day-to-day REP occurrence with high probability. This model is potentially useful, however, for understanding long-term changes in the REP frequency associated with changes to these underlying physical variables, as we show below.

3.4.5.2. SIMULATION RESULTS

The results of our conditional simulation based on the GCM-derived atmospheric indices and the reanalysis-observation coefficient estimates for the historical record are shown in Figure 3-8. When the model is estimated based on the full historic reanalysis-observed record, the regression coefficient estimates for b_{Z_L}, b_{Z_H}, b_{Z_p}, $b_{Z_{HUM}}$ and $b_{Z_{OMG}}$ have means of -0.72, 0.65, 0.41, 0.90, and -1.11, and standard deviations of 0.18, 0.30, 0.18, 0.25, and 0.21. All coefficients are of their expected sign and the HUM and OMG indices have a strongest effect on the probability of REP occurrence. The bias in the REP frequency is substantially reduced through the use of this simulation model (compare Figure 3-8 to the top of Figure 3-3), while the persistence bias is still significant. The bottom row of supporting information Tables S1 and S2 quantitatively illustrate (based on Wilcoxon rank sum tests) that the distributions of GCM-index-based simulated REP days per year and observed REP days per year are more similar than the distributions of GCM-precipitation-based REP days per year and observed REP days per year.

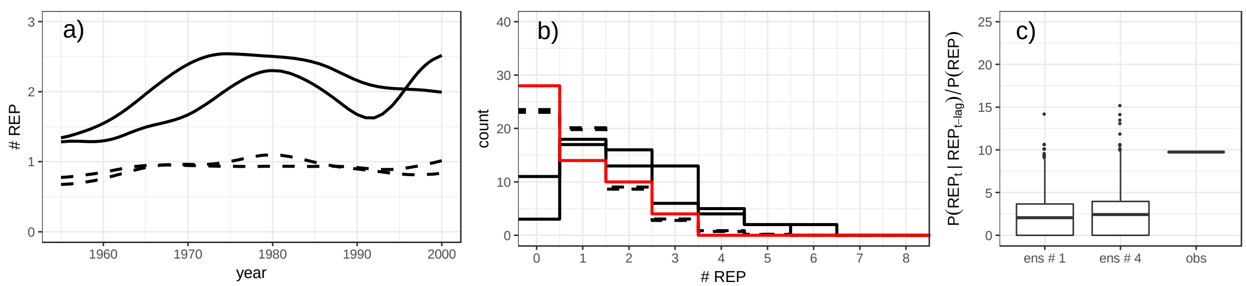


Figure 3-8 (a) The number of MAM REP days by year based on the two GFDL CM3 ensemble member's precipitation fields (black solid lines), the mean of the simulated REP counts obtained via the regression on the indices derived from the two GFDL CM3 ensemble member's Z_{700}, Q_{700}, and ω_{700} fields (black dashed lines), and the 50th and 95th percentile prediction intervals based on the 1,000 simulations (dark and light shaded regions, respectively). All data has been Gaussian kernel smoothed (bandwidth = 10 years) before the mean and prediction intervals are computed. The first and last 5 years of the smooths have been truncated from the figure to avoid edge effects. (b) The counts for the number of MAM REP days by year for the observed record (solid red line), the record derived from the GFDL GCM CM3 precipitation fields (solid black lines), and the mean of the simulations for each ensemble member (dashed black lines). (c) Probability of a MAM REP day on a day given that a REP day occurred the day prior divided by the marginal probability of a REP day for the observed record and the REP simulated records for the two ensemble members and the observed record. The boxplot whiskers extend to points within 1.5 times the interquartile range above the 75th percentile, and any observation outside of this range is shown as a point.

We use a future simulation of the CM3 GCM under the RCP 8.5 forcing scenario to simulate daily REP records via our conditional simulation model for the years 2006 to 2100 and compare these projections against future daily REP records estimated directly from the GCM's precipitation field (Figure 3-9). The standardization of the indices was still based on the historical mean and standard deviation. We also compare our future simulations against projections based on a linear bias-corrected version of the GCM REPs where we assume that the past frequency bias in the GCM REP record is multiplicative and representative of GCM REP frequency bias in the future. Our simulation model projects a significant increasing trend throughout much of the 21st century similar to that projected by the GCM precipitation fields, although the index-based projections show lower absolute REP frequency. The rescaled GCM precipitation field derived projection (blue line in Figure 3-9), i.e., the bias-corrected GCM REP projection, deviates substantially from the mean index simulation projections in the late period of the 21st century.

However, the rescaled projections do lie within the 95th percentile prediction interval of the simulation model projections. The observation that a positive, albeit weaker trend exists even after our conditional simulation provides some evidence that an increasing trend may occur. However, we emphasize restraint in this interpretation, since both approaches assume the RCP 8.5 forcing scenario and that the large-scale circulation patterns in the future are well represented by the CM3 model physics.

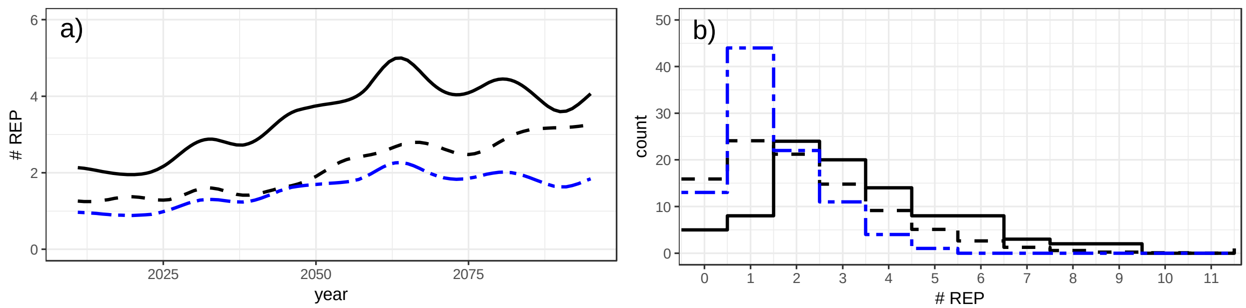


Figure 3-9 (a) The projected number of MAM REP days by year based on the GFDL CM3 RCP 8.5 ensemble member precipitation field (black solid line), the mean of the simulated REP counts obtained via the regression on the GCM-based indices (black dashed lines), and the 50th and 95th percentile prediction intervals based on the 1,000 simulations (dark and light shaded regions, respectively). The blue dashed line is the projected MAM REP record when we assume that the historical bias between the GCM and observed REP frequency is multiplicative and stationary and we rescale the projection based on the GCM precipitation field. In this case, this amounts to dividing the solid black line by about 2.2. All data has been Gaussian kernel smoothed (bandwidth = 10 years) before the mean and prediction intervals are computed. The first and last 5 years of the smooths have been truncated from the figure to avoid edge effects. (b) The counts for the number of MAM REP days by year with corresponding line colors and types as in Figure 3-9a.

3.4.5.3. MOISTURE TREND CONTRIBUTION

It is notable that the increase in REP frequency estimated by our conditional sampling model is driven by a positive shift in the probability distribution of the HUM as well as other indices. To explore the relative contribution of the moisture changes (HUM) versus changes in the other indices, we performed additional simulations using the last 30 years of GCM output from each of the 20th and 21st centuries (1970–1999 and 2070–2099). We first compute the mean change in all GCM-derived indices between these two time periods (using the GCM ensemble mean for the historic period). We find that the mean MAM HUM increases by about 0.6 (i.e., about half a standard deviation). Then we use the regression estimates from the full observed historical record, but simulate REPs using three sets of predictors: (1) using the GCM indices for the 2070–2099 time period; (2) removing the trend in the HUM index by subtracting 0.6 from all HUM index values from 2070 to 2099 and then simulating the REPs for the 2070–2099 time period; (3) using the GCM indices for the 1970–1999 time period. We retain 1,000 simulations for each of these scenarios and plot the resulting REP incidence in Figure 3-10. The median increase in the GCM simulations using our procedure from 1970–1999 to 2070–2099 is about 200% when all index trends are included. It is only 60% when the trends in the HUM are removed. These results suggest that, given our model, about two thirds of the future increase in MAM REPs is due to a humidity increase.

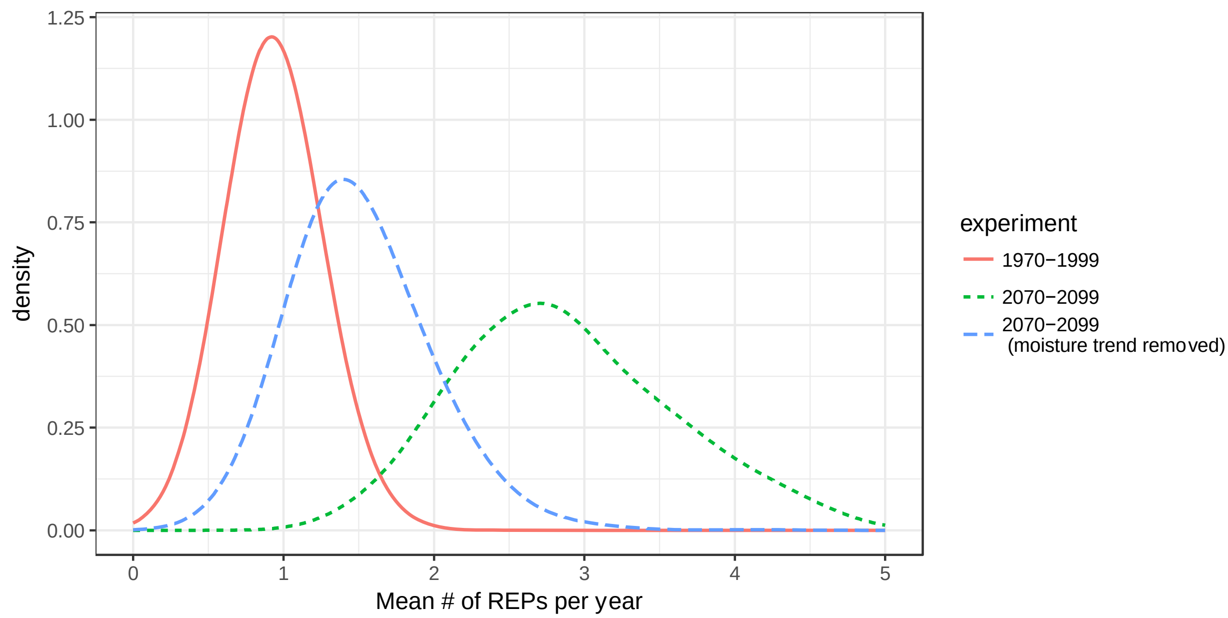


Figure 3-10 Kernel density smoothed probability density functions showing the mean number of simulated MAM REP days over the 30 year periods of 1970–1999 (red line) and 2070–2099 (short-dashed green line) and 2070–2099 after the trend in the HUM index has been removed (long-dashed blue line). Each curve is composed from 1,000 points that represent the mean # of REPs per year in a 30-year simulation.

3.5. CONCLUSION

3.5.1. SUMMARY

Precipitation is the primary climate input into the modeling of extreme riverine floods. Consequently, hydrologists need to consider how to best use future predictions of regional climate in GCMs, given that many factors contribute to the well-documented biases in GCM-based precipitation simulation. We were interested in an approach that provided a diagnostic of the physical factors associated with such biases. Next we were interested whether these factors could be used to achieve a better representation of the causal factors associated with extreme precipitation, and especially with regional extreme precipitation in a large river basin (the Ohio as the example), such that future GCM simulations could be used to statistically assess potential changes.

We began by defining a regional extreme precipitation index, illustrating its relationship to extreme streamflows in the study region, and investigating the dominant atmospheric circulation patterns associated with the precipitation events. Next we showed that the frequency and persistence properties of this regional extreme precipitation index are not well simulated by a GCM, but that the large-scale atmospheric circulation indices (defined by large-scale geopotential height, moisture, and vertical velocity fields) that are strongly associated with the extreme precipitation are credibly simulated by the same GCM. Then we constructed a logistic regression model to simulate the regional extreme precipitation index at the daily scale based on five atmospheric indices. This simulation framework greatly reduced the frequency bias in the historic record of the GCM REP days. Using this model for future projections we found that future GCM simulations likely overestimate the total number of regional extreme precipitation events out to the year 2100. However, an increasing trend in REP occurrence in the 21st century, attributed to trends in both the moisture index and other circulation indices, is still evident in our simulations. We acknowledge that our approach still relies on the assumptions that the relationship between the large-scale climate indices and the REP occurrence is stationary into the future and that our regression is valid over the ranges of the indices in the future GCM runs.

3.5.2. Relationship to Bias-Correction and Downscaling Approaches

Similarly to many bias-correction and downscaling techniques, we assume that the GCM is deficient in its simulation of processes that link the global-synoptic scale circulations and the grid-scale processes that determine precipitation over a specific river basin which may represent just a few grid cells of the GCM. We developed our approach with the following common limitations of bias-correction and downscaling approaches in mind. Using most bias-correction techniques (e.g., Friederichs & Hense, 2007; Goly et al., 2014; Gutmann et al., 2014; Piani et al., 2010a, 2010b; Pierce et al., 2015; Yang et al., 2005) for extrapolation into the future projections is uncertain given that most approaches do not explicitly identify the underlying model deficiencies (Ehret et al., 2012; Dittes et al., 2018). Many statistical downscaling schemes to recover precipitation estimates from large-scale circulation features (e.g., Wilby et al., 2002) have been proposed, including many tailored for use in future climate projection (see Maraun et al., 2010, and references therein). However, it is often unclear how to adapt weather generator (e.g., Thorndahl et al., 2017) and weather typing approaches (e.g., Jacobeit et al., 2003; Muñoz et al., 2015) in a nonstationary climate. Dynamical downscaling (e.g., Schmidli et al., 2007) is another option but is computationally expensive (Wilby et al., 2002) and is often sensitive to precipitation-related parameterizations and the size of the embedded domain used (Leduc & Laprise, 2009; Liu et al., 2011). Regression downscaling (e.g., Wilby et al., 2002) is computationally cheap and is more able to deal with nonstationary conditions. However, the regressions often do not represent the extremes well and explain only a relatively small portion of the variance in the data (Wilby et al., 2002). The latter point is particularly problematic if a goal of the downscaling is to estimate future precipitation conditions since it may be that the model sensitivity to future regional forcing is below the level of the noise (i.e., a signal in the precipitation may simply be an artifact of the model parameterization and estimation).

3.5.3. Caveats and Further Discussion

A shortcoming of our model is that it does not fully capture the serial correlation in the REP process, as represented by supporting information Figure S8 and Figure 3-8. The negative persistence bias in the reconstruction of the observed-reanalysis record suggests that our model could be improved through the incorporation of other variables that inform the temporal clustering of the REP days. While the persistence bias can be partially mitigated by including lagged REP days as predictors, we chose not to include a lagged REP predictor because the predictor was not significant given the presence of the other predictors and the absence of a lagged REP predictor greatly reduces the computational cost of the simulation model.

As previously noted, our simulation method does not avoid a reliance on the assumption that circulation (and associated moisture) changes are well simulated into the future by the GCM. The frequency bias in the regional extreme precipitation record appears to be a manifestation of inflated spatial correlation of high intensity precipitation. The precipitation event persistence bias appears to be a manifestation of a strong and southerly displaced springtime jet in the GCM that results in faster moving storms and lower autocorrelation in the humidity field over our study region. We were able to limit our simulation model's sensitivity to the southerly displacement bias by using standardized indices (i.e., a form of bias correction to translate the mean to be ≈0 and the rescale the variance to be ≈1), but we did not fully address the persistence bias. Other approaches to handling biases in GCM circulation fields have been proposed when credible precipitation fields are the desired outcome; Eden et al. (2012) advocate for the approach of nudging GCM fields toward observed fields and then letting the GCM simulate the precipitation fields.

Two deficiencies of this approach, however, are the reliance on the convective parameterization scheme of the GCM (which can be particularly problematic during summer), and an inability to project future precipitation events because there exists no future reanalysis field to nudge toward. Thus, it is difficult to avoid a reliance on GCM circulation fields when it comes to projecting regional scale precipitation events, and difficult to estimate the validity of the GCM under warming and other related and relevant changes such as changing midlatitude meridional temperature gradients due to Arctic Amplification (Barnes & Screen, 2015). Finally, the simulation model presented in this paper has been shown to better predict the REP event frequency than do the GCM precipitation fields and is therefore plausibly useful for understanding the future trends in REP frequency. Having said that, the simulation model does not necessarily provide daily time sequences that are appropriate for impacts models given supporting information Figure S9 and Figure 3-8.

3.6. REFERENCES

Barnes, E. A., & Screen, J. A. (2015). The impact of Arctic warming on the midlatitude jet-stream: Can it? Has it? Will it? *Wiley Interdisciplinary Reviews: Climate Change*, 6(3), 277– 286. http://doi.org/10.1002/wcc.337 Carpenter, B., Gelman, A.,

Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). *Stan*: A probabilistic programming language. *Journal of Statistical Software*, 76(1). http://doi.org/10.18637/jss.v076.i01

Cioffi, F., Conticello, F., & Lall, U. (2016). Projecting changes in Tanzania rainfall for the 21st century. *International Journal of Climatology*, 36(13), 4297– 4314. http://doi.org/10.1002/joc.4632

Cioffi, F., Conticello, F., Marotta, L., & Telesca, V. (2017). Large scale climate and rainfall seasonality in a Mediterranean Area: Insights from a non-homogeneous Markov model applied to the Agro-Pontino plain. *Hydrological Processes*, 31(3), 668–686. http://doi.org/10.1002/hyp.11061

Dai, A. (2006). Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, 19(18), 4605– 4630. http://doi.org/10.1175/JCLI3884.1

Dirmeyer, P. A., & Kinter, J. L. (2010). Floods over the U.S. Midwest: A regional water cycle perspective. *Journal of Hydrometeorology*, 11(5), 1172– 1181. http://doi.org/10.1175/2010JHM1196.1

Dittes, B., Špačková, O., Schoppa, L., & Straub, D. (2018). Managing uncertainty in flood protection planning with climate projections. *Hydrology and Earth System Sciences Discussions*, 22(4), 2511– 2526. http://doi.org/10.5194/hess-22-2511-2018

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., et al. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *Journal of Climate*, 24(13), 3484– 3519. http://doi.org/10.1175/2011JCLI3955.1

Durman, C. F., Gregory, J. M., Hassell, D. C., Jones, R. G., & Murphy, J. M. (2001). A comparison of extreme European daily precipitation simulated by a global and a regional climate model for present and future climates. *Quarterly Journal of the Royal Meteorological Society*, 127(573), 1005– 1015. http://doi.org/10.1002/qj.49712757316

Eden, J. M., Widmann, M., Grawe, D., & Rast, S. (2012). Skill, correction, and downscaling of GCM-simulated precipitation. *Journal of Climate*, 25(11), 3970– 3984. http://doi.org/10.1175/JCLI-D-11-00254.1

Ehret, U., Zehe, E., Wulfmeyer, V., & Liebert, J. (2012). Should we apply bias correction to global and regional climate model data? *Hydrology and Earth System Sciences*, 16, 3391– 3404. http://doi.org/10.5194/hess-16-3391-2012

Friederichs, P., & Hense, A. (2007). Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly Weather Review*, 135(6), 2365– 2378. http://doi.org/10.1175/MWR3403.1

Gimeno, L., Drumond, A., Nieto, R., Trigo, R. M., & Stohl, A. (2010). On the origin of continental precipitation. *Geophysical Research Letters*, 37, L13804. http://doi.org/10.1029/2010GL043712

Goly, A., Teegavarapu, R. S. V., & Mondal, A. (2014). Development and evaluation of statistical downscaling models for monthly precipitation. *Earth Interactions*, 18, 18. http://doi.org/10.1175/EI-D-14-0024.1

Grasso, L. D. (2000). The differentiation between grid spacing and resolution and their application to numerical modeling abstract. *Bulletin of the American Meteorological Society*, 81, 579– 580.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen-Skaugen, T. (2012). Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations—A comparison of methods. *Hydrology and Earth System Sciences*, 16(9), 3383– 3390. http://doi.org/10.5194/hess-16-3383-2012

Gutmann, E., Pruitt, T., & Clark, M. (2014). An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resources Research*, 50, 7167– 7186. http://doi.org/10.1002/2014WR015559

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., et al. (2013). Global flood risk under climate change. *Nature Climate Change*, 3(9), 816– 821. http://doi.org/10.1038/nclimate1911

Holsclaw, T., Greene, A. M., Robertson, A. W., & Smyth, P. (2015). A Bayesian Hidden Markov Model of daily precipitation over South and East Asia. *Journal of Hydrometeorology*, 17, 3– 25. http://doi.org/10.1175/JHM-D-14-0142.1

Hughes, J. P., Guttorp, P., & Charles, S. P. (1999). A non-homogeneous hidden Markov model for precipitation occurrence.

Journal of the Royal Statistical Society: Series C (Applied Statistics), 48(1), 15– 30. http://doi.org/10.1111/1467-9876.00136

IPCC. (2012). In C. B. Field et al. (Eds.), Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of Working Groups I and II of the Intergovernmental Panel on Climate Change (582 pp.). Cambridge, UK: Cambridge University Press.

Jacobeit, J., Wanner, H., Luterbacher, J., Beck, C., Philipp, A., & Sturm, K. (2003). Atmospheric circulation variability in the North-Atlantic-European area since the mid-seventeenth century. *Climate Dynamics*, 20, 341– 352. http://doi.org/10.1007/s00382-002-0278-0

Jongman, B., Ward, P. J., & Aerts, J. C. J. H. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, 22(4), 823– 835. http://doi.org/10.1016/j.gloenvcha.2012.07.004 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), 437– 471. http://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Karl, T. R., Wang, W.-C., Schlesinger, M. E., Knight, R. W., & Portman, D. (1990). A methodof relating general circulation model simulated climate to the observed local climate. Part I: Seasonal statistics. *Journal of Climate*, 3(10), 1053– 1079. http://doi.org/10.1175/1520-0442(1990)003<1053:AMORGC>2.0.CO;2

Kendon, E. J., Roberts, N. M., Senior, C. A., & Roberts, M. J. (2012). Realism of rainfall in a very high-resolution regional climate model. *Journal of Climate*, 25(17), 5791– 5806. http://doi.org/10.1175/JCLI-D-11-00562.1

Knippertz, P., & Wernli, H. (2010). A Lagrangian climatology of tropical moisture exports to the Northern Hemispheric extratropics. *Journal of Climate*, 23(4), 987– 1003. http://doi.org/10.1175/2009JCLI3333.1

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., et al. (2013). Flood risk and climate change: Global and regional perspectives. *Hydrological Sciences Journal*, 59(1), 1– 28. http://doi.org/10.1080/02626667.2013.857411

Kundzewicz, Z. W., Lugeri, N., Dankers, R., Hirabayashi, Y., Döll, P., Pińskwar, I., et al. (2010). Assessing river flood risk and adaptation in Europe—Review of projections for the future. *Mitigation and Adaptation Strategies for Global Change*, 15(7), 641– 656. http://doi.org/10.1007/s11027-010-9213-6

Kwon, H.-H., Lall, U., & Obeysekera, J. (2009). Simulation of daily rainfall scenarios with interannual and multidecadal climate cycles for South Florida. *Stochastic Environmental Research and Risk Assessment*, 23, 879– 896. http://doi.org/10.1007/s00477-008-0270-2

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

Leduc, M., & Laprise, R. (2009). Regional climate model sensitivity to domain size. *Climate Dynamics*, 32(6), 833– 854. http://doi.org/10.1007/s00382-008-0400-z

Lehner, B., Döll, P., Alcamo, J., Henrichs, T., & Kaspar, F. (2006). Estimating the impact of global change on flood and drought risks in Europe: A continental, integrated analysis. *Climatic Change*, 75(3), 273– 299. http://doi.org/10.1007/s10584-006-6338-4

Liang, X. Z., Pan, J., Zhu, J., Kunkel, K. E., Wang, J. X. L., & Dai, A. (2006). Regional climate model downscaling of the U.S. summer climate and future change. *Journal of Geophysical Research*, 111, D10108. http://doi.org/10.1029/2005JD006685

Liu, C., Ikeda, K., Thompson, G., Rasmussen, R., & Dudhia, J. (2011). High-resolution simulations of wintertime precipitation in the Colorado Headwaters Region: Sensitivity to physics parameterizations. *Monthly Weather Review*, 139(11), 3533– 3553. http://doi.org/10.1175/MWR-D-11-00009.1

Lu, M., Lall, U., Schwartz, A., & Kwon, H. (2013). Precipitation predictability associated with tropical moisture exports and circulation patterns for a major flood in France in 1995. *Water Resources Research*, 49, 6381– 6392. http://doi.org/10.1002/wrcr.20512

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

Maraun, D., Wetterhall, F., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., et al. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48, RG3003. http://doi.org/10.1029/2009RG000314

Muñoz, Á. G., Goddard, L., Robertson, A. W., Kushnir, Y., & Baethgen, W. (2015). Crosstime scale interactions and rainfall extreme events in Southeastern South America for the Austral Summer. Part I: Potential predictors. *Journal of Climate*, 28(19), 7894– 7913. http://doi.org/10.1175/JCLI-D-14-00693.1

Nakamura, J., Lall, U., Kushnir, Y., Robertson, A. W., & Seager, R. (2012). Dynamical structure of extreme floods in the U.S. Midwest and the UK. *Journal of Hydrometeorology*, 14(2), 485– 504. http://doi.org/10.1175/JHM-D-12-059.1

NWS Internet Services Team. (2015). *Hydrologic Information Center—Flood loss data*.

Perry, C. A. (2000). *Significant floods in the United States during the 20th century—USGS measures a century of floods* (Fact Sheet 024-00). Reston, VA: U.S. Geological Survey.

Piani, C., Haerter, J. O., & Coppola, E. (2010a). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology*, 99(1–2), 187– 192. http://doi.org/10.1007/s00704-009-0134-9

Piani, C., Weedon, G. P., Best, M., Gomes, S. M., Viterbo, P., Hagemann, S., et al. (2010b). Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *Journal of Hydrology*, 395(3–4),

199– 215. http://doi.org/10.1016/j.jhydrol.2010.10.024
Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., & Hegewisch, K. C. (2015). Improved bias correction techniques for hydrological simulations of climate change. *Journal of Hydrometeorology*, 16(6), 2421– 2442. http://doi.org/10.1175/JHM-D-14-0236.1
Robertson, A. W., & Smyth, S. K. P. (2003). *Hidden Markov models for modeling daily rainfall occurrence over Brazil* (Tech. Rep. UCI-ICS 03–27). Irvine: University of California, Irvine.
Schmidli, J., Goodess, C. M., Frei, C., Haylock, M. R., Hundecha, Y., & Ribalaygua, J. (2007). Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps. *Journal of Geophysical Research*, 112, D04105. http://doi.org/10.1029/2005JD007026
Smith, A., Lott, N., Houston, T., Shein, K., Crouch, J., & Enloe, J. (2016). *U.S. billion-dollar weather & climate disasters 1980–2016* (Technical report). Asheville, NC: NOAA National Centers for Environmental Information.
Steinschneider, S., & Lall, U. (2016). Spatiotemporal structure of precipitation related to tropical moisture exports over the eastern United States and its relation to climate teleconnections. *Journal of Hydrometeorology*, 17(3), 897– 913. http://doi.org/10.1175/JHM-D-15-0120.1
Stephens, G. L., L'Ecuyer, T., Forbes, R., Gettlemen, A., Golaz, J. C., Bodas-Salcedo, A., et al. (2010). Dreary state of precipitation in global models. *Journal of Geophysical Research*, 115, D24211. http://doi.org/10.1029/2010JD014532
Thorndahl, S., Andersen, A., & Larsen, A. (2017). An event-based stochastic point rainfall series resampler for statistical replication and climate projection of historical rainfall series. *Hydrology and Earth System Sciences*, 21(9), 4433– 4448. http://doi.org/10.5194/hess-21-4433-2017
Wehner, M. F., Smith, R. L., Bala, G., & Duffy, P. (2010). The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model. *Climate Dynamics*, 34(2), 241– 247. http://doi.org/10.1007/s00382-009-0656-y
Wilby, R. L., Dawson, C. W., & Barrow, E. M. (2002). Sdsm—A decision support tool for the assessment of regional climate change impacts. *Environmental Modelling & Software*, 17, 145– 157.
Winsemius, H. C., Aerts, J., Beek, L., Bierkens, M., Bouwman, A., Jongman, B., et al. (2015). Global drivers of future river flood risk. *Nature Climate Change*, 6, 1– 5. http://doi.org/10.1038/nclimate2893
Xie, P., Chen, M., & Shi, W. (2010). *CPC unified gauge-based analysis of global daily precipitation*. Paper presented at the 24th Conference on Hydrology, Atlanta, GA.
Yang, D., Kane, D., Zhang, Z., Legates, D., & Goodison, B. (2005). Bias corrections of long-term (1973–2004) daily precipitation data over the northern regions. *Geophysical Research Letters*, 32, L19501. http://doi.org/10.1029/2005GL024057

4. CHARACTERIZATION AND CLIMATE INFORMED PROJECTIONS STREAMFLOW EXTREMES IN THE OHIO RIVER BASIN

This chapter is published in Water Resources Research*:*

Schlef, K.E., François, B., Robertson, A.W., Brown, C., 2018. A General Methodology for Climate-Informed Approaches to Long-Term Flood Projection-Illustrated With the Ohio River Basin. Water Resour. Res. 54, 9321–9341. https://doi.org/10.1029/2018WR023209

4.1. EXECUTIVE SUMMARY

Estimating future hydrologic floods under non-stationary climate is a key challenge flood management. Climate informed approaches to long-term flood projection are an appealing alternative to traditional modeling chains. This work formalizes climate informed approaches into a general methodology consisting of four steps: (1) selection of predictand representing extreme events, (2) identification of credible large-scale predictors which mechanistically control the occurrence and magnitude of the predictand, (3) development of a statistical model relating the predictors to the predictand, and (4) projection of the predictand by forcing the model with predictor projections. These four steps, developed from a review of the current literature, are demonstrated for multiple gages in the northwest Ohio River Basin in the United States Midwest as a case study. Floods are defined as annual maximum series events in January through April and are linked to geopotential height and soil moisture predictors in a Bayesian linear regression model. The projections generally show a slight decrease in future flood magnitude and demonstrate the transparency of the climate informed approach. An initial step for more general application across the United States as well as remaining challenges associated with climate informed flood projection are discussed.

4.2. TECHNICICAL APPROACH

Previous literature has applied climate informed approaches to streamflow extremes in the East, Daqinghe, and Mekong River basins in China (Delgado et al., 2014; Li & Tan, 2015; Zhang et al., 2015), Spain and France in Europe (López & Francés, 2013; Renard & Lall, 2014), various states in the western and central U.S. (Bracken et al., 2018; Condon et al., 2015; Jain & Lall, 2001; Kwon et al., 2008; Mallakpour et al., 2017; Sankarasubramanian & Lall, 2003; Villarini et al., 2013), the Mono River in West Africa (Tramblay et al., 2014) and the Negro River in Brazil (Lima et al., 2015). These approaches have also been applied to precipitation extremes in California in the U.S. (Ouarda & El-Adlouni, 2011; Shang et al., 2011; Steinschneider & Lall, 2015) and Southern Queensland in Australia (Sun et al., 2014). Based on a review of this literature, we have formalized the variety of climate informed approaches into a four-step general methodology (Table 4-1).

Table 4-1: The four-step methodology and associated methods. References are not intended to be exhaustive (indicates a relevant method despite not specifically a climate informed approach).*

Step	Key Idea	Primary Methods and Selected References
1. Select predictand	The predictand should be useful to stakeholders and enable identification of predictors	Predictand definition - Annual maximum series (López & Francés, 2013) - Peaks over threshold (Renard & Lall, 2014)
		Special predictand cases - Seasonal (Sun et al., 2014) - Impaired (López & Francés, 2013) - Regional (Steinschneider & Lall, 2015)

2. Identify credible large-scale predictors	The predictors should (a) mechanistically control the occurrence and magnitude of predictand, (b) be robust under climate change, and (c) be well-simulated in GCMs	Identify mechanistic control - Literature review (Delgado et al., 2014) - Time series correlation (Kwon et al., 2008) - Composite analysis (Jain & Lall, 2001) - Weather typing (Robertson et al., 2015)* - Simulation experiments (Cook, 1999)* - Bayesian identification (Renard & Lall, 2014)
		Assess robustness - First principles - Thermodynamic and dynamic drivers (Greene et al., 2011)*
		Assess GCM performance - Literature review - Calculate performance metrics
3. Formulate, calibrate, and validate statistical model	The model should represent the link between the predictand and the predictors	Form of model - Simple linear regression (Lima et al., 2015) - GAMLSS (Villarini et al., 2012) - Quantile regression (Sankarasubramanian & Lall, 2003) - Bayesian model (Renard & Lall, 2014)
		Calibration by maximum likelihood optimization - Optimization algorithm (Delgado et al., 2014) - Markov Chain Monte Carlo (Steinschneider & Lall, 2015)
4. Project predictand into future	The projections should be credible and useful to stakeholders	Combine projections - Multi-model mean (Delgado et al., 2014) - Performance-based weighting (Stocker et al., 2010)
		Calculate return periods - Separate value in each time step (Delgado et al., 2014) - Assume stationarity within selected time period - Use design life level concept (Condon et al., 2015)

4.2.1. Step 1: Select Predictand

The first step is to select the predictand. A necessary step to any flood frequency analysis, here the key idea is to define extreme events in such a way that is both useful to stakeholders and for which a relationship to large-scale predictors either exists or can be identified. Thus, streamflow data is preferentially from unimpaired stations, although Condon et al. (2015) use unregulated flows simulated from a hydrologic model and López & Francés (2013) and Li & Tan (2015) develop methods to account for the impact of reservoirs at impaired sites. While some studies analyze only one gage (e.g., Delgado et al., 2014) or fit a unique model to each gage within a region (e.g., López & Francés, 2013), a regional analysis allows for better identification of climate effects (Sun et al., 2014) and is more informative for emergency preparedness given that extremes are often not isolated events (Shang et al., 2011). Regionalization requires identification of a hydro-climatologically homogeneous region (Sun et al., 2014) and can be accomplished using techniques such as Bayesian modeling (Renard & Lall, 2014; Steinschneider & Lall, 2015), copulas (Sun et al., 2014), and max-stable processes (Shang et al., 2011).

Once the data is acquired, extreme events are often defined as the annual maximum series (AMS) events (e.g., López & Francés, 2013). In some cases, extreme events are restricted to a particular season to enable identification of a clear link to large-scale ocean-atmospheric patterns (e.g., the summer season in Sun et al., 2014). Alternatively, peaks over threshold (POT) methods have been used to capture both number of occurrences and magnitude (e.g., Renard & Lall, 2014; Steinschneider & Lall, 2015). The choice of AMS or POT will be influenced by what information is useful for decision-making (e.g., POT allows frequency to be modeled separately from magnitude) and whether predictors can be identified (e.g., Mallakpour et al. (2017), Renard & Lall (2014), and Villarini et al. (2013) apply a climate informed approach to the frequency of flood events from POT, but do not model magnitude).

4.2.2. Step 2: Identify Credible Large-Scale Predictors

The second step is to identify credible large-scale predictors. The key idea is that the identified predictors (a) mechanistically control the occurrence and magnitude of extreme events in the region of interest, (b) are robust under climate change, and (c) are relatively well-simulated by GCMs. At the catchment scale, the causative or proximate mechanisms of floods (Merz & Blöschl, 2003) are ultimately generated by ocean-atmospheric patterns, such as extratropical cyclones and sea surface temperature anomalies, operating at much larger spatiotemporal scales, as classically described by Hirschboeck (1988).

In climate informed approaches, ultimate mechanisms to be used as predictors are often identified through review of the hydro-climatology literature or historic reports of flooding (e.g., Delgado et al., 2014). Relationships in the literature can then be tested by comparing the performance of models with different subsets of predictors (e.g., Villarini et al. 2013). In addition to literature review, a simple and often-used method of identifying predictors is correlation of time series of the extreme events to time series of pre-defined indices or gridded fields (e.g., Kwon et al. 2008); however, correlations are "not necessarily optimal for heavily non-Gaussian data" (Renard & Lall, 2014) and may be falsely assumed to indicate causation if the underlying physical processes are poorly understood. Another method is composite analysis, which compares the climate patterns associated with the highest events to climatology (e.g., Jain & Lall 2001). Similarly, in weather typing, the atmospheric circulation patterns associated with extreme events are clustered into types that can be related to large-scale patterns (e.g., Robertson et al. 2015). Another technique often used in the climate sciences literature is that of simulation experiments (e.g., Cook 1999). Finally, Renard & Lall (2014) provide a unique approach to identification of predictors through a Bayesian model that uses maximum likelihood estimation to identify spatial patterns probabilistically related to floods.

Once ultimate mechanisms associated with the predictand are identified as possible predictors, they should be further evaluated for robustness under climate change and how well they are simulated by GCMs. Robustness is important because the predictors are often based on climate variability, but, since the goal is long-term projection, are also intended to be appropriate under changes in mean climate. Given the lack of a good reference, since GCM performance is highly biased, as previously discussed, and observed climate changes are much smaller than projected (IPCC, 2013), a simple beginning point is to roughly estimate expected climate change impacts on floods from first principles. Specifically, the Clausius-Clapeyron equation indicates that increased temperature leads to increased moisture holding capacity of the atmosphere which leads to increased precipitation extremes which leads to greater floods. Similarly, increased temperature will cause more precipitation to fall as rain rather than snow and consequently alter flooding. However, first principles cannot provide a definitive projection of future floods because there are many feedback mechanisms that may accentuate or dampen their effect (Collins et al., 2013; Held & Soden, 2000; O'Gorman & Schneider, 2009). Consequently, we conclude that robustness under climate change can be expected to improve when predictors account for both thermodynamic and dynamic processes (e.g., as used in the downscaling study of Greene et al., 2011), rather than only a change in dynamics (e.g., as in Delgado et al. 2014).

Relatively good simulation by GCMs is important because climate informed approaches are motivated by GCMs' poor simulation of extreme precipitation. The challenge is determining what "relatively good simulation” means. Qualitatively, first-order variables (e.g., temperature) can be expected to be more skillfully simulated than derived or second order variables (e.g., precipitation). Similarly, GCM performance can be expected to increase, to a certain extent, with increasing spatiotemporal scale (e.g., daily data for a grid cell compared to annual data for a region). Quantitatively, many studies have assessed GCM simulation of large-scale patterns (e.g., Bellenger et al., 2014; Fuentes-Franco et al., 2016; Lee & Black, 2013; Ning & Bradley, 2016; Polade et al., 2013; Sheffield et al., 2013; Taschetto et al., 2014; Yim et al., 2015) and may sufficiently indicate performance for highly studied patterns (e.g., ENSO) or performance metrics can be calculated directly.

4.2.3. Step 3: Formulate, Calibrate, and Validate Statistical Model

The third step is to formulate, calibrate, and validate a statistical model. The key idea is that the model is representative of the link between the identified predictors and the extreme events. For a thorough discussion of the statistics underlying non-stationary models of flood events, see the recent review by Salas et al. (2018). The form of the model is often as simple as linear regression of the location and/or scale parameter of the extreme value distribution on the predictor(s) (e.g., Lima et al., 2015). More complex model formulations include using the generalized additive models for location, scale, and shape (GAMLSS) as demonstrated by Villarini et al. (2009) (e.g., López & Francés, 2013), quantile regression techniques (Sankarasubramanian & Lall, 2003), and Bayesian modeling (Renard & Lall, 2014) with the possibility to include copulas that account for interdependent variables (Bracken et al., 2018). While the identifying climate indices are often used directly as predictors in a model, López & Francés (2013) perform dimension reduction on the identified predictors. Models are calibrated through optimization of likelihood functions using techniques such as the shuffled complex evolutionary algorithm (e.g., Delgado et al. 2014), or in a Bayesian context, Monte Carlo sampling methods (e.g., Steinschneider & Lall, 2015). Model performance can be evaluated in a variety of ways; those employed in climate informed approaches include but are not limited to deviance statistics (Delgado et al., 2014), the Bayesian and Akaike Information Criterions (Lima et al., 2015; Zhang et al., 2015), assessment of residuals using worm plots and quantile-quantile plots (López & Francés, 2013; Zhang et al., 2015), and leave-one-out cross validation (Lima et al., 2015; Renard & Lall, 2014; Sankarasubramanian & Lall, 2003).

4.2.4. Step 4: Project Predictand into Future

The fourth step is to project the predictand into the future. The key idea is that the projections should be credible and useful to stakeholders. Projections of the predictors used to force the statistical model can be stochastically generated time series or short-term forecasts (e.g., Kwon et al., 2008; Lima et al., 2015), but here the focus is on long-term projection using GCM simulations, which surprisingly has only been accomplished by Delgado et al. (2014) and Tramblay et al. (2014). Credibility primarily rests on the choice of credible predictors and the performance of the statistical model, discussed previously. However, the projections can also be compared to what is expected from first principles and to projections obtained from model chains; determining the source of discrepancies among the different types of projections can improve assessment of credibly and knowledge of flood-generating processes.

Creating projections useful for stakeholders requires combining projections forced by different GCMs and calculating return periods. A common method for combining projections is the multi-model mean, which Delgado et al. (2014) use for a subset of GCMs selected according to performance. The multi-model mean performs better than individual models on average, but lack of independence between models leads to small sample sizes (Edwards, 2011; Knutti et al., 2010; Weigel et al., 2010). Alternatively, models may be weighted based on performance metrics such as the climate prediction index (Murphy et al., 2004), reliability ensemble averaging (Giorgi & Mearns, 2002), a variable convergence score (Johnson & Sharma, 2009) and error metrics (Gleckler et al., 2008; Pierce et al., 2009). However, no commonly accepted weighting scheme exists (Stocker et al., 2010) and projections obtained using performance-based weighting may be only minimally different from those obtained using the multi-model mean (Chen et al., 2017). For calculation of return periods, most of the statistical models used in climate informed approaches provide a distinct flood distribution for each time step (e.g., Delgado et al., 2014). Furthermore, available statistical techniques for calculating non-stationary returns periods (Cooley, 2013; Salas & Obeysekera, 2014) are not viable because they require summing to infinity. The available options are either to assume stationarity of the projections within a chosen time period and follow traditional flood frequency analysis techniques (England et al., 2015) or to employ the concept of design life level developed by Rootzén & Katz (2013) and demonstrated by Condon et al. (2015).

4.3. DATA AND METHODS

The Ohio River Basin in the Midwest U.S. periodically experiences devastating floods; the most recent occurred in 2015, but records of floods and extreme river stages date back to 1773 (Horton & Jackson, 1913a; NWS, 2017a). Here, the Ohio River Basin is used as a case study to illustrate the application of the four-step general methodology to climate informed flood projection.

4.3.1. SELECT PREDICTAND IN THE OHIO RIVER BASIN

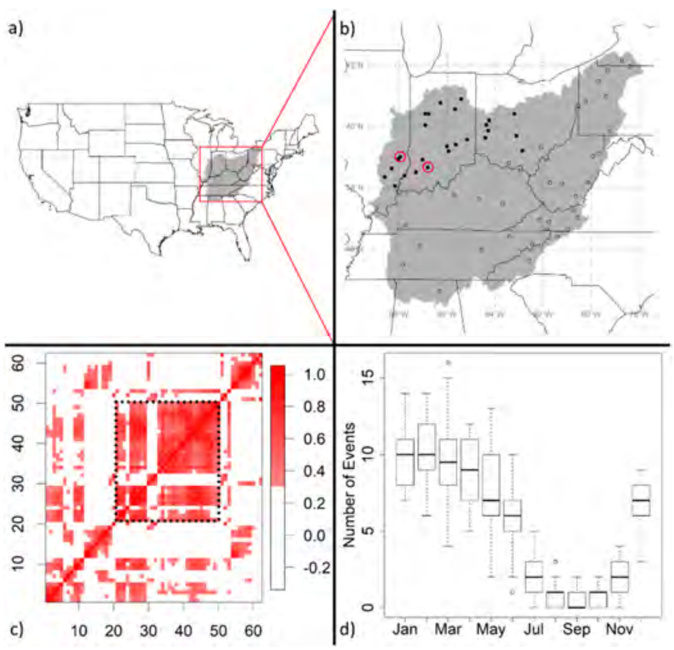


Figure 4-1 Diagnostic information about flood events. (a) The Ohio River Basin (USGS hydrologic units 5 and 6); (b) the Ohio River Basin with the HCDN gages plotted as dots (filled dots indicate gages in the northwest region and the encircled dots are the example gages used subsequently); (c) correlations between the JFMA maximum event of each gage (represented by a number) to the other gages (the dashed box indicates high correlations associated with the northwest region); (d) the number of AMS events in each month for gages in the northwest region.

Daily streamflow data for the basin was obtained from the Hydro-Climatic Data Network (HCDN) (Landwehr & Slack, 1992). HCDN Gages are designated as unimpaired or reference gages based on analysis of data up to 1988. A total of 62 gages were identified that have a basin area greater than approximately 500 km^2 (200 square miles) and have less than 0.1% data missing between 1950 and 2015, the chosen analysis period (Figure 4-1a-b). Through exploratory diagnostics, including correlation (Figure

4-1c) and empirical orthogonal functions (not shown), maximum flood events in January through April (JFMA) for the 26 gages in the northwest region of the basin were found to be strongly related. For the 26 gages, as expected from historic records (see supporting information), JFMA maximum flood events capture between 50% - 71% of annual maximum series floods (Figure 4-1d) and are likely to have winter teleconnections (see subsequent section). For these reasons, all subsequent analysis was performed on JFMA maximum flood events for the 26 northwest region gages. Based on the Mann-Kendall test, only 3 of the 26 gages show a significant (positive) trend in JFMA maximum flood events. Here and throughout the paper, significance is reported at the 5% level unless noted otherwise. To regionalize the analysis, principal component analysis was performed on standardized JFMA maximum event time series for the 26 gages. The first and second principal components, which comprise 66% and 11% of the total variance, respectively, were retained for further analysis. Across all gages, the correlation between the observed time series and the time series reconstructed from the first two principal components is significant, ranging from approximately 0.77 to 0.95. Based on the Shapiro-Wilk test for normality, the residuals of the reconstructed time series relative to the observed fail to reject the null hypothesis of a normal distribution for all but four gages.

4.3.2. Identify Credible Large-Scale Predictors in the Ohio River Basin

We begin our identification process with a literature review. At the local scale, Berghuijs et al. (2016) found that AMS flood events in the region are primarily caused by rainfall in excess of soil moisture storage capacity. Historic reports also note the importance of antecedent soil moisture (see supporting information). At the daily synoptic scale, Schwarz (1961) identifies two typical atmospheric flow patterns, a quasi-stationary front and an occluding low, that can cause heavy winter or spring rains in the region. Both patterns are characterized by a low-pressure trough to the west and a high pressure ridge to the east, which draws warm moist sub-tropical air into the region, often associated with the phenomena known as an atmospheric river or tropical moisture export (Knippertz & Wernli, 2017; Nayak & Villarini, 2017). This pressure configuration has been explicitly linked to extreme floods in the region by composite analysis and weather typing (Nakamura et al., 2013; Robertson et al., 2015) and is related to the negative phase of the Pacific/North American (PNA) teleconnection pattern (Roller et al., 2016). Its converse, which is related to the positive phase of the PNA, causes cyclonic circulation that inhibits tropical moisture transport and results in drier conditions during the winter season (Ning & Bradley, 2014). The PNA is an intrinsic mode of intra-seasonal atmospheric variability which is strongly impacted, through Rossby wave propagation, by inter-annual tropical climate variability, particularly ENSO (Horel & Wallace, 1981; Wallace & Gutzler, 1981). The PNA is also impacted by inter-decadal variability associated with the Pacific Decadal Oscillation (PDO) (Yu & Zwiers, 2007). These mechanisms explain the significant correlations observed between winter rainfall or streamflow in the region and PNA (Coleman & Rogers, 2003; Mallakpour & Villarini, 2017), ENSO (Gershunov & Barnett, 1998a, 1998b; Higgins et al., 2007; Montroy, 1997; Rogers & Coleman, 2003) and PDO (Higgins et al., 2007; Mallakpour & Villarini, 2017; Mantua & Hare, 2002).

The relationships identified in the literature were tested using correlation maps. Gridded data was obtained for global monthly sea surface temperatures (Rayner, 2003), global monthly geopotential heights at the 500 mbar pressure level (Kalnay et al., 1996) which is the pressure level used to calculate PNA, and U.S. monthly soil moisture (Fan & van den Dool, 2004). Each grid cell of each data set was converted from a monthly to annual time series by taking the maximum value within either the concurrent months of JFMA for soil moisture, or the preceding months of December through February (DJF) for sea surface temperatures and geopotential heights at the 500 mbar pressure level. The correlation value between the first and second principal components of flood events and the 1950 through 2015 time series at every grid cell for every data set was calculated.

Maps of the correlation values reveal significant relationships that corroborate what is expected from the literature (Figure 4-2). The first principal component (PC1) is significantly and negatively correlated to the winter Nino3 region and has significant correlation to a winter geopotential height pattern similar to the PNA with pronounced centers over central Canada and the North Pacific. This is expected because the positive phases of winter ENSO and PNA are associated with drier conditions due to cyclonic circulation inhibiting moisture in the Gulf of Mexico from reaching the basin (Ning & Bradley, 2014). PC1 is also significantly correlated to concurrent soil moistures over the northwest region of the basin, reflecting the importance of soil moisture noted by historic reports (see supporting information) and Berghuijs et al. (2016). The second principal component (PC2) is not significantly correlated with winter sea surface temperatures, but is positively correlated to geopotential heights over the eastern Atlantic, which corresponds to the eastern component of the pressure pattern identified by Nakamura et al. (2013). The second principal component is also positively correlated to soil moistures over the Mississippi River Valley to the west, reflecting the importance of moisture transport from the Gulf of Mexico as discussed in Schwarz (1961).

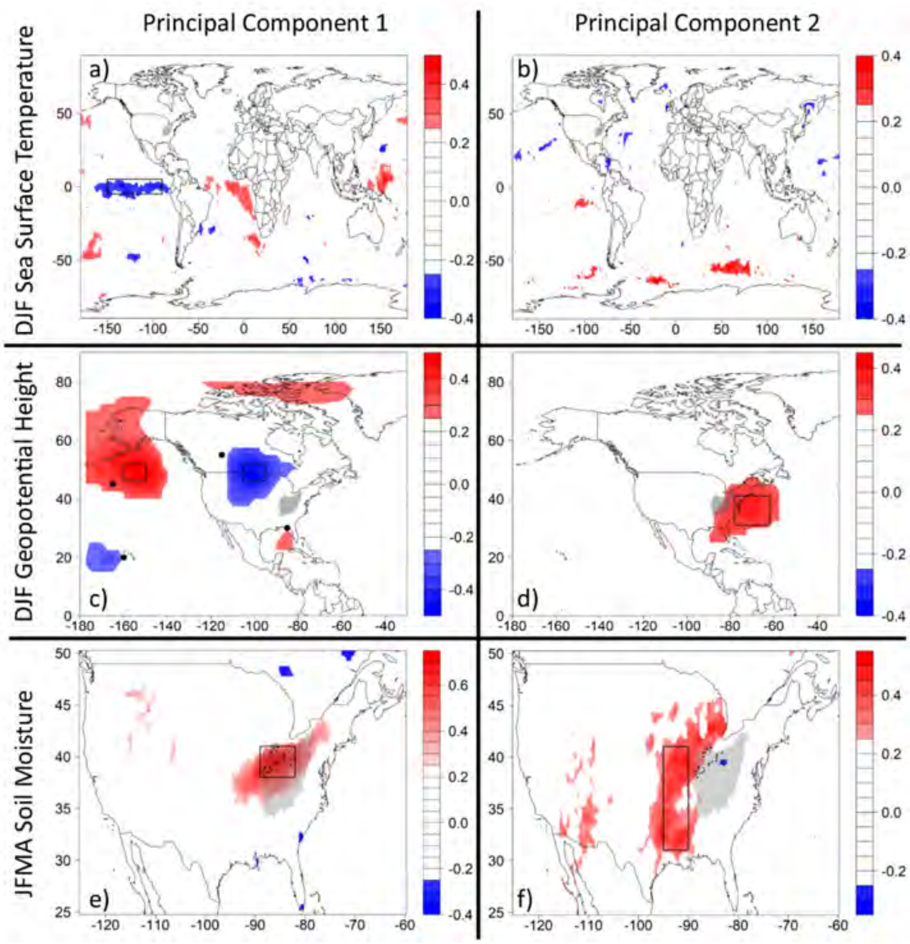


Figure 4-2 Correlation maps of the principal components (a, c, and e are PC1 while b, d, and f are PC2) to the climate fields (a-b are DJF sea surface temperatures, c-d are DJF geopotential heights at the 500 mbar level, and e-f are JFMA soil moistures). The regions used to define predictors are outlined by rectangles and in (c) the dots indicate the PNA centers. The scale indicates the magnitude of the correlation (white areas are insignificant). The basin is shaded grey and in (e-f) the dots indicate the gages in the northwest region. The x- and y-axis labels are longitude and latitude (degrees), respectively.

From the correlation maps, the following predictors were developed and standardized:

- $sst_{Nino3}^{DJF}\mathrm{sst}_{\mathrm{Nino3}}^{\mathrm{DJF}}$ is the DJF sea surface temperatures averaged over the Nino3 region (5S – 5N, 150W – 90W), correlation to $PC1$ is -0.289 (p-value < 0.05)
- hgt_{CC-NP}^{DJF} is the difference in DJF geopotential heights at the 500 mbar level averaged over central Canada (46N – 52N, 160W – 150W) and averaged over the North Pacific (46N – 52N, 160W – 150W), correlation to $PC1$ is -0.530 (p-value < 0.001)
- $soil_{basin}^{JFMA}$ is the JFMA soil moisture averaged over the northwest region of the basin (38N – 41N, 89W – 81W), correlation to $PC1$ is 0.706 (p-value < 0.001)
- hgt_{EA}^{DJF} is the DJF geopotential heights at the 500 mbar level averaged over the eastern Atlantic (31N – 41N, 78W – 62W), correlation to $PC2$ is 0.375 (p-value < 0.01)
- $soil_{west}^{JFMA}$ is JFMA soil moisture averaged over the Mississippi River Valley to the west of the basin (31N – 41N, 95W – 90W), correlation to $PC2$ is 0.505 (p-value < 0.001).

4.3.3. Formulate, Calibrate, and Validate Statistical Model for the Ohio River Basin

From among the possible model formulations, Bayesian modeling was chosen for its ability to clearly represent parameter uncertainty. Given the multiple predictors identified, multiple models for each principal component were developed (Table 4-2). The models were fit over the time period 1950 through 2015 by JAGS in R (Plummer, 2016; Yu-Sung & Yajima, 2015) using three model chains each having 2000 samples with 1000 samples discarded as burn-in. Sufficiently vague priors were placed on the variances (a uniform distribution from zero to 10) and on the coefficients (a normal distribution with mean zero and variance 25). For all models, both the potential scale reduction factor, also known as Gelman's R, and the effective sample size were well within accepted rules of thumb (less than 1.1 and greater than 300, respectively). Predictors are deemed to be significant if the 95% credible interval of the coefficient does not include zero. Model performance is judged by the coefficient of determination, R^2, between the simulated and observed principal components and by the deviance information criterion (DIC) which accounts for parameter uncertainty and is appropriate even when the prior is non-informative or improper (Spiegelhalter et al., 2002; Sun et al., 2014).

The sign of coefficients of the fitted models match what is expected from the correlation maps and the literature; the coefficients for sst_{Nino3}^{DJF} and hgt_{CC-NP}^{DJF} are negative, while the remaining coefficients are positive. The intercept, β, is essentially zero for all models, which is expected given that the mean of the principal components is zero. As evaluations of model performance, R^2 and DIC are inversely related and as model performance improves, the variance decreases. For models with only one predictor, performance improves as proximity increases; for example, models based on soil are better than models based on geopotential height. In the models that use all available predictors (PC1all3 and PC2all2), the 95% credible interval of the coefficient on the least proximate predictor (sst_{Nino3}^{DJF} and hgt_{EA}^{DJF}, respectively) contains zero, indicating that the predictor is not significant. Based on this result, an alternate model for $PC1$ (PC1hgtsl) and the soil-based model for $PC2$ (PC2soil) were chosen for use in all subsequent analysis. Based on the Shapiro-Wilk test for normality, the residuals of the PC1hgtsl and PC2soil models are normal for more than 96% and 93%, respectively, of the 3000 model runs.

Simulated data for each gage based on observed climate can be obtained by (1) sampling from the models to stochastically generate the principal components, (2) back-transforming the new principal components using the loadings, (3) de-standardizing, and (4) taking the exponent. To find a quantile of interest for a given gage, l-moments are used to fit the simulated data to a log Pearson Type 3 (LP3) distribution, chosen based on an l-moments diagram (not shown). Model performance can be further assessed by visual comparison (Figure 4-3) and through statistical tests comparing the empirical cumulative distribution function of the observed data to the data simulated from the model when forced with observed climate. Across all gages, the p-value of the Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) tests for distribution similarity between the observed data and the median of the simulated data ranges from 0.57 to 1.0 and from 0.41 to 1.0 respectively, indicating failure to reject the null hypothesis that the distributions are the same. Across all gages, the correlation of the model median to observed data ranges from 0.4 to 0.72 and the percent of observed data which fall outside the 95% credible interval of the simulated data (i.e., a "miss" rate) ranges from 0% to 17%. For the 100 year flood, the observed magnitude falls within the simulated 95% credible interval for all except three gages (Figure 4-3b). Gage 23 has two anomalously high peaks that the model cannot capture, and gages 39 and 40, which are in close spatial proximity, each have an anomalously low peak and no high peaks, which skew the distribution. Finally, the sensitivity of the model to the predictors was tested by setting the predictors to zero. The results exhibited degraded performance, both visually (not shown) and quantitatively. Overall, based on the tests of model performance described above, the model was deemed satisfactory.

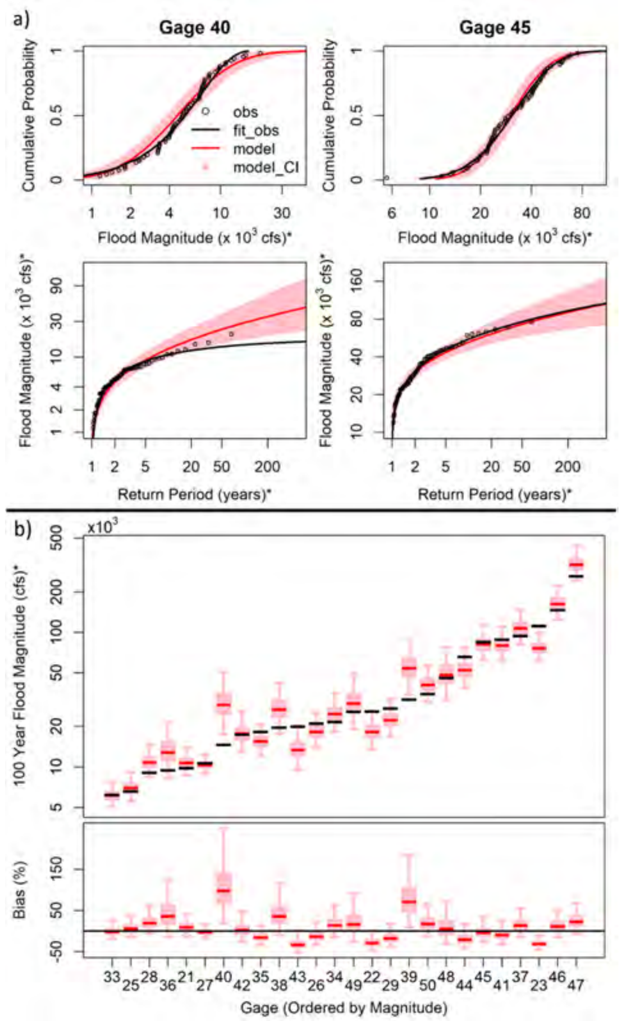


Figure 4-3 Performance of statistical model for (a) two example gages and (b) the magnitude and bias of the 100 year flood. "obs" is the empirical cumulative distribution function based on the Weibull plotting position of the observed data, "fit_obs" is the LP3 fit to the observed data, and "model" and "model_CI" are the median and 95% credible intervals of the LP3 fit to the simulated data from the model forced with observed climate.

Table 4-2: Model form and associated parameters and performance. $N()$ indicates the normal distribution. Values are given as the mean (standard deviation). Chosen models are bolded.

Model	Model Equation	$\boldsymbol{\alpha_1}$	$\boldsymbol{\alpha_2}$	$\boldsymbol{\alpha_3}$	$\boldsymbol{\alpha_4}$	$\boldsymbol{\alpha_5}$	$\boldsymbol{\beta}$	$\boldsymbol{\sigma}$	$\boldsymbol{R^2}$	DIC
PC1sst	$PC1 \sim N\left(\alpha_1 sst_{Nino3}^{DJF} + \beta, \sigma^2\right)$	-1.19 (0.5)	-	-	-	-	-0.01 (0.5)	4.07 (0.37)	0.02 (0.03)	374
PC1hgt	$PC1 \sim N\left(\alpha_2 hgt_{CC-NP}^{DJF} + \beta, \sigma^2\right)$	-	-2.18 (0.45)	-	-	-	-0.01 (0.44)	3.61 (0.33)	0.09 (0.06)	358
PC1soil	$PC1 \sim N\left(\alpha_3 soil_{basin}^{JFMA} + \beta, \sigma^2\right)$	-	-	2.91 (0.38)	-	-	-0.01 (0.38)	3.01 (0.27)	0.25 (0.08)	335
PC1all3	$PC1 \sim N\left(\alpha_1 sst_{Nino3}^{DJF} + \alpha_2 hgt_{CC-NP}^{DJF} + \alpha_3 soil_{basin}^{JFMA} + \beta, \sigma^2\right)$	-0.15 (0.42)	-1.18 (0.45)	2.43 (0.37)	-	-	-0.01 (0.35)	2.79 (0.26)	0.32 (0.08)	328
PC1hgtsl	$PC1 \sim N\left(\alpha_2 hgt_{CC-NP}^{DJF} + \alpha_3 soil_{basin}^{JFMA} + \beta, \sigma^2\right)$	-	-1.27 (0.37)	2.44 (0.38)	-	-	0.00 (0.35)	2.78 (0.25)	0.33 (0.08)	325
PC2hgt	$PC2 \sim N\left(\alpha_4 hgt_{EA}^{DJF} + \beta, \sigma^2\right)$	-	-	-	0.62 (0.21)	-	-0.01 (0.21)	1.66 (0.15)	0.03 (0.04)	256
PC2soil	$PC2 \sim N\left(\alpha_5 soil_{west}^{JFMA} + \beta, \sigma^2\right)$	-	-	-	-	0.87 (0.19)	0.00 (0.19)	1.53 (0.15)	0.08 (0.06)	245
PC2all2	$PC2 \sim N\left(\alpha_4 hgt_{EA}^{DJF} + \alpha_5 soil_{west}^{JFMA} + \beta, \sigma^2\right)$	-	-	-	0.30 (0.20)	0.75 (0.21)	0.00 (0.19)	1.52 (0.14)	0.08 (0.06)	245

A sensitivity analysis was used to assess the implications of assuming a stationary model. Since the Bayesian linear regression model is implemented with 3000 model runs, there are correspondingly 3000 parameter sets, where the mean and standard deviation of each parameter across all sets is given in Table 4-2. For each parameter of the $PC1$ model (α_2, α_3, β, and σ), a delta change of -2, -1, 0, 1, or 2 times the standard deviation was applied to the 3000 member set. For example, for the α_2 parameter, this procedure results in five levels of means -2.01, -1.64, -1.27, -0.9, and -0.53. The analysis used full factorial design (i.e., each level tested with all other levels, 5 (α_2 levels) x 5 (α_3 levels) x 5 (β levels) x 5 (σ levels) = 625 combinations), using the GCM historic and future values as forcing data (discussed in the subsequent section). Only the parameter of the $PC1$ model were included in the sensitivity analysis because $PC1$ accounts for a much larger portion of the variance than $PC2$ and because of increasing computational cost.

4.3.4. Project Predictand into the Future for the Ohio River Basin

To create projections of future flood events, projections of the predictors were obtained from GCM simulations. Specifically, monthly gridded historical runs from 1950 through 2005 and projections from 2006 through 2100 of 500 mbar geopotential heights and soil moisture were obtained from the fifth generation of GCM experiments (CMIP5) directed by the Intergovernmental Panel on Climate Change (Taylor et al., 2012; Van Vuuren et al., 2011). This study used simulations from 10 GCMs (CanESM2, CCSM4, CNRM-CM5, CSIRO-Mk3.6.0, GFDL-CM3, GISS-E2-H, HadGEM2-AO, IPSL-CM5A-MR, MPI-ESM-LR, and NorESM1-M) associated with the historical run and the representative concentration pathway (RCP) 8.5 scenario that had initialization condition r1i1p1. RCP 8.5 was chosen for illustrative purposes and because it is the most extreme climate change scenario in CMIP5. The predictors were calculated in the same way as described for observed data, except that standardization was performed using the GCM historical data.

According to the literature, while CMIP5 GCMs generally replicate the spatial pattern and magnitude of PNA, the slight errors have a large influence on storm track variability (Lee & Black, 2013; Ning & Bradley, 2016). CMIP5 GCM performance in simulating seasonal persistence of soil moisture over North America is poor, likely due to biases in precipitation (Sheffield et al., 2013). In the warm season in particular, CMIP5 GCMs can capture the seasonal variability of soil moisture, but show biases in magnitude which vary by region and by model (Yuan & Quiring, 2017).

The GCMs do not necessarily preserve correlations between the specific predictors used in this study, based on the historical runs. While GCMs correctly simulate the lack of correlation between hgt_{CC-NP}^{DJF} and $soil_{west}^{JFMA}$, they underestimate the correlation between hgt_{CC-NP}^{DJF} and $soil_{basin}^{JFMA}$ (observed is -0.379 but GCMs range from -0.27 to 0.33 with only two significant) and overestimate the correlation between $soil_{basin}^{JFMA}$ and $soil_{west}^{JFMA}$ (observed is 0.573, but GCMs range from 0.54 to 0.81). In contrast, the empirical quantiles of the historical runs, which remove temporal issues, generally match observations, with the largest deviances observed in the distribution tails (not shown, see supporting information). Specifically, GCMs uniformly under- (over-) estimate the lowest (highest) quantiles of hgt_{CC-NP}^{DJF}, and show both positive and negative bias at the lowest (uniformly underestimate the highest) quantiles of $soil_{basin}^{JFMA}$ and $soil_{west}^{JFMA}$.

According to the literature, CMIP5 GCMs show that future intensification of ENSO and PDO will likely increase PNA variability (Fuentes-Franco et al., 2016), but the spatial patterns and amplitude are highly uncertain (Ning and Bradley, 2016). CMIP5 GCMs also show a general consensus of decreasing soil moisture but that there will be increased land-atmospheric coupling driven by soil moisture variations (Dirmeyer et al., 2013).

For the predictors used in this study, the projections exhibit greater inter-model spread than the historical runs (not shown, see supporting information), which is expected. The projected increase in hgt_{CC-NP}^{DJF}, which will cause flood magnitude to decrease, is accentuated at higher quantiles. With the exception of GFDL-CM3, the extreme quantiles of $soil_{basin}^{JFMA}$ do not change, but the spread of the average values increases; thus, the impact on flood magnitude is uncertain. The variable $soil_{west}^{JFMA}$ exhibits similar tendencies as $soil_{basin}^{JFMA}$, with the exception of IPSL-CM5A-MR. Extreme precipitation shows a dramatic projected increase for all but CSIRO-Mk3.6.0.

When the observed predictors in the statistical model are replaced with GCM historical data, there is generally good model performance based on visual inspection of plots similar to Figure 4-3a. When the model is forced with GCM projections, quantiles of interest are obtained by assuming stationarity within a given time period and using l-moments to fit the LP3 distribution. The time period is set using a 61 year moving window ending on every decade from 2010 through 2100; the first moving window, covering 1950 through 2010, is representative of the historical period, although 2006 through 2010 are technically projected by GCMs. Projections of flood events at each gage from the GCMs are combined using a simple multi-model mean or median.

4.4. RESULTS AND DISCUSSION

The results show two key outcomes of the general methodology for climate informed approaches as applied to the Ohio River Basin. The first is the change in flood event distribution between past and future time periods and the second is the attribution of change to various predictors.

4.4.1. CHANGE IN FLOOD EVENT DISTRIBUTION

Figure 4-4 shows projected change in flood magnitude between the last and first moving window. The results for two example gages and two examples GCMs (Figure 4-4a), show that for a given GCM, there is consistency across gages regarding the direction of change (e.g., CanESM2 projects a slight decrease for both gages), likely due to the high correlations observed between the gages and the regionalized model. The performance of GCMs in the historic period relative to the observed data largely follows the model performance when forced with observed predictors; gage 40 is poorly represented, while gage 45 is skillfully represented. However, for a given gage, there is variation among GCMs; CanESM2 is more skillful than GFDL-CM3. CanESM2 shows a slight decrease in flood magnitude between the last and first moving window (2040 through 2100 and 1950 through 2010, respectively), which is representative of most of the GCMs; GFDL-CM3 is an outlier and shows an increase, due to its projected increase in soil moisture (discussed subsequently). The multi-model median is more robust to outliers than the mean, as demonstrated by results for gage 45 (Figure 4-4b); for the future period, the multi-model median is lower than the multi-model mean but almost the same as the multi-model mean without GFDL-CM3. While the performance of GFDL-CM3 over the historic time period is similar to other GCMs, its projected increase in soil moisture is different from the general consensus of decreasing soil moistures (Dirmeyer et al., 2013). Across all gages (Figure 4-4c), the median percent change in the multi-model median is relatively homogeneous (i.e., most gages tend towards the same direction and relative magnitude of change, as expected from using principal components and the flood diagnostics) and increases with increasing return period, but remains negative across the return periods shown. The magnitude of the projected percent change is not clearly linked to a spatial relationship nor to the catchment area (not shown).

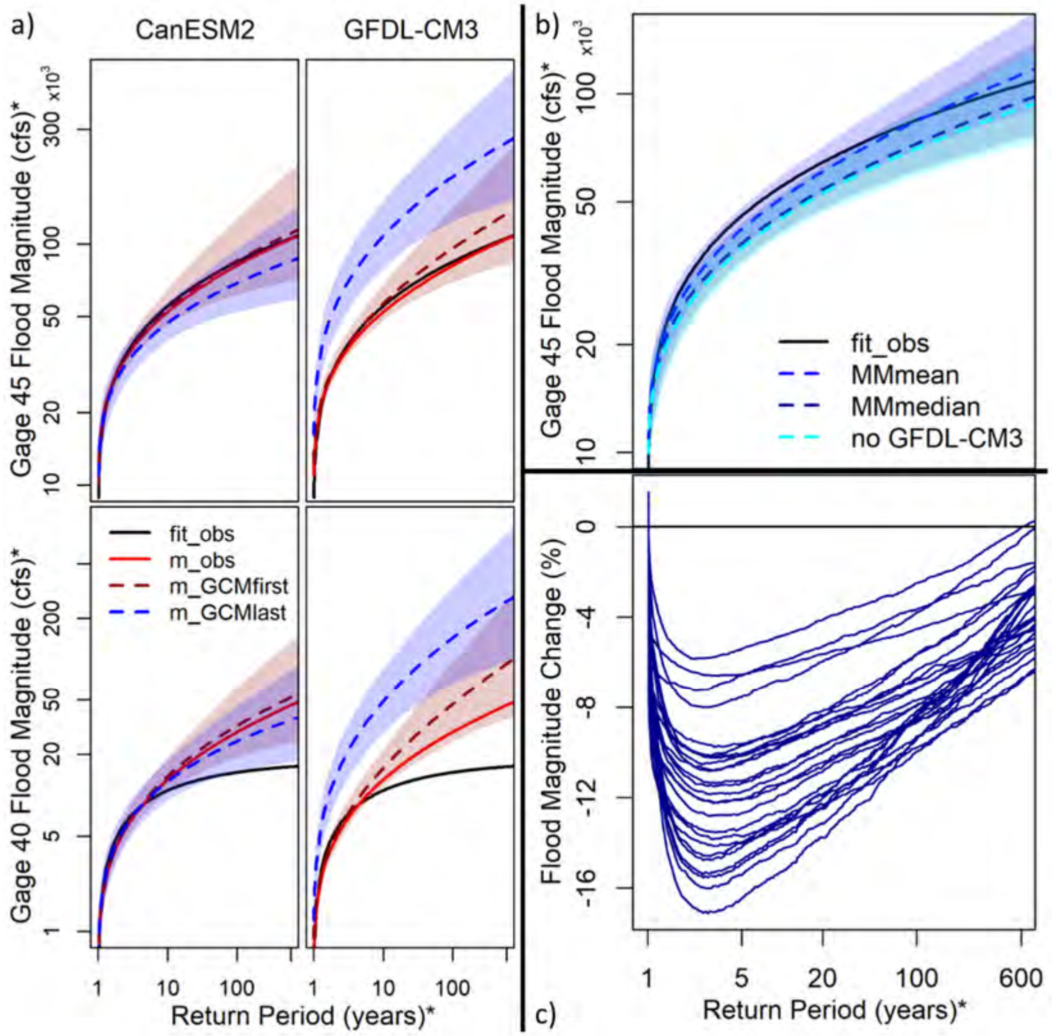


*Figure 4-4 (a) flood magnitude at two example gages as a function of return period for two GCMs (CanESM2, GFDL-CM3) for the median of the LP3 distribution fit to the observed data "fit_obs", to the model output when forced with observed predictors "m_obs", and to the model output when forced with GCM predictors from the first (1950 through 2010, "m_GCMfirst") and last (2040 through 2100, "m_GCMlast") moving windows. The shaded areas are 95% credible intervals (not shown for "m_obs"). (b) the same as (a) except for the multi-model mean with and without GFDL-CM3 (MMmean and no GFDL-CM3, respectively) and the multi-model median (MMmedian) (x-axis is the same as c). (c) the multi-model median of the median percent change in flood magnitude for each gage. * indicates the axis is log scale.*

4.4.2. Attribution of Projected Change to Predictors

What is driving the projected change in flood magnitude for each GCM? The increase associated with GFDL-CM3 is likely driven by the increase in $soil_{basin}^{JFMA}$, but the cause of the decrease associated with CanESM2 is less clear. To answer this question, the effect of individual predictors or subsets of predictors on the projection results was isolated by subtracting the 31 year moving average from all remaining predictors, thus removing any trend, and forcing the statistical model with the modified time series. For illustrative purposes, results are only shown for the 100 year flood for gage 45 for the two GCMs used previously and CSIRO-Mk.3.6.0 (Figure 4-5).

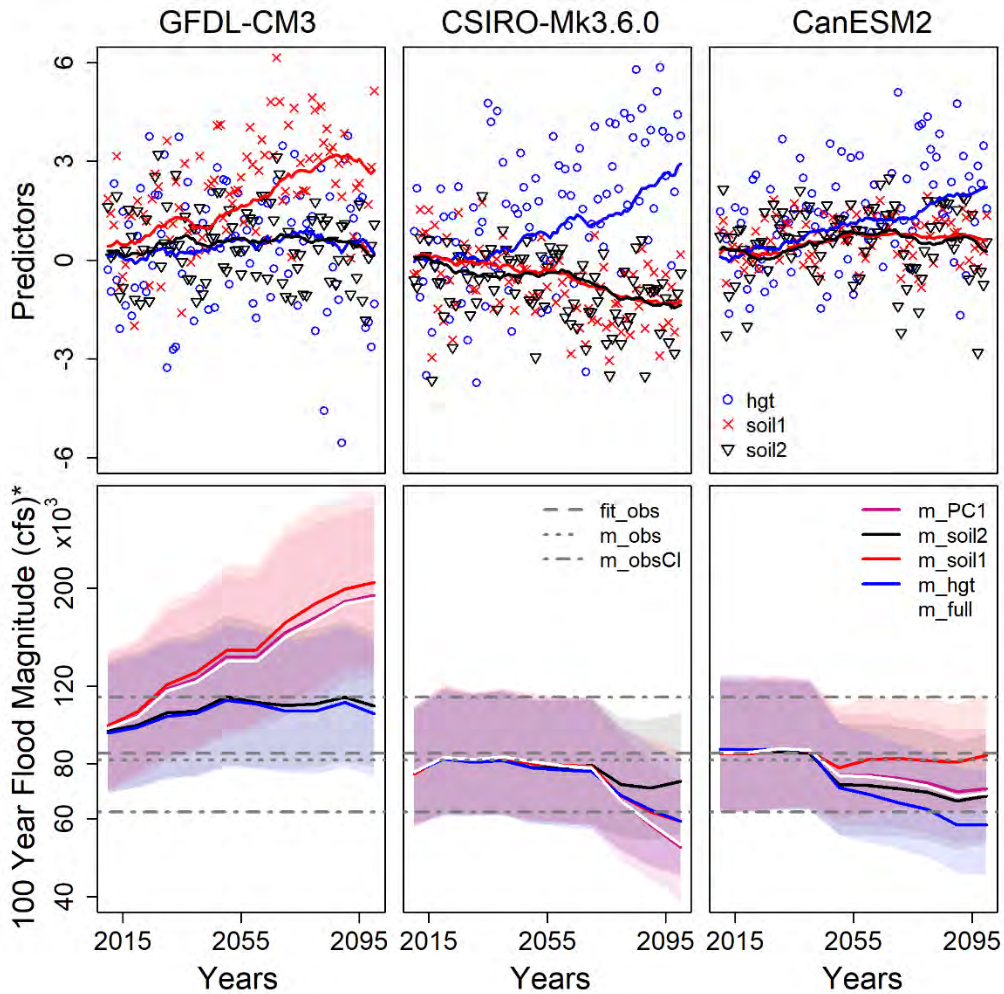


*Figure 4-5 Projections of the predictors and the 100 year flood magnitude for gage 45 from three representative GCMs (GFDL-CM3, CSIRO-Mk3.6.0, and CanESM2). For the predictors (which are unit-less), "hgt", "soil1", and "soil2" indicate hgt_{CC-NP}^{DJF}, $soil_{basin}^{JFMA}$, and $soil_{west}^{JFMA}$, respectively and the lines indicate the 31 year moving average. For the 100 year flood, the values are from the LP3 distribution where "fit_obs" is the observed data, "m_obs" is the model forced with observed predictors and "m_obsCI" is the associated credible intervals, "m_hgt", "m_soil1", "m_soil2", "m_PC1", and "m_full" are the models forced with GCM predictors where only the trend on the indicated predictor or subset of predictors has been. The shaded areas indicate credible intervals. * indicates the axis is log scale.*

For GFDL-CM3, the increase in $soil_{basin}^{JFMA}$ causes flood magnitude to increase (m_soil1), while the relatively negligible trends in hgt_{CC-NP}^{DJF} and $soil_{west}^{JFMA}$ result in relatively negligible trends in flood magnitude (m_hgt and m_soil2). Even under the influence of multiple predictors (m_PC1 and m_full), flood magnitude still follows an increasing trend, indicating that $soil_{basin}^{JFMA}$ is driving the GFDL-CM3 projected increase. For CSIRO-Mk3.6.0, the increase in hgt_{CC-NP}^{DJF}, though nearly two times the absolute magnitude of the decrease in $soil_{basin}^{JFMA}$, causes an approximately similar decrease in flood magnitude (m_hgt versus m_soil1). When the opposing trends of hgt_{CC-NP}^{DJF} and $soil_{basin}^{JFMA}$ are combined, the decrease in flood magnitude is even larger (m_PC1) and is nearly matched by the full model (m_full), indicating that both hgt_{CC-NP}^{DJF} and $soil_{basin}^{JFMA}$ are driving the CSIRO-Mk3.6.0 projected decrease. For CanESM2, the increase in hgt_{CC-NP}^{DJF} causes a large decrease in flood magnitude (m_hgt), similar to CSIRO-Mk3.6.0. In contrast to CSIRO-Mk3.6.0 however, even though the trend in $soil_{basin}^{JFMA}$ closely follows that of $soil_{west}^{JFMA}$, the associated decrease in flood magnitude is much smaller (m_soil1 versus m_soil2). This seems counterintuitive given that $soil_{basin}^{JFMA}$ is a more significant predictor than $soil_{west}^{JFMA}$ but is explained by the high variability of $soil_{west}^{JFMA}$, which includes some very negative outliers, in comparison to $soil_{basin}^{JFMA}$. Thus, under the influence of multiple predictors (m_PC1 and m_full), the decrease in flood magnitude is similar to that caused by $soil_{west}^{JFMA}$. These results illustrate that all predictors are important, that both outliers and mean change in the predictors influence the change in flood magnitude, and that the predictors driving the projected sign of change can differ widely among the GCMs.

4.4.3. Sensitivity of Model to Stationarity Assumption

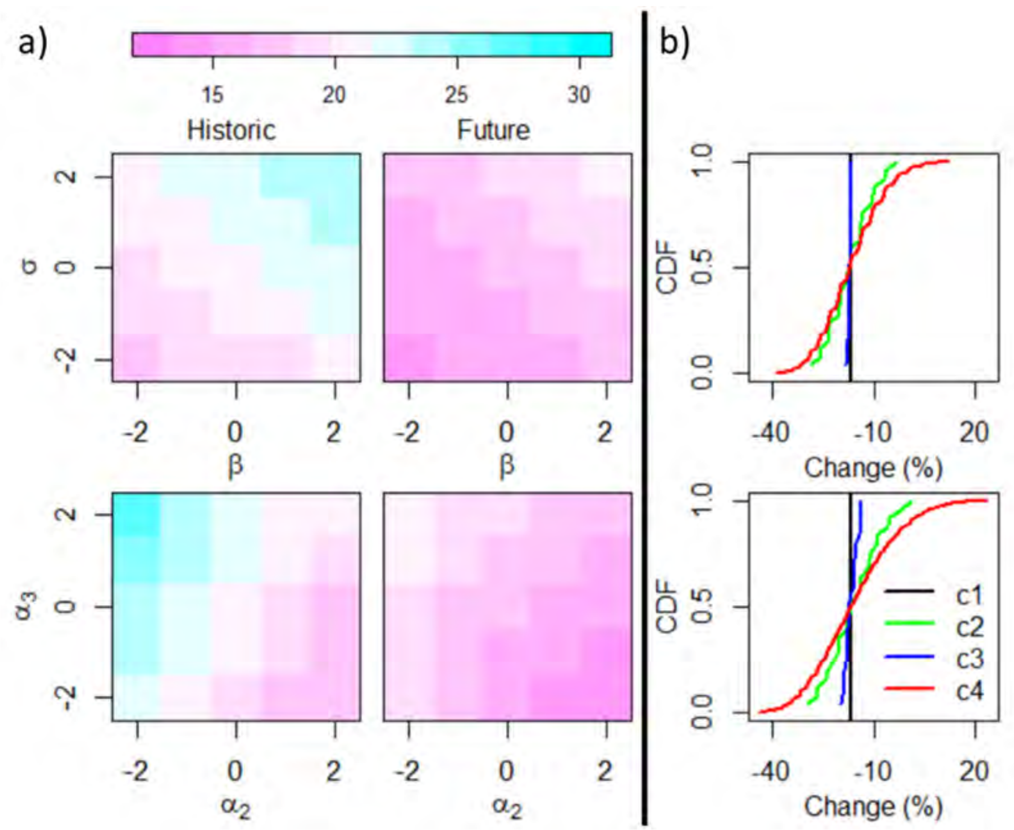


Figure 4-6 Sensitivity analysis results for the multi-model median of the median 100 year flood at gage 45. (a) is the magnitude in units of 10,000 cfs for historic (1950 through 2010) and future (2040 through 2100), the x- and y-axes are the delta change factor applied to the parameter set and, for a given plot, the two covariates not shown are at their original values (i.e., the delta change factor is 0). (b) is the empirical cumulative distribution function (CDF) of the percent change between future and historic values (where the legend meaning is described in the text).

The results of the sensitivity analysis, which assesses the influence of assuming stationarity in the relationship between the covariates and the flood events and of assuming a stationary scale parameter, are shown in Figure 4-6 for the multi-model median of the median 100 year flood across all 3000 parameter sets at gage 45. From the historic and future period, it is clear that the parameters influence the 100 year flood magnitude as expected; α_2 is inversely related while α_3, β, and σ are directly related. Furthermore, all four parameters are important for determining the flood magnitude, based on the approximately diagonal alignment of the contours. The variability across the different levels is greater for the historic period, in which the magnitude differs by up to nearly a factor of two, than for the future period. Percent change between the future and historic period is assessed assuming model stationarity ("c1"; calculated between the future and historic parameter sets for a delta change factor of zero), assuming that the historic model parameters could change in the future ("c2"; calculated between the future parameter sets for all delta change factors relative to the historic parameter set for a delta change factor of zero), assuming that model parameters are stationary, but that the original model is limited by short historic records ("c3"; calculated between the future and associated historic parameter set for each delta change factor), and assuming that the original model is limited and its parameters could change in the future ("c4"; calculated between all future parameter sets and all historic parameters sets across all delta change factors). All cases ("c1" through "c4") have the same median of approximately -16%; however, when the assumption of stationarity is relaxed, the percent change could range anywhere from approximately -40% to 20% in the most extreme case of "c4". Interestingly, "c3" is nearly the same as "c1", whereas "c2" is between "c1" and "c4", indicating that a limited model has less impact on the magnitude of change compared to non-stationarity in model parameters. Obviously, this analysis does not address the fact that the distribution itself may change, or that other covariates may become important which are not represented, or that stationarity is assumed within the historic and future time periods; however, it does provide a quantitative assessment of the impacts of certain stationarity assumptions.

4.4.4. Discussion

The case study results show a projected increase in sea surface temperatures and a generally projected decrease in soil moisture. While the projected change in flood events is relatively homogeneous across gages, for a given gage, the performance and projected changes vary widely among GCMs. Additionally, even if two GCMs project the same sign of change, the underlying cause of that change from the predictors can be very different. Finally, the sign of change projected by the multi-model mean is significantly affected by GFDL-CM3, which alone among the GCMs projects a large increase in soil moisture, and thus an increase in floods. Without GFDL-CM3, the multi-model mean projects a decrease in floods for all but the highest quantiles.

How do the results obtained with the climate informed approach compare to those from the model chain approach? As observed in this study, CMIP5 GCMs consistently project an increase in normal and extreme precipitation in the region (Easterling et al., 2017; Maloney et al., 2014; Wuebbles et al., 2014), which would likely contribute to an increase in flood events. In a national analysis based on regressions between flood discharge and localized extreme climate indices, projections of those indices from 10 GCMs forced by the SRES A2, A1b, and B1 scenarios (associated with the previous generation of GCM experiments, CMIP3) cause a projected increase in the multi-model and multi-scenario mean of the 100 year flood by 2100 over the whole United States (AECOM, 2013). Idealized carbon dioxide quadrupling forcing of one GCM causes both an increase in magnitude and frequency of exceedance of the 100 year flood in the basin (Milly et al., 2002). In a global analysis of 21 GCMs forced with the SRES A1b scenario, between one and two thirds of the GCMs project an increase in the magnitude of the 100 year flood by 2050 in the basin (as estimated from a global map) (Arnell & Gosling, 2016). In a global analysis of 11 GCMs forced with RCP 8.5, the 21st century multi-model median seems to indicate either no change or a slight increase in the frequency of the 100 year flood in the basin (as estimated from a global map), but consistency in the projected sign of change among GCMs is low (Hirabayashi et al., 2013). Finally, in a global analysis using 5 GCMs forced by RCP 8.5, the multi-model mean projects minimal change (i.e., absolute value less than 10%) or some decrease in the 30 year 5 day average peak flow by the end of the century (as estimated from a global map) (Dankers et al., 2014).

In summary, it is difficult to draw a conclusion regarding the direction and magnitude of change projected by model chain studies. Not only are there significant discrepancies among GCMs and various studies, most of which are on national or global scales, but attribution of change cannot be easily diagnosed. If anything, this is ample motivation for regional-scale flood projection studies based on credible predictors and simplified modeling frameworks where attribution of change can be easily diagnosed, as demonstrated here for the climate informed methodology.

4.4.5. GENERALIZATION TO THE UNITED STATES

Having developed climate informed flood projections for the Ohio River Basin following the general methodology, the next challenge is to demonstrate broad applicability across hydro-climatologically diverse basins. As a preliminary step, we assess ENSO, PNA and soil moisture as potential predictors for JFMA floods across the contiguous U.S. Streamgages from the Hydro-Climatic Data Network with less than 0.1% data missing between 1950 and 2015 and with catchment area greater than 500 km^2 (200 square miles) were identified, but only gages with more than 50% of AMS events occurring in JFMA were retained for further analysis, resulting in 255 gages. Similar to the approach used for the Ohio River Basin, JFMA maximum flood events were calculated for each gage. Monthly Nino3 and PNA indices were obtained from NOAA (2012, 2017) and processed into annual indices by taking the maximum value within DJF. Gridded monthly soil data was obtained from Fan & van den Dool (2004) and, for each gage, the four closest soil moisture grid points were averaged and then processed into an annual index by taking the maximum value in JFMA.

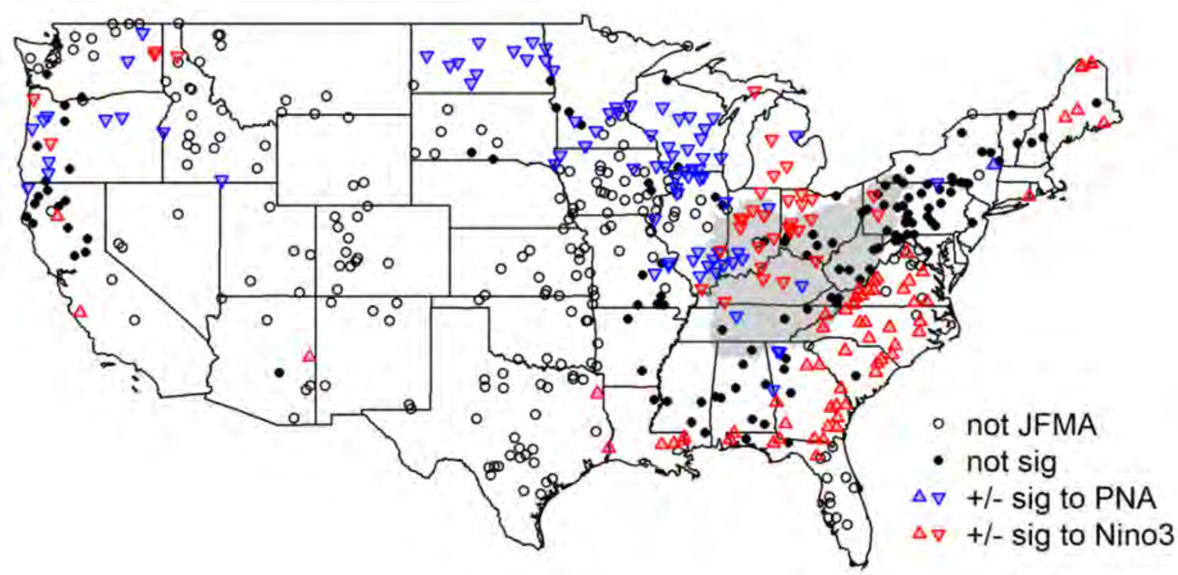


Figure 4-7 Correlation results between JFMA flood events and DJF PNA and DJF Nino3. "not JFMA" indicates that less than 50% of the annual maximum events occur in JFMA, "not sig" indicates no significant correlation to either index, and "+/- sig to PNA" or "+/- sig to Nino3" indicates positive/negative significant correlation (if a location is significantly correlated to both indices, the stronger correlation is plotted). The Ohio River Basin is shaded grey.

The correlation between the soil moisture index and JFMA flood events is significantly positively correlated for all but 2% of gages; however, the correlation magnitude does not exhibit a clear spatial pattern (not shown). The correlation results for the DJF Nino3 and DJF PNA indices to the JFMA flood events are shown in Figure 4-7. Most gages in the Midwest and Southeast are not included in the analysis because less than 50% of AMS events occur in JFMA. Of gages included in the analysis, gages in the northwest region of the Ohio River Basin are significantly correlated to PNA and Nino3, which corroborates the case study diagnostics. Most gages in the northeastern Midwest are significantly negatively correlated to PNA, while on the eastern side of the Appalachians, most gages in the coastal south are significantly positively correlated to Nino3. These results generally align with the literature on relationships between extreme precipitation and ENSO across the contiguous U.S. (Gershunov & Barnett, 1998a, 1998b; Higgins et al., 2007; Zhang et al., 2010). For gages in major mountain ranges (i.e., the Sierra Nevada, Cascades, and the Appalachians) and most of the Northeast, the correlations are not significant or are site-specific, likely due to the influence of orthography, snow, and multiple climate mechanisms. For example, in the Northeast, snow is a dominant flood generating mechanism (Berghuijs et al., 2016) and the influence of PNA and PDO on precipitation is modulated by ENSO (Ning & Bradley, 2014). While more work is needed to extend climate informed approaches to the whole U.S. and for all seasons of the year, this simple analysis shows potential applicability in the northeastern Midwest and coastal south.

4.5. CONCLUSION

Climate informed approaches, now formalized into a general methodology and demonstrated in entirety in a regional analysis, are a promising and useful alternative to traditional model chain approaches for long-term flood projection. Specifically, GCMs more skillfully simulate large-scale ocean-atmospheric patterns, used to force the climate informed model, in comparison to local temperature and precipitation fields, used to force hydrologic models. Furthermore, the simplicity and transparency of the statistical model allows projected changes in flood events to be easily attributed to changes in the predictors.

However, the climate informed approach is not without its own limitations, as demonstrated in this work. The explained variability of flood events is limited by the degree to which flood generating processes are understand and representation of the underlying physical processes is limited to only primary drivers; additional restrictions may also occur when indices of the primary drivers are not calculable from GCMs (e.g., the Southern Oscillation Index, a measure of ENSO variability, is calculated as the difference between sea level pressures at two point locations, which cannot be resolved from a coarse GCM grid). Stationarity is assumed in the relationships between the covariates and the flood events, in distribution parameters which are not conditioned on a covariate (e.g., in this study, the scale parameter), and in the distribution itself. In this study, the simple sensitivity analysis, performed to assess the implications of relaxing a subset of these stationarity assumptions, showed that allowing distribution parameters to be non-stationary has a much greater impact on projected percent change than assuming stationary model parameters in a model that is limited by a short observational record. Additionally, in this study, return periods were calculated assuming stationarity within user-defined historic and future periods, although other alternatives exist as discussed in section 4.3.4. Finally, robustness under climate change is influenced by the stationarity of the relationship between covariates and flood events, and whether both thermodynamic and dynamic processes are represented, regardless of performance over the historic period. For example, in the Ohio River Basin, a possible shift in the location and type of ENSO (Taschetto et al., 2014; Yeh et al., 2009) would have downstream effects on teleconnections (e.g., the PNA), conceivably causing moisture transport from the Gulf of Mexico to be less frequently directed over the region, which, though not represented in the statistical model, would impact flooding.

What are the remaining challenges associated with the general methodology? While there are well accepted methods for defining a predictand in flood frequency analysis (step one), the challenge of the climate informed approach is to find a predictand related to suitable predictors. For example, this case study was restricted to a sub-region of the full basin and the JFMA season. For identification of credible large-scale predictors (step 2), there is an extensive climate sciences literature on teleconnections. However, such studies are usually written for climate scientists rather than hydrologists or engineers and often focus on precipitation, which may not translate to floods, especially when the proximate mechanism is not rainfall. For example, many of the articles cited in this study on Ohio River Basin teleconnections focus on extreme precipitation and were not always replicable for floods. Furthermore, predictors used for one region may not be generalizable. For example, PNA or Nino3 are possibly suitable predictors for JFMA flood events in only certain regions of the U.S. While the national analysis presented here and the global analysis of well-recognized climate patterns' influence on seasonal peak flow by Lee et al. (2018) are steps forward, until knowledge about flood teleconnections is better synthesized, steps one and two will likely require trying different predictands and lengthy investigation in the climate literature coupled with in-depth knowledge of climate processes.

Another challenge is determining a generalizable model formulation (step 3) that correctly represents the relationship between the predictand and the predictors, since currently there is a wide variety of model forms in the literature. The multi-site Bayesian linear regression model used in this case study could easily be generalized by substituting appropriate predictors, but its applicability would need to be demonstrated for many hydro-climatologically diverse basins. Yet such correlation-based relationships may not be appropriate for highly non-linear or phase-based systems which may be better represented by regime-based distributions similar to that described by Salas and Obeysekera (2014) or may be difficult to identify in regions where climate and hydrology are complexly coupled (Renard & Lall, 2014). Finally, given the limited number of studies which have actually developed climate informed projections (this case study and Condon et al., 2015; Delgado et al., 2014; Tramblay et al., 2014) the challenges associated with assessing projection credibility and creating outputs useful for decision-making have been only cursorily investigated. In this study, credibility is based on the predictor characteristics and model performance while the results are summarized with the multi-model mean of return periods calculated assuming a stationary window. Ideally, methods for calculating non-stationary return periods (Cooley, 2013; Salas & Obeysekera, 2014) could be adapted for limited-horizon projections such as those from climate informed approaches. Finally, the uncertainty in the projections stemming from the GCMs (Kundzewicz et al., 2017), highlights the need for analyses of uncertainty attribution and reduction, using techniques such as global sensitivity analysis (Razavi & Gupta, 2015; Song et al., 2015), and the need to integrate climate informed projections into robustness-based decision-making under uncertainty paradigms (e.g., Spence & Brown, 2016).

4.6. REFERENCES

AECOM. (2013). The Impact of Climate Change and Population Growth on the National Flood Insurance Program Through 2100, (June), 1–257.

Alfieri, L., Burek, P., Feyen, L., & Forzieri, G. (2015). Global warming increases the frequency of river floods in Europe. *Hydrology and Earth System Sciences*, *19*(5), 2247–2260. https://doi.org/10.5194/hess-19-2247-2015

Arnell, N. W., & Gosling, S. N. (2016). The impacts of climate change on river flood risk at the global scale. *Climatic Change*, *134*(3), 387–401. https://doi.org/10.1007/s10584-014-1084-5

Austin, G., Rizzo, K., Matte, A., & Finnerty, B. (1998). *Ohio River Valley Flood of March 1997*. National Weather Service Assessment.

Beaber, H., & Rostvedt, J. (1965). *Floods of March 1964 Along the Ohio River*. Geological Survey Water-Supply Paper 1840-A.

Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M., & Vialard, J. (2014). ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dynamics*, *42*(7–8), 1999–2018. https://doi.org/10.1007/s00382-013-1783-z

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

Bloschl, G., & Montanari, A. (2010). Climate change impacts - throwing the dice? *Hydrological Processes*, *24*, 374–381.

Bracken, C. (2016). Downscaled CMIP3 and CMIP5 Climate Projections - Addendum: Release of Downscaled CMIP5 Clmate Projections (LOCA) and Comparison with Preceding Information.

Bracken, C., Holman, K.D., Rajagopalan, B., & Moradkhani, H. (2018). A Bayesian Hierarchical Approach to Multivariate Nonstationary Hydrologic Frequency Analysis. *Water Resources Research, 54,* 243-255. https://doi.org/10.1002/2017WR020403

Breslin, S. (2015). Flooding in the Ohio River Valley: Louisville Man Missing after being swept away by floodwaters. The Weather Channel. Retrieved from https://weather.com/news/news/flooding-midwest-south-latest-news

Chen, J., Brissette, F. P., Lucas-Picher, P., & Caya, D. (2017). Impacts of weighting climate models for hydro-meteorological climate change studies. *Journal of Hydrology*, *549*, 534–546. https://doi.org/10.1016/j.jhydrol.2017.04.025

Coleman, J., & Rogers, J. (2003). Ohio River Valley Winter Moisture Conditions Associated with the Pacific – North American

Teleconnection Pattern. *Journal of Climate*, *16*, 969–981.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., … Wehner, M. (2013). Long-term Climate Change: Projections, Commitments and Irreversibility. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, … P. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1029–1136). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.024

Condon, L.E., Gangopadhyay, S., & Pruitt, T. (2015). Climate change and non-stationary flood risk for the upper Truckee River basin. *Hydrol. Earth Syst. Sci. 19*, 159-175. https://doi.org/10.5194/hess-19-159-2015

Congress. (1939). The Chattanooga Flood Problem. (76th Congress, 1st Session, House Document No. 91). Retrieved from http://docplayer.net/4690177-The-east-tennessee-flood-of-1867.html

Cook, K. H. (1999). Generation of the African easterly jet and its role in determining West African precipitation. *Journal of Climate*, *12*(5), 1165–1184. https://doi.org/10.1175/1520-0442(1999)012<1165:GOTAEJ>2.0.CO;2

Cooley, D. (2013). Extremes in a Changing Climate, *65*. https://doi.org/10.1007/978-94-007-4479-0

Cretat, J., Vizy, E. K., & Cook, K. H. (2014). How well are daily intense rainfall events captured by current climate models over Africa? *Climate Dynamics*, *42*(9–10), 2691–2711. https://doi.org/10.1007/s00382-013-1796-7

Cross, W., & Brooks, H. (1959). *Floods of January-February 1959 in Ohio*. Geological Survey Circular 418.

Dai, A. (2006). Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, *19*(18), 4605–4630. https://doi.org/10.1175/JCLI3884.1

Dankers, R., Arnell, N. W., Clark, D. B., Falloon, P. D., Fekete, B. M., Gosling, S. N., … Wisser, D. (2014). First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. *Proceedings of the National Academy of Sciences*, *111*(9), 3257–3261. https://doi.org/10.1073/pnas.1302078110

Delgado, J. M., Merz, B., & Apel, H. (2014). Projecting flood hazard under climate change: An alternative approach to model chains. *Natural Hazards and Earth System Sciences*, *14*(6), 1579–1589. https://doi.org/10.5194/nhess-14-1579-2014

Dirmeyer, P. A., Jin, Y., Singh, B., & Yan, X. (2013). Trends in Land–Atmosphere Interactions from CMIP5 Simulations. *Journal of Hydrometeorology*, *14*(3), 829–849. https://doi.org/10.1175/JHM-D-12-0107.1

Easterling, D. R., Kunkel, K. E., Arnold, J. R., Knutson, T., LeGrande, A. N., Leung, L. R., … Wehner, M. (2017). Precipitation change in the United States. In D. J. Wuebbles, D. W. Fahey, K. Hibbard, D. J. Dokken, B. C. Stewart, & T. K. Maycock (Eds.), *Climate Science Special Report: Fourth National Cilmate Assessment, Volume I* (pp. 207–230). Washington, DC, USA: U.S. Global Change Research Program. https://doi.org/10.7930/J0H993CC

Edelen, G., & Miller, J. (1976). *Floods of March-April 1973 in Southeastern United States*. U.S. Geological Survey Professional Paper 998.

Edwards, P. N. (2011). History of climate modeling. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(1), 128–139. https://doi.org/10.1002/wcc.95

England, J. J. F., Cohn, T. A., Faber, B. A., Stedinger, J. R., Thomas Jr., W. O., Veilleux, A. G., … Mason, R. R. (2015). *Guidelines for Determining Flood Flow Frequency - Bulletin 17C*.

EO. (2017). Flooding Along the Ohio River. Retrieved from https://earthobservatory.nasa.gov/IOTD/view.php?id=85539

Fan, Y., & van den Dool, H. (2004). Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present. *Journal of Geophysical Research D: Atmospheres*, *109*(10), 1–8. https://doi.org/10.1029/2003JD004345

Fuentes-Franco, R., Giorgi, F., Coppola, E., & Kucharski, F. (2016). The role of ENSO and PDO in variability of winter precipitation over North America from twenty first century CMIP5 projections. *Climate Dynamics*, *46*(9–10), 3259–3277. https://doi.org/10.1007/s00382-015-2767-y

Gershunov, A., & Barnett, T. P. (1998a). ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Implications for Long-Range Predictability. *Journal of Climate*, *11*, 3192–3203. https://doi.org/10.1175/1520-0442(1998)011<1575:eioier>2.0.co;2

Gershunov, A., & Barnett, T. P. (1998b). ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results. *Journal of Climate*, *11*, 3062–3065. https://doi.org/10.1175/1520-0442(1998)011<1575:eioier>2.0.co;2

Giorgi, F., & Mearns, L. (2002). Calculation of Average, Uncertainty Range, and Reliability of Regional Climate Changes from AOGCM Simulations via the "'Reliability Ensemble Averaging'" (REA) Method. *Journal of Climate*, *15*, 1141–1158. https://doi.org/10.1175/1520-0442(2002)015<1141

Glatfelter, D., & Chin, E. (1988). *Floods of March 1982 in Indiana, Ohio, Michigan, and Illinois*. U.S. Geologcal Survey Professional Paper 1467.

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research Atmospheres*, *113*(6), 1–20. https://doi.org/10.1029/2007JD008972

Greene, A. M., Robertson, A. W., Smyth, P., & Triglia, S. (2011). Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model. *Quarterly Journal of the Royal Meteorological Society*, *137*(655), 347–359. https://doi.org/10.1002/qj.788

Held, I. M., & Soden, B. J. (2000). Water Vapor Feedback and Global Warming. *Annual Review of Energy and the Environment*, *25*, 441–475.

Higgins, R. W., Silva, V. B. S., Shi, W., & Larson, J. (2007). Relationships between climate variability and fluctuations in daily precipitation over the United States. *Journal of Climate*, *20*(14), 3561–3579. https://doi.org/10.1175/JCLI4196.1

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., … Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, *3*(9), 816–821. https://doi.org/10.1038/nclimate1911

Hirschboeck, K. K. (1988). Flood hydroclimatology. *Flood Geomorphology*, 27–49.

Horel, J. D., & Wallace, J. M. (1981). Planetary-scale phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.* https://doi.org/10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2

Horton, A., & Jackson, H. (1913a). The Ohio Valley Flood of March-April, 1913.

Horton, A., & Jackson, H. (1913b). *The Ohio Valley Flood of March-April, 1913*. US Geological Survey, Water Supply Paper, (334).

IACWD. (1982). Guidelines for Determining Flod Flow Frequency: Bulletin 17B (revised and corrected). Washington D.C.

IPCC. (2013). Summary for Policymakers. In T. Stocker, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, … P. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.004

Jain, S., & Lall, U. (2001). Floods in a changing climate: Does the past represent the future? *Water Resources Research*, *37*(12), 3193–3205. https://doi.org/10.1029/2001WR000495

Johnson, F., & Sharma, A. (2009). Measurement of GCM skill in predicting variables relevant for hydroclimatological assessments. *Journal of Climate*, *22*(16), 4373–4382. https://doi.org/10.1175/2009JCLI2681.1

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., … Woollen, J. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471.

Knippertz, P., & Wernli, H. (2010). A Lagrangian Climatology of Tropical Moisture Exports to the Northern Hemispheric Extratropics. *Journal of Climate, 23,* 987-1003. https://doi.org/10.1175/2009JCLI3333.1

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. *Journal of Climate*, *23*(10), 2739–2758. https://doi.org/10.1175/2009JCLI3361.1

Kundzewicz, Z. W., Krysanova, V., Dankers, R., Hirabayashi, Y., Kanae, S., Hattermann, F. F., … Schellnhuber, H. J. (2017). Differences in flood hazard projections in Europe–their causes and consequences for decision making. *Hydrological Sciences Journal*, *62*(1), 1–14. https://doi.org/10.1080/02626667.2016.1241398

Kwon, H.-H., Brown, C., & Lall, U. (2008). Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling. *Geophysical Research Letters*, *35*(5), L05404. https://doi.org/10.1029/2007GL032220

Landwehr, J. M., & Slack, J. R. (1992). Hydro-Climatic Data Network: A U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1874-1988. Retrieved from https://pubs.usgs.gov/wri/wri934076/1st_page.html

Langenbrunner, B., & Neelin, J. D. (2013). Analyzing enso teleconnections in cmip models as a measure of model fidelity in simulating precipitation. *Journal of Climate*, *26*(13), 4431–4446. https://doi.org/10.1175/JCLI-D-12-00542.1

Lee, D., Ward, P., & Block, P. (2018). Attribution of Large-Scale Climate Patterns to Seasonal Peak-Flow and Prospects for Prediction Globally. *Water Resources Research*, 916–938. https://doi.org/10.1002/2017WR021205

Lee, Y. Y., & Black, R. X. (2013). Boreal winter low-frequency variability in CMIP5 models. *Journal of Geophysical Research Atmospheres*, *118*(13), 6891–6904. https://doi.org/10.1002/jgrd.50493

Leng, G., Tang, Q., Huang, S., Zhang, X., & Cao, J. (2016). Assessments of joint hydrological extreme risks in a warming climate in China. *International Journal of Climatology*, *36*(4), 1632–1642. https://doi.org/10.1002/joc.4447

Li, J., & Tan, S. (2015). Nonstationary Flood Frequency Analysis for Annual Flood Peak Series, Adopting Climate Indices and Check Dam Index as Covariates. *Water Resources Management, 29*(15), 5533-5550. https://doi.org/10.1007/s11269-015-1133-5

Lima, C., Lall, U., Troy, T. J., & Devineni, N. (2015). A climate informed model for nonstationary flood risk prediction: Application to Negro River at Manaus, Amazonia. *Journal of Hydrology*, *522*, 594–602. https://doi.org/10.1016/j.jhydrol.2015.01.009

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., … Lettenmaier, D. P. (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

López, J., & Francés, F. (2013). Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. *Hydrology and Earth System Sciences*, *17*(8), 3189–3203. https://doi.org/10.5194/hess-17-3189-2013

Lott, G., & Myers, V. (1956). Meteorology of Flood-Producing Storms in the Mississippi River Basin. Hydrometeorological Report No. 34.

Luke, A., Vrugt, J. A., & Sanders, B. F. (2017). Water Resources Research, 5469–5494. https://doi.org/10.1002/2016WR019676.Received

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., & Kjeldsen, T. R. (2014). Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *Journal of Hydrology*, *519*, 3634–3650.

Mallakpour, I., & Villarini, G. (2017). Analysis of changes in the magnitude, frequency, and seasonality of heavy precipitation over the contiguous USA. *Theor. Appl. Climatology, 130,* 345-363. https://doi.10.1007/s00704-016-1881-z

Mallakpour, I., Villarini, G., Jones, M. & Smith, J. (2017). On the use of Cox regression to examine the temporal clustering of flooding and heaving precipitation across the central United States. *Global and Planetary Change, 155,* 98-108. https://dx.doi.org/10.1016/j.gloplacha.2017.07.001

Maloney, E. D., Camargo, S. J., Chang, E., Colle, B., Fu, R., Geil, K. L., … Zhao, M. (2014). North American climate in CMIP5 experiments: Part III: Assessment of twenty-first-century projections. *Journal of Climate*, *27*(6), 2230–2270. https://doi.org/10.1175/JCLI-D-13-00273.1

Mantua, N., & Hare, S. (2002). The Pacific Decadal Oscillation. *Journal of Oceanography*, *58*, 35–44.

Mehran, A., Aghakouchack, A., & Phillips, T. J. (2014). Evaluation of CMIP5 continental precipitation simulations relative ot satellite-based gauge-adjusted observations. *Journal of Geophysical Research Atmospheres*, *119*, 1695–1707. https://doi.org/10.1002/2013JD021152

Merz, R., & Blöschl, G. (2003). A process typology of regional floods. *Water Resources Research*, *39*(12), 1–20. https://doi.org/10.1029/2002WR001952

Milly, P. C. D., Wetherald, R. T., Dunne, K. A., & Delworth, T. L. (2002). Increasing risk of great floods in a changing climate. *Nature*, *415*(6871), 514–517.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., … Krysanova, V. (2015). On Critiques of "Stationarity is Dead: Whither Water Management?" *Water Resources Research*, *51*, 7785–7789. https://doi.org/10.1002/2015WR017408

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: whither water management? *Science*, *319*(5863), 573–574. https://doi.org/10.1126/science.1151915

Montroy, D. (1997). Linear Relation of Central and Eastern North American Precipitation to Tropical Pacific Sea Surface Temperature Anomalies *. *Journal of Climate*, *10*, 541–558.

Moore, G. (2016). The flood of 1917. Times Free Press. Retrieved from http://www.timesfreepress.com/news/opinion/columns/story/2016/jan/24/moore-flood-1917/345947/

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., & Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, *430*(7001), 768–772. https://doi.org/10.1038/nature02771

Nakamura, J., Lall, U., Kushnir, Y., Robertson, A. W., & Seager, R. (2013). Dynamical Structure of Extreme Floods in the U.S. Midwest and the United Kingdom. *Journal of Hydrometeorology*, *14*(2), 485–504. https://doi.org/10.1175/JHM-D-12-059.1

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*, 1144-1166. https://doi.org/10.1002/2016WR019033

Ning, L., & Bradley, R. (2014). Winter precipitation variability and corresponding teleconnections over the northeastern United States. *Journal of Geophysical Research: …*, *119*, 7931–7945. https://doi.org/10.1002/2014JD021591.Received

Ning, L., & Bradley, R. S. (2016). NAO and PNA influences on winter temperature and precipitation over the eastern United States in CMIP5 GCMs. *Climate Dynamics*, *46*(3–4), 1257–1276. https://doi.org/10.1007/s00382-015-2643-9

NOAA. (2012). Pacific/North American (PNA). NOAA National Weather Service Climate Prediction Center. Retrieved from http://www.cpc.ncep.noaa.gov/data/teledoc/pna.shtml

NOAA. (2017). Nino3. NOAA Climate Prediction Center. Retrieved from https://www.esrl.noaa.gov/psd/data/correlation/nina3.data

NWS. (2017a). Flooding in Kentucky. Retrieved from http://www.floodsafety.noaa.gov/states/ky-flood.shtml

NWS. (2017b). The Great Flood of 1937. Louisville KY Weather Forecast Office The National Weather Service. Retrieved from http://www.weather.gov/lmk/flood_37

O'Gorman, P. A., & Schneider, T. (2009). The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proceedings of the National Academy of Sciences*, *106*(35), 14773–14777. https://doi.org/10.1073/pnas.0907610106

Osment, T. (2008). Floods of 1916 and 1940. The Digital Heritage Project. Retrieved from https://digitalheritage.org/2010/08/floods-of-1916-and-1940/

Ouarda, T. B. M. J., & El-Adlouni, S. (2011). Bayesian nonstationary frequency analysis of hydrological variables. *Journal of the American Water Resources Association*, *47*(3), 496–505. https://doi.org/10.1111/j.1752-1688.2011.00544.x

Pierce, D. W., Barnett, T. P., Santer, B. D., & Gleckler, P. J. (2009). Selecting global climate models for regional climate change studies. *Proceedings of the National Academy of Sciences*, *106*(21), 8441–8446. https://doi.org/10.1073/pnas.0900094106

Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., & Hegewisch, K. C. (2015). Improved Bias Correction Techniques for Hydrological Simulations of Climate Change*. *Journal of Hydrometeorology*, *16*(6), 2421–2442. https://doi.org/10.1175/JHM-D-14-0236.1

Pierce, D. W., Cayan, D. R., & Thrasher, B. L. (2014). Statistical Downscaling Using Localized Constructed Analogs (LOCA)*. *Journal of Hydrometeorology*, *15*(6), 2558–2585. https://doi.org/10.1175/JHM-D-14-0082.1

Plummer, M. (2016). rjags: Bayesian Graphical Models using MCMC. R package version 4-6. Retrieved from https://cran.r-project.org/package=rjags

Polade, S. D., Gershunov, A., Cayan, D. R., Dettinger, M. D., & Pierce, D. W. (2013). Natural climate variability and teleconnections to precipitation over the Pacific-North American region in CMIP3 and CMIP5 models. *Geophysical Research Letters*, *40*(10), 2296–2301. https://doi.org/10.1002/grl.50491

Rayner, N. A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. https://doi.org/10.1029/2002JD002670

Razavi, S., & Gupta, H. V. (2015). What do we mean by sensitivity analysis? The need for comprehensive characterization of "'global'" sensitivity in Earth and Environmental systems models. *Water Resources Research*, *51*(5), 3070–3092.

https://doi.org/10.1002/2014WR016527.Received

Renard, B., & Lall, U. (2014). Regional frequency analysis conditioned on large-scale atmospheric or oceanic fields. *Water Resources Research*, *50*, 9536–9554. https://doi.org/10.1002/ 2014WR016277

Robertson, A. W., Kushnir, Y., Lall, U., & Nakamura, J. (2015). Weather and Climatic Drivers of Extreme Flooding Events over the Midwest of the United States. *Extreme Events: Observations, Modeling, and Economics*, 113–124. https://doi.org/10.1002/9781119157052.ch9

Rocheta, E., Sugiyanto, M., Johnson, F., Evans, J., & Sharma, A. (2014). How well do GCMs represent low-frequency rainfall variability? *Water Resources Research*, 2–31. https://doi.org/10.1002/acr.22212

Rogers, J. C., & Coleman, J. S. M. (2003). Interactions between the Atlantic Multidecadal Oscillation, El Niño/La Niña, and the PNA in winter Mississippi Valley stream flow. *Geophysical Research Letters*, *30*(10), n/a-n/a. https://doi.org/10.1029/2003GL017216

Roller, C. D., Qian, J. H., Agel, L., Barlow, M., & Moron, V. (2016). Winter weather regimes in the northeast United States. *Journal of Climate*, *29*(8), 2963–2980. https://doi.org/10.1175/JCLI-D-15-0274.1

Rootzén, H., & Katz, R.W. (2013). Design Life Level: Quantifying risk in a changing climate. *Water Resources Research, 49,* 5964-5972. https://doi.org/10.1002/wrcr.20425

Rostvedt, J. O. (1968). *Summary of Floods in the United States During 1963*. Geological Survey Water-Supply Paper 1830-B.

Rostvedt, J. O. (1970a). *Summary of Floods in the United States During 1965*. Geological Survey Water-Supply Paper 1850-E.

Rostvedt, J. O. (1970b). *Summary of Floods in the United States During 1966*. Geological Survey Water-Supply Paper 1870-D.

Rostvedt, J. O. (1972). *Summary of Floods in the United States During 1967*. Geological Survey Water-Supply Paper 1880-C.

Roudier, P., Andersson, J. C. M., Donnelly, C., Feyen, L., Greuell, W., & Ludwig, F. (2016). Projections of future floods and hydrological droughts in Europe under a +2°C global warming. *Climatic Change*, *135*(2), 341–355. https://doi.org/10.1007/s10584-015-1570-4

Salas, J., Rajagopalan, B., Saito, L., & Brown, C. (2012). Special Section on Climate Change and Water Resources: Climate Nonstationarity and Water Resources Management. *Journal of Water Resources Planning and Management*, *138*(5), 385–388. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000279

Salas, J., & Obeysekera, J. (2014). Revisiting the Concepts of Return Period and Risk for Nonstationary Hydrologic Extreme Events. *Journal of Hydrologic Engineering*, *19*(3), 554–568.

Salas, J., Obeysekera, J., & Vogel, R. (2018). Techniques for assessing water infrastructure for nonstationary extreme events: a review. *Hydrological Sciences Journal, 63*(3), 325-352.

Sankarasubramanian, A., & Lall, U. (2003). Flood quantiles in a changing climate: Seasonal forecasts and causal relations. *Water Resources Research*, *39*(5), 1–12. https://doi.org/10.1029/2002WR001593

Schlef, K. (2018). Ohio Paper Code and Data. HydroShare.

Schwarz, F. (1961). Meteorology of flood-producing storms in the Ohio River Basin (No. 38). US Department of Commerce, Weather Bureau.

Seidou, O., Ramsay, A., & Nistor, I. (2012). Climate change impacts on extreme floods I: Combining imperfect deterministic simulations and non-stationary frequency analysis. *Natural Hazards*, *61*(2), 647–659. https://doi.org/10.1007/s11069-011-0052-x

Service Assessment Team. (2010). Record Floods of Greater Nashville: Including Flooding in Middle Tennessee and Western Kentucky, May 1-4, 2010.

Shang, H., Yan, J., & Zhang, X. (2011). El Nino-Southern Oscillation influence on winter maximum daily precipitation in California in a spatial model. *Water Resources Research*, *47*(11), 1–9. https://doi.org/10.1029/2011WR010415

Sheffield, J., Barrett, A. P., Colle, B., Fernando, D. N., Fu, R., Geil, K. L., … Yin, L. (2013). North American Climate in CMIP5 experiments. Part I: Evaluation of historical simulations of continental and regional climatology. *Journal of Climate*, *26*(23), 9209–9245. https://doi.org/10.1175/JCLI-D-12-00592.1

Sheffield, J., Camargo, S. J., Fu, R., Hu, Q., Jiang, X., Johnson, N., … Zhao, M. (2013). North American Climate in CMIP5 experiments. Part II: Evaluation of historical simulations of intraseasonal to decadal variability. *Journal of Climate*, *26*(23), 9209–9245. https://doi.org/10.1175/JCLI-D-12-00592.1

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research Atmospheres*, *118*(4), 1716–1733. https://doi.org/10.1002/jgrd.50203

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., & Xu, C. (2015). Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *Journal of Hydrology*, *523*(225), 739–757. https://doi.org/10.1016/j.jhydrol.2015.02.013

Spence, C. M., & Brown, C. M. (2016). Nonstationary decision model for flood risk decision scaling. *Water Resources Research*, *52*(11), 8650–8667. https://doi.org/10.1002/2016WR018981

Spiegelhalter, D. J., Best, N., Carlin, B. P., & Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society. Series B*, *64*(4), 583–639.

Steinschneider, S., & Lall, U. (2015). A hierarchical Bayesian regional model for nonstationary precipitation extremes in Northern California conditioned on tropical moisture exports. *Water Resources Research*, *51*, 9127–9140. https://doi.org/10.1002/2014WR016259

Stocker, T., Dahe, Q., Plattner, G.-K., Tignor, M., & Midgley, P. (2010). *IPCC Expert Meeting on Assessing and Combining Multi*

Model Climate Projections - Meeting Report. *National Center for* …. Retrieved from https://www.ipcc-wg1.unibe.ch/publications/supportingmaterial/IPCC_EM_MultiModelEvaluation_MeetingReport.pdf

Sun, X., Thyer, M., Renard, B., & Lang, M. (2014). A general regional frequency analysis framework for quantifying local-scale climate effects: A case study of ENSO effects on Southeast Queensland rainfall. *Journal of Hydrology*, *512*, 53–68. https://doi.org/10.1016/j.jhydrol.2014.02.025

Swenson, B. (1937). Rivers and floods. *Monthly Weather Review*, 71–77.

Taschetto, A. S., Gupta, A. Sen, Jourdain, N. C., Santoso, A., Ummenhofer, C. C., & England, M. H. (2014). Cold tongue and warm pool ENSO Events in CMIP5: Mean state and future projections. *Journal of Climate*, *27*(8), 2861–2885. https://doi.org/10.1175/JCLI-D-13-00437.1

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Tramblay, Y., Amoussou, E., Dorigo, W., & Mahé, G. (2014). Flood risk under future climate in West Africa: Linking extreme value models and flood generating processes, *Journal of Hydrology*, *519*, 549-558. http://dx.doi.org/10/1016/j.jhydrol.2014.07.052

TVA. (1961). Flood on Piney River, November 18-19, 1957, in vicinity of Spring City, Tennessee. Tennessee Valley Authority, Division of Water Control Planning. Retrieved from https://hdl.handle.net/2027/coo.31924005012921

USGS. (1961). Floods on the Kokosing River, Dry Creek and Center Run, at Mount Vernon, Ohio in 1959,1961. USGS Hydrologic Atlas 40.

USGS. (1964). Floods of January-February 1957 in Southeastern Kentucky and Adjacent Areas. US Geological Survey Water-Supply Paper 1652-A.

Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D., & Krajewski, W. F. (2009). Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Advances in Water Resources*, *32*(8), 1255–1266. https://doi.org/10.1016/j.advwatres.2009.05.003

Villarini, G., Smith, J. A., Serinaldi, F., Ntelekos, A. A., & Schwarz, U. (2012). Analyses of extreme flooding in Austria over the period 1951-2006. *International Journal of Climatology, 32*, 1178-1192. https://doi.org/10.1002/joc.2331

Villarini, G., Smith, J. A., Vitolo, R., & Stephenson, D. B. (2013). On the temporal clustering of US floods and its relationship to climate teleconnection patterns. *International Journal of Climatology*, *33*(3), 629–640. https://doi.org/10.1002/joc.3458

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

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., … Lamarque, J.-F. (2011). The representative concentration pathways: an overview. *Climatic Change*, *109*, 5–31.

Wallace, J. M., & Gutzler, D. S. (1981). Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. *Monthly Weather Review*, *109*(4), 784–812.

Weigel, A. P., Knutti, R., Liniger, M. A., & Appenzeller, C. (2010). Risks of model weighting in multimodel climate projections. *Journal of Climate*, *23*(15), 4175–4191. https://doi.org/10.1175/2010JCLI3594.1

Winsemius, H. C., Aerts, J. C. J. H., van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., … Ward, P. J. (2016). Global drivers of future river flood risk. *Nature Climate Change*, *6*(4), 381–385. https://doi.org/10.1038/nclimate2893

Wuebbles, D., Meehl, G., Hayhoe, K., Karl, T. R., Kunkel, K., Santer, B., … Sun, L. (2014). CMIP5 climate model analyses: Climate extremes in the United States. *Bulletin of the American Meteorological Society*, *95*(4), 571–583. https://doi.org/10.1175/BAMS-D-12-00172.1

Yeh, S.-W., Kug, J.-S., Dewitte, B., Kwon, M.-H., Kirtman, B. P., & Jin, F.-F. (2009). El Niño in a changing climate. *Nature*, *461*(7263), 511–514. https://doi.org/10.1038/nature08316

Yim, B. Y., Kwon, M. H., Min, H. S., & Kug, J. S. (2015). Pacific Decadal Oscillation and its relation to the extratropical atmospheric variation in CMIP5. *Climate Dynamics*, *44*(5–6), 1521–1540. https://doi.org/10.1007/s00382-014-2349-4

Yu-Sung, S., & Yajima, M. (2015). R2jags: Using R to Run "JAGS". R package version 0.5-7. Retrieved from https://cran.r-project.org/packages=R2jags

Yu, B., & Zwiers, F. W. (2007). The impact of combined ENSO and PDO on the PNA climate: A 1,000-year climate modeling study. *Climate Dynamics*, *29*(7–8), 837–851. https://doi.org/10.1007/s00382-007-0267-4

Yuan, S., & Quiring, S. M. (2017). Evaluation of soil moisture in CMIP5 simulations over the contiguous United States using in situ and satellite observations. *Hydrology and Earth System Sciences*, *21*(4), 2203–2218. https://doi.org/10.5194/hess-21-2203-2017

Zhang, Q., Gu, X., Singh, V. P., Xiao, M., & Chen, X. (2015). Evaluation of flood frequency under non-stationarity resulting from climate indices and reservoir indices in the East River basin, China. *Journal of Hydrology*, *527*(August), 565–575. https://doi.org/10.1016/j.jhydrol.2015.05.029

Zhang, X., Wang, J., Zwiers, F. W., & Groisman, P. Y. (2010). The influence of large-scale climate variability on winter maximum daily precipitation over North America. *Journal of Climate*, *23*(11), 2902–2915. https://doi.org/10.1175/2010JCLI3249.1

5. HYDROLOGICALLY RELEVANT CLIMATE PROJECTIONS IN THE UPPER MISSOURI RIVER BASIN

5.1. EXECUTIVE SUMMARY

In the Upper Missouri River Basin spring and summer flooding is often correlated with high antecedent winter snowpack conditions. In a changing climate, snow accumulation and melt dynamics will change as regional temperatures warm over time, and will likely shift the magnitude and seasonality of flood risk in this region.

The goal of this study is to examine how climate change may influence snowpack characteristics over the Upper Missouri River Basin using a suite of dynamically downscaled, high-resolution (25km) climate change simulations from the North America Coordinated Regional Climate Downscaling Experiment (NA-CORDEX). Examination of future changes in snow may provide important insights into how streamflow management practices need to change in the future if runoff characteristic associated with snowmelt shift in timing and intensity.

The key findings from this study show that there is considerable uncertainty in the characteristics of observed snow over the Upper Missouri River Basin and that the region could benefit from enhanced surface observations to help inform and constrain gridded snow products. This study also found that the baseline climate simulations from NA-CORDEX have a cold, wet bias in the cool season over the Upper Missouri River Basin. These biases result in more than observed snow water equivalent (SWE; the liquid water content of snow). These biases likely influence the future changes in SWE, as the role of surface radiation and surface temperatures will be influenced by excess SWE in the baseline climate.

Midcentury projections from regional climate models for the Upper Missouri River Basin project increases in cold-season temperatures and precipitation over the domain. Rising temperatures result in decreases in snow in two ways: they reduce the proportion of precipitation that falls as snow and they increase snow loss via ablation. In the absence of any temperature changes, changes in snowfall would be positively correlated with precipitation and increases in precipitation correspond with increases in snowfall and SWE. However, these two variables are changing together. SWE over the domain is projected to decrease over almost all regions in all simulations, across all seasons. Peak snow water volume will decrease over all sub-domains of the Upper Missouri River Basin and the timing of peak snow volume is projected to shift to earlier in the season. These changes in snowpack characteristics will alter streamflow and are important to consider when updating water resource management practices.

5.2. TECHNICICAL APPROACH

The Missouri River Basin (MBR) is the largest watershed within the United States (U.S.) and has an area of over 500,000 square miles, including about 9,700 square miles located in Canada (Figure 5-1a). The river systems in the basin are regulated to serve eight purposes: flood control, water supply, water quality control, navigation, hydroelectric power generation, irrigation, recreation, and well-being of fish and wildlife, including threatened and endangered species. The Basin is dominated by mountains in the west and plains to the east (Figure 5-1). This study examines how climate change may influence snow in the Upper Missouri River Basin (UMBR, Figure 5-1), a critical variable which influences runoff, baseline streamflow, and flooding.

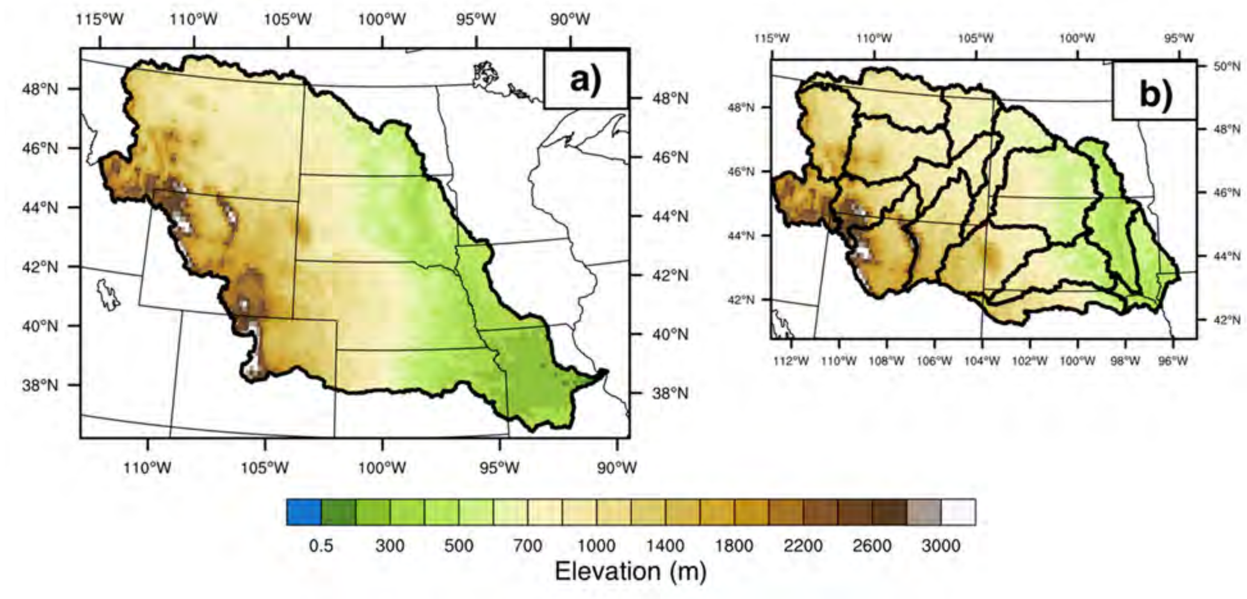


Figure 5-1 Surface elevation for the entire Missouri River Basin (a) and the Upper Missouri River Basin (b), which is the focus of this study. Secondary watersheds are shown in (b).

Because of its mid-continent location, the MRB experiences large temperature fluctuations and extremes. Winters are relatively cloudy and cold over much of the basin, while summers are fair and hot. Daily temperature extremes range from winter lows of -50 ºC in Montana to summer highs of 45 ^{o}C in the lower basin. Precipitation is highest in the spring and early summer and lowest in mid-winter (corresponding with cold-dry airmasses that impact the region). The basin experiences significant year-to-year variability in precipitation and runoff, with 10 of the 13 highest runoff years occurring since 1970 (Livneh et al., 2016).

In the UMRB spring and summer flooding is often correlated with high antecedent winter snowpack conditions. In fact, most of the major historic flood events have corresponded with rapid snow melt or rain-on-snow events. For example, the co-occurrence of rapid spring snowmelt and above average spring precipitation resulted in the record 2011 flood event that was extremely costly. While much of the snowmelt in the UMRB is generated in the mountains to the west, plains snowpack can also influence streamflow.

In a changing climate, snow accumulation and melt dynamics are almost certainly going to change as regional temperatures warm over time, shifting the magnitude and seasonality of flood risk in this region. This study explores how climate change is projected to influence snowpack characteristics (timing, amount, and distribution) using a suite of dynamically downscaled, high-resolution (25km) climate change simulations from the North America Coordinated Regional Climate Downscaling Experiment (NA-CORDEX). Analysis includes an evaluation of the historical climate simulations as well as estimation of future changes for snowpack dynamics and rain-on-snow events.

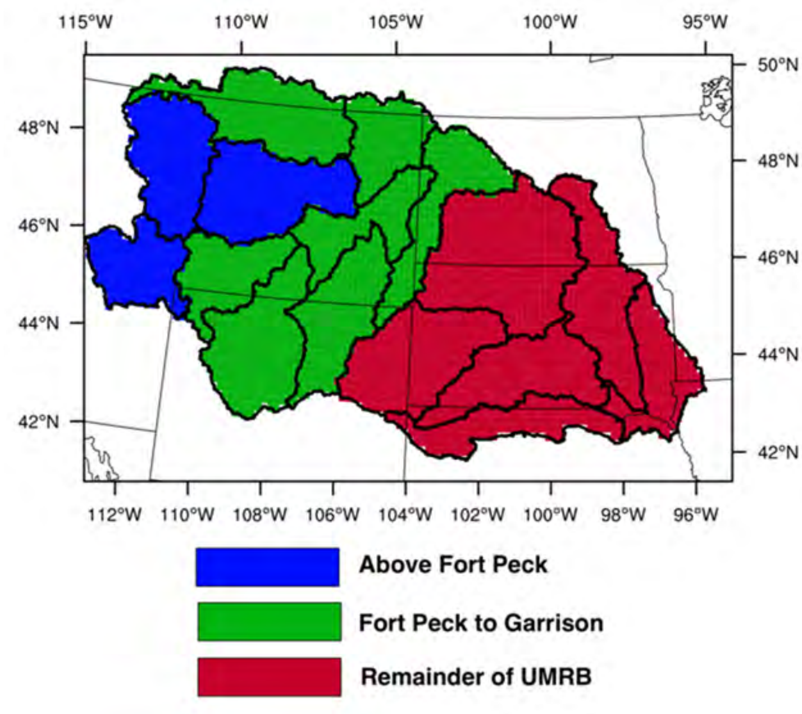


Figure 5-2 Sub-basins explored in this study.

We focus our analysis on the UMRB as snow plays a key role in streamflow and water management there. We examine results for the entire UMRB as well as three sub-basins. The reach above Fort Peck Dam, the reach between Fort Peck Dam and Garrison Dam, and the remaining part of the UMRB. These sub-basins are highlighted in Figure 5-2.

5.3. DATA AND METHODS

5.3.1. OBSERVATIONAL-BASED DATASETS

Multiple observational datasets are used to examine the climate of the UMB and evaluate the skill of the climate models in capturing the climate of the region. We use multiple datasets to examine uncertainty in the observational datasets, as gridded observations must be interpolated from in situ surface observations. These datasets are summarized in Table 5-1.

Table 5-1 Observation-based gridded datasets used in this study.

Dataset	Resolution	Timestep	Time period	Variables	Citation
UDEL	0.5ºlat × 0.5ºlon	Monthly	1901-2012	T, P	Willmott and Matsuura (2001)
Maurer	0.125ºlat × 0.125ºlon	Daily	1949-2010	T,P	Maurer et al. (2002)
PRISM	4km	Monthly	1981-2018	T,P	PRISM Climate Group
Livneh	$1/16^{th\,o}$ (~6km)	Daily	1950-2013	T, P, SWE	Livneh et al. (2013); Livneh et al. (2015)
GLDAS	0.25ºlat × 0.25ºlon	3hrly	1948-2010	SWE	Rodell et al. (2004)
MERRA-Land	0.5ºlat × 0.67ºlon	Daily	1980-2016	SWE	Reichle et al. (2011); Reichle (2012)

The University of Delaware Air Temperature and Precipitation dataset version 4 (UDEL) is the coarsest dataset used to examine the climate of the basin, but we include this dataset because it uses independent methods to calculate temperature (T) and precipitation (P) from surface observations. The Parameter-elevation Regression on Independent Slopes Model (PRISM) and similar products (Maurer and Livneh) are also used in this study. These datasets are high resolution and interpolate data onto a finer grid using assumptions relating to topography.

A challenge for SWE and evaluating SWE in climate models is the lack of long-term, high-resolution (spatial and temporal), well-vetted observations (Brown et al. 2003). Long-term records of SWE are often available from single points (e.g., SNOTEL); however, SWE is heterogeneous and measurements from a single point may not adequately represent a basin or region of interest. While a few long-term, gridded, purely observational datasets do exist over North America (e.g., Armstrong et al. 2005; Dyer and Mote 2006), on their own these products contain significant uncertainties. To overcome the poor spatial resolution of SWE observations, many studies choose to use SWE from models, either atmospheric reanalysis products (Kapnick and Delworth 2013) or blended observational–model surface products (Frei et al. 2005). Unfortunately, model-derived SWE also has large uncertainties in data-sparse areas where SWE is heavily influenced by model parameters.

Rather than attempting to identify the best SWE product, we use an ensemble of gridded observation-based SWE products for North America to capture the uncertainty in SWE observations. In this study we include SWE from the Global Land Data Assimilation System (GLDAS), SWE from the MERRA land surface reanalysis product (MERRA-land), and SWE from Livneh which was generated from the VIC land surface model. See McCrary et al. (2017) and McCrary et al. (2019) for more details on this type of method.

5.3.2. Regional Climate Models

Simulations from two regional climate models (RCMs) from the North American Coordinated Regional Downscaling Experiment (NA-CORDEX) are used in this study to explore future climate changes for the MRB. These RCMs are:

- The Weather Research and Forecasting model as run at the National Center for Atmospheric Research (NCAR) (WRF; Skamarock et al. 2005)
- The International Centre for Theoretical Physics RCM version 4 as run at NCAR and Iowa State (RegCM4; Giorgi et al. 1993a, Giorgi et al. 1993b, Pal and Coauthors 2007)

These two RCMs were forced with boundary conditions from three global climate models (GCMs) that were part of CMIP5 (Taylor et al., 2012). These three models were chosen as they span the equilibrium climate sensitivity of the CMIP5 ensemble and they had the necessary data output to drive an RCM. These GCMs are:

- The Global Fluid Dynamics Laboratory Earth System Model (Dunne et al., 2012), GFDL-ECM2M
- The Max Plank Institute for Meteorology Earth System Model (Giorgetta et al., 2013), MPI-ESM-LR
- The Met Office Hadley Center Climate Prediction Model (The HadGEM2 Development Team, 2011), HadGEM2-ES

Each RCM downscales each GCM at two resolutions, 50km and 25km resulting in a $3 \times 2 \times 2$ matrix of simulations (3GCMs, 2RCMs, and 2 resolutions) for a total of 12 climate simulations. This matrix allows us to explore multiple types of uncertainty in future climate change; uncertainty inherited from the driving GCM (GCM model uncertainty) , uncertainty based on the choice of RCM (RCM model uncertainty), and uncertainty associated with the resolution of the simulations (resolution uncertainty). Table 5-2 summarizes the RCM-GCM matrix used in this study.

Table 5-2 *Matrix of RCM simulations used in this study.*

RCM-GCM	GFDL-ECM2M		MPI-ESM-LR		HadGEM2-ES	
	50km	25km	50km	25km	50km	25km
WRF	×	×	×	×	×	×
RegCM4	×	×	×	×	×	×

Simulations from within this sub-set of NA-CORDEX span from 1950-2100. The period 1950-2005 is called the "historical" or "baseline" climate period, and the GCMs are driven by historical greenhouse gas and aerosol concentrations. The period 2006-2100 is called the "future" climate time period, and simulations from CMIP5 that use RCP8.5 to force anthropogenic climate changes to greenhouse gas emissions are used for the future.

This study focuses primarily on the 25km simulations from NA-CORDEX, although in a few locations we highlight differences between the 50km and 25km runs.

5.3.3. Methods

In this present study we evaluate the historical or baseline climate simulations over the time period 1970-2005. We then look at end-of-century changes over the UMRB for the time period 2060-2100. Future changes are calculated as the difference between the future climate and historical climate. The RCMs' ability to capture observed precipitation (P), temperature (T), and snow water equivalent (SWE) are examined over the region. We examine both the spatial distribution of these variables and their annual cycle. Basin averages are examined for each sub-basin described in Figure 5-3.

5.4. RESULTS AND DISCUSSION

5.4.1. Climate Model Evaluation

We start by identifying biases in the NA-CORDEX models by comparing the simulated current climate conditions with observations. This is a critical step, as model bias may influence climate change response of each model. We evaluate the seasonal cycle of temperature, precipitation and SWE.

Figure 4-3 shows the observed climatological seasonal cycle of temperature over the UMRB from four gridded observation products. As discussed in section 5.4.2 observed temperatures are cooler in the mountains than the plains and cooler in the northern half of the basin than the southern half. Winters are cold, summers are hot, and fall and spring have moderate temperatures. The annual cycle of surface temperature averaged over the entire UMBR as well as the three sub-basins highlighted in Figure 5-2 are shown in the top panel of Figure 5-4. From this figure we see winter temperatures are much more variable than summer temperatures.

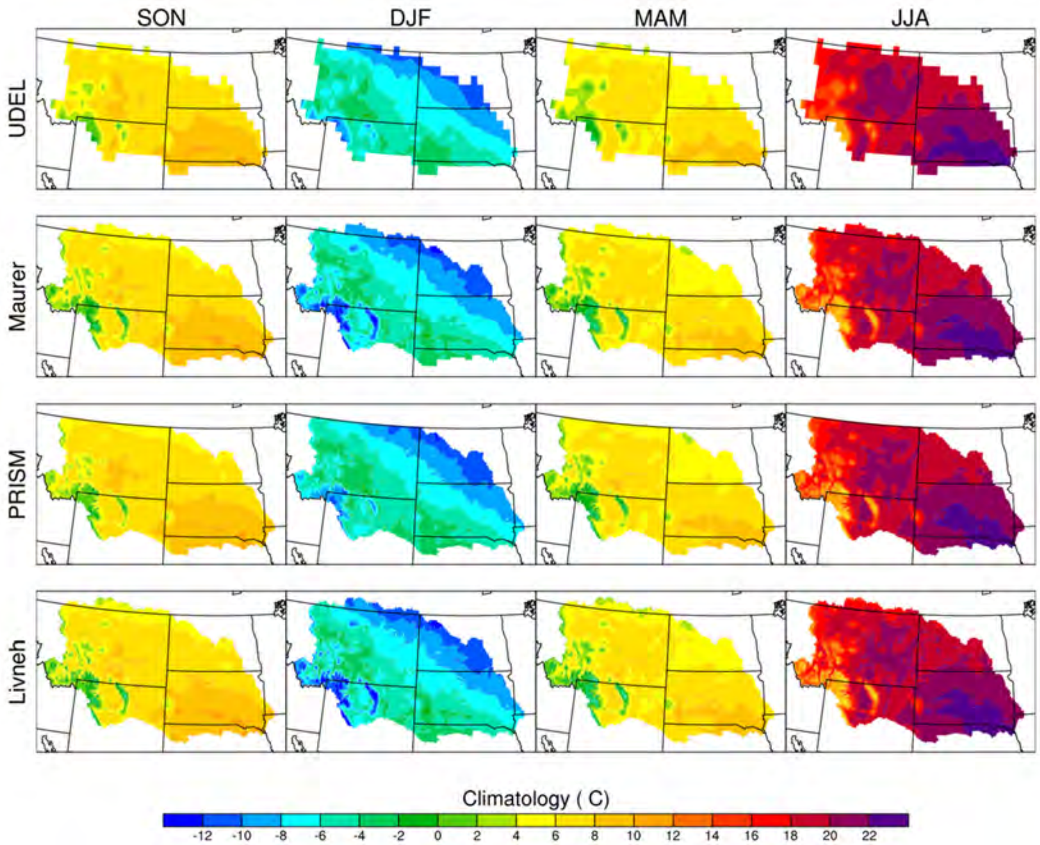


Figure 5-3 The observed climatological seasonal cycle of average daily mean temperature over 1970-2009 (ºC). As for precipitation temperature data is from four sources, UDEL, Maurer, PRISM, and Livneh.

The simulated climatology of temperature from the baseline RCM simulations is shown in Figure 5-5 and the remaining panels in Figure 5-4. The RCM simulations are generally biased cold over the region. This is especially true for the WRF simulations in DJF and MAAM and the RegCM4 simulations in SON and MAM. The only simulation/season with a warm bias is the wrf-hadgem simulation in JJA. However, this will have little influence on winter snowpack. As with observations, winter temperatures in the simulations have much more variability than summer temperatures, however this variability is larger than observed. Colder than observed temperatures in Fall/Winter/Spring will impact snowfall, snow accumulation, and snowmelt and are part of the reason these models have more snow than observed (see below).

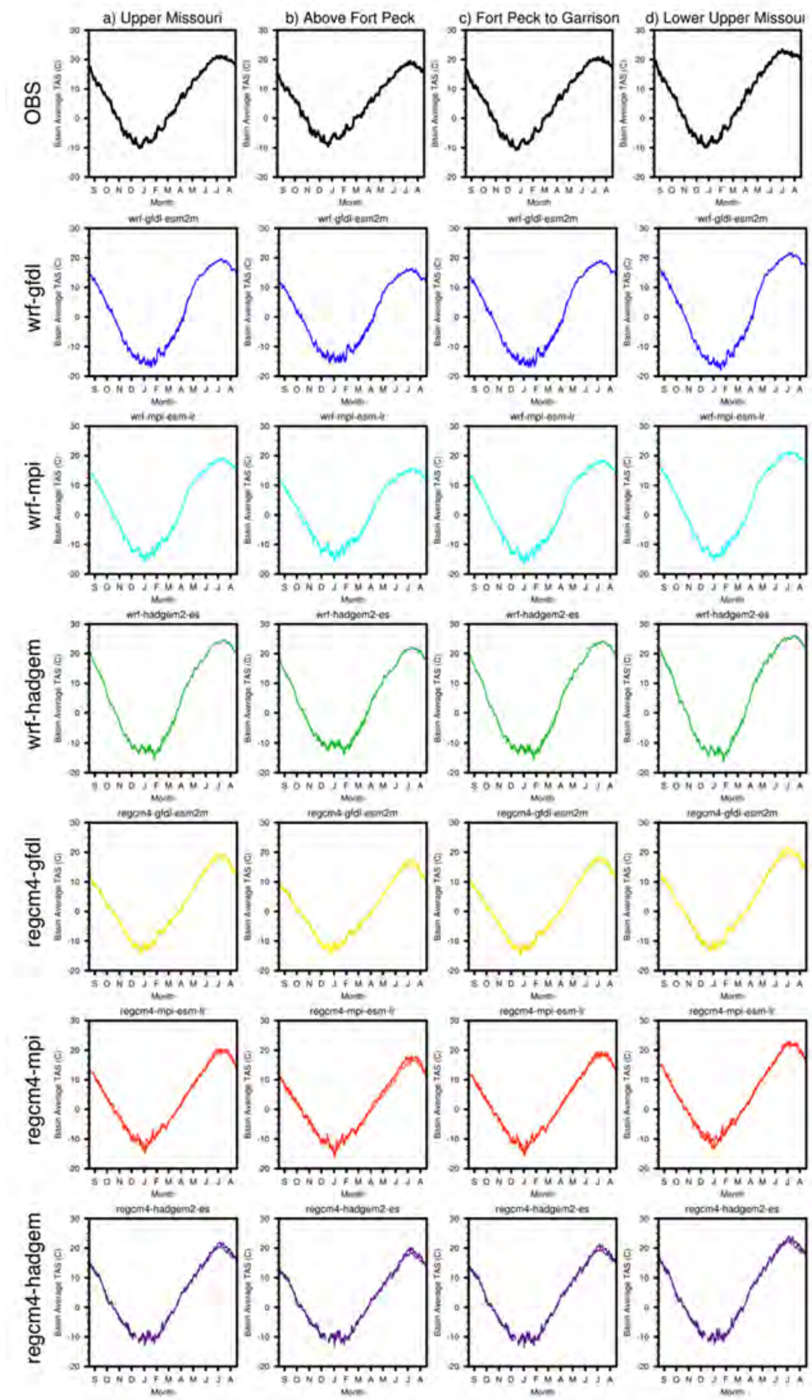


Figure 5-4 The annual cycle of temperature averaged over different sub-basins of the UMRB from observations (top, black) and the different RCM-GCM pairs. Column a) is averaged over the entire UMRB, column b) is the region above Fort Peck, column c) is the region between Fort Peck and Garrison, and column d) is the remainder of the basin. The time period is from 1970-2009 in the observations, and 1970-2005 in the models.

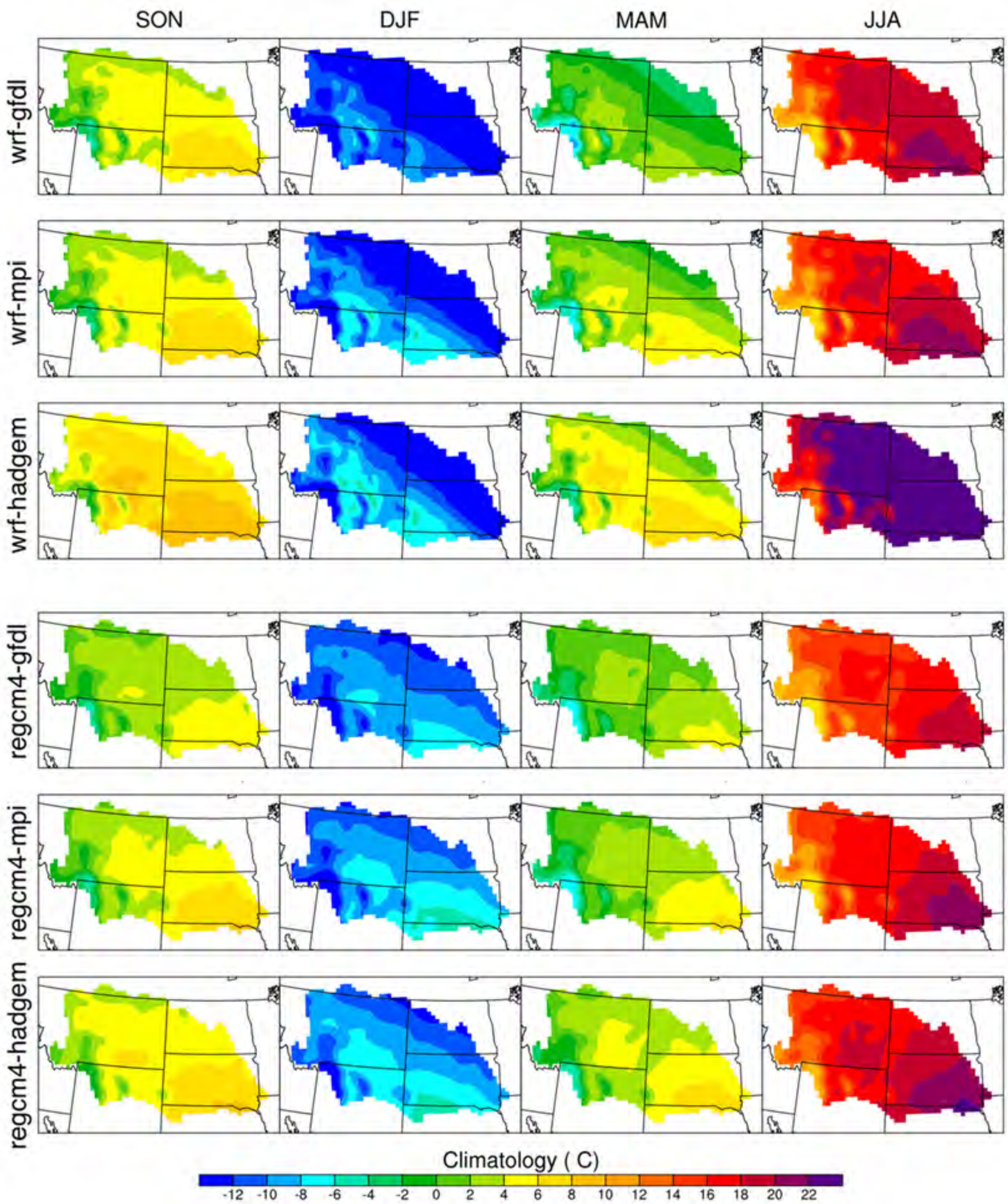


Figure 5-5 The simulated climatological seasonal cycle of average daily mean temperature over 1970-2005 (ºC) from the 25km NA-CORDEX experiments. The top three panels are the simulations from WRF, the bottom three panels are the simulations from RegCM4. Each RCM is driven by three GCMs, which are labeled on the left.

Figure 5-6 shows maps of the observed climatological seasonal cycle of precipitation over the UMRB from four gridded observation products, and Figure 5-7 shows the seasonal cycle of precipitation averaged over the entire UMBR as well as the three sub-basins. Across the entire basin, precipitation is a maximum in spring and summer and lowest in winter. Precipitation is generally low in the winter in this region, as the air masses that influence the region are very cold, and cold air holds little moisture. The mountains in the west have the highest regional precipitation amounts due to orographic forcing. Although winter precipitation is low in the plains, this moisture and snowfall can still play a key role in streamflow and flood dynamics in the region.

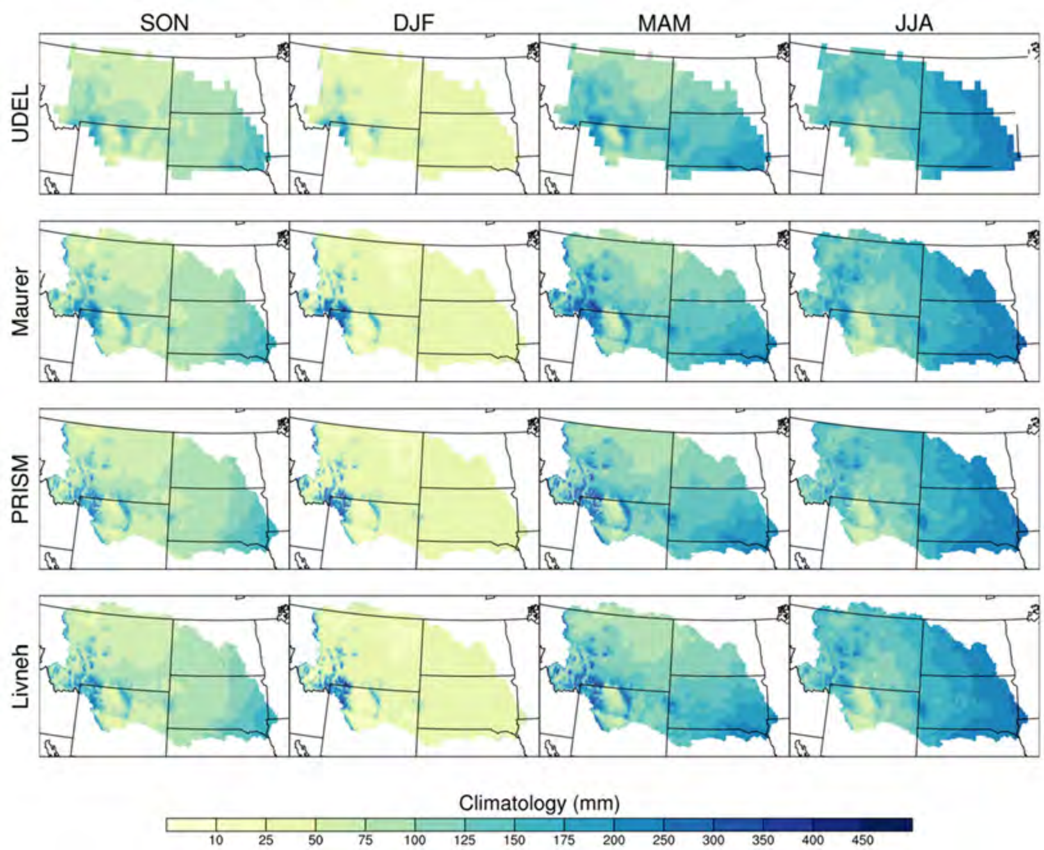


Figure 5-6 The observed climatological seasonal cycle of average accumulated precipitation over 1970-2009 (mm). Precipitation data from four sources, UDEL, Maurer, PRISM, and Livneh.

The simulated climatology of precipitation from the baseline RCM simulations is shown in Figure 5-8 and the panels in Figure 5-7. Although the phasing of the annual cycle of precipitation is captured by the RCMs, precipitation is overestimated in all of the RCMs, especially in late spring and early summer. Winter precipitation is overestimated in the mountains and the plains, which will influence snowfall and SWE.

The observed climatology of SWE from the observational products is shown in Figure 5-9 and the top panel of Figure 5-10. It is clear from these figures that the observational products differ significantly from one another, and there is much uncertainty in our observational knowledge of snow over this region. In all three of these products SWE is derived from a land-surface model being driven by observed meteorology to generate snow. Our understanding of snow processes is limited because snow is so difficult to observe; therefore these models capture snow in different ways. The Livneh product has the highest resolution, and SWE is much higher in the mountains in this dataset than the other two. However Livneh has almost no SWE across the plains, which is likely incorrect. In the other two products, the topography of the mountains are smoothed and SWE is lower. However there is more snow in the plains. Across all of the examined basins, the timing and magnitude of maximum snow volume (or peakSWE) varies across the datasets.

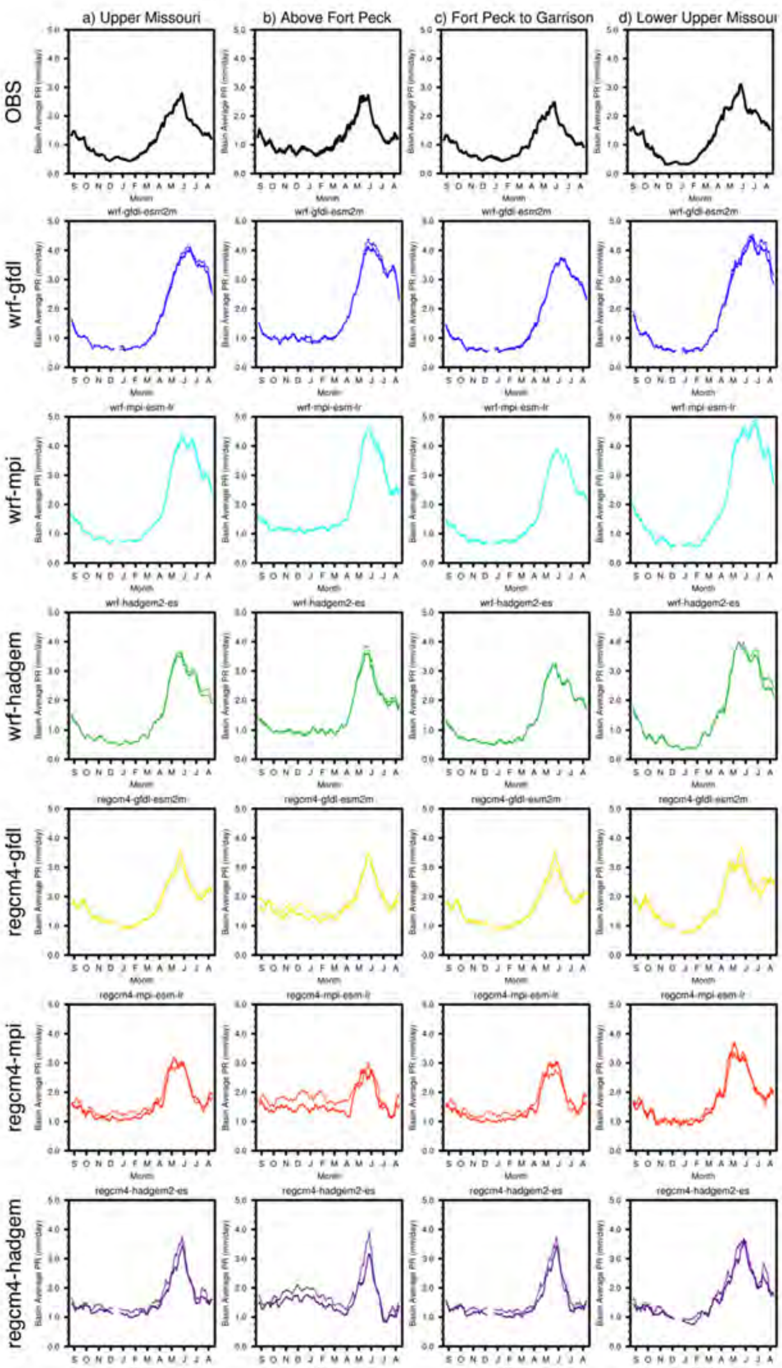


Figure 5-7 Same as Figure 5-4, but for daily precipitation. Precipitation timeseries have had a 15-day running average applied to smooth the field. The breaks in winter are due to the method we calculated the running average.

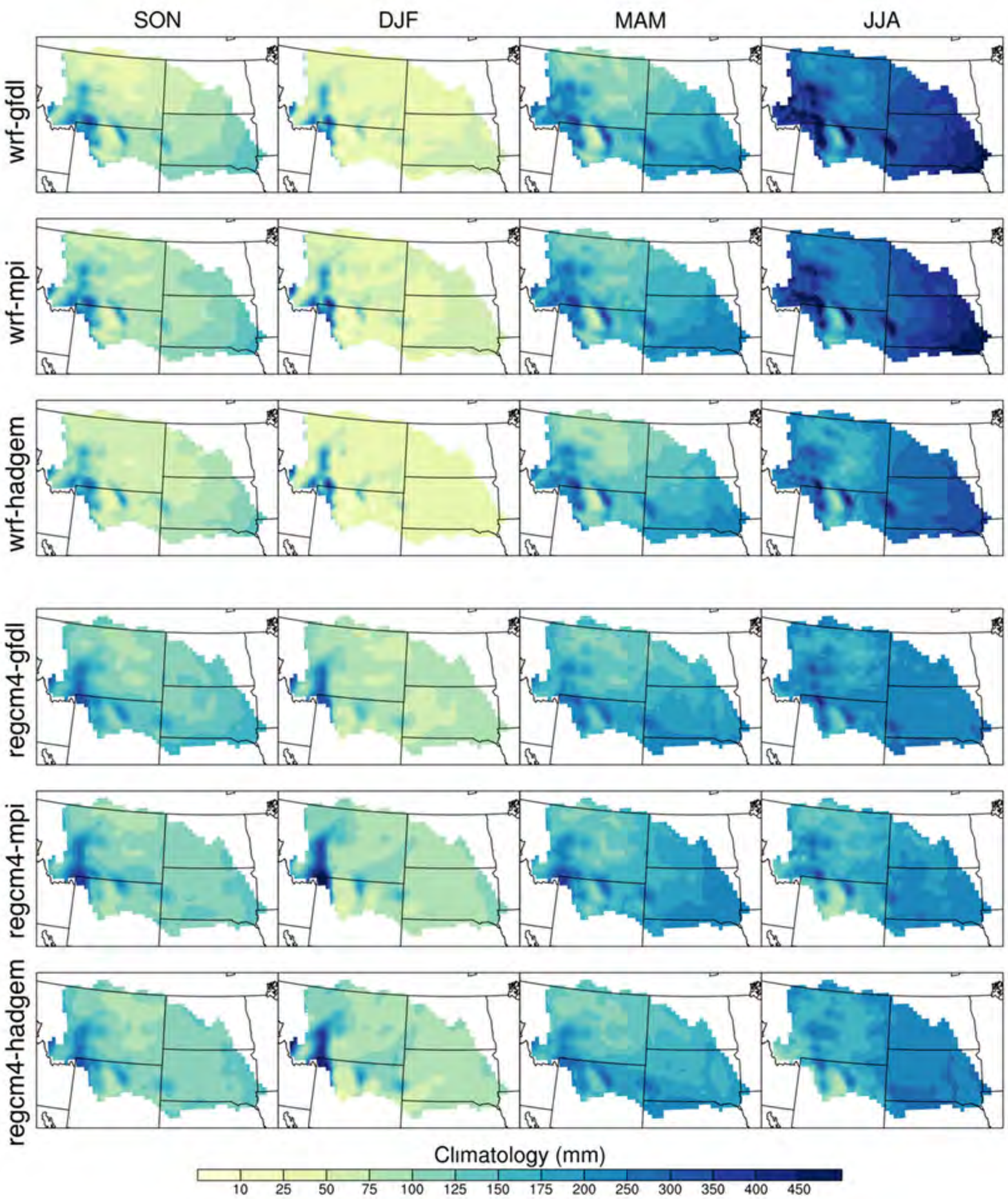


Figure 5-8 The same as Figure 5-5 except for seasonal precipitation.

When examining the seasonal cycle of SWE in the RCMs (Figure 5-10 and Figure 5-11) we see that all of the simulations explored here overestimate SWE in the mountains and the plains, and across all seasons. In some models (regcm4-mpi) SWE is 3 times greater than any of the observed products. These biases are due to the temperature and precipitation biases described above (both models are too cold and have too much precipitation in the winter). However, how the land-surface model used in each RCM parameterizes snow processes also influences these biases (McCrary et al., 2017). Figure 5-10 also shows differences between the 50km and 25km simulations. When mountains are present, the higher resolution simulations result in more SWE in the RegCM4 simulations, but resolution has little influence on SWE in the WRF runs.

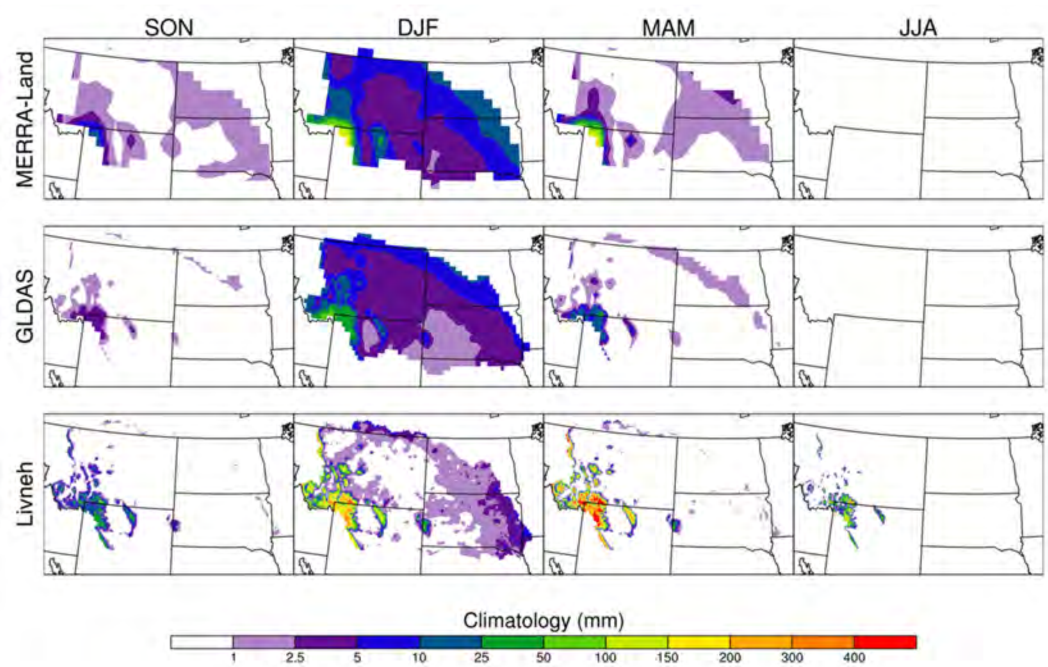


Figure 5-9 The observed climatological seasonal cycle of average SWE over 1970-2009 (ºC). SWE data is from three sources, MERRA-Land, GLDAS, and Livneh.

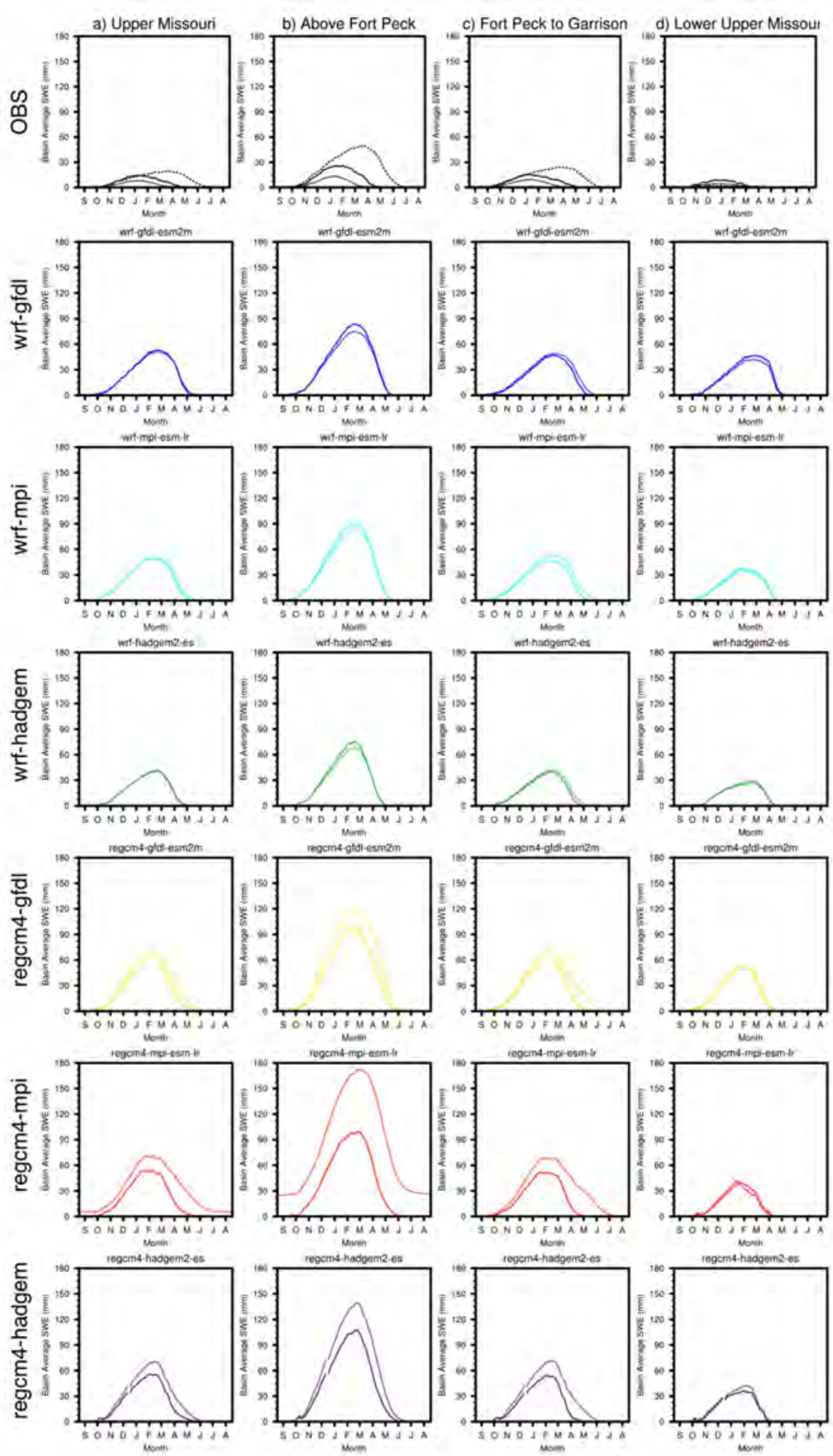


Figure 5-10 Same as Figure 5-4, but for SWE. The three observations from Figure 5-9 are plotted on the top panel. For each RCM-GCM panel, the solid line is the 50km simulation, the dashed line is the 25km panel.

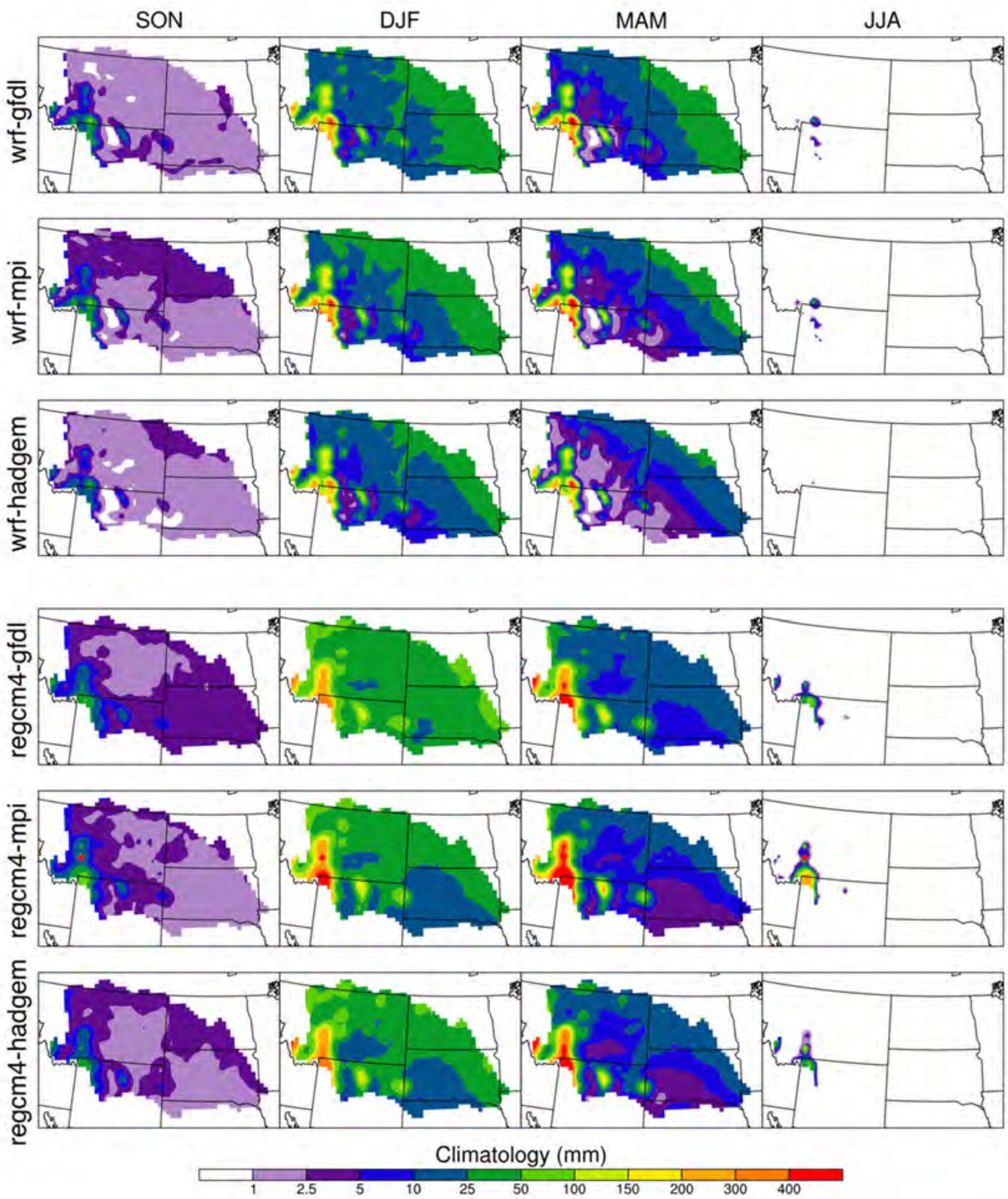


Figure 5-11 The same as Figure 5-5 except for seasonal average SWE.

5.4.2. FUTURE CHANGES

In this section we examine future changes in snow over the UMRB and place them in the context of changes in precipitation and temperature. Figure 5-12 shows the change in surface temperature climatology from the RCM simulations. Here we are looking at the difference between seasonal temperatures in the future climate and the baseline climate. Across the board, in all seasons, temperatures will rise over the UMRB. The pattern of these changes is dependent on the RCM, driving GCM, and season. As one might expect, the largest temperature increases are found in the simulations driven by the HadGEM2-ES model. This global model has a very high equilibrium climate sensitivity (meaning global mean temperatures increase the most in this model corresponding with increasing greenhouse gas concentrations). In general, temperature increases are highest in the southern and eastern half of the domain in the WRF simulations in all Fall/Winter/Spring (the seasons important for snow hydrology). The greatest spatial variability in the change is found in DJF, possibly related to changes in snow and the snow-albedo-feedback. The warming signal is lower in RegCM4 compared to WRF in DJF and MAM, but higher in summer.

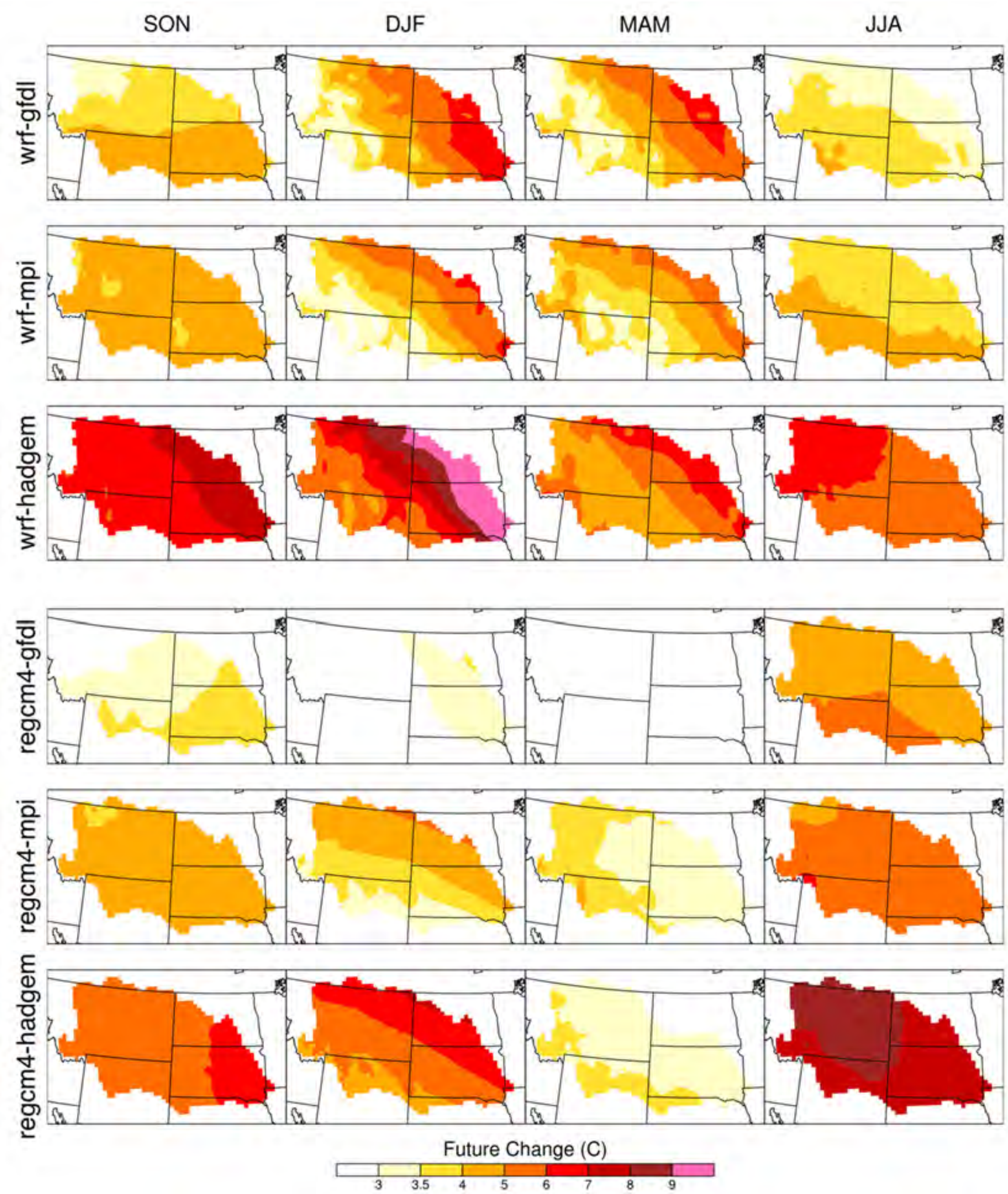


Figure 5-12 End of century changes in the simulated climatological seasonal cycle of average daily mean temperature in (ºC) from the 25km NA-CORDEX experiments. The top three panels are the simulations from WRF, the bottom three panels are the simulations from RegCM4. Each RCM is driven by three GCMs, which are labeled on the left.

Precipitation patterns are also expected to change in the future. Warmer air is capable of holding more moisture, and warmer fall, winter and spring temperatures over the region correspond with significant increases in precipitation. In these seasons there is a correlation between increasing temperatures and increasing precipitation patterns. Winter precipitation increases are highest in the WRF simulations, corresponding with their greater warming. While the simulations driven by RegCM4 show a drying in summer, the WRF simulations show precipitation increasing in summer. In both the WRF and RegCM4 simulations driven by the GFDL-ESM2M models, precipitation in SON is shown to increase in the northwestern part of the domain, but decrease over the remainder of the domain.

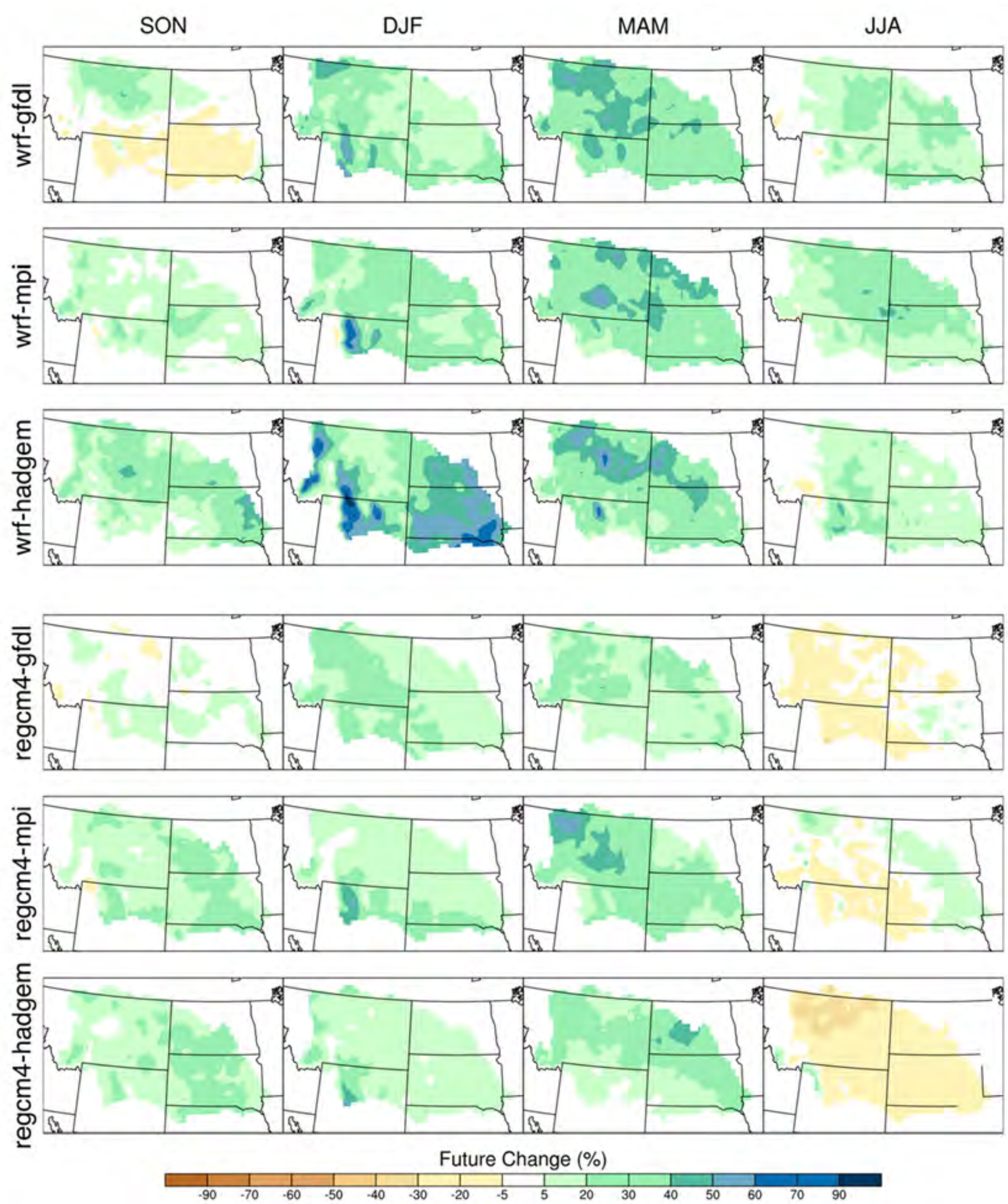


Figure 5-13 *The same as* Figure 5-12*, but for the percent change in precipitation.*

In the future, changes in snow and snow related variables will be due to the complex interaction between increasing temperatures and changing precipitation patterns. Rising temperatures result in decreases in snow in two ways: they reduce the proportion of precipitation that falls as snow and they increase snow loss via ablation. In the absence of any temperature changes, changes in snowfall would be positively correlated with precipitation and increases (decreases) in precipitation would correspond with increases (decreases) in snowfall. In the future, temperature and precipitation will interact with each other in complex ways, resulting in regional variations in the sign and magnitude of changes in snowfall, SWE, and snow cover.

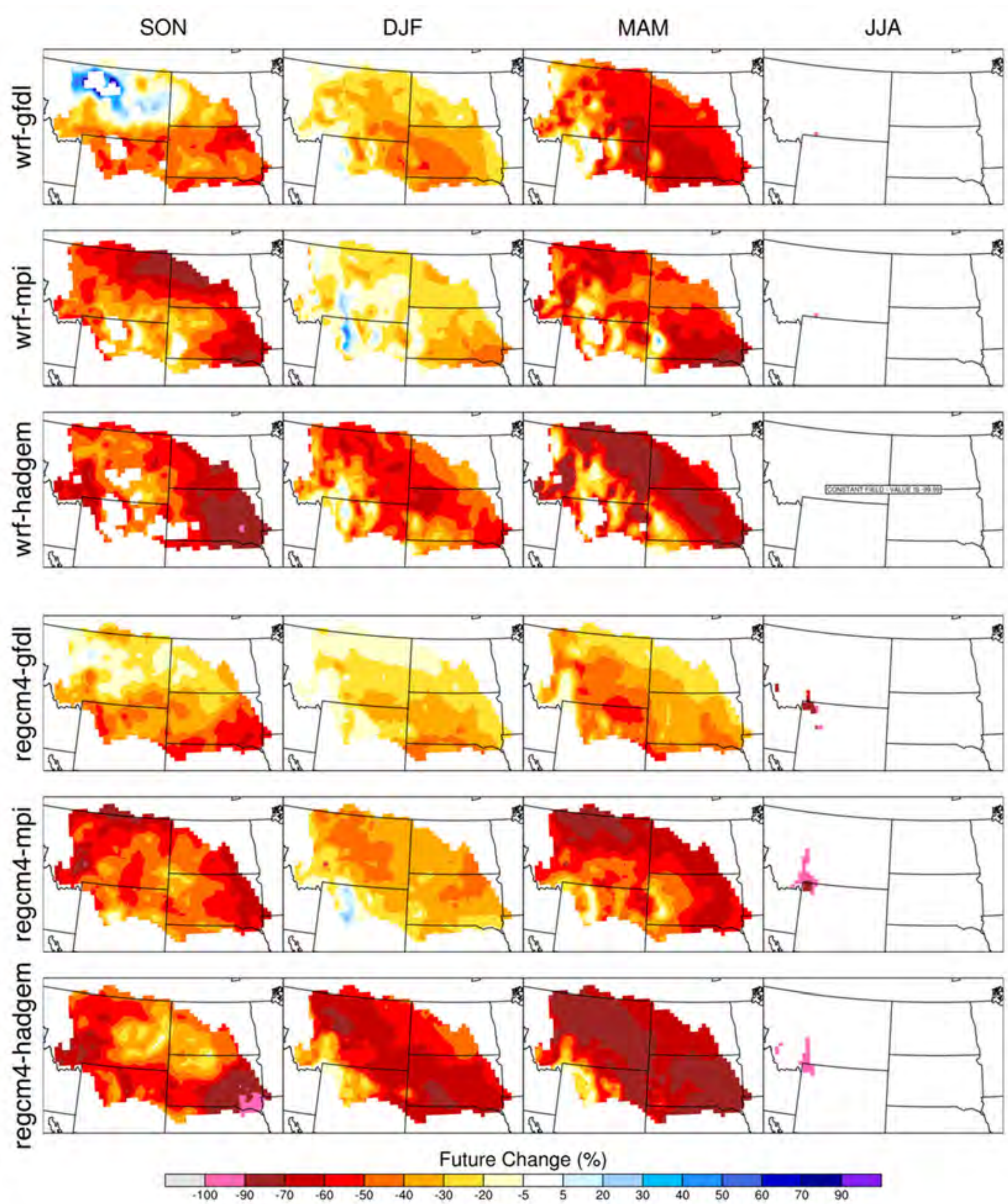


Figure 5-14 The same as Figure 5-12, but for the percent change in SWE.

Figure 5-14 shows the spatial patterns of the seasonal changes in SWE over the region. With the exception of three of the twelve simulations explored here, SWE is projected to decrease at all locations during all seasons. The exceptions to this are the WRF-GFDL (SON), WRF-MPI (DJF), and RegCM4-MPI (DJF) simulations/seasons. In all three exceptions, increases are found in regions where initial SWE is very low in the baseline climate and small increases are found in the future. The losses in SWE found everywhere else are primarily driven by the warming signal over the region. While increases in precipitation can mitigate snow losses (especially in winter when temperatures remain below freezing), total snow accumulation is reduced as less precipitation falls as snow in fall and spring, more mid-season melt events occur, and melting initiates earlier.

Future changes in the annual cycle of SWE are further examined in Figure 5-15. Here we can see that in all the simulations, when averaged over the entire basin and/or sub basins, total snow volume is decreased. The timing of annual maximum SWE is also shifted to earlier in the season in all of the basins indicating the onset of the snowmelt seasons will shift earlier in the seasons, impacting streamflow and stream management.

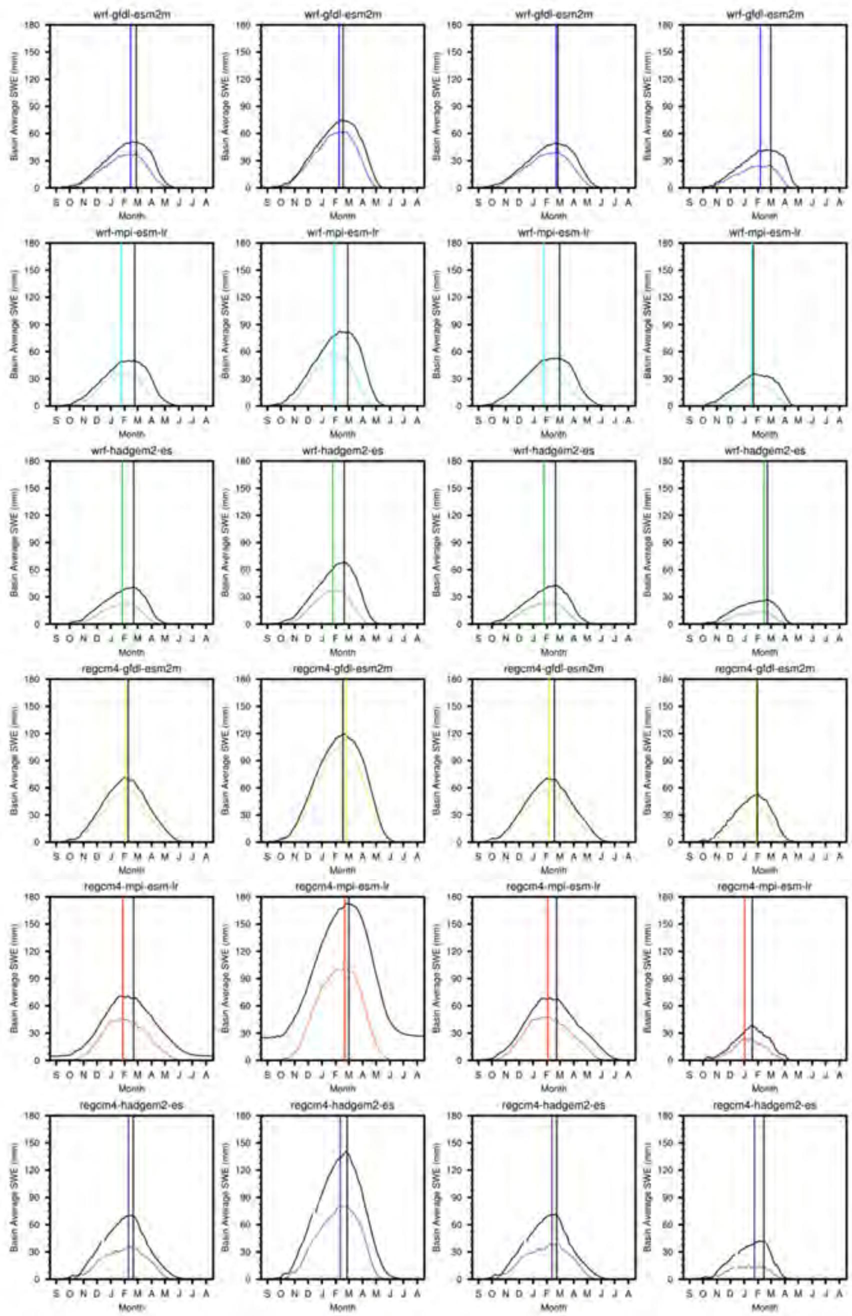


Figure 5-15 Plots of the annual cycle of SWE from the NA-CORDEX RCM simulations averaged over our four study regions (from left to right, Upper Missouri, above Fort Peck, Fort Peck to Garrison, Lower Missouri). The historical-baseline SWE climatology is shown in black in each panel. The future SWE climatology is shown as a colored line on each panel. Each color represents a different RCM-GCM combination. The vertical lines mark the annual peak SWE volume in the historical (black) and future (colored) simulations.

5.5. CONCLUSION

This study evaluated the climate of the Upper Missouri River Basin in regional climate simulations from NA-CORDEX and examined future changes in snow and their climatological drivers. In the UMRB spring and summer flooding is often correlated with high antecedent winter snowpack conditions. Therefore, it is critical that we understand how snowpack conditions are projected to change in the region in the future.

We demonstrated that the four commonly used observed gridded datasets of temperature and precipitation used in climate and hydrology studies are in agreement with each other over the UMRB. While small spatial differences do occur with regard to how topography is treated and the complexity of how in situ surface observations are interpolated, these differences are small. We also demonstrated, however, that observations of SWE over the domain have significant uncertainties and are largely in disagreement. We will continue to investigate other SWE datasets to include in our studies to improve our knowledge of observed SWE.

In their baseline climate simulations, the 12 25km simulations from WRF and RegCM4 have large cold biases in Fall, Winter, and Spring. These models are also too wet during all months of the ear. The cold, wet biases in the models result in much larger snow volumes than were observed in any of the observational datasets. These biases in snow will impact surface radiation, which can feed back onto temperature in the models. Furthermore, this positive bias in SWE may play a role in the magnitude of the future changes in SWE over the region. However, a quantitate analysis of the role bias plays was beyond the scope of this study.

As for the future, the models show large increases in cold season temperatures and increases in precipitation. Almost everywhere and in every simulation, SWE is projected to decrease throughout the year. Warmer temperatures decrease the fraction of precipitation falling as snow and increase the melt rate of snow on the surface. While increases in cold-season precipitation may mitigate some of the SWE losses, temperature increases dominate the signal. We also found that the timing of peak snow volume shifts to earlier in the season which will have implications for streamflow management practices.

5.6. REFERENCES

Armstrong, R., M. Brodzik, K. Knowles, and M. Savoie, 2005: Global monthly EASE-grid snow water equivalent climatology. Subset used: 1979–2003, NASA National Snow and Ice Data Center Distributed Active Archive Center, accessed 5 November 2012, doi:10.5067/KJVERY3MIBPS.

Brown, R. D., B. Brasnett, and D. Robinson, 2003: Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. *Atmosphere-Ocean*, **41**, 1–14, https://doi.org/10.3137/ao.410101.

Dunne, J.P., J.G. John, A.J. Adcroft, S.M. Griffies, R.W. Hallberg, E. Shevliakova, R.J. Stouffer, W. Cooke, K.A. Dunne, M.J. Harrison, J.P. Krasting, S.L. Malyshev, P.C. Milly, P.J. Phillipps, L.T. Sentman, B.L. Samuels, M.J. Spelman, M. Winton, A.T. Wittenberg, and N. Zadeh, 2012: GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics. *J. Climate*, **25**, 6646–6665.

Dyer, J. L., and T. L. Mote, 2007: Trends in snow ablation over North America. *International Journal of Climatology*, **27**, 739–748.

Frei, A., R. Brown, J. A. Miller, and D. A. Robinson, 2005: Snow Mass over North America: Observations and Results from the Second Phase of the Atmospheric Model Intercomparison Project. *Journal of Hydrometeorology*, **6**, 681–695.

Giorgetta, M. A., et al. (2013), Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, *J. Adv. Model. Earth Syst.*, 5, 572– 597.

The HadGEM2 Development Team, 2011: The HadGEM2 family of Met Office Unified Model climate configurations, Geosci. Model Dev., 4, 723-757.

Kapnick, S. B., and T. L. Delworth, 2013: Controls of Global Snow under a Changed Climate. *Journal of Climate*, **26**, 5537–5562.

Livneh B., E.A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K.M. Andreadis, E.P. Maurer, and D.P. Lettenmaier, 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, 9384–9392.

Livneh B., T.J. Bohn, D.S. Pierce, F. Munoz-Ariola, B. Nijssen, R. Vose, D. Cayan, and L.D. Brekke, 2015: A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and southern Canada 1950-2013, Nature Scientific Data, 5:150042, doi:10.1038/sdata.2015.42.

Livneh, B., M. Hoerling, A. Badger, and J. Eischeid, 2016: Causes for Hydrologic Extremes in the Upper Missouri River Basin, NOAA Cimate Assessment Report.

Maurer, E.P., A.W. Wood, J.C. Adam, D.P. Lettenmaier, and B. Nijssen, 2002, A Long-Term Hydrologically-Based Data Set of

Land Surface Fluxes and States for the Conterminous United States, J. Climate 15(22), 3237-3251.
McCrary, R. R., S. McGinnis and L. O. Mearns, 2017: Evaluation of Snow Water Equivalent in NARCCAP Simulations, Including Measures of Observational Uncertainty. *J, Hydromet.*, **18**, 28
McCrary, R. R and L. O. Mearns, 2019: Quantifying and Diagnosing Sources of Uncertainty in Mid-Century Changes in North American Snowpack from NARCCAP. *J. Hydromet*, In Review.
McCrary and Mearns, 2019 PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 15 Mar 2015.
Rodell, M., P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann, and D. Toll, 2004: The Global Land Data Assimilation System, Bull. Amer. Meteor. Soc., 85(3), 381-394.

6. IMPROVED UNDERSTANDING OF FLOOD MECHANISMS IN THE UPPER MISSOURI BASIN: A MODEL COMPARISON APPROACH

6.1. EXECUTIVE SUMMARY

Understanding the effects of climate change on hydrological extreme drivers across the Missouri River is challenging. In this region, floods are mainly triggered by snow melting, either when temperatures get mild in spring/summer, or when rain falls over snow in early spring and fall. In this work, we compare future flood peak estimates from three different methods to gain insight on the relative role of various hydro-meteorological process driving flood peak intensity. We consider two conceptual hydrological models of different complexities (i.e., a lumped and a distributed version of the Sacramento – Soil Moisture Accounting). We also consider a Deep Artificial Neural Networks (DANNs) that uses additional inputs than hydrological models do (e.g, soil moisture and snow in addition to precipitation and temperature) to predict monthly maximum peak discharge. We drive the above models with perturbed hydro-meteorological time series within the Decision Scaling approach and subsequently assess potential changes in peak flows. Climate projections from the NARCCAP experiment used to infer potential changes in peak flow within the Decision Scaling approach. By comparing the estimates of future peak discharges from the three modeling framework, the objective is to improve our understanding of the relative weight of the main hydrological drivers of extreme streamflow in the Upper Missouri Basin. Three snowmelt-dominated catchments are considered in this study; Yellowstone River basin at Billings (MT; USGS 06214500), Powder River near Locate (Montana; USGS 06326500) and the James River near Scotland (South Dakota; USGS 06478500). Results show that the role of precipitation and temperature spatial variability is key in the flood generation mechanism and that decreasing snowpack is likely to decrease the snowmelt peak flow in the area.

6.2. TECHNICICAL APPROACH

The objective of this study is to improve understanding of the role of various hydro meteorological processes in peak discharge change under climate change. Three modeling approaches are considered, a lumped and a distributed version of the Sacramento Soil-Moisture Accounting (SACSMA) model and a Deep Artificial Neural Network. The use of both lump and distributed SACSMA models allows gaining insight of the role of the spatial variability in the meteorological variables driving streamflow generation. DANNs are used to explore the role of other hydro meteorological variables such as soil-moisture and snow. The estimate of change in annual peak discharge is obtained following the Decision Scaling approach. Perturbed time series of precipitation, temperature, soil moisture and snow are obtained following the change factor approach. Projections for NARCCAP climate experiment are used to infer future changes in hydro meteorological variables leading to a climate-informed assessment of change in peak discharges.

6.3. DATA AND METHODS

6.3.1. CASE STUDY

This study focuses on the cold and dry Upper Missouri Basin (Figure 6-1). This region regularly experiences major flooding events. For instance, in March 2010, a record river stage has been observed across the James River catchment (South Dakota, Figure 6-1). Such an extreme flooding resulted from an important snowmelt that followed from very wet fall and winter seasons. Another example is the 2011 flood across the Yellowstone River Basin that led to the largest peak flow ever recorded at the station near Livingston (Figure 6-1). The current study focuses on three catchments (Figure 6-1); the Yellowstone River at Billings (Montana; USGS 06214500), the Powder River near Locate (Montana; USGS 06326500) and the James River near Scotland (South Dakota; USGS 06478500).

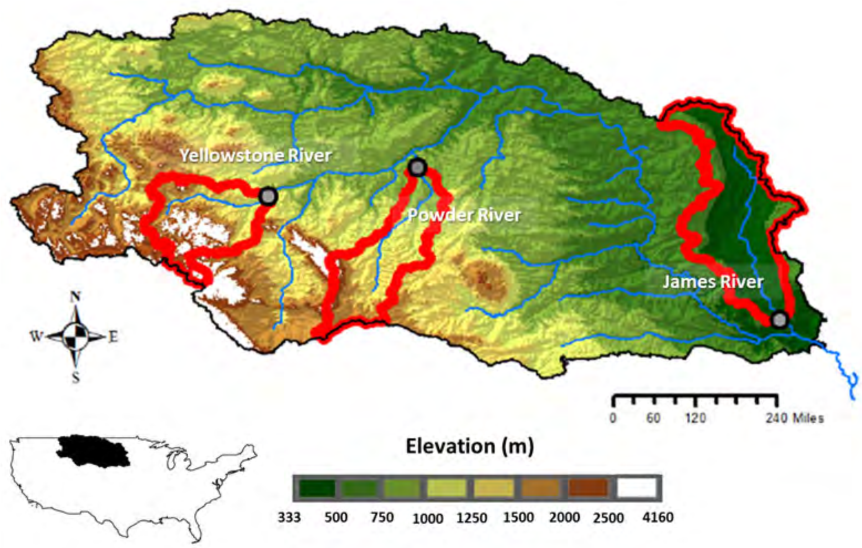


Figure 6-1Map of the Upper Missouri basin. Considered catchments are highlighted with red boundaries.

Yellowstone River's catchment is located at high altitudes (Table 6-1) and shows logically low average temperature and relatively high average precipitation. As illustrated on Figure 6-2, average precipitation peaks in the middle of May – early June. Figure 6-2 also gives insight on precipitation variability by means of the grey shaded area showing the deviation between the 10^{th} and the 90^{th} percentiles. Low temperature during winter associated with rather high precipitation in winter leads to significant snowpack accumulation. Soil-moisture follows rather well the evolution of the other variables; it increases during snowmelt and decreases in summertime due to high temperatures and low average precipitation. As result, the hydrological regime of the Yellowstone River presents a marked seasonality with low flow in winter and high flow during spring (Figure 6-3). The date of the annual maximum discharge is rather consistent every year and appears to be around mid-June.

The Powder River catchment is located at lower altitude (Table 6-1) than Yellowstone's and consequently shows slightly higher average temperature. The catchment is also significantly drier as median precipitation from October to early March is nearly null (Figure 6-2). Precipitation variability is low but shows a high peak in early June and a second peak in October. Snowpack accumulation is rather low compare to Yellowstone River Basin, which results from the rather low precipitation cumuli during winter season. The soil-moisture variability is low compare to other catchments in the region, although it shows a small peak during the rainy period in June. The average discharge cycle presents an interesting average cycle because, contrary to Yellowstone, two peaks are observed (Figure 6-3). The highest peak in June is correlated with the precipitation peak at this season. The second peak is in the middle of March. This peak cannot be explained by the precipitation at this period. A possible explanation is a significant contribution of snowmelt or soil water content after rainy events at this period.

The James River near Locate is the lowest elevation catchment we focus on (Table 6-1). Winters are very dry with the median precipitation nearly equal to zero. However, precipitation in spring shows an important peak with tremendous rainfall intensity occurring from early spring to the end of the summer season. Snow accumulation is slightly higher than for the Power River but much lower than for the Yellowstone River. River flows are very low during fall and winter seasons. At the beginning of spring, peak discharges result from both snowmelt and precipitation. Later in spring, only precipitation events trigger peak flow events.

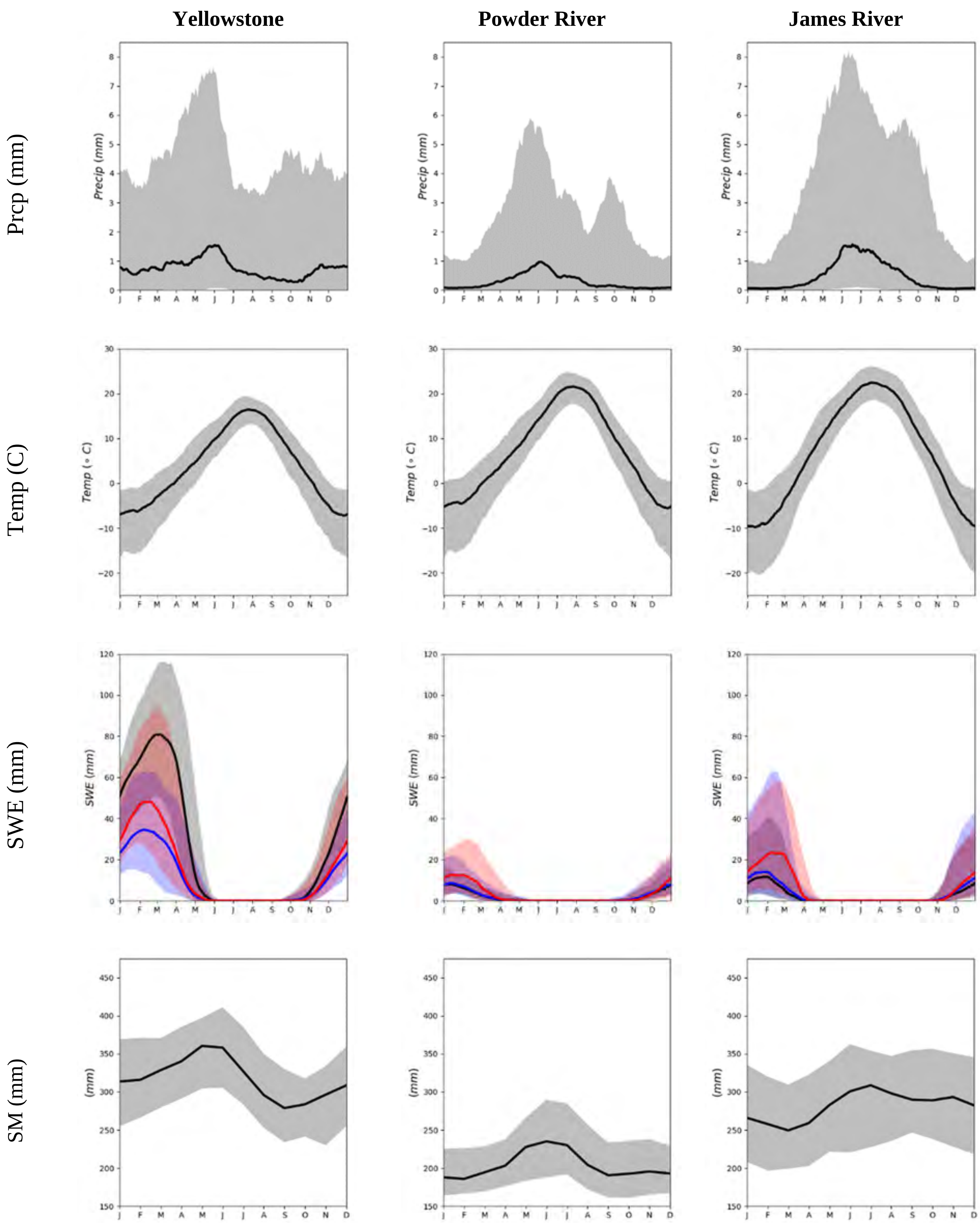


Figure 6-2 Climatology of the three considered catchments (columns). Bold curves give the median cycles and the shaded areas give the distance between 10th and 90th percentiles. For the snow water equivalent, black, red and blue colors correspond to three different databases (MERRA, ERAIL and BrownERA, respectively) (1980-2010).

Table 6-1 Description of the considered catchments. Temperature and precipitation values are annual (average for the temperature variable and total for precipitation.

Gauge	USGS ID	Surface (km^2)	Elevation (m) mean [min – max]	Temperature (°C) mean [min – max]	Precipitation (mm) mean [min – max]
Yellowstone R.	06214500	30,580	1967 [944-3870]	3.2 [1.6 – 4.4]	641 [476 – 852]
Powder River	06326500	33,825	1386 [734-4006]	6.5 [4.7 – 7.9]	361 [200 - 537]
James River	06478500	53,540	464 [360-693]	6.4 [4.0 - 8.9]	511 [270 – 702]

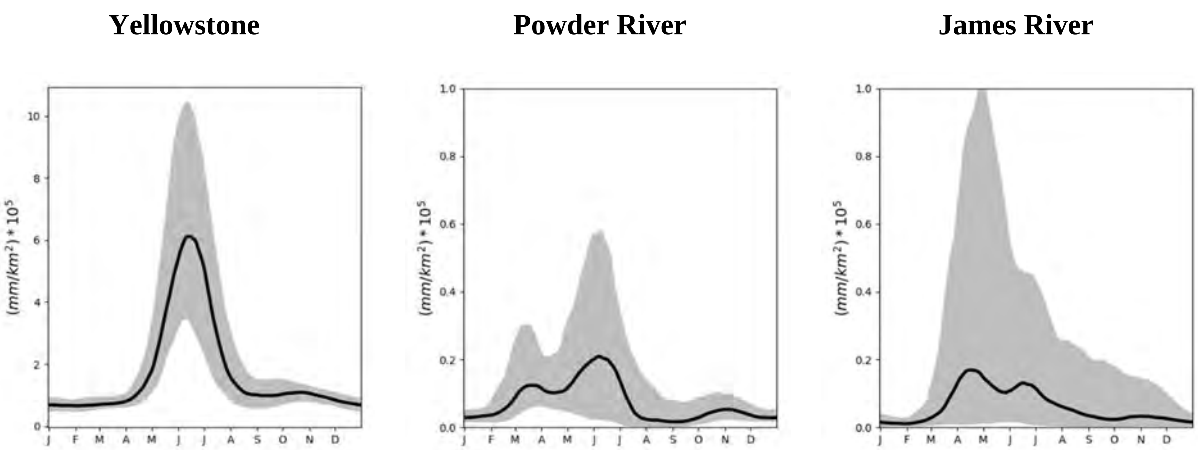


Figure 6-3 Annual average cycles of streamflow for the considered catchment. Grey shaded areas give the distance in-between the 10th and 90th percentiles.

6.3.2. Database

In this study, precipitation and temperature variables are used as input of the SAC-SMA hydrological model. The DANNs can used different combinations of inputs that can include precipitation, temperature, soil moisture and snow water equivalent (the choice of the inputs is actually a part of their calibration; see section 6.3.4).

Precipitation and Temperature data at 1/8th resolution are obtained from the meteorological database developed by Livneh et al. (2015) for continental United States. Temperature and precipitation data are available for the period 1949-2010. Snow Water Equivalent data (hereafter denoted as SWE) are candidate as input for the DANNs. Three different SWE reanalysis datasets are considered; MERRA-Land (Reichle et al. 2011), ERAI-Land (Balsamo et al. 2015) and Brown ERA (Brown et al. 2003; McCrary et al. 2017). The three datasets were found to be rather well correlated but present significant differences in average (Figure 6-2). Soil Moisture data is also candidate as input for the DANNs. They are simulated monthly mean values at 0.5-degree resolution. They are provided by the Climate Prediction Center of the National Oceanic & Atmospheric Administration (NOAA)(Fan and van den Dool, 2004) over a period spanning from 1948 to 2017. Observed streamflow time series are provided by the United States Geological Survey (USGS) (https://www.usgs.gov/). We only used unpaired discharges time series (i.e., no major dam is located upstream the considered gauges) whom data are available, with no significant gap, over the period 1980-2010. This period being the longest period where all input data described above are available.

6.3.3. SACSMA HYDROLOGICAL MODEL

The Sacramento- Soil Moisture Accounting (SACSMA) is a conceptual hydrological model (Anderson and McDonnell, 2005) that models soil moisture accounting and evapotranspiration. SACSMA is here coupled with the temperature index snowmelt model SNOW-17 that ensures a good representation of the snowpack dynamic (i.e., accumulation and melt)(Anderson, 2006). Flow routing across the basins is performed using the Lohmann model (Lohmann et al. 1998). Inputs are precipitation and temperature time series. More details on the actual SACSMA model set-up are given in Brown et al. 2016.

The National Weather Service has used the lumped SACSMA model historically for flood forecasting. In addition to the lumped version of SACSMA, we also consider a distributed version. Distributed hydrological models allow accounting for both spatial variability of surface characteristics and subsequent hydrological processes, and for spatial variability of precipitation and temperature driving streamflow generation. Variability of weather inputs is often shown as having the highest influence on peak discharge modeling (see discussions in Andréassian et al., 2004 and in Lobligeois et al. 2014).

SACSMA models for the three considered catchments are calibrated using a Genetic Algorithm (Wang, 1991) over the 1980-2000 period and validated over the 2001-2010 period. The considered goodness-to-fit criteria is the Nash-Sutcliffe Efficiency (Nash and Sutcliffe, 1970) that is estimated at daily time scale.

6.3.4. DEEP ARTIFICIAL NEURAL NETWORK (DANN)

In this study, we use Deep Artificial Neural Networks (DANNs) to predict the maximum daily flow values for each month of the simulation period. Below, we describe the basics of how DANNs work and more specifically how we use ANN for our application. For a comprehensive description of DANNs, the reader is invited to head to the review by LeCun et al. (2015). In this study, we considered sequential, dense and feed forward networks (Figure 6-4).

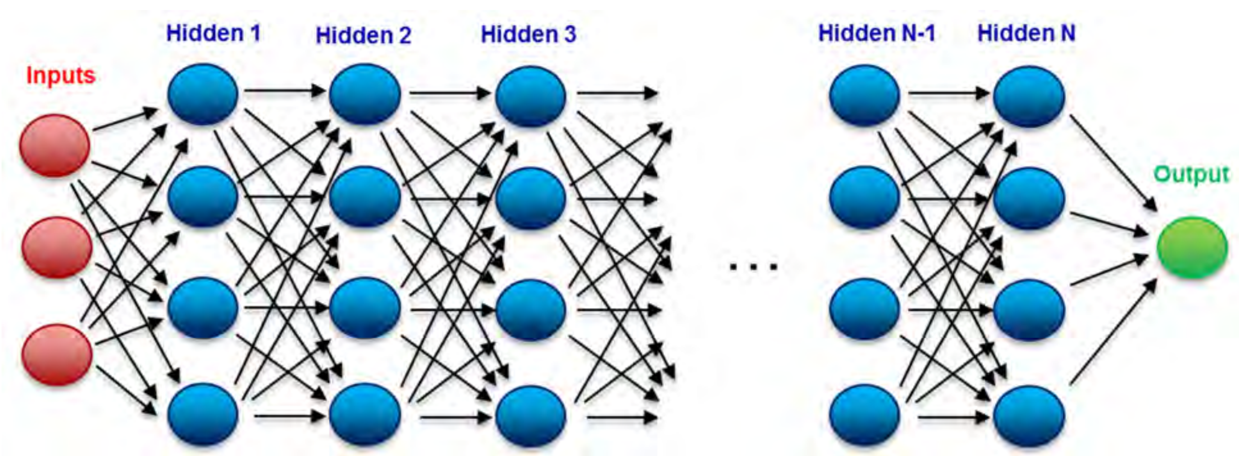


Figure 6-4 Illustration of a deep, sequential, dense and feed forward artificial neural network.

The input $H_{i,j}$ to the neuron *i* at the hidden layer *j* is obtained by applying an activation function f_j to the weighed sum of the output from neurons of the previous layer *j-1*:

$$H_{i,j} = \sum_{k=1}^{K_{j-1}} f_j(w_{k,i}H_{k,j-1} + \theta_j) \qquad 6.1$$

with K_j is the number of neurons within the hidden layer j; $w_{k,i}$ is the weight that is applied to the output of the neuron k of the previous layer and θ_j is a bias correction factor. DANNs' prediction results from the last transfer through the output layer (Figure 6-4).

DANNs need training (i.e., calibration). During DANN's training, observed hydrometeorological variables are used as inputs to predict a set of known events, which is here a time series of observed maximum daily streamflow within each month of the considered period. Once the prediction is done, an objective function (also denoted *loss function*) is calculated. For the sake of consistency with SACSMA calibration, the negative of the NSE is used as loss function that is meant to be minimized during the network training procedure. Note that for DANNs, the NSE is calculated at monthly time scale and considering the maximum daily flow for each month. Throughout training, DANN parameters (weights $w_{k,i}$ and bias θ_j, equation 6.1) are corrected by mean of a gradient descent algorithm through a backpropagation procedure that aims at minimizing the loss function. Once the DANN' parameters are updated, a new prediction can be made and a new evaluation of the loss function is done. This procedure is repeated until convergence of the loss function. As suggested by Lecun et al. (1998), inputs to the DANNs are first normalized (i.e. subtracting the mean and then dividing by the variance) in order to improve the convergence of the backpropagation procedure. In practice, the number of loss function updates (also termed as 'epochs') is meant to be small to avoid overfitting of the network. To reduce the risk of overfitting we also include a dropout rate of 20 %, which consists at selecting randomly 20 % of the neurons from each hidden layer and disregarding them during the training (Srivastava et al., 2014). For the activation function, we use the Rectified Linear Unit function as activation function (denoted as *ReLU* function):

$$ReLU(x) = \max(0, x). \qquad 6.2$$

ReLU function is known to avoids the vanishing gradient problem for deep networks, which is the issue that weights of the first hidden layers remain almost unchanged comparing to the weights of the last hidden layers, and so whatever the number of epochs. To optimize the network weights, we use the stochastic gradient descent algorithm Adam (Adaptive Moment Estimate; Kingma and Lei Ba, 2015). Stochastic gradient descent algorithms are efficient because they perform frequent weight updates with high variance, which may lead to significant fluctuations in the loss function across the epochs but in practice allows the parameter sets to jump out of potential local optimal.

During DANN training, the network structure (i.e., number of hidden layers and number of nodes within each hidden layer) must be defined. So must the hydromoeteorological variables that will serve as inputs to the network. These decisions are commonly made following a trial-and-error procedure (cf. section 6.4.2.2). Candidate inputs to DANNs are monthly precipitation, temperature, soil moisture and snow water equivalent for the month of the prediction. Note that the values for months prior the prediction can also be input to the network if they reveal to be pertinent during the trial-and-error procedure. Accounting for months prior the prediction allows the DANNs to get knowledge on the temporality of the relationship between the hydro meteorological drivers and the peak discharges at the outlet of the catchments. Note that hydro-meteorological inputs are averaged over the considered basins and then normalized following the procedure discussed above.

6.3.5. Climate projections

We use climate projections from the NARCCAP experiment (Mearns et al., 2012) to infer the likelihood of changes in peak discharges from projected changes in hydrometeorological variables. In total, projections from four GCMs have been downscaled by four Regional Climate Models (RCMs). Although twelve GCM/RCM combinations are available from NARCCAP experiment (https://www.narccap.ucar.edu/results/index.html#climate-change), only eight include soil moisture and snow water equivalent (Table 6-2). A detail presentation of the considered GCMs and RCMs is available on the NCAR website (https://www.earthsystemgrid.org/project/NARCCAP.html). Downscaled projections for variables of interest were then averaged over the considered catchments. Projected changes are calculated between the historic period (1970-2000) and the future period (2040-2070).

Table 6-2 Matrix of GCM-RCM projections considered from NARCCAP experiment

GCM - RCM	CRCM	ECP2	MM5I	RCM3
CCSM	x		x	
CGCM3	x			x
GFDL		x		x
HADCM3		x	x	

6.3.6. Peak discharge model fitting

In this study, we focus on the changes in annual maximum streamflow (AMS) distribution. The annual maximum daily streamflow values are first extracted from SACSMA and DANN simulated time series. Then, a Generalized Extreme Value distribution (Jenkinson, 1955) is fitted to the extracted AMS time series. Distribution fitting is performed using the L-moment method (Hosking, 1990).

6.4. RESULTS AND DISCUSSION

6.4.1. Model calibration and validation

6.4.1.1. SACSMA

SACSMA calibration results are shown on Table 6-3 for the three considered basins. The lumped and distributed versions of SACSMA are denoted as SACSMA-L SACSMA-D respectively. SACSMA-D simulation results range from quite good (Yellowstone River) to fair performance (Powder and James Rivers). Not surprisingly, SACSMA-L has lower performance. For the Yellowstone River Basin, for instance, SACSMA–L simulation results are good during the calibration period (NSE=0.7) but collapses for the validation period (NSE=0.53). This is also true for the James River for which NSE value decreases to value lower than 0.4 for the validation period. Performance obtained for the Powder River is low, with NSE values lower than 0.5 for both calibration and validation periods.

Table 6-3 Evaluation of SAC-SMA hydrological model over the three considered catchments. NSE stands for Nash-Sutcliffe Efficiency. NSE values are given for calibration (cal) and validation (val) periods and for daily and monthly time step.

	NSE(cal); NSE(val)		
	Yellowstone	**Powder**	**James**
SACSMA-D	0.92 ; 0.91	0.59 ; 0.63	0.66 ; 0.46
SACSMA-L	0.70 ; 0.53	0.29; 0.46	067 ; 0.36

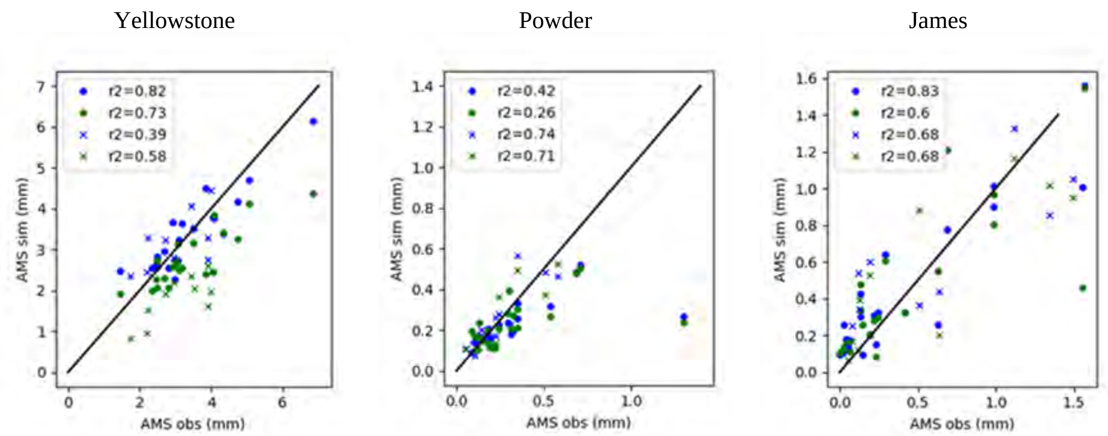


Figure 6-5 : Scatter plots between observed and modeled annual maxima of streamflow (AMS). AMS for years during calibration and validation periods are shown with 'dot' and 'cross' symbols respectively. Colors separate the distributed (blue) and lumped (green) version of SAC-SMA. (Calibration=1980-200; Validation=2001-2010).

Focusing on the Annual Maximum Streamflow (noted AMS; see scatterplots on Figure 6-5); we note a high correlation between observed and simulated AMS, which highlights that SACSMA-L and –D capture rather well the inter-annual variability of annual peak discharges. Beside, we note a bias in simulated AMS values, especially for the largest events, which is the well-known issue for hydrology models (cf. discussion in Chapters 2 and 9, for instance). Bias in SACSMA-L model is usually larger than SACSMA-D bias. Overall, however, the models' performance is satisficing for doing a climate change impact assessment.

6.4.1.2. DANNs

Regarding DANN, in addition to estimating the model parameters, an important part of the calibration relies on the choice of a correct network structure. For the considered DANNs, this translate to choosing the number of hidden layers and the number of neurons (nodes) in each layers. This decision is made by carrying out a trial-and-error procedure that consists at training DANNs for a range of hidden layers and nodes. This procedure is illustrated on Figure 6-6. Note that throughout this procedure, we tested several input combinations among temperature, precipitation, and snow water equivalent and soil moisture variables. The combination that performed best is using monthly precipitation, temperature, soil moisture and snow water equivalent for the month of the peak discharge prediction plus the values for the two months prior the prediction. Note also that all three SWE dataset are considered as inputs.

Figure 6-6 shows the evolution of the NSE criteria during the training process for the James River and for a large range of network structure. Note that only 50 epochs are represented for the sake of clarity but 250 epochs have actually been used. As illustrated on Figure 6-6, we note that a simple network (e.g., one hidden layer and 10 nodes by layer) is not capable of prediction the peak discharges of the James River. However, the training procedure for more complex networks (i.e., with larger number of hidden layers or larger number of nodes) is successful as the NSE values increase with the number of epochs, for both calibration and validation set, highlighting the fact that a more complex network structure allows to use advanced knowledge of the potential relationship among inputs and peak discharges. Similar to what is commonly obtained for hydrology models; NSE values for calibration periods are higher than for the validation period. The network parameters and network structure that are eventually retained for the analysis are the one for which the maximum NSE value for the validation period is obtained (red curves on Figure 6-6). NSE for each catchment and for calibration and validation periods are given Table 6-4. Similar to SACSMA, the performance decreases significantly for the validation periods when compared to the calibration period.

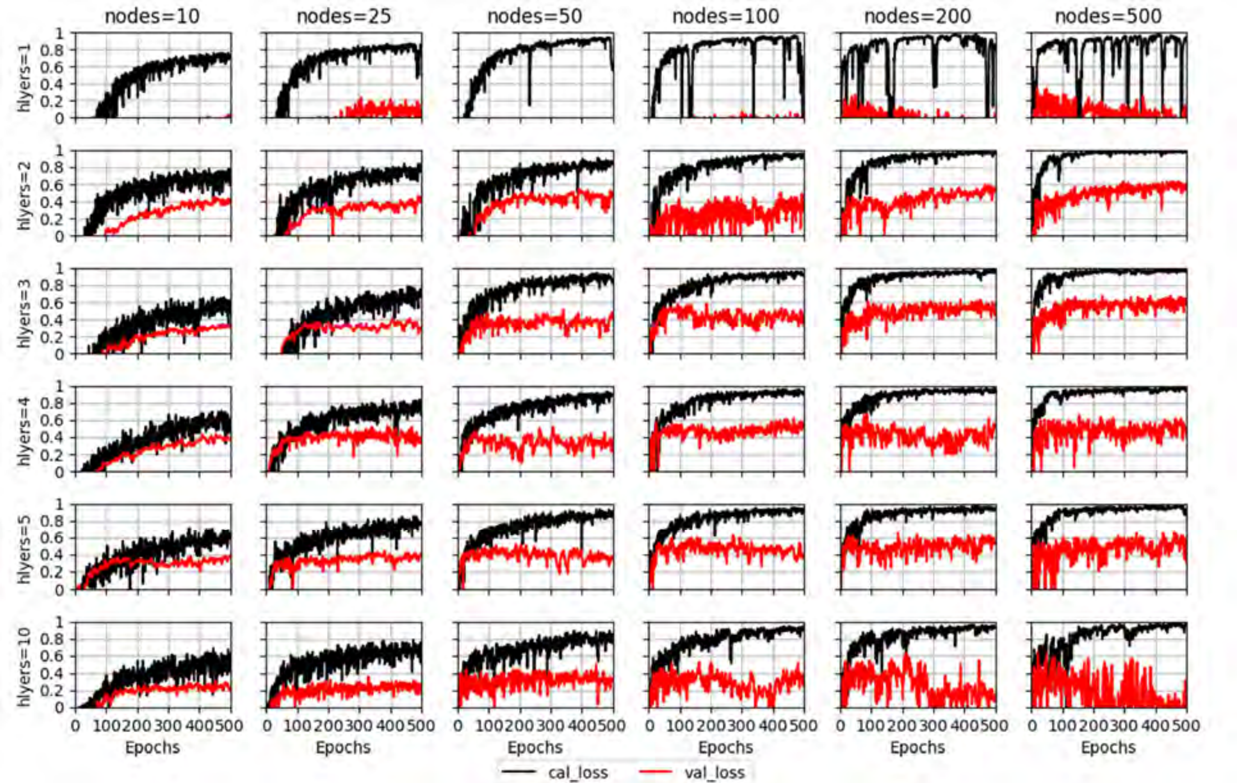


Figure 6-6 Example of trial-and-error board that illustrates the decision process of the DANN structure. Columns show different numbers of nodes (neurons) within each hidden layers (hlyers) while each row shows different numbers of hidden layers. Black and red curves show the evolution of the NSE criteria with the number of evaluation of the loss function (epochs). The current board illustrates the case of the James River Basin. (NB: some NSE values for the top-left panels are negative and, as such; do not appear on the figure). Note that overfitting can be observed when NSE keeps increasing for the calibration period but starts decreasing for the validation period (e.g., hlyers=10 and nodes=200 or 500).

Table 6-4 Evaluation of of the DANNs over the three considered catchments. NSE stands for Nash-Sutcliffe Efficiency. NSE values are given for calibration (cal) and validation (val) periods and for the maximum daily streamflow values at a monthly time scale. The chosen network structure (i.e., number of hidden layers and number of nodes within each hidden layers) is also given.

	#hidden layers	#nodes	NSE (Cal)	NSE (Val)
Yellowstone R	5	500	0.97	0.86
James R	2	500	0.95	0.63
Power R	5	50	0.75	0.61

The relative contribution of each DANNs' inputs to the predicted peak flow values can be assessed by means of the Profile method (Lek, 1996; Shojaeefard et al. 2013). The Profile method is a sensitivity analysis that allows the assessment of the contribution of each input variable to the predicted value. The contribution profile illustrated on Figure 6-7 is the outcome of the method. The profile shows the relationship between the predicted peak flow values and each individual input, while the other input are hold at constant values (e.g., min, 10th percentile, 20th percentiles... 90th percentiles and max). For instance, Figure 6-7 shows for Yellowstone River Basin that largest values of peak discharges follow from large precipitation for the month of the prediction (bold blue curve); the second largest contributor being the SWE value two month prior the prediction (dotted black curve). For the James River Basin, the largest contributor to large peak flow value is the SWE two month prior the prediction, and then come SWE for the month of the prediction of the month prior the prediction. Fourth contributor is the precipitation during the month of the prediction. For Powder River, precipitation and SWE variables are also the main contributors to the predicted peak flow values (not shown).

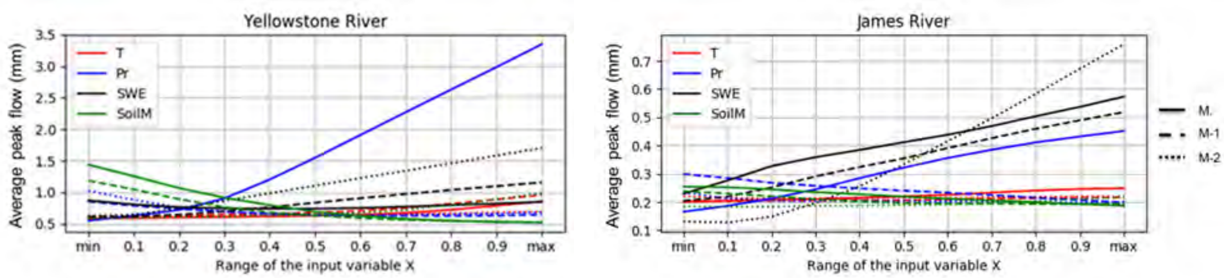


Figure 6-7 Contribution profile of each independent variable to the prediction of the peak flow discharge. Contribution of Temperature (red), Precipitation (blue) SWE (black) and Soil Moisture (green) are given for a range of values ranging from the minimum to the maximum observed in the historic period. Bold, dashed and dotted curves show the contribution of the input variable X for the month of the prediction (bold), the month prior the prediction (dashed) and two months prior the prediction (dotted). For a given input variable, the value on the y-axis is the average peak flow value obtained when holding the other input variables all equal and subsequently equal to the historical minimum values, the 10th percentile, 20th percentiles... 90th percentiles and its historic maximum.

6.4.2. Change in Peak Flow Discharge

This section describes the assessment of changes in peak discharges for the considered catchments. This assessment is done following the Decision Scaling approach (Brown et al. 2016). The focus is here on the peak discharge that has a return period of 100 years, mainly because its importance in infrastructure design. We also focus on the 2-yr return event to show the potential evolution of common flow peaks.

6.4.2.1. SACSMA

This section describes the potential evolution of the peak discharges as obtained from SACSMA for the Yellowstone River basin. Similar results are obtained for the Janes River Basin and the Powder River Basin (not shown). Following the Decision Scaling approach, a climate stress test has been considered as historical precipitation and temperature time series have been perturbed following a delta change approach. Considered changes in precipitation ranges from -30% to +30% and temperature changes range from 0 to +6C.

For both SACSMA-L and SACSMA-D model, the sensitivity map illustrated on Figure 6-8 show that an increase in precipitation leads to higher peak discharges. The relative increase in peak discharges is larger for small return period (2-yr) than for extreme peak discharge (100-yr), which suggests that precipitation alone is not the main driver of extreme peak flow across this catchment. We also note that increasing precipitation leads to a larger increase in peak flow for SACSMA-L than for SACSMA-D, which suggests the precipitation variability matters.

The effect of temperature rise is, however, significantly different when comparing the results obtained from SACSMA-L and SACSMA-D. First, we notice that a small increase in temperature (with no change in precipitation) leads to a slight decrease in bog floods (100-yr) for SACSMA-L while the big floods decrease when considering SACSMA-D. This suggests that temperature-related processes (e.g., snowpack dynamic) is sensitive to spatial variability, and especially sensitive to the range in elevation that is not accounted for in the lumped version SASSMA-L. This assumption is confirmed when considering a combined increase in precipitation and temperature (i.e., top-right corner of the response surfaces on Figure 6-8). SACSMA-D indeed simulates a decrease in both common (2-yr) and extreme (100-yr) floods in this configuration while SACSMA-D simulates an increase (i.e., the decrease in peak flow due to higher temperature is offset by the increase in precipitation).

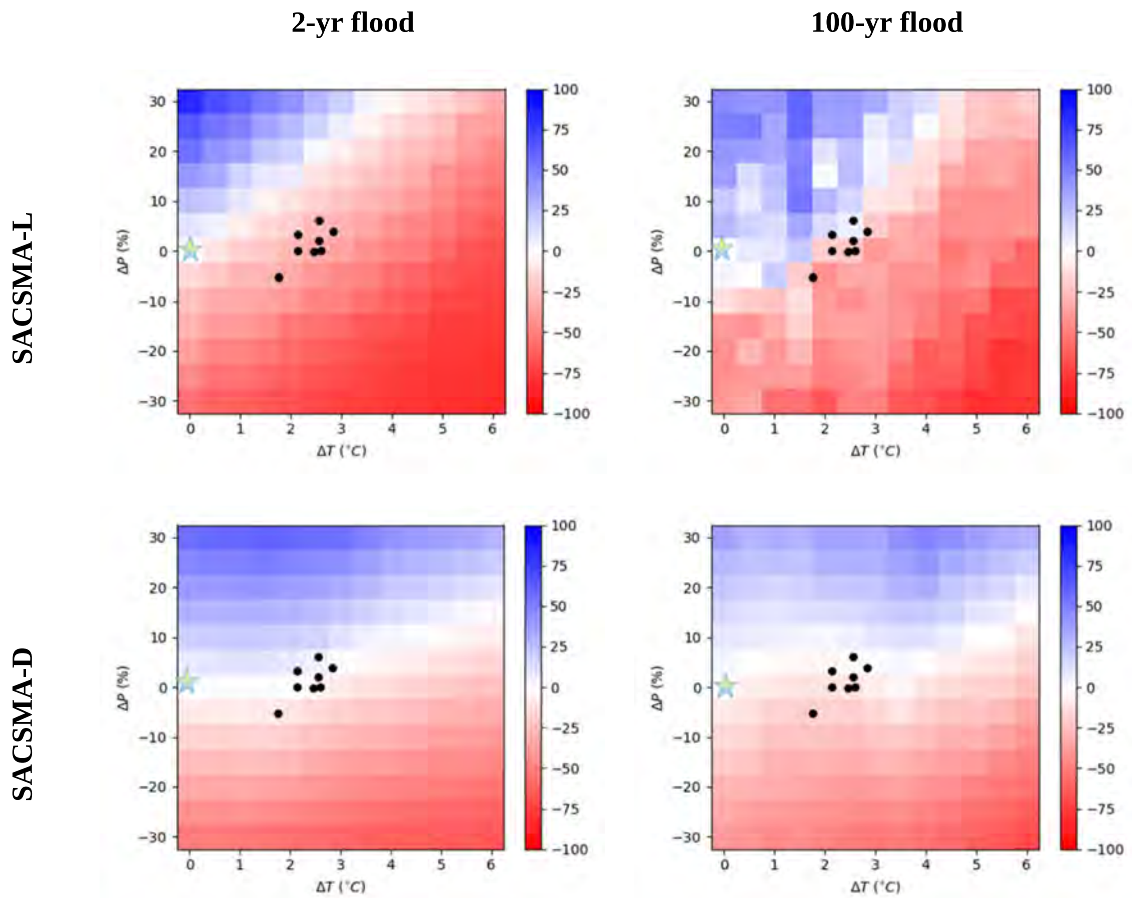


Figure 6-8 Climate Response Surface showing the sensitivity of peak discharge to change in precipitation (ΔP,%) and temperature (ΔT,C). The heat maps show the relative change (%) in 2-yr flood (left column) and 100-yr flood (left column). The 'star' shows the no change scenario (i.e., no change in precipitation and no change in temperature). The black dots show the projected changes in temperature and precipitation as obtained from the considered NARCCAP projections (Table 6-2). Results are shown for the Yellowstone River Basin.

6.4.2.2. DANNs

This section describes the stress test results obtained with the DANNs. Contrary to the stress test performed for SACSMA in the previous section, here the stress test focuses on the sensitivity to change in precipitation and SWE variables since they are revealed as the primary divers of peak discharges for the considered catchments (cf. section 6.4.1.2, Figure 6-6). Note that in this case, the Δ change factor (%) applied to either P or SWE variables is the same for all months (i.e., for the month of the peak discharge prediction and for the two months prior the prediction). The climate response functions on Figure 6-9 highlight different sensitivity for the different case studies.

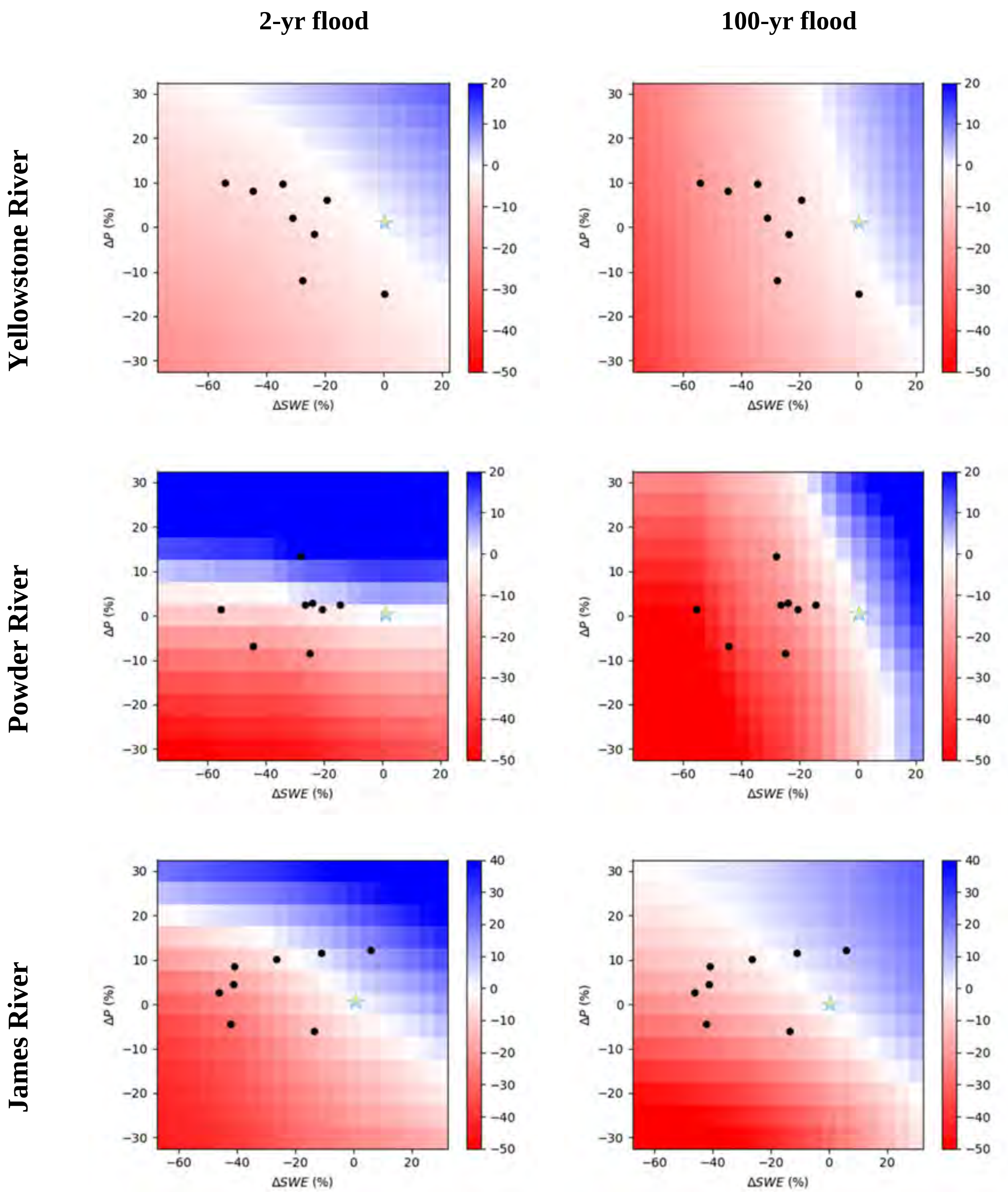


Figure 6-9 Climate Response Surface showing the sensitivity of peak discharge to change in snow (ΔSWE,%) and precpitation (ΔT,%). The heat maps show the relative change (%) in 2-yr flood (left column) and 100-yr flood (left column). The 'star' shows the no change scenario (i.e., no change in snow water equivalent and no change in precipitation). The black dots show the projected changes in temperature and precipitation as obtained from the considered NARCCAP projections (Table 6-2). Contrary to Figure 6-8, changes in SWE and P are for the three months prior the flood occur.

For common floods (i.e., 2-yr return period, left column on Figure 6-9), we note that an increase in SWE is likely to increase the peak intensity. This increase can, however, be offset by a significant decrease in precipitation for James and Yellowstone River basins while even a small decrease in precipitation will cancel the increase in flooding for the Power River. On the other hand, a decrease in snowpack (SWE) during the flooding season can only be offset by a large increase in precipitation at this period for the James River (i.e., from +20 to +30% precipitation depending on the decrease in snow). For Yellowstone, we note that even an increase by 30% of the precipitation during the flooding season does compensate with a decrease in SWE. For the Power River basin, results show that any precipitation increase would probably make up for any decrease in snow (at least for a decrease down to 60%). These results highlight different mechanisms across the three catchments. For Powder River, change in the common floods appear to be driven by precipitation changes only while for James and Yellowstone River basins change in common peak flow seem to be resulting from a combination of change in SWE and P.

For the large floods (i.e., 100-yr return period, right column on Figure 6-9), the mechanisms appear to be slightly different depending on the basin. For instance, we note that an increase in precipitation across either James or Yellowstone catchment will hardly offset a decrease in SWE, contrary to what has been observed for the 2-yr flood. This highlights that a significant snow pack during the flooding season is required to trigger a large flood event. This is also true for the Power River basin for which we note that even a decrease in precipitation by 30% during the flooding season does not cancel the increase in peak flow that would result from a larger snow pack across the basin.

6.4.2.3. CLIMATE PROJETIONS

Projected changes in precipitation, temperature and SWE are shown on the climate response function on Figure 6-8 and Figure 6-9 with black dot symbols. These can be used to infer potential changes in peak flow. This is illustrated on Figure 6-10 for the Yellowstone River. We note that potential changes in flood as obtained for DANN and SACSMA-D are close when compared with SACSMA-L. For DANNs and SACSMA-D, future changes in common flood (2-yr flood) range from -10% to +5% of the historic intensity and for the larger flood events (100-yr) from -20% to no change. However, inference of likelihood of changes using climate information from only eight climate projections (i.e., eight combinations of GCM and RCM, Table 6-2) is likely not robust and the use of additional projections would be valuable.

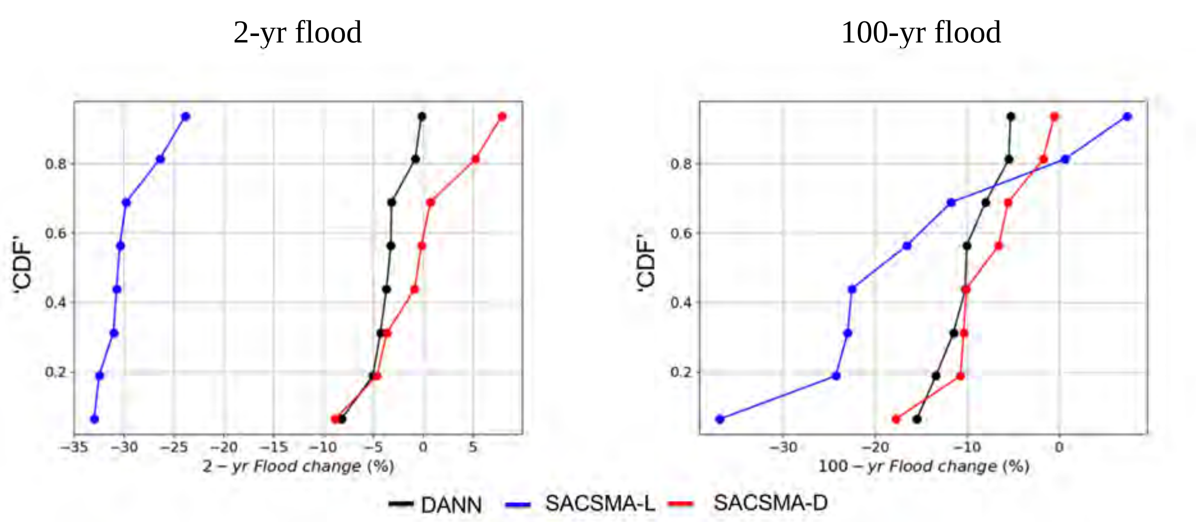


Figure 6-10 Likelihood of change in peak discharges as obtained by combining the climate response function (Figure 6-8 and Figure 6-9) with the NARCCAP climate projections.

6.5. CONCLUSION

The current study has compared three different approaches to predict the sensitivity of peak discharges to change in temperature, precipitation, soil moisture and snow water equivalent via a modeling comparison approach. A lumped and a distributed version of SACSMA model and artificial neural networks were used to predict change in peak flows using the Decision Scaling approach.

The results of the model comparison has allowed discussing different behaviors for three catchments in the dry and cold Upper Missouri basin. Results show that the contribution to peak flow intensity from the different hydro-meteorological drivers vary across the basin. Results also show that for a given scenario of change, for instance a given change in precipitation and snow cover, the response in terms of common flood or large flood events can be significantly different, highlighting different mechanisms for these events and potentially different evolutions.

Future research should extend the number of case studies; explore more the uncertainty stemming from the model structure (for both DANNs and for hydrology models, for which only one model has been considered, although a lumped and a distributed version were used). Also, one important limitation for inferring changes in peak flows from the NARCCAP experiment is the limited number of projections. As such, future research should also include other projections such as the CMIP5 experiment projections and projections for CORDEX Africa.

6.6. REFERENCES

Anderson, E., 2006. Snow Accumulation and Ablation Model - SNOW-17. NOAA.

Anderson, M.G., McDonnell, J.J. (Eds.), 2005. Sacramento Soil Moisture Accounting Model (SAC-SMA), in: Encyclopedia of Hydrological Sciences. John Wiley & Sons, Ltd, Chichester, UK.

Andréassian, V., Oddos, A., Michel, C., Anctil, F., Perrin, C., Loumagne, C., 2004. Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall-runoff models: A theoretical study using chimera watersheds: CHIMERA WATERSHEDS. Water Resour. Res. 40. https://doi.org/10.1029/2003WR002854

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T., Vitart, F., 2015. ERA-Interim/Land: a global land surface reanalysis data set. Hydrol. Earth Syst. Sci. 19, 389–407. https://doi.org/10.5194/hess-19-389-2015

Brown, C.M., Steinschneider, S., Wi, S., Weatherly, J., Case, M., Hayden, T., Koster, A., Mearns, L.O., Bukovsky, M.S., McCrary, R., 2016. Climate Risk Assessment: Technical Guidance Manual for DoD Installations and Built Environment (Technical Report No. AD1029408). University of Massachusetts Amherst United States.

Brown, R.D., Brasnett, B., Robinson, D., 2003. Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. Atmosphere-Ocean 41, 1–14. https://doi.org/10.3137/ao.410101

Fan, Y., van den Dool, H., 2004. Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present. J. Geophys. Res. 109. https://doi.org/10.1029/2003JD004345

Hosking, J.R.M., 1990. L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics. J. R. Stat. Soc. Ser. B 105–124.

Jenkinson, A. F. (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society, 81(348), 158–171.

Kingma, D.P., Lei Ba, J., 2015. ADAM: a method for stochastic optimization. Presented at the 3rd International Conference for Learning Representations, San Diego, p. 15. https://doi.org/arXiv:1412.6980

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Bottou, L., Orr, G.B., Müller, K.-R., 1998. Efficient BackProp, in: Neural Networks: Tricks of the Trade. Springer, p. 44pp.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90, 39–52.

Livneh, B., Bohn, T.J., Pierce, D.W., Munoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D.R., Brekke, L., 2015. A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. Sci. Data 2, 150042. https://doi.org/10.1038/sdata.2015.42

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., Loumagne, C., 2014. When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. Hydrol. Earth Syst. Sci. 18, 575–594. https://doi.org/10.5194/hess-18-575-2014

Lohmann, D., Raschke, E., Nijssen, B., Lettenmaier, D.P., 1998. Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. Hydrol. Sci. J. 43, 131–141. https://doi.org/10.1080/02626669809492107

McCrary, R.R., McGinnis, S., Mearns, L.O., 2017. Evaluation of Snow Water Equivalent in NARCCAP Simulations, Including Measures of Observational Uncertainty. J. Hydrometeorol. 18, 2425–2452. https://doi.org/10.1175/JHM-D-16-0264.1

Mearns, L.O., Arritt, R., Biner, S., Bukovsky, M.S., McGinnis, S., Sain, S., Caya, D., Correia Jr, J., Flory, D., Gutowski, W., others, 2012. The North American regional climate change assessment program: overview of phase I results. Bull. Am. Meteorol. Soc. 93, 1337–1362.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Shojaeefard, M.H., Akbari, M., Tahani, M., Farhani, F., 2013. Sensitivity Analysis of the Artificial Neural Network Outputs in Friction Stir Lap Joining of Aluminum to Brass. Adv. Mater. Sci. Eng. 2013, 1–7. https://doi.org/10.1155/2013/574914

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Wang, Q. J. (1991) The Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-Runoff Models. Water Resources Research, 27(9), 2467–2471.

7. REGIONAL CLIMATE MODEL PRECIPITATION CREDIBILITY IN THE EAST-SOUTH-CENTRAL UNITED STATES

7.1. EXECUTIVE SUMMARY

Midcentury projections from climate model ensembles for the East-South-Central (ESC) U.S. indicate an increase in mean precipitation during the cool-season (November-March), particularly in the north half of the region, over the Ohio-Tennessee River Basin. However, given the pervasiveness of a dry bias in mean precipitation in many model simulations of this region, the credibility of these projections, despite general model agreement, requires further assessment. The bias in mean precipitation points to biases in the underlying processes behind the precipitation and may affect the credibility of projections in not only mean precipitation but also extreme precipitation.

The goal of this study, therefore, is to examine the drivers of precipitation in the ESC during the cool season to better understand, at a process-level, the causes of the bias in regional climate models (RCMs). Particular attention is paid to reanalysis-driven simulations from the North American Regional Climate Change Assessment Program (NARCCAP) to better understand the bias relative to the RCMs only, without inherited bias from Global Climate Models (GCMs), although some of the most important GCM-driven biases are also detailed.

The ESC precipitation bias is found to have several sources. These are related to moisture flux into the region, transient, synoptic-scale low-pressure systems, Gulf of Mexico and Caribbean SSTs, and ENSO-related teleconnections. Until additional analysis is completed to examine the evolution of these factors in the future simulations, we would recommend caution using the projections from the NARCCAP simulations in the ESC region, given their now known problems in simulating the processes that drive the baseline precipitation climate in this region.

7.2. TECHNICICAL APPROACH

The purpose of this study is to assess the dynamical, process-level credibility of the precipitation from the regional climate model (RCM) simulations produced for the North American Regional Climate Change Assessment Program (NARCCAP) in the context of CMIP3 and CMIP5, for cool-season (November-March) precipitation over the Ohio-Tennessee River Basin (OTRB). The OTRB is situated just north of and within a cool-season maximum in precipitation that is located over the Deep South. Therefore, examining the credibility of the simulations for the OTRB requires us to examine model simulations over the basin and upstream (from an atmospheric perspective) of that basin. Therefore, our analysis focuses on a region we are calling the East-South-Central U.S. (hereafter the ESC), named after the census region centered in the region, and its surroundings. The ESC generally encompasses most of the OTRB and the region to its south to the Gulf of Mexico, including part of the lower-Mississippi River Basin. Our core ESC analysis region is outlined in Figure 7-1.

The ESC region is particularly interesting because it has not received much attention from the climate analysis community, and models, even those running high resolution convection permitting simulations (Liu 2017, Mearns et al. 2012), have great difficulty reproducing this cool-season maximum in precipitation. As this maximum in precipitation is strongly, dynamically forced by transient low pressure systems [extratropical cyclones (ETCs) and inverted troughs (which for the purposes of this study we will combine with the ETCs)], and the dominant moisture source for precipitation is adjacent to the region (i.e., the Gulf of Mexico), it is puzzling as to why models have a difficult time reproducing it. As the problem exists across a variety of resolutions, a lack of resolution does not appear to be the main driver (at least in RCMs and at resolutions at or below 50km; that is, we are not including coarser resolution global climate models (GCMs) in this statement). The dry bias is known to exist in both simulations that are driven with reanalyses and with GCMs; therefore, the bias is not solely caused by biased GCM boundary conditions, although they can exacerbate the problem.

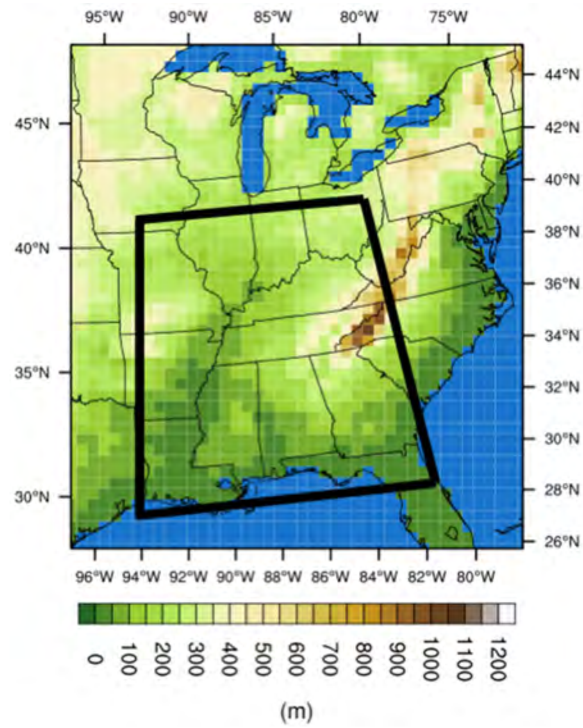


Figure 7-1 Surface elevation (m) from HRM3, with the core of the East-South-Central analysis region for this study outlined in black.

Given the pervasiveness of this problem in mean precipitation, it warrants further understanding before we can determine the credibility of model simulations for the future or for other moments of the precipitation distribution (e.g., extremes). Our goal, therefore, is to assess the causes of the low precipitation bias in the ESC during the cool-season in a variety of regional models. Herein, we will mostly focus on the potential drivers of bias in reanalysis-driven simulations, but also detail important factors contributing to bias in some of the GCM-driven simulations. Our analysis includes an evaluation of factors that both affect moisture and force precipitation; that is, moisture flux, regional sea-surface temperatures, ETCs, and global sea-surface temperature feedbacks.

7.3. DATA AND METHODS

7.3.1. REANALYSES AND OBSERVATION-BASED DATASETS

- North American Regional Reanalysis (NARR; Mesinger et al. 2006): 32-km horizontal resolution, 3-hour temporal resolution, and data from 1980 onward. Unlike most other reanalyses (including the next two listed below), precipitation is, in essence, assimilated where observations are available, making this reanalysis particularly useful for verification of the full atmosphere over the contiguous U.S. (CONUS; problems do exist elsewhere, particularly over oceans and outside of CONUS in precipitation, as noted in Bukovsky and Karoly 2007, Ruane 2010a, Ruane 2010b).

- National Center for Environmental Prediction (NCEP)/ Department of Energy (DOE) Reanalysis II (hereafter NCEP: Kanamitsu et al. 2002): a global, T62 resolution (210-km resolution at the Equator) reanalysis used as boundary conditions in the NARCCAP reanalysis-driven simulations.

- The European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim (hereafter ERA-Int; Dee et al. 2011): a global, approximately 80-km (T255 spectral) resolution reanalysis.

- Livneh daily CONUS near-surface gridded meteorological and derived hydrometeorological data set (Livneh; Livneh et al. 2013): a gridded observation-based dataset at a 1/16-degree resolution that spans 1915 to 2011. Used for comparison with NARR, but also for years not included in the NARR dataset (i.e., any time pre-1980).

- The University of Delaware ½ degree resolution, monthly mean, gridded, observation-based dataset (UDEL; Willmott and Matsuura 1995)

- The Hadley Centre Global Sea Ice and Sea Surface Temperature dataset (HadISST; Rayner et al. 2003) provides 1-degree resolution, monthly mean, global fields for SST and sea-ice derived from observations (both in-situ and satellite-derived post 1982).

7.3.2. NARCCAP

Six RCMs from NARCCAP are used for this analysis (Mearns et al. 2007). They are the:

- Canadian RCM (CRCM[1]; Caya and Laprise 1999)
- Experimental Climate Prediction Center's version of the Regional Spectral Model (ECP2; Juang et al. 1997)
- Third-generation Hadley Centre RCM (HRM3; Jones et al. 2003)
- Fifth-generation Pennsylvania State University - National Center for Atmospheric Research Mesoscale Model as run by the Iowa State University modeling team (MM5I; Grell et al. 1993)
- International Centre for Theoretical Physics RCM version 3 (RCM3; Giorgi et al. 1993a, Giorgi et al. 1993b, Pal and Coauthors 2007)
- Weather Research and Forecasting model as run by the Pacific Northwest National Lab (PNNL) modeling team (WRFG; Skamarock et al. 2005)

All RCMs have a horizontal resolution of 50 km. The CRCM and ECP2 are the only two models that include some form of interior nudging, a 'relaxation' toward the large-scale driving conditions within the interior of the domain, which constrains these regional models to more closely follow their driving reanalysis or GCM. The impact of the nudging is illustrated in Mearns et al. (2012), where these two models more closely reproduce observed temperature and precipitation variability and average magnitude in the regions selected for analysis when forced with a reanalysis product. Further details on NARCCAP and the configuration of the models can be found in Mearns et al. (2012), and online at www.narccap.ucar.edu.

We focus mainly on the set of simulations from NARCCAP that acquire their initial and boundary conditions from the NCEP reanalysis. These simulations are for the period of 1980-2004; however, most of the included analysis covers 1980-2003, as data for all variables is not available for 2004 from all of the

---

[1] All RCM acronyms are as used in the NARCCAP model archive.

simulations.

We also use the GCM-driven simulations from NARCCAP for some analyses. This set is comprised of simulations forced by four GCMs for the period of 1971-2000 (or 1999, depending on the GCM). Three out of the four are simulations that are included in the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) archive. The GCMs, their horizontal resolution, and their ensemble member number, where applicable, are:

- NCAR CCSM version 3.0 (CCSM; Collins et al. 2006): T85 (1.4° x 1.4°), run 5.
- Canadian Global Climate Model version 3 (CGCM; Flato et al. 2000): T47 (1.9° x 1.9°), run 4.
- GFDL Climate Model version 2.0 (GFDL; Anderson et al. 2004): 2.0° x 2.5°, run 2.
- Hadley Centre Climate Model version 3Q0 (HadCM3Q0, hereafter HADCM; Gordon et al. 2000, Pope et al. 2000): 2.5° x 3.75°, not included in the CMIP3 archive. This version differs from the version used to perform the CMIP3 simulations in that it includes a flux adjustment and an aerosol cycle.

Twelve of the possible 24 RCM+GCM pairings were completed as part of NARCCAP. Each RCM uses boundary conditions from two different GCMs and each GCM drives 3 RCMS, with combinations chosen using a balanced fractional factorial design. When in combination, an RCM and its driver will be referred to as, for example, RCM3-gfdl, with the forcing simulation or reanalysis in lower case. When not in combination, all acronyms will be in standard upper case. Future projections are based on the Special Report on Emissions Scenarios (SRES; Nakicenovic et al. 2000) A2 scenario.

7.3.3. Global Model Ensembles

For comparison purposes, we also include projections of precipitation from simulations produced for the Coupled Model Intercomparison Projects version 3 and 5 (CMIP3 and CMIP5). Single realizations from 17 CMIP3 simulations and 35 CMIP5 simulations are used, and the models are listed in the Supplementary Material in Bukovsky et al. (2017). For the future projections, the CMIP3 simulations used are based on the SRES A2 scenario and the CMIP5 simulations are based on representative concentration pathway 8.5 (RCP8.5; Moss et al. 2008). For the purpose of computing ensemble calculations only, all CMIP3 simulations were interpolated to a common 2x2 degree grid and all CMIP5 to a 1x1 degree grid.

7.3.4. Extratropical Cyclone Activity

Extratropical storm locations are calculated using an approach that builds off of Wallace et al. (1988) and Chang et al. (2012 and 2016). To differentiate pressure differences due to transient low-pressure systems with synoptic time-scale variability from local, short-term variances in near-surface pressure, a modified 24-hour variance filter is applied to a three-to-six hourly sea-level pressure field. That is:

$$\begin{aligned}\{psl(t+24) - psl(t) > 0 \rightarrow Pvar = [psl(t+24) - psl(t)]^2\}\\ \wedge \{psl(t+24) - psl(t) \leq 0 \rightarrow Pvar = 0\}\,,\end{aligned} \tag{7.1}$$

where *psl* is the sea-level pressure at a given time, *t*, in hours. The resulting *Pvar* field highlights locations of approaching and passing synoptic-scale pressure systems. To not double-count systems (on approach *and* retreat), and to only count cyclones and not anti-cyclones, the conditional where the 24-hour-difference field must be positive was added to the original methodology of Chang et al. (2016). The average of the *Pvar* field with time has been shown to correspond to average cyclone activity from other tracking systems (Chang et al. 2012). While we could use the average *Pvar* field to examine average cyclone activity, examining the rainfall associated with each system requires tracking each system. Therefore, we used the

Pvar field to track the storms by locating the storm centers. Local maxima were identified in the Pvar field, and filtered. The local maxima were first filtered to include maxima with *Pvar* > 50 hPa^2 only (thus, the systems with a 24-hr pressure drop of at least ~7 hPa). The remaining maxima were then filtered to find the location of the largest maxima no less than 2000-km from the next largest maxima. As this method does not require closed pressure contours, it does count systems, like inverted troughs, that reach these specified thresholds. As these other systems are also important in forcing precipitation in this study region though, this is considered to be a benefit of this approach. Also, in the terminology used herein, "ETC" is meant to encompass all identified transient, synoptic-scale low-pressure systems. In examining the location of ETC that may be associated with precipitation in study area outlined in Figure 7-1, only ETC within 2000km of the box center were considered.

Figure 7-2 demonstrates the ETC tracking system over a period of two days in NARR during a heavy precipitation event in the ESC focus region. Precipitation is shown in the color contours, the sea-level pressure field is contoured in black, the center of the low-pressure system as identified by our tracking technique is marked with a large, red "L", and grey shading shows where the *Pvar* field is greater than 50 hPa^2. The ETC that is directly affecting the ECS region is first identified by this method over southern Texas in the second panel of Figure 7-2, and propagates directly over the region over the course of the 2 days shown.

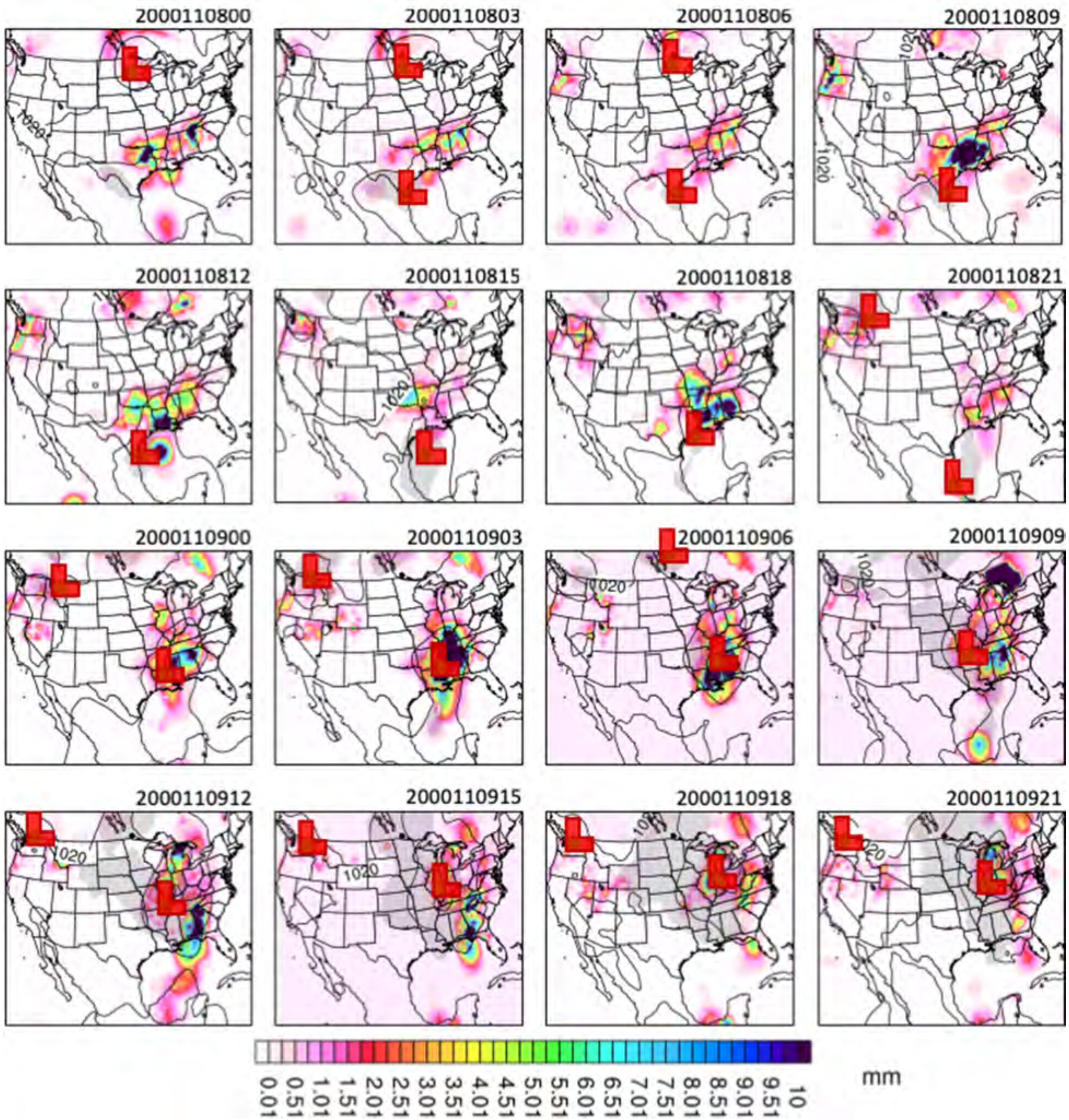


Figure 7-2 Two day evolution of ETCs in NARR from 8 Nov. 2000 at 00UTC to 9 Nov. 2000 at 21UTC (time indicated in upper right corner of each panel in YYYYMMDDHH format). ETC center are marked with a bold, red "L". Sea-level pressure contoured in black every 10 hPa. Areas of Pvar greater than or equal to 50 hPa^2 shaded in light grey. Three hour total precipitation in mm (color fill) also shown.

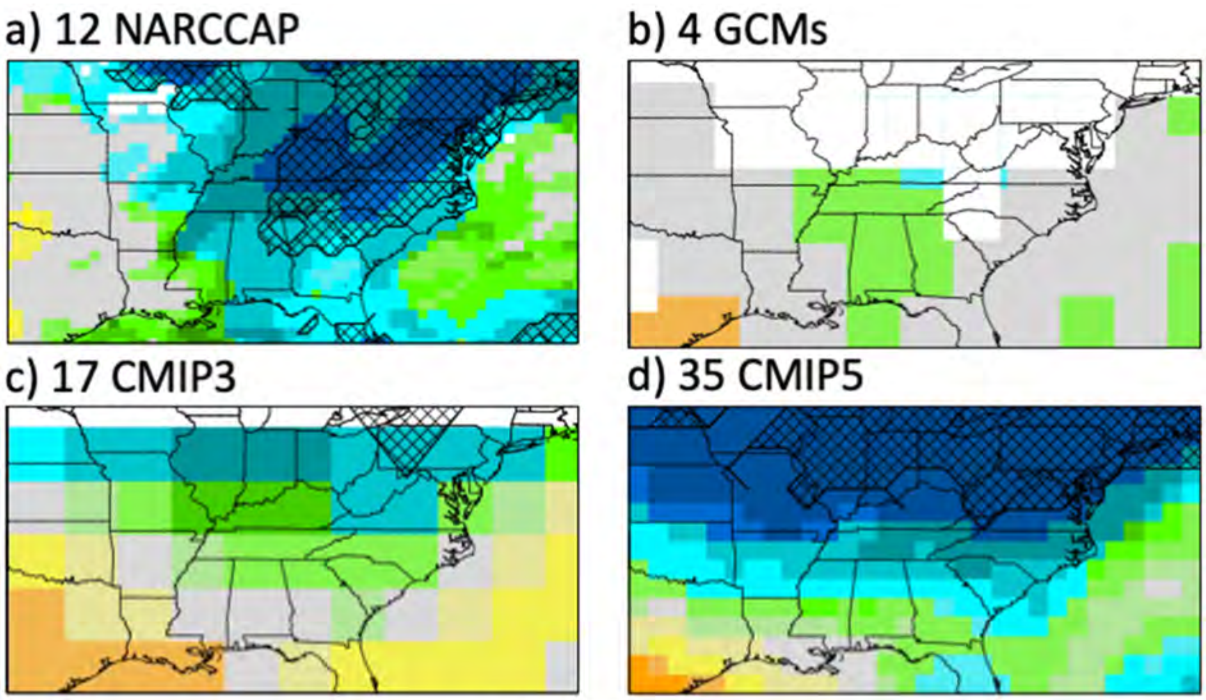
7.3.5. Significance Testing and Agreement

Methodology for statistical significance testing of the climate changes, and model ensemble agreement follow that described in Bukovsky et al. (2017), in their section 2.d.1. In brief, statistical significance is tested at the 0.1 level using bootstrapping with bias correction and acceleration unless otherwise noted.

Ensemble mean agreement on climate change is presented with the intensity of the color scale adjusted based on the level of percent agreement on the sign of change in the ensemble. The percent of agreement is scaled by the likelihood of agreement.

7.4. RESULTS AND DISCUSSION

7.4.1. PRECIPITATION PROJECTIONS



	# of Simulations			
% Agreement	12 NARCCAP	4 GCM	17 CMIP3	35 CMIP5
90-100	11-12	4	16-17	32-35
80-90	10		14-15	29-31
70-80	9		13	25-28
60-70	8	3	11-12	22-24
50-60			10	20-21
<50	0-7	0-2	0-9	0-19

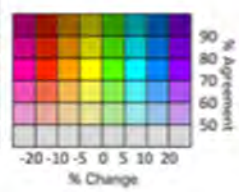


Figure 7-3 November - March mean precipitation change (%) from the baseline (1971-2000) to the future (2041-2071) period for a) the 12-simulation NARCCAP ensemble, b) the 4 GCMs used to force the NARCCAP suite, c) 17 CMIP3 GCMs, and d) 35 CMIP5 GCMs. Precipitation change is presented following methodology proposed by Tebaldi et al. (2011), with some modification: hatching indicates where more than 50% of the simulations show change that is significant at the 0.10 level (as determined by a Student's t test) and where more than 75% of the simulations agree on the sign of change (thus, where the majority agree on significance and sign). White grid cells indicate where more than 50% of the simulations show change that is significant but also where 75% of the simulations or less agree on the sign of the change (thus indicating true disagreement and little information). Additionally, the percent of simulations that agree on the sign of the change is indicated by the color saturation and value (the vertical axis on the color bar). The percent agreement on sign of change is not straight observed percent agreement but is scaled across the ensembles to adjust for differences in likelihood of agreement given differences in ensemble sizes; therefore, a table is also provided to indicate how many simulations are needed in each ensemble for a given level of agreement. To facilitate creating this ensemble average, all models were regridded to common grids of 0.5°x0.5° latitude/longitude for the RCMs, 2°x2° for the CMIP3 GCMs, and 1°x1° for the CMIP5 GCMS.

The cool-season precipitation projections in Figure 7-3 over the ESC motivate our in-depth analysis of mean precipitation in the region. As, without knowing what is driving the precipitation bias we know exists in our baseline simulations, and which will be detailed in the next section, it is difficult to have confidence in these projections.

Despite the mixed appearance of the changes, the NARCCAP, CMIP3, and CMIP5 ensembles generally agree on an increase in mean precipitation over the ESC, but they do not necessarily agree on the significance of that increase or the magnitude. The CMIP5 simulations project a greater increase, with more statistical confidence than the NARCCAP and CMIP3 ensembles, and the NARCCAP ensemble, a stronger and more significant increase than the CMIP3 ensemble. There is no good consensus for change between the four GCMs used to force the NARCCAP simulations, but this is not mirrored in the 12 NARCCAP simulations. There is more certainty on a stronger increase in the northern-half of the ESC region, than in the southern-half, which lies in or near the transition zone between increasing and decreasing future precipitation. The pattern of precipitation change may be due to greater precipitation efficiency in the north and may include a northward expansion of the winter maximum in precipitation, given the greater increase in precipitation in the ESC's northern reaches. It may also be caused by a northward shift in storm track.

7.4.2. Baseline precipitaiton credibility

Figure 7-4 and Figure 7-5 illustrate the baseline cool-season mean precipitation bias that causes us to question the credibility of at least the magnitude and spatial patterns of the projections in section 5.4.1. In Figure 7-4, the RCMs driven by reanalysis, regardless of resolution, are almost all too dry, especially in the southern part of the maximum, occasionally place the maximum too far east (e.g., RCM3), and often extend the maximum too far north, such that there is a wet bias in, for example, Iowa, in some runs.

When driven by GCMs (Figure 7-5), these biases remain fairly consistent to those in the NCEP-driven simulations, but they do change. When driven by the CGCM3, for example, the WRFG has a strong wet bias near the Gulf coast, unlike the other simulations driven by the CGCM3, and unlike when the RCM is forced by NCEP. And in the MM5I, when driven by the HadCM3, the dry bias in the ESC is worse, and the maximum in precipitation appears to have moved to the east coast. These changes in bias suggest additional factors are at play that are inherited from the GCMs, aside from those driving the dry bias in the NCEP-driven simulations. While the biases introduced by the GCMs are consistent across some of the simulations with the same driving GCMs, this is not the case in all sub-sets, with the WRFG-cgcm3 and MM5I-hadcm3 being particularly obvious outliers.

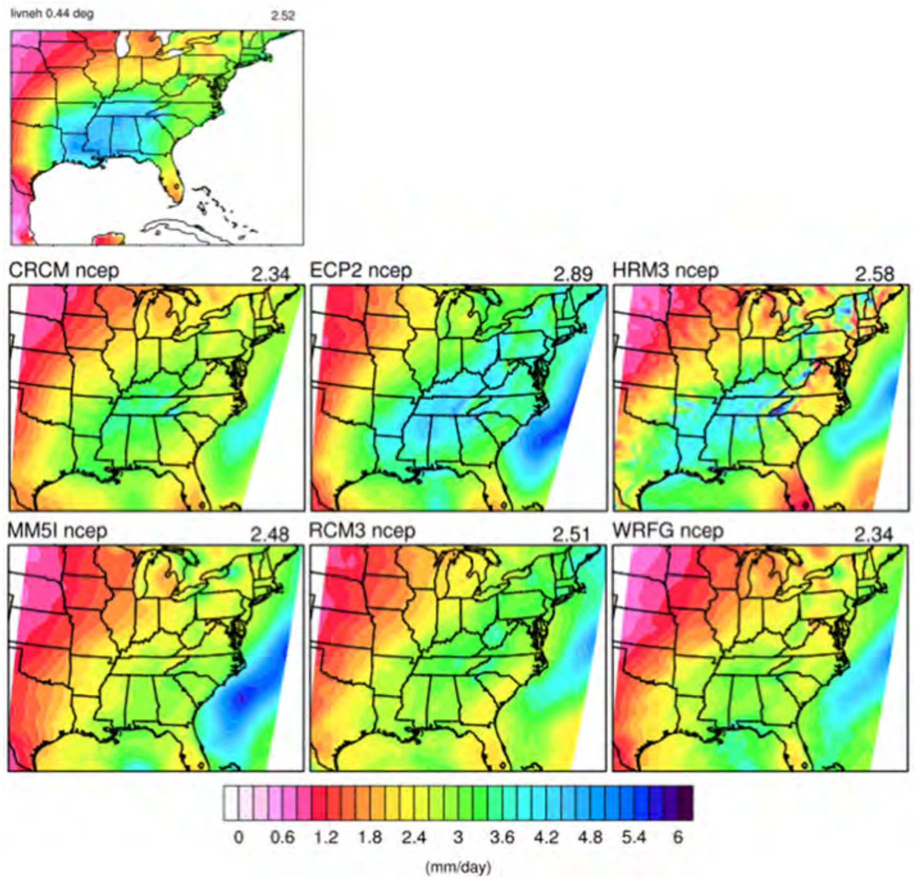


Figure 7-4 1980-2003 November – March monthly mean precipitation (mm/day) from the Livneh gridded, observation-based dataset scaled up to 50km and the 50km NCEP-driven NARCCAP simulations. The spatial average of the precipitation (mm/day) over the region plotted in the panel is given in the upper right corner of each panel.

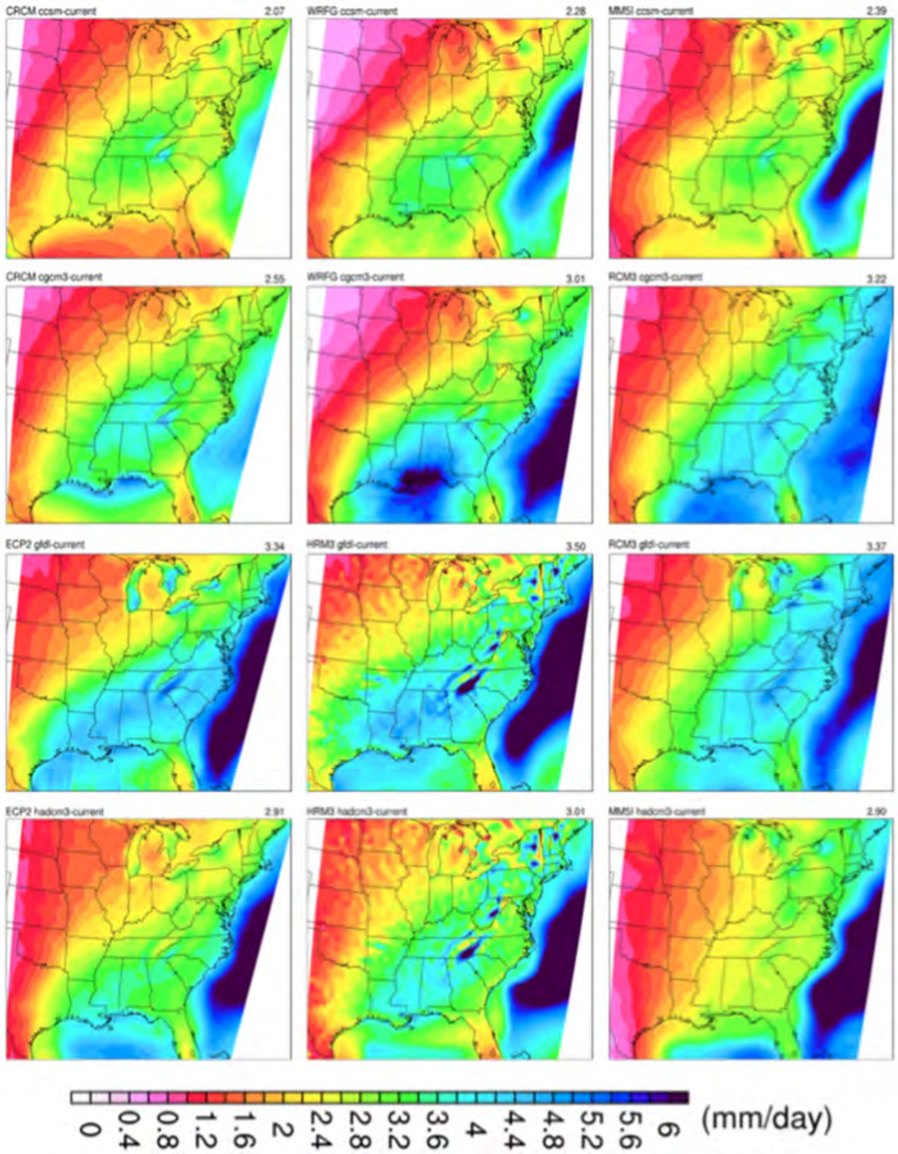


Figure 7-5 As in the previous figure, but for 1971-2000 November – March mean precipitation from the NARCCAP GCM-driven baseline climate simulations.

7.4.3. Physical Mechanisms driving Precipitation Biases

7.4.3.1. MOISTURE

Cool-season mean moisture flux in the NCEP-driven RCMs, compared to NARR, is presented in Figure 7-6. While the flux of moisture from the Gulf of Mexico into the ESC is captured in all simulations, it is more robust in some models than others. Note, for example, the lack of a northward flux component over the Gulf of Mexico just south of the ESC in the MM5I, and a weak and almost non-existent northward flux component over the southern part of the region. This likely explains why this simulation is the driest biased of the set over the ESC. Interestingly, while part of the MM5I bias in moisture flux is due to a weak onshore transient moisture flux component, it is largely driven by a much stronger bias in the stationary component of the flux, which is strongly offshore in the MM5I, with a circulation suggesting a stationary anticyclone near the Big Bend region of Texas, instead of a stationary flux that is parallel to the Gulf Coast in the NARR, south of anticyclonic flow centered over the far southern border between Mississippi and Alabama (not shown). The RCM3 is also lacking a northward flux component over the Gulf of Mexico near the ESC, but has a somewhat stronger northward flux over land than the MM5I. While the RCM3 has a similar bias in its stationary moisture flux component, it is much smaller, and the driver of the bias in total moisture flux is due to a much weaker onshore transient flux component, suggesting too few occurrences of or too weak moisture flux associated with transient eddies (e.g., ETC) (not shown). WRFG has similar bias to RCM3, but the southerly bias just off the Gulf Coast is not as strong, and the flux is more parallel to the coast. There is still a bias in the strength of the northerly flux over the region though, particularly nearer the coast, driven by small biases in both the transient and stationary flux components. The CRCM has the least biased moisture flux of the set, in terms of both magnitude and pattern, followed by the ECP2 and HRM3, which likely explains why these three have less of a dry bias over the ESC than the others. However, it does not explain why they are still dry biased, particularly the CRCM. This also does not explain why these three, like the others, are drier biased near the coast than further inland.

One likely suspect for the latter problem is the representation of SSTs over the Gulf of Mexico near the coast where the driest part of the precipitation bias lies. However, in the RCM simulations, SSTs are a lower boundary condition (LBC) and are nudged toward their driver ever six hours or so, depending on the RCM configuration. Therefore, in the NCEP-driven simulations (using the NCEP-reanalysis observationally-based SSTs as LBC), SSTs that are biased relative to observations do not appear to be a problem (Figure 7-7).

However, there are numerous problems in the SST field in the GCM-driven runs that are likely contributing to some of the differences seen between the precipitation fields from the NCEP-driven versus GCM-driven simulations (Figure 7-8). Many of the SST problems shown in Figure 7-8 are firstly inherited from the GCMs, and secondly caused by poor interpolation of the coarse GCM skin-temperature field onto the RCM grid. The HadCM3 coastline, for instance, is clearly showing up over water in the 50km RCM simulations, and the grid box that represents Florida in the CGCM3, appears as cold-biased water in the WRF-cgcm3. The CGCM3 has warmer water near the coast than the observations, and that is also reflected in the CRCM and WRFG simulations driven by the CGCM3. This may be responsible for some of the wet bias in this WRFG simulation, as it might be responding to the warmer water and additional likely atmospheric moisture with more precipitation. There is also a notable cold-bias in SSTs over the Caribbean in the GCMs in Figure 7-8. While this is not directly mimicked in most of the RCMs, as their domains do not extend that far south, this problem would be inherited in the GCM-provided atmospheric boundary conditions as a low atmospheric moisture bias. This cold SST bias is strongest in the CCSM, and this is likely contributing to the RCMs forced by the CCSM having the strongest dry bias (Figure 7-5).

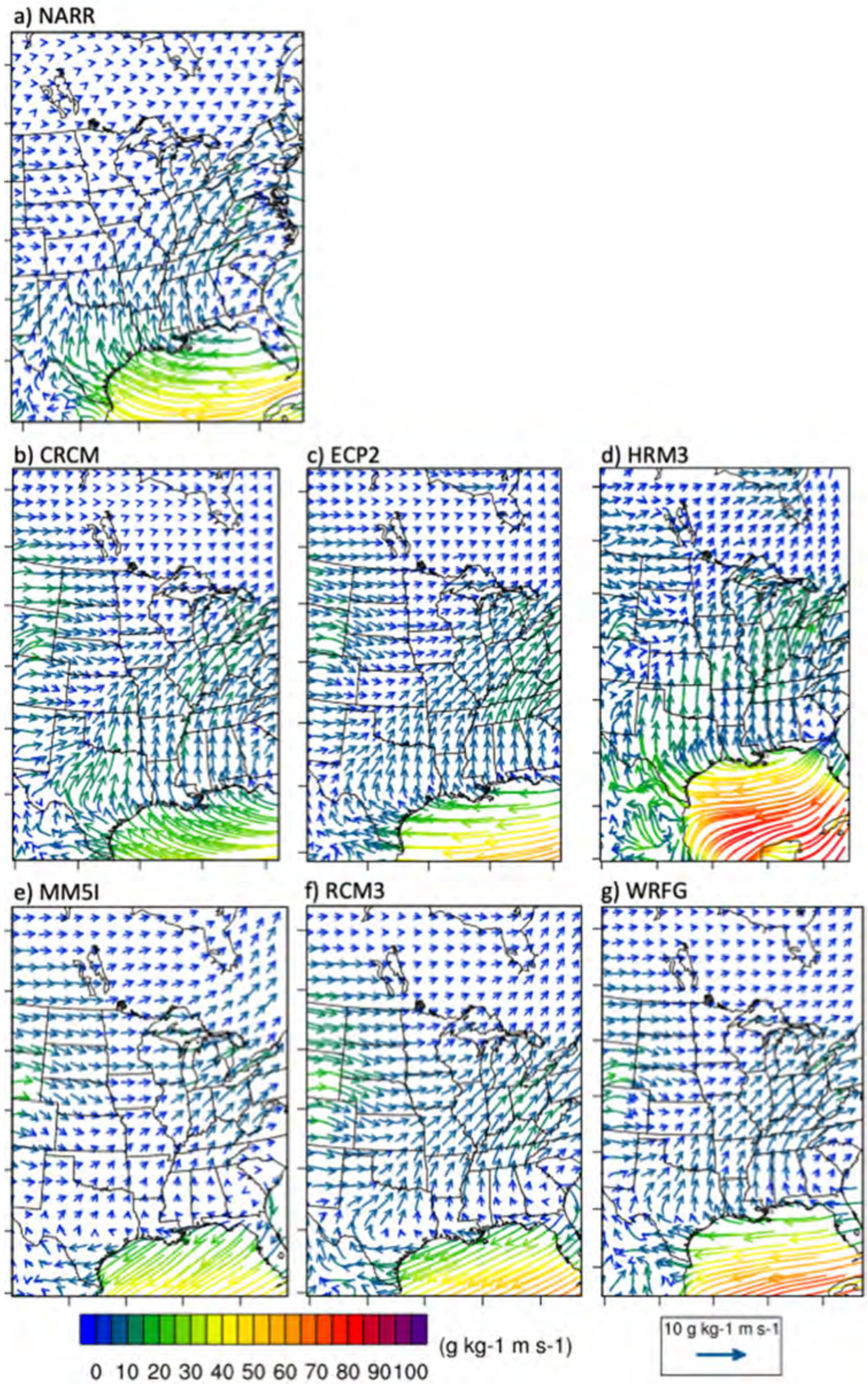


Figure 7-6 1980-2003 November-March average near-surface specific humidity flux from the NARCCAP NCEP-driven RCMs and NARR. For the CRCM, near-surface (2m) specific humidity is not available, so specific humidity from the lowest model was used in the calculation instead.

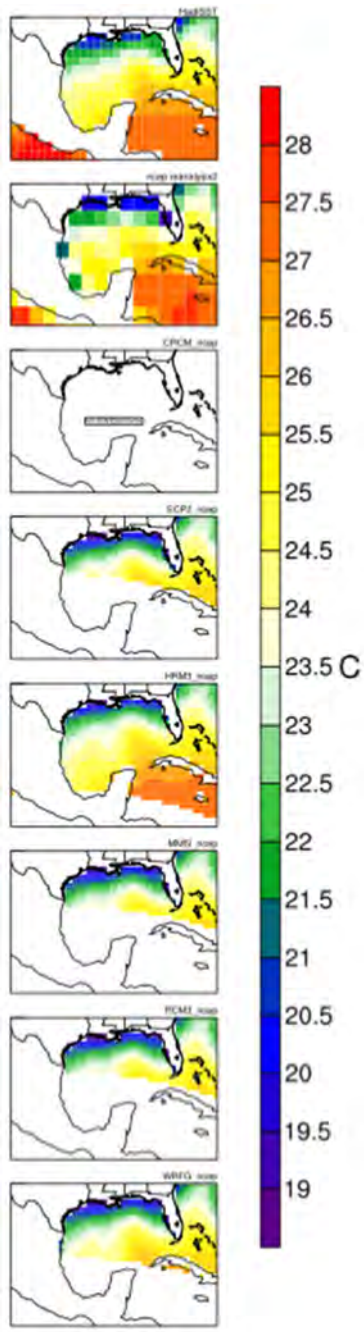


Figure 7-7 1980-2003 November– April average sea-surface temperature from the HadISST observation-based dataset, the NCEP reanalysis, and the NARCCAP NCEP-driven RCMs. Data over land is masked. Values in the RCMs cut off at the southern boundary of the simulation domain (minus the relaxation zone).

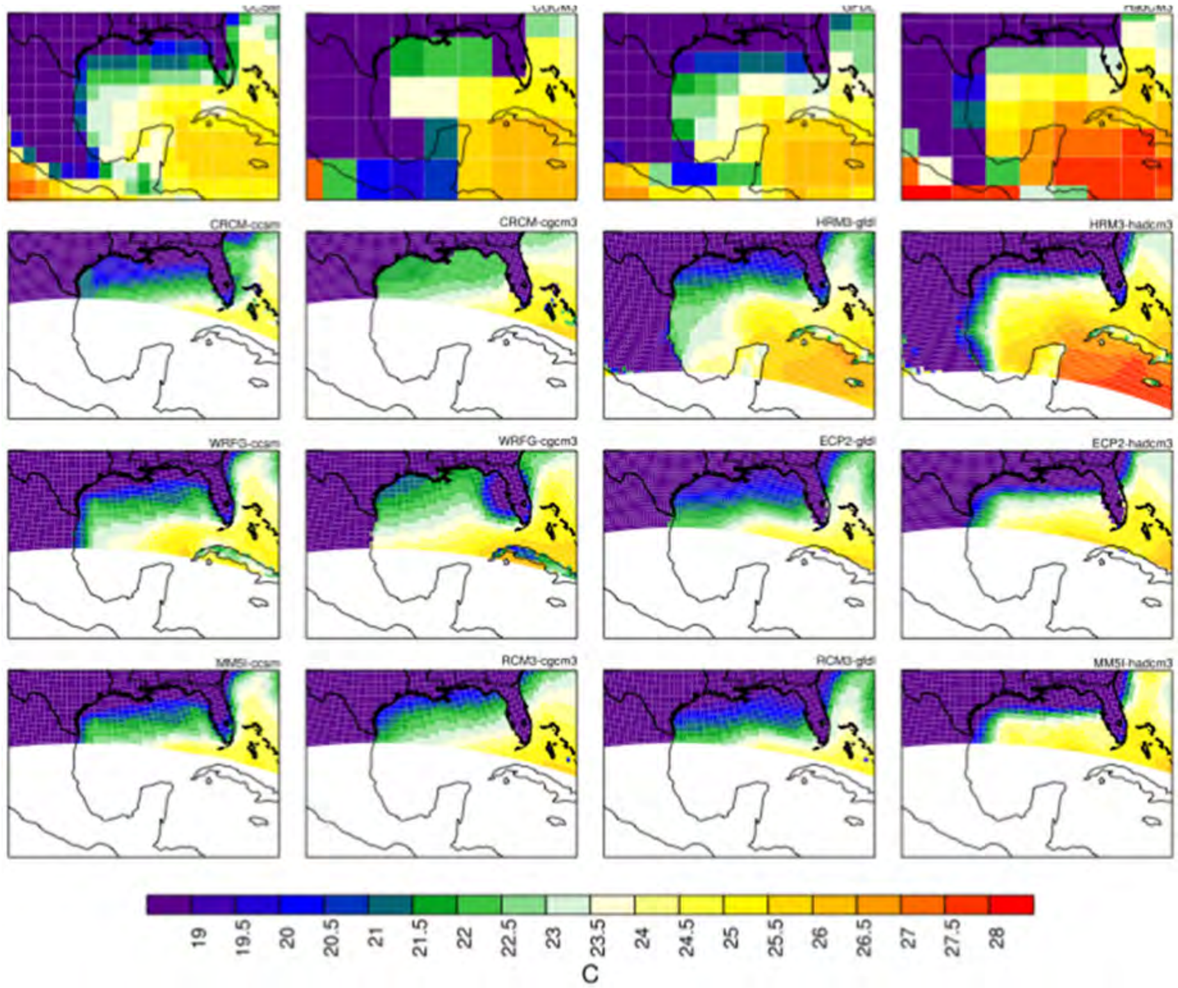


Figure 7-8 As in the previous figure, but for 1971-1998 November– April average sea-surface temperature from the NARCCAP GCM-driven RCMs, and their forcing GCMs. Data over land is not masked. Cell size represents the model resolution.

7.4.3.2. FORCING

ETCs that are most likely to force precipitation over ESC during the cool-season form on the eastern slope of the Rocky Mountains (or redevelop there after crossing the mountains) or the Sierra Madre Oriental. They then usually propagate East and/or North towards the Great Lakes and New England. This is seen in the ETC density maps in Figure 7-9 in three different reanalyses. Note that in Figure 7-9 NCEP and ERA-Int are considerably coarser in resolution than NARR; therefore, some of the differences between them and NARR are likely due to resolution (particularly in the column of half-boxes over the far Western part of the image), as shifts in how a cyclone center is identified may occur based on the center of placement of cyclone with a data grid box on a coarse versus finer grid. Difference may also be due to the configuration of the different reanalyses. Because NARR more closely matches the spatial and temporal resolution of the RCMs, we will use it in our comparisons in this section.

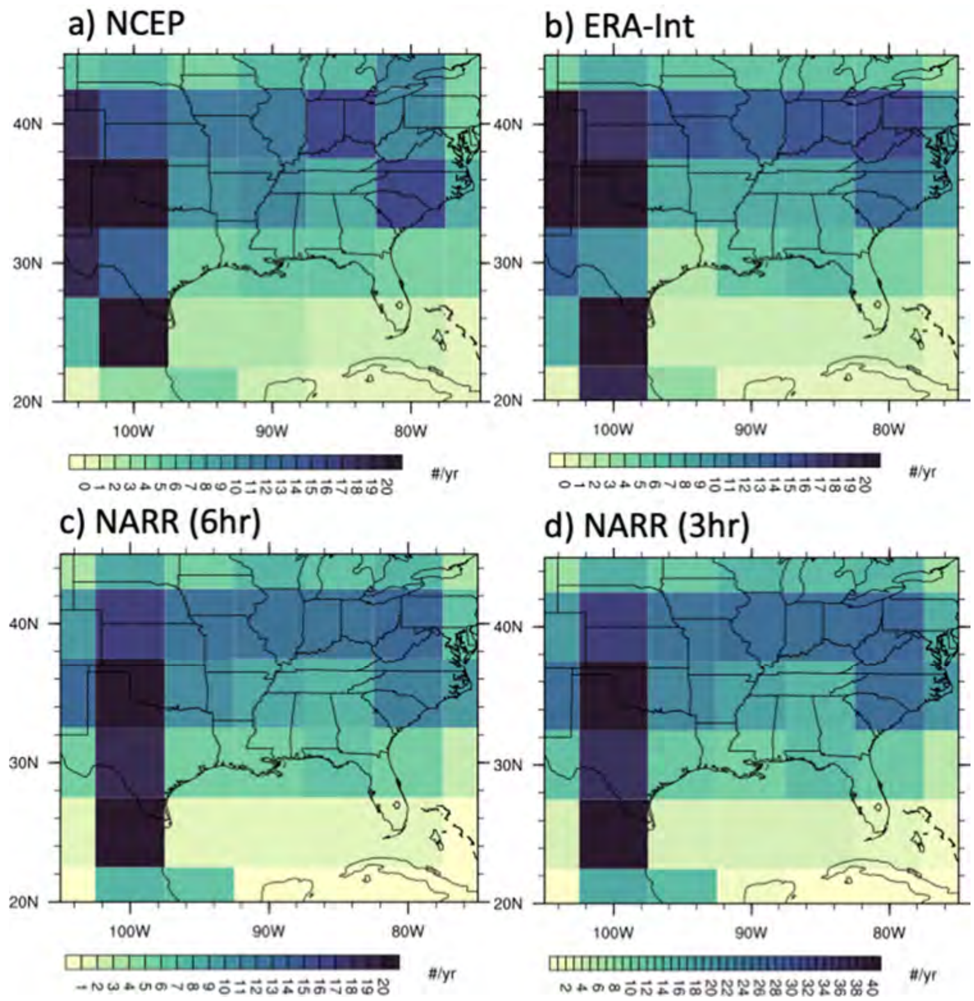


Figure 7-9 1980-2004 November – March average number of ETC per season to pass through a given 5x5 degree box from a-c) from 6-hour average reanalysis data or d) from 3-hour average reanalysis data.

Figure 7-10 illustrates the differences between the NARCCAP NCEP-driven RCMs and NARR. While the RCMs generally do a good job capturing the spatial features of the density field, there are some subtle differences in the density fields along the lee side of the continental divide that are important. Chiefly, there are low biases in the RCMs at different locations along the eastern slope of the mountains. Some of the RCMs have two few over the central high plains, while some are more biased over the southern high plains and Mexico. Either way, too few storms upstream of the ESC suggests less transient moisture flux into the region and fewer forcing events for precipitation over the region. Too few over the southern high plains and Mexico would likely bias precipitation most in the southern part of the ESC.

To further illustrate how this may bias precipitation in the ESC, Figure 7-11 illustrates the density of ETC associated with precipitation in the ESC focus region. Again, while the spatial pattern of this field is reasonable, there is a definite low bias in the RCMs in the number of ESC precipitation-producing ETCs upstream of the region, particularly in the southern high plains and Mexico, highlighting that a lack of forcing is likely contributing to the dry bias in the RCMs, particularly in the southern half of the ESC. This may be due to the close proximity of this cyclogenesis zone to the southern boundary of many of the RCMs; however, additional simulations would be required to test this hypothesis.

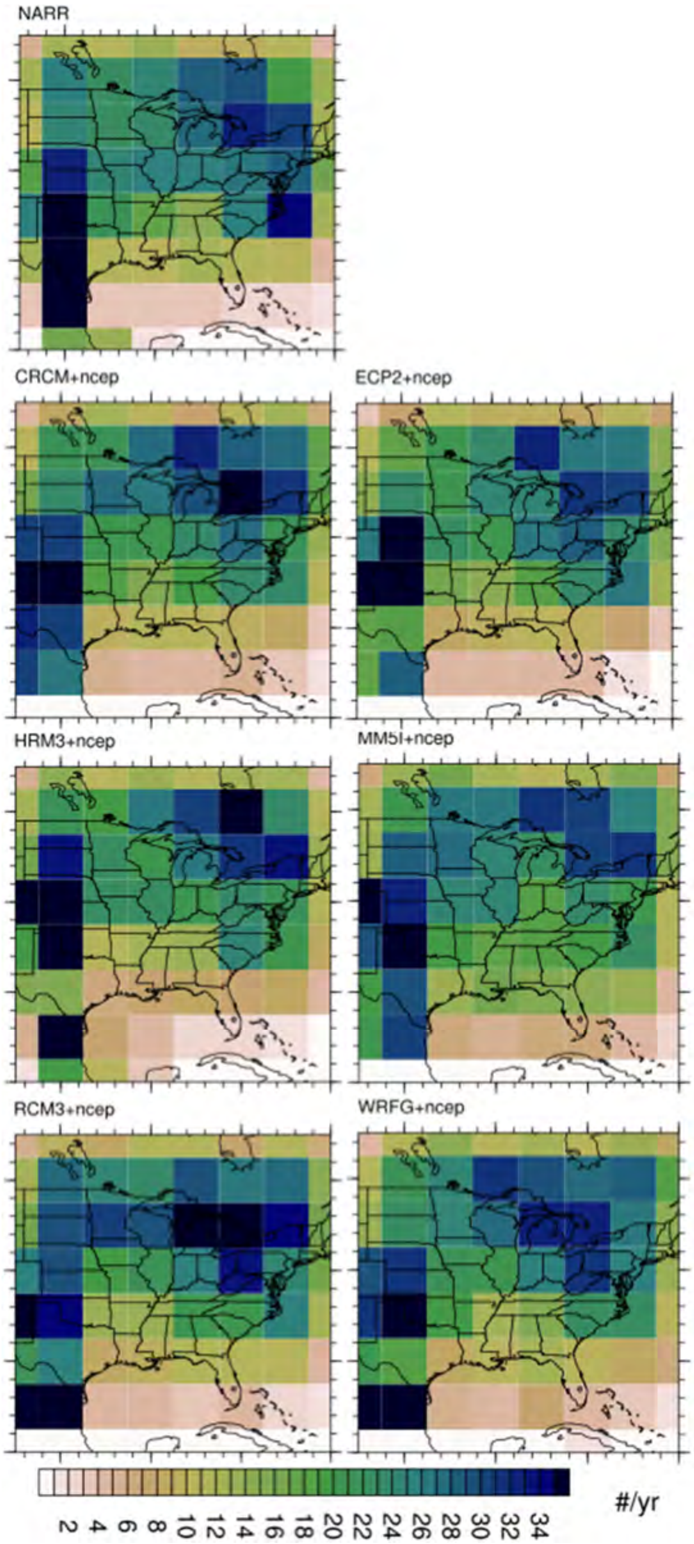


Figure 7-10 1980-2003 November – March average number of ETC per season to pass through a given 5x5 degree box as calculated from 3-hourly fields from the NARR and the NARCCAP NCEP-driven simulations.

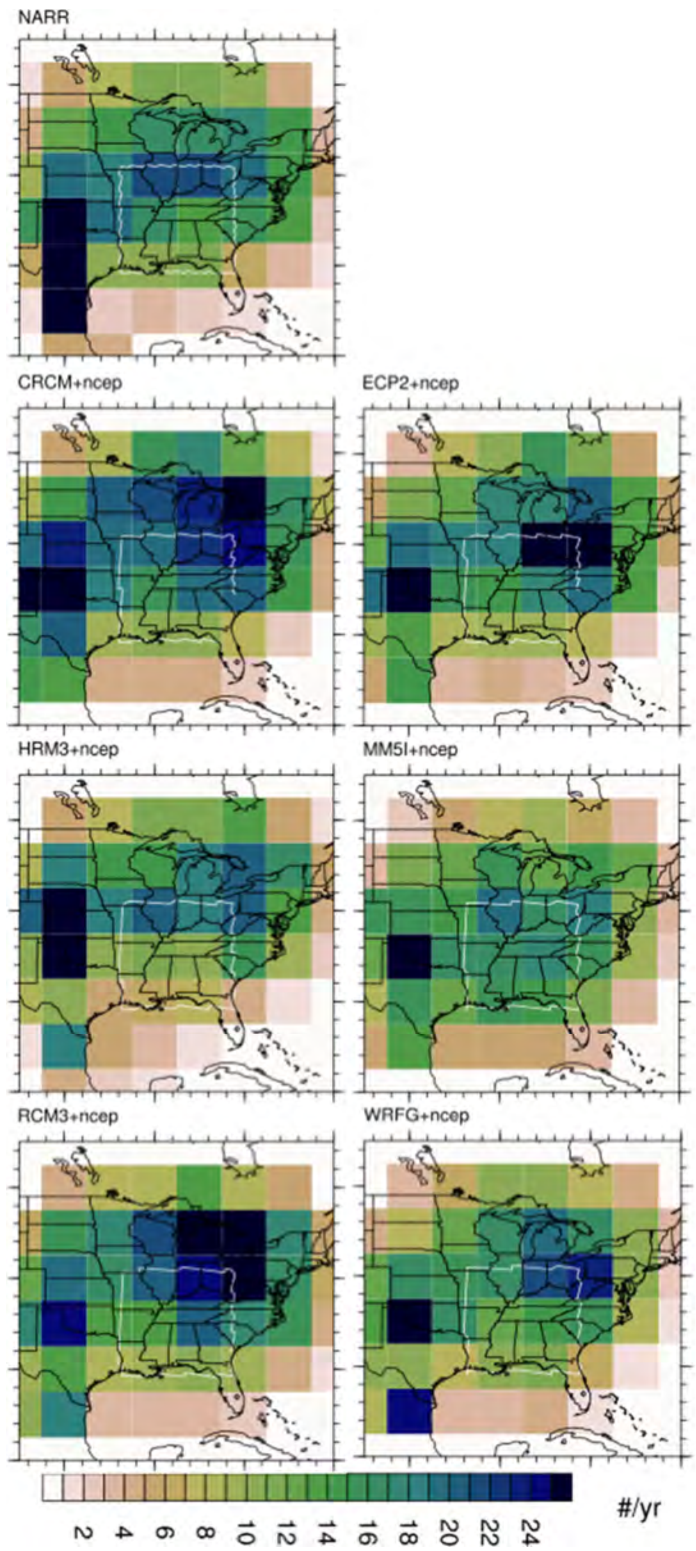


Figure 7-11 1980-2003 November – March average number of ETC associated with precipitation in the area of interest (outlined in white) per season to pass through a given 5x5 degree box from the NARR and the NARCCAP NCEP-driven simulations.

Some of the RCMs, like MM5I, also have too few ETC associated with rain passing over the northern part of the ESC region (Figure 7-11). Not only do the lows that pass over and just north of the region help to form precipitation, but so do their associated fronts. With too few lows in the northern and northeastern parts of the analysis domain, MM5I probably does not contain enough cold-front forced precipitation events in the southern part of the ESC, and enough warm-front and low-center forced precipitation over the OTRB. This forcing problem combined with the moisture problems seen in MM5I help explain why this RCM has the strongest dry bias out of the full NARCCAP set over the full ESC region.

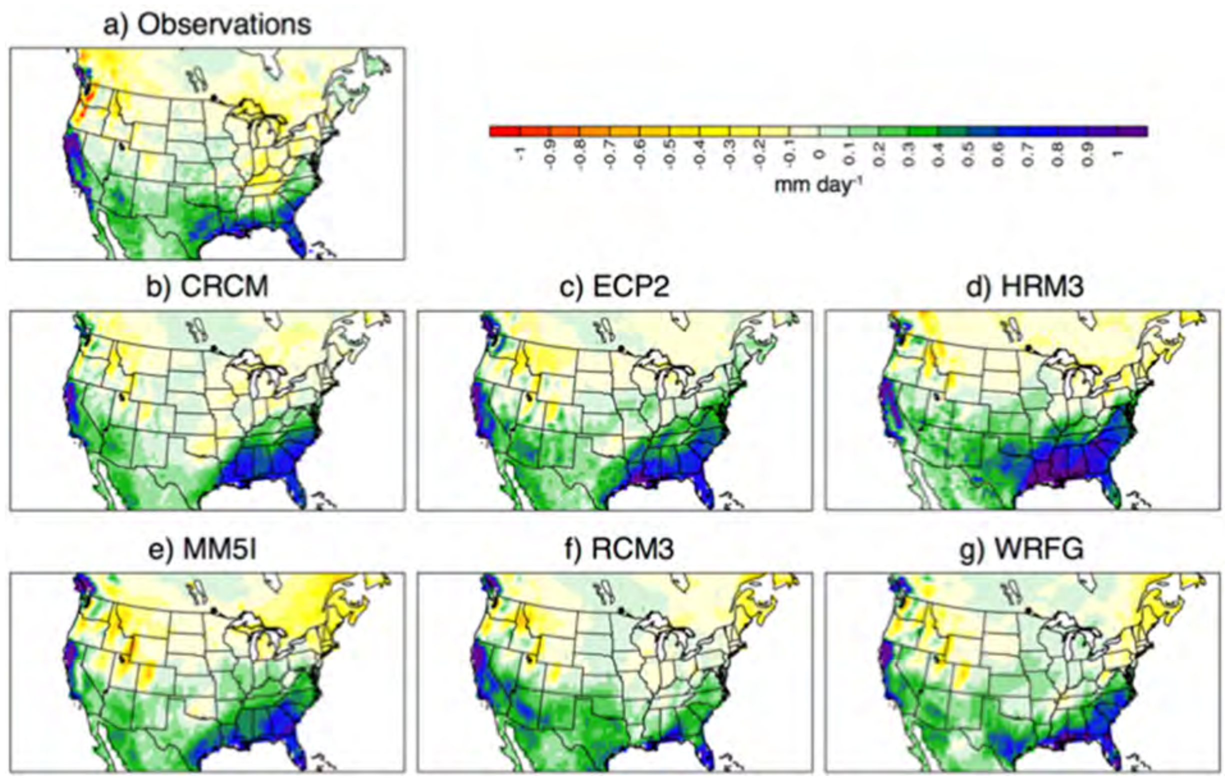


Figure 7-12 1980-2003 December – February precipitation teleconnections from the NCEP-driven NARCCAP RCMs and Observations. Teleconnections were calculated by linear regression analysis of precipitation against the Niño 3. 4 SST index. I.e., the change in precipitation rate per degree change of SSTs in the Niño-3.4 region.

7.4.3.3. ADDITIONAL MECHANISMS

The precipitation response to drivers of interannual climate variability, such as the El Niño Southern Oscillation (ENSO), is a challenge for climate models to capture. In order to examine the effect ENSO teleconnections of precipitation may have on mean precipitation bias, we have implemented the techniques used in Langenbrunner and Neelin (2013). Figure 7-12 shows the teleconnections calculated by linear regression analysis of precipitation against the Niño-3.4 SST index. That is, it illustrates the change in precipitation per degree of change in the SSTs in the Niño-3.4 region. El Niño corresponds with wetter than normal winter conditions over much of the Southern U.S., and drier conditions over the Ohio River Basin. When forced by reanalysis, the NARCCAP RCMs (Figure 7-12) broadly capture the patterns of these winter precipitation teleconnections; however, the intensity of the precipitation response varies across the RCMs. Many of them dramatically overestimate the wet response over the Southeast U.S. This leads them to underestimate the drying response over the OTRB, and overestimate the wet response over the southern half of the ESC. If this plays a role in the precipitation bias in the ESC, in the southern half of the region, it is overcome by other factors forcing the dry bias.

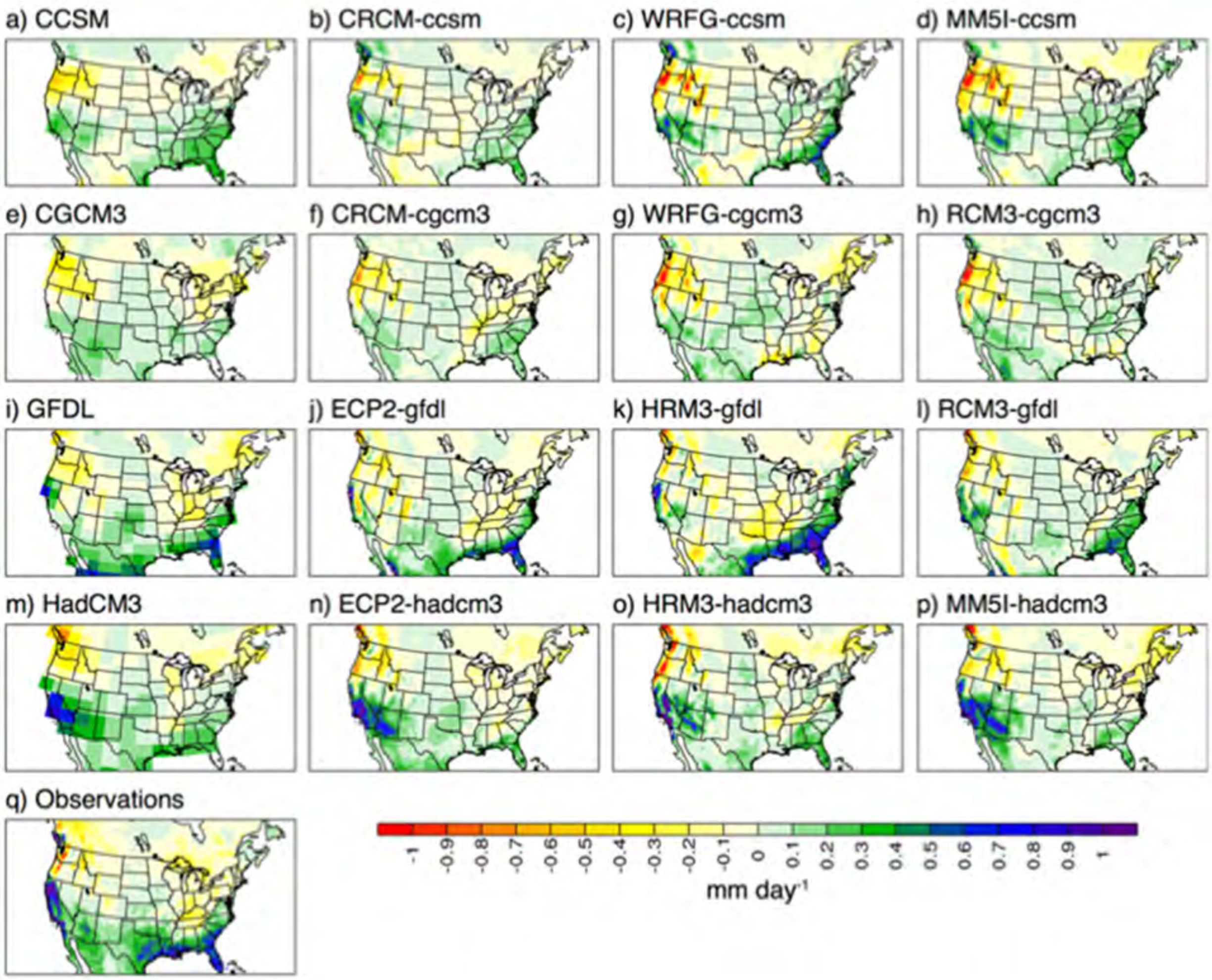


Figure 7-13 As in the previous figure, but for precipitation teleconnections from the GCM-driven NARCCAP RCMs, the GCMs that provided forcing data, and Observations.

In the GCM-driven NARCCAP simulations (Figure 7-13), it is clear that the RCMs follow the teleconnection patterns from their drivers. Generally, the amplitude of the teleconnections is too low in all but the HRM3-gfdl. Also, in the WRFG and RCM3 CGCM3-driven simulations, the correlation is reversed in the ESC. If this was contributing to a dry response in precipitation in these simulations, given their precipitation fields, it is likely being overcome by the warm SSTs in these runs and their resulting increase in available moisture for ESC precipitation. The teleconnection biases in Figure 7-13 could, however, affect the credibility of the precipitation projections from these simulations, as the simulations move towards increasingly El Niño- (like many GCMs) or La Niña- (as in CGCM3) like states in the future.

7.5. CONCLUSION

Midcentury projections from the NARCCAP RCMs for the ESC indicate an increase in mean precipitation during the cool-season (November-March), particularly in the north half of the ESC region, over the OTRB. This projection is consistent with, although not of the same magnitude or significance as, the CMIP5 and CMIP3 simulations. However, given the pervasiveness of a dry bias in mean precipitation in many model simulations of this region, including those from NARCCAP, the credibility of these projections, despite general model agreement, requires further assessment. The bias in mean precipitation points to biases in the underlying processes behind the precipitation and may affect the credibility of projections in not only mean precipitation but also extreme precipitation in the ESC.

Therefore, this study assessed drivers of precipitation in the ESC during the cool season to better understand, at a process-level, the causes of bias in the models. Particular attention was paid to the reanalysis-driven NARCCAP simulations to better understand the bias relative to the RCMs only, without inherited bias from the GCMs, although some of the most important GCM-driven biases were also detailed.

In summary, precipitation bias in the RCMs has several sources. Moisture flux into the region is biased in some of the RCMs (both in response to biases in stationary circulations and transient features like ETCs), and forcing for precipitation, in the form of transient, synoptic-scale low-pressure systems, in the ESC is also lacking in some of the RCMs (too few systems forcing precipitation in the ESC, particularly those that have the most influence on the southern half of the region). Additional problems in GCM-driven simulations related to Gulf of Mexico SSTs and ENSO-related teleconnections contribute to additional bias in the GCM-driven simulations. However, in a number of the RCM+GCM simulations, compensating biases contribute to many of the simulations having less of a dry bias.

Given all of these different competing factors, it is difficult at this time to assess the overall credibility of the projections from the NARCCAP set. More analysis will need to be done on the evolution of these factors in the future simulations first. However, we would recommend caution using the projections from these simulations in the ESC region, given their now known problems in simulating the processes that drive the baseline climate in this region. Future analysis will also include additional RCM simulations driven by CMIP5 simulations to increase our sample-size.

7.6. REFERENCES

Anderson, J. L., and Coauthors, 2004: The new GFDL global atmosphere and land model AM2-LM2: Evaluation with prescribed SST simulations. J. Climate, 17, 4641–4673, doi:10.1175/JCLI-3223.1.

Bukovsky, M. S., and D. J. Karoly, 2011: A regional modeling study of climate change impacts on warm-season precipitation in the central United States. J. Climate, 24, 1985–2002, doi:10.1175/ 2010JCLI3447.1.

Bukovsky, M.S., R.R. McCrary, A. Seth, L.O. Mearns, 2017: A mechanistically credible, poleward shift in warm-season precipitation projected for the U.S. Southern Great Plains? J. Climate, 30, 8275-8298, doi: 10.1175/JCLI-D-16-0316.1.

Caya, D., and R. Laprise, 1999: A semi-implicit semi-Lagrangian regional climate model: The Canadian RCM. Mon. Wea. Rev., 127, 341–362.

Chang, E.K., Guo, Y. and Xia, X., 2012. CMIP5 multimodel ensemble projection of storm track change under global warming. Journal of Geophysical Research: Atmospheres, 117(D23).

Chang, E.K., Ma, C.G., Zheng, C. and Yau, A.M., 2016. Observed and projected decrease in Northern Hemisphere extratropical cyclone activity in summer and its impacts on maximum temperature. Geophysical Research Letters, 43(5), pp.2200-2208.

Collins, W. D., and Coauthors, 2006: The Community Climate System Model: CCSM3. J. Climate, 19, 2122–2143.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, D.P. and Bechtold, P., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137(656), pp.553-597.

Flato, G. M., G. J. Boer, W. G. Lee, N. A. McFarlane, D. Ramsden, M. C. Reader, and A. J. Weaver, 2000: The Canadian Centre for Climate Modeling and Analysis global coupled model and its climate. Climate Dyn., 16, 451–467.

Giorgi, F., M. R. Marinucci, and G. T. Bates, 1993a: Development of a second-generation regional climate model (RegCM2). Part I: Boundary-layer and radiative transfer processes. Mon. Wea. Rev., 121, 2794–2813.

——, ——, G. de Canio, and G. T. Bates, 1993b: Development of a second-generation regional climate model (RegCM2). Part II: Convective processes and assimilation of lateral boundary conditions. Mon. Wea. Rev., 121, 2814–2832.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Climate Dyn., 16, 147–168, doi:10.1007/s003820050010.

Grell, G. A., 1993: Prognostic evaluation of assumptions used by cumulus parametrizations. Mon. Wea. Rev., 121, 764–787.

Jones, R. G., D. C. Hassell, D. Hudson, S. S. Wilson, G. J. Jenkins, and J. F. B. Mitchell, 2003: Workbook on generating high-

resolution climate change scenarios using PRECIS. UNDP Tech. Rep., 32 pp.
Juang, H. M., S. Y. Hong, and M. Kanamitsu, 1997: The NCEP regional spectral model: An update. Bull. Amer. Meteor. Soc., 78, 2125–2143.
Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). Bull. Amer. Meteor. Soc., 83, 1631–1643.
Langenbrunner, B. and Neelin, J.D., 2013. Analyzing ENSO teleconnections in CMIP models as a measure of model fidelity in simulating precipitation. Journal of Climate, 26(13), pp.4431-4446.
Liu, C., 2017: Update on the High-Resolution CONUS Climate Simulations. Presentation. 2017 Water Systems Retreat, NCAR, Boulder, CO.
Livneh B., E.A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K.M. Andreadis, E.P. Maurer, and D.P. Lettenmaier, 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, 9384–9392.
Mearns, L. O., and Coauthors, 2007: The North American Regional Climate Change Assessment Program dataset. National Center for Atmospheric Research Earth System Grid data portal, Boulder, CO, digital media. [Available online at http://www.earthsystemgrid.org/project/NARCCAP.html.]
——, and Coauthors, 2012: The North American Regional Climate Change Assessment Program: Overview of phase I results. Bull. Amer. Meteor. Soc., 93, 1337–1362.
Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. Bull. Amer. Meteor. Soc., 87, 343–360, doi:10.1175/BAMS-87-3-343.
Moss, R., and Coauthors, 2008: Towards New Scenarios for Analysis of Emissions. Intergovernmental Panel on Climate Change, 132 pp.
Nakicenovic, N., and Coauthors, 2000: Special Report on Emissions Scenarios. Cambridge University Press, 599 pp.
Pal, J. S., and Coauthors, 2007: Regional climate modeling for the developing world: The ICTP RegCM3 and RegCNET. Bull. Amer. Meteor. Soc., 88, 1395–1409.
Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton, 2000: The impact of new physical parameterizations in the Hadley Centre climate model: HadAM3. Climate Dyn., 16, 123–146, doi:10.1007/s003820050009.
Rayner, N.A.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C. and Kaplan, A., 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. Journal of Geophysical Research: Atmospheres, 108(D14).
Ruane, A. C., 2010a: NARR's atmospheric water cycle components. Part I: 20-year mean and annual interactions. J. Hydrometeor., 11, 1205–1219.
——, 2010b: NARR's atmospheric water cycle components. Part II: Summertime mean and diurnal interactions. J. Hydrometeor., 11, 1220–1233.
Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF version 2. NCAR Tech. Note NCAR/TN-4681STR, 88 pp.
Wallace, J.M., Lim, G.H. and Blackmon, M.L., 1988. Relationship between cyclone tracks, anticyclone tracks and baroclinic waveguides. Journal of the atmospheric sciences, 45(3), pp.439-462.
Willmott, C.J. and Matsuura, K., 1995. Smart interpolation of annually averaged air temperature in the United States. Journal of Applied Meteorology, 34(12), pp.2577-2586.

8. PRECIPITATION-INFORMED SAFETY FACTORS FOR DESIGNING URBAN DRAINAGE SYSTEMS – CASE STUDIES ACROSS DIVERSE CLIMATIC AREAS IN U.S.

8.1. EXECUTIVE SUMMARY

Design of urban drainage systems is often based on the concept of Intensity-Duration-Frequency curves (IDF). IDF curves illustrate the probability of occurrence of a given rainfall intensity for a given duration. Applying a safety factor to the precipitation intensity obtained for a given probability of occurrence is common engineering practice to account for epistemic and aleatory uncertainties (cf. Chapter 2.6). The concept of deep uncertainty introduced by climate change, however, often calls for applying an additional safety factor denoted as '*climate factor*'. Although climate factor values are usually based on GCM simulated changes in precipitation extremes between a future and the historic periods, a well-defined and broadly accepted methodology for assessing these factors is still missing. For instance, climate factors are sometimes obtained using a single GCM (cf. the example given in Chapter 2.6.2), which is somehow counterintuitive given their purpose to account for deep uncertainty. The objective of this study is to advance the field of urban drainage management by demonstrating how to estimate climate factors that account for deep uncertainty. The considered technical approach combines frequency analysis of extreme rainfall with Decision Scaling. The approach is demonstrated over a set of nine watersheds located in various climatic conditions in continental U.S. for the annual maxima daily precipitation with and return period of 100 years. However, the presented analysis framework could be extended to other temporal scales, frequency of occurrence or climatic areas. Results indicate that robust climate factors accounting for the uncertainty stemming from the climate projections range from 30% to 80% depending on the considered catchment. Results also show that the choice of the downscaling method to estimate distribution parameters for projected precipitations can have a significant influence on the climate factor estimate.

8.2. TECHNICICAL APPROACH

The developed analysis framework aims at combining frequency analysis of extreme rainfall with Decision Scaling, a robust decision analysis framework developed to account for deep uncertainty. For a considered system, Decision Scaling allows assessing the system response to change in a set of factors, most often to changes in precipitation and temperature variables. Future projections obtained from Global Circulation Models (GCMs) for the key climate drivers are used for ex-post assessment of the likelihood of the climate states that could potentially lead to significant failures for the considered system. We invite the reader to head to the final report of the SERDP Project RC-2204 for a more comprehensive description of the Decision Scaling approach (Brown et al., 2016). Decision Scaling has already been considered for the assessment of climate change impact on flood risk (e.g., Steinchneider et al., 2015; Spence and Brown 2016). In this study, the 'climate stress test' within the decision scaling analysis is implemented by varying the parameters for the precipitation extreme distribution. Changes in the distribution parameters from climate projections are used ex-post to infer the likelihood of changes in precipitation extreme distribution.

8.3. DATA AND METHODS

8.3.1. CASE STUDIES

The proposed methodology is demonstrated for a set of nine basins located in various climate areas in continental U.S. All considered catchments are larger than 1,000 square miles (Table 8-1). Their average elevation ranges from 35 feet (CA) to almost 3,500 feet (AZ). These various elevations and geographic locations (e.g., continental vs. coastal; high vs. low elevation) lead to climate regimes likely to drive various types of hydrological extremes. Based on the classification from Kottek et al. (2006), two catchments are arid (AZ, MT), four are warm temperate (CA, GA, IN, TX) and three are significantly influenced by snow (CO, ME, WA) (Table 8-1). Precipitation regime is fully humid for all catchments but for CA and MT for which rainfall regime is categorized as 'summer dry' and AZ that is said 'Steppe'. Temperature regime is considered as 'warm summer' in MT, CO, ME and WA, while it is 'hot summer' for GA, IN and TX. On the other side of the spectrum, temperature regime for AZ and MT is classified as 'cold arid' basin. The reader is invited to head to the study by Kottek et al. (2006) for a more comprehensive description of the climate classification.

Table 8-1: List of considered basins with their main climate characteristics (from Kottek et al. 2006). Considered basins are named from the name of the U.S. state where their outlet is located. They are Arizona (AZ), Montana (MT), California (CA), Georgia (GA), Indiana (IN), Texas (TX), Colorado (CO), Maine (ME) and Washington State (WA).

Basin	USGS id	Area miles2	Elev (ft)	Main Climate	Precipitation	Temperature	Number of rain gages
AZ	9444500	2766	3436	Arid	Steppe	Cold Arid	6
MT	6214500	11807	3081	Arid	Summer Dry	Cold Arid	22
CA	11477000	3113	36	Warm Temperate	Summer Dry	Warm Summer	2
GA	2228000	2790	15	Warm Temperate	Fully Humid	Hot Summer	5
IN	3374000	11125	388	Warm Temperate	Fully Humid	Hot Summer	34
TX	8085500	3988	1164	Warm Temperate	Fully Humid	Hot Summer	9
CO	9119000	1061	7639	Snow	Fully Humid	Warm Summer	3
ME	1010500	2680	590	Snow	Fully Humid	Warm Summer	2
WA	12442500	3550	1138	Snow	Fully Humid	Warm Summer	3

8.3.2. Precipitation Data

We used daily precipitation ground measurements from the Global Historical Climatology Network Daily (GHCND) made available by NOAA (https://www.ncdc.noaa.gov/ghcnd-data-access). For each catchment, all rain gages with at least 50 years of data and located within the catchment boundaries were considered to construct a model that represent the distribution of the extreme rainfall at the basin scale. The number of rain gages considered for each catchment is given in Table 8-1.

8.3.3. Climate Projections

We consider 32 GCMs from the 5th generation of the Coupled Model Intercomparison Project (CMIP5, Taylor et al., 2012) of the Intergovernmental Panel on Climate Change (IPCC), which has been bias-corrected and downscaled following the locally constructed analogs approach (LOCA, Pierce et al., 2014; 2015). Only the representative concentration pathway 8.5 W/m^2 (RCP 8.5) is considered in the analysis.

In order to initiate a discussion about the contribution to the uncertainty stemming from the choice of the downscaling approach, we compare the results obtained from a subset of GCM projections downscaled with LOCA method with projections from the same GCMs but downscaled with two RCMs available from the North-America CORDEX experiment (Mearns et al., 2017). The three GCMs for which dynamically downscaled projections are available are GFDL-ESM2M (Dunne et al., 2012; 2013), HadGEM2-ES (Collins et al., 2011) and MPI-ESM-LR (Giorgetta et al., 2013). The two RCMs used for downscaling the three above-mentioned GCMs are the Weather and Research Forecast model (WRF; Powers et al., 2017) and the Regional Climate Model 4 (RegCM4, Giorgi et al., 2012). The number of GCM/RCM combinations (3 GCMs x 2 RCMs) is too small to provide a robust claim regarding the relative performance of the RCMs compared to LOCA. However, this preliminary analysis provides motivation for further research to assess the uncertainty following from the use of downscaling method within a robust decision making approach such as Decision Scaling.

8.3.4. Regional Precipitation Extreme Distribution

We here focus on the distribution of the annual maximum daily precipitation. The Generalized Extreme Value (GEV, Jenkinson, 1955) distribution has been considered to represent the annual max precipitation at each rain gage (Equation 8.1):

$$F_{Y_i}(\mathbf{y}_i|\xi_i,\alpha_i,\kappa_i) = exp\left\{-\left[1-\kappa_i\left(\frac{\mathbf{y}_i-\xi_i}{\alpha_i}\right)\right]^{1/\kappa_i}\right\}, \tag{8.1}$$

where $\xi_i > 0$, $\alpha_i > 0$ and κ_i are the location, scale and shape parameters and $\mathbf{y}_i$ of the vector of annual maximum daily precipitation at the gage 'i' (or grid cell 'i' if using GCM or RCM outputs). When applied to precipitation extreme values (i.e., with positive values), the shape parameter κ is negative and the distribution is said to be heavy tailed. If the shape parameter κ equals zero, the GEV distribution corresponds to the well-known Gumbel distribution, and is said to be thin tailed. The GEV parameters have been estimated for each rain gages using the L-moment method (Hosking, 1990). We applied the Kolmogorov-Smirnov (KS) test to check whether the hypothesis that the observed annual maxima record come from the fitted GEV distribution. Considering 10% confidence level, this hypothesis cannot be rejected for all considered rain gages but for two gages within the Montana basin. These gages have been consequently disregarded for the rest of the analysis.

Following Hosking and Wallis (1997), the parameters of a regional distribution $F_Y(\xi^R,\alpha^R,\kappa^R)$ can be obtained from the parameters of the distributions F_{Y_i}:

$$\xi^R = \frac{1}{N}\sum_{i=N}^{N}\xi_i, \tag{8.2}$$

$$\alpha^R = \frac{1}{N}\sum_{i=N}^{N}\alpha_i, \tag{8.3}$$

$$\kappa^R = \frac{1}{N}\sum_{i=N}^{N}\kappa_i, \tag{8.4}$$

where ξ^R,α^R,κ^R are regional distribution parameters and N the number of rain gages within the considered catchment. Note that different weights could be assigned to different gages, especially if records at some gages have much shorter duration than others (long records would get higher weights). For the sake of simplicity, and because all selected gages have at least 50 years of data, we assume an equal weight for all available gages (i.e., equal to $1/N$ as shown in equations 8.2, 8.3 and 8.4).

8.4. RESULTS AND DISCUSSION

8.4.1. Precipitation-informed climate factors

The climate stress test is here implemented by varying the location and scale parameters of the regional distribution obtained for each basin. The considered analysis does not account for any potential change in the shape parameter since it is acknowledged that significant change in shape parameter cannot be detected with the length of the available records. Figure 8-1 illustrates the sensitivity of the 100-yr annual precipitation maxima (i.e., frequency of occurrence equals 0.01) to change in location and scale parameters for the WA and MT basins. For both catchments, we note that an increase in either location or scale parameter will tend to increase the intensity of precipitation extreme. Conversely, a decrease in these parameters will tend to decrease the intensity. For both catchments, we note that an increase in precipitation extreme intensity that would follow from a larger location parameter can be offset by a relative lower decrease in scale parameter. Such an observation agrees with the well-accepted assumption that extremes are more sensitive to changes in variability (i.e., scale parameter) than to changes in average (i.e., location parameter).

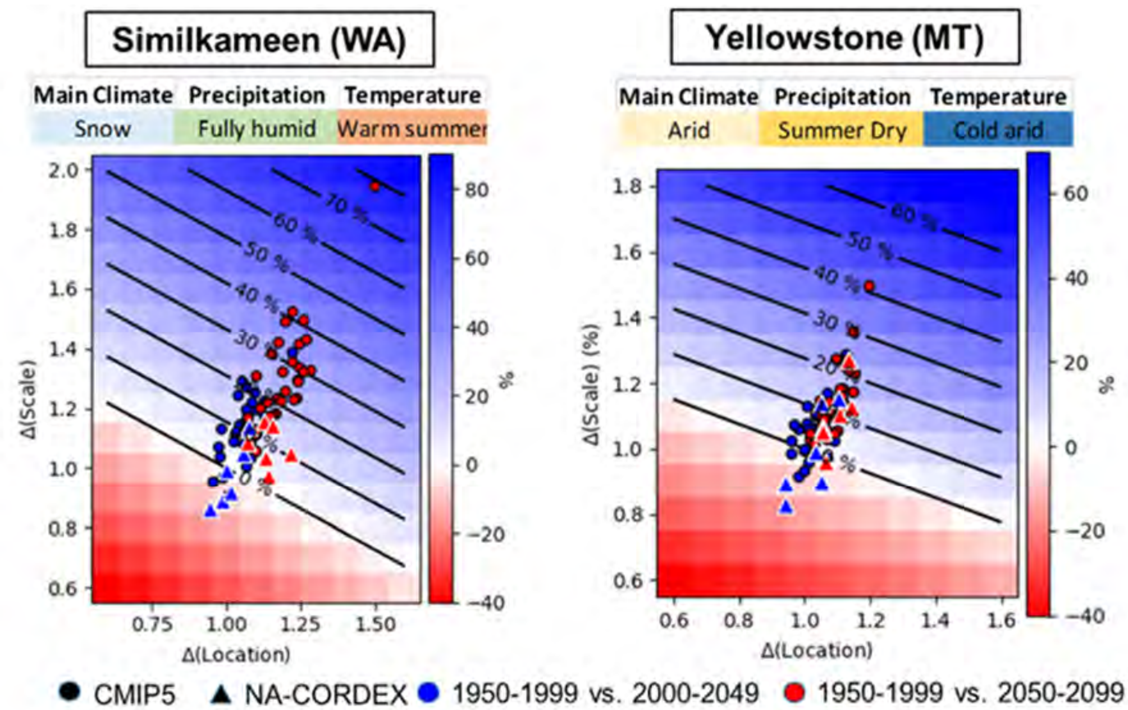


Figure 8-1 Response functions showing the change in annual maximum precipitation with a probability of occurrence of 0.01 (i.e., 100-yr return period). The x- and y-axis show relative changes in location (ξ^R) and scale (α^R) parameters of the regional distribution (i.e., a delta change Δ of 2 means that considered parameters has been multiply by 2 compared to its historic value). The heat map shows the percentage change in the annual daily maximum rainfall with a return period of 100 years. White color indicates the current value of the 100-yr return rainfall. Black lines show increment by 10% of the change in the 100-yr return rainfall. Blue and red symbols show the change in ξ^R and α^R obtained from the climate projections. Dots are for GCM projections downscaled with LOCA method and pyramid symbols are for projections downscaled via a RCM.

Red and blue symbols on Figure 8-1 show the estimated changes in location and scale parameters of annual maximum precipitation as obtained from the climate projections for two future periods. Blue color shows the expected change for the first half of the 21st century while red color shows change for the second half of the century; the historical period being 1950-1999. For both catchments, we note that estimated changes in location parameter range from roughly -5% to +30%, which suggests that precipitation extreme could increase on average at these locations. The change in precipitation extreme variability suggested by the projection ensemble is more uncertain as it ranges roughly from -20% to +50% for both catchments. We note the presence of an outlier in the case of the WA catchment for which one climate projection indicates an increase in scale parameter by +90% of its historical value. When combining the information of the sensitivity to change in location and scales parameters with the projected changes from climate models, we note for both catchments that precipitation extremes will likely increase as time goes on.

From the results illustrated on Figure 8-1, it is possible to estimate what would be the required climate factor to use for any new infrastructure to provide a robust climate factor that accounts for a chosen likelihood of changes in precipitation extremes and for each considered location. We can indeed combine the information provided by the sensitivity of the precipitation extreme to change in location and scale parameters (i.e., heat map on Figure 8-1) with the information of the likelihood of change in location and scale parameters obtained from the climate projections. This process is illustrated on Figure 8-2. Let us assume that a decision maker is willing to apply a climate factors on a design value to account for deep uncertainty stemming from climate change. However, he/she does not want to plan for the worst scenario. Let us thus assume this decision maker is risk averse in a way that he/she agrees to disregard 10% of the GCMs leading to the worst-case scenarios, which would result to account for 90% of the available climate projections. In such a case, Figure 8-2 illustrates that a climate factor of 25% for MT and 40% for WA would be required.

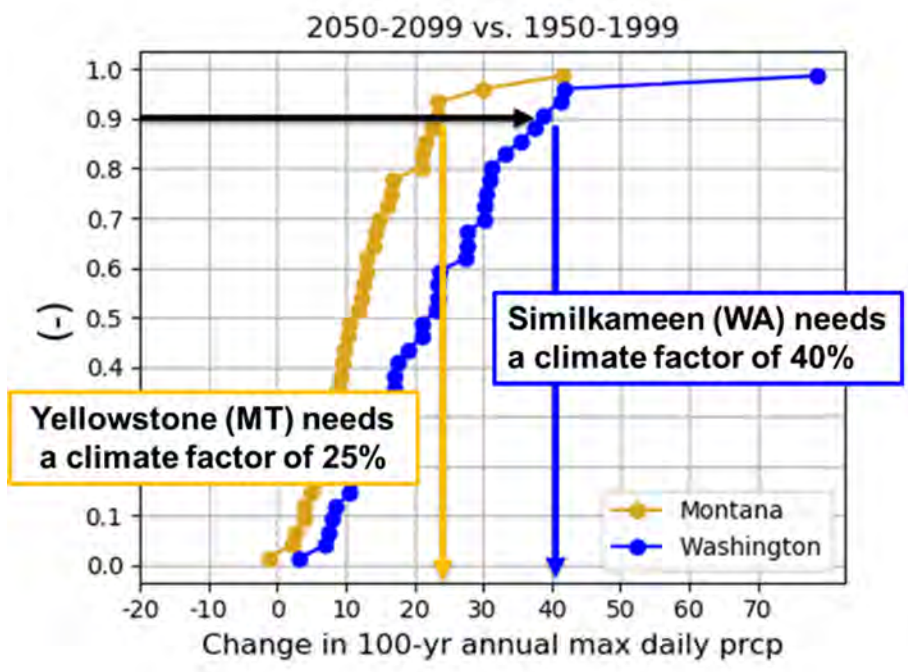


Figure 8-2 Likelihood of change in 100-yr annual maxima daily precipitation as obtained from the climate projection ensemble. For both catchments, each dot corresponds to one symbol on Figure 8-1. The y-axis is obtained from a plotting position formula. The figure shows an example of decision for which the decision-maker wants to account for 90% of the climate projection. In this case, a climate factor of 25% is required for MT while a climate factor of 40% is required for WA.

Following the above-described methodology, Table 8-2 shows for different risk aversion levels the climate factors that should be consider to account for climate change deep uncertainty for all considered basins. We note that required climate factors vary significantly from one catchment to another, especially for the high levels of risk aversion (i.e., P>90%). We also observed that significantly different climate factors are obtained for basins from similar climatic areas. This result could follow from either the small sample size of basins for each climate conditions or from the climate categories themselves that could reflect more the average climate conditions than the in-situ climate and weather processes driving precipitation extreme events.

Table 8-2 Basin scale climate factors for annual maxima precipitation and for different risk aversion level (defined as the percentage P of climate projections accounted for to estimate the climate factors) and for 2050-2099 period.

State	P>50%	P>75%	P>90%	P>95%	P=100%
AZ	15%	30%	40%	45%	75%
MT	15%	20%	25%	25%	45%
CA	25%	30%	35%	40%	55%
GA	15%	25%	35%	45%	50%
IN	25%	30%	40%	40%	55%
TX	15%	30%	40%	40%	50%
CO	15%	20%	25%	25%	30%
ME	25%	35%	45%	45%	55%
WA	25%	35%	40%	45%	80%

8.4.2. SENSITIVITY TO DOWNSCALING

Table 8-3 shows the influence of the choice of a specific downscaling approach on the estimated change in annual maximum daily rainfall (100-yr return). Changes are obtained by combining the climate stress test and the projected location and scale parameters of the precipitation extreme distribution.

Table 8-3 Uncertainty on the estimated change in 100-yr annual maximum daily rainfall following the considered downscaling approach. The changes are estimate for the period 2050-2099 compared to 1950-1999.

GCM	GFDL-ESM2M			HadGEM2-ES			MPI-ESM-LR		
Downscaling	**LOCA**	**RegCM4**	**WRF**	**LOCA**	**RegCM4**	**WRF**	**LOCA**	**RegCM4**	**WRF**
AZ	38.2%	9.8%	6.8%	12.4%	16.6%	22.4%	35.9%	11.6%	1.9%
MT	11.6%	6.4%	22.9%	12.2%	-0.9%	10.4%	16.7%	5.3%	13.1%
CA	39.7%	16.4%	17.9%	35.2%	9.3%	-2.1%	27.9%	32.4%	20.9%
GA	22.8%	41.1%	30.6%	21.1%	46.6%	43.9%	-7.5%	11.4%	-0.8%
IN	25.4%	13.7%	19.9%	18.3%	26.2%	6.8%	36.3%	55.2%	37.7%
TX	7.6%	12.9%	1.9%	41.4%	39.1%	12.2%	22.2%	29.1%	-3.6%
CO	7.0%	3.7%	7.9%	14.2%	11.1%	14.5%	19.9%	14.7%	7.26%
ME	33.2%	15.8%	37.9%	44.6%	21.1%	27.6%	25.4%	9.0%	20.5%
WA	7.4%	3.2%	7.9%	37.4%	10.6%	14.5%	23.5%	14.7%	7.0%

The uncertainty coming from the use of different downscaling approaches appears to be significant for most basins. Overall, we notice that the ranking of the different downscaling approaches (i.e., which one leads to the large increase and which leads to the lowest increase, for instance) is not consistent neither across GCMs nor across basins. For some catchments, the use of different downscaling approaches to infer change in precipitation extreme may even lead to different directions of change. This is for instance the case for the TX catchment and the MPI-ESM-LR climate model for which inference using WRF downscaled time series gives a reduction of about 3.6% while RegCM4 gives an increase by nearly 30%. Another example is the CA catchment for which the HadGEM2-ES projection downscaled with LOCA leads to an increase by 35.2% while the downscaled projection via WRF leads to slight decrease by 2.1%.

8.5. CONCLUSION

Climate factors are used to increase the design value of new infrastructure to account for climate change deep uncertainty. This study has presented a new methodology define climate factors accounting for deep uncertainty. This methodology combines frequency analysis of precipitation extreme and Decision Scaling, a robust decision-making approach. The sensitivity to change in precipitation extreme model parameters is mapped through a stress test approach and climate projections were used to infer likelihood of change of model parameters as seen in the projections. Climate factors can be defined for a given aversion of risk (i.e., accounting for either all or a sub-sample of the available projections). Results show that climate factors vary significantly from one catchment to another, even when located in similar climatic conditions, which highlight the need for in-situ estimate when planning for new infrastructure. We also explored the sensitivity of the climate factor estimates to the considered downscaling approach. The climate factor estimates have revealed to be significantly influenced by the choice of the downscaling approach, which is source of motivation for using an ensemble of downscaling approaches. Further work should focus on applying this methodology to larger set of case studies, extending the ensemble of downscaling approach and to compare the climate factor estimates with factors that are used already and directly obtained from the raw climate model outputs.

8.6. REFERENCES

Brown, C.M., Steinschneider, S., Wi, S., Weatherly, J., Case, M., Hayden, T., Koster, A., Mearns, L.O., Bukovsky, M.S., McCrary, R., 2016. Climate Risk Assessment: Technical Guidance Manual for DoD Installations and Built Environment (Technical Report No. AD1029408). University of Massachusetts Amherst United States.

Collins, W.J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C.D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., Woodward, S., 2011. Development and evaluation of an Earth-System model – HadGEM2. Geosci. Model Dev. 4, 1051–1075. https://doi.org/10.5194/gmd-4-1051-2011

Dunne, J.P., John, J.G., Adcroft, A.J., Griffies, S.M., Hallberg, R.W., Shevliakova, E., Stouffer, R.J., Cooke, W., Dunne, K.A., Harrison, M.J., Krasting, J.P., Malyshev, S.L., Milly, P.C.D., Phillipps, P.J., Sentman, L.T., Samuels, B.L., Spelman, M.J., Winton, M., Wittenberg, A.T., Zadeh, N., 2012. GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics. J. Clim. 25, 6646–6665. https://doi.org/10.1175/JCLI-D-11-00560.1

Dunne, J.P., John, J.G., Shevliakova, E., Stouffer, R.J., Krasting, J.P., Malyshev, S.L., Milly, P.C.D., Sentman, L.T., Adcroft, A.J., Cooke, W., 2013. GFDL's ESM2 global coupled climate–carbon earth system models. Part II: carbon system formulation and baseline simulation characteristics. J. Clim. 26, 2247–2267.

Giorgetta, M.A., Jungclaus, J., Reick, C.H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K.D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., Stevens, B., 2013. Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5: Climate Changes in MPI-ESM. J. Adv. Model. Earth Syst. 5, 572–597. https://doi.org/10.1002/jame.20038

Giorgi, F., Coppola, E., Solmon, F., Mariotti, L., Sylla, M., Bi, X., Elguindi, N., Diro, G., Nair, V., Giuliani, G., Turuncoglu, U., Cozzini, S., Güttler, I., O'Brien, T., Tawfik, A., Shalaby, A., Zakey, A., Steiner, A., Stordal, F., Sloan, L., Brankovic, C., 2012. RegCM4: model description and preliminary tests over multiple CORDEX domains. Clim. Res. 52, 7–29. https://doi.org/10.3354/cr01018

Hosking, J.R.M., 1990. L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics. J. R. Stat. Soc. Ser. B 105–124.

Hosking, J.R.M., Wallis, J.R., 1997. Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge ; New York.

Jenkinson, A. F. (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society, 81(348), 158–171.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15, 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Mearns, L.O., et al., 2017: The NA-CORDEX dataset, version 1.0. NCAR Climate Data Gateway, Boulder CO, accessed [6-11-2019], https://doi.org/10.5065/D6SJ1JCH

Pierce, D. W., Cayan, D. R., & Thrasher, B. L. (2014). Statistical Downscaling Using Localized Constructed Analogs (LOCA)*. Journal of Hydrometeorology, 15(6), 2558–2585. https://doi.org/10.1175/JHM-D-14-0082.1

Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., & Hegewisch, K. C. (2015). Improved Bias Correction Techniques for Hydrological Simulations of Climate Change*. Journal of Hydrometeorology, 16(6), 2421–2442. https://doi.org/10.1175/JHM-D-14-0236.1

Powers, J.G., Klemp, J.B., Skamarock, W.C., Davis, C.A., Dudhia, J., Gill, D.O., Coen, J.L., Gochis, D.J., Ahmadov, R., Peckham, S.E., Grell, G.A., Michalakes, J., Trahan, S., Benjamin, S.G., Alexander, C.R., Dimego, G.J., Wang, W., Schwartz, C.S., Romine, G.S., Liu, Z., Snyder, C., Chen, F., Barlage, M.J., Yu, W., Duda, M.G., 2017. The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions. Bull. Am. Meteorol. Soc. 98, 1717–1737. https://doi.org/10.1175/BAMS-D-15-00308.1

Spence, C.M., Brown, C.M., 2016. Nonstationary decision model for flood risk decision scaling. Water Resour. Res. 52, 8650–8667. https://doi.org/10.1002/2016WR018981

Steinschneider, S., Wi, S., Brown, C., 2015b. The integrated effects of climate and hydrologic uncertainty on future flood risk assessments. Hydrol. Process. 29, 2823–2839. https://doi.org/10.1002/hyp.10409

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An Overview of CMIP5 and the Experiment Design. Bull. Am. Meteorol. Soc. 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

9. COMPARING FLOOD PROJECTION APPROACHES ACROSS HYDRO-CLIMATOLOGICAL DIVERSE RIVER BASINS

A slightly different version of this chapter has been submitted to Water Resources Research.

Schlef, K.E., François, B., Brown, C., submitted. Comparing Flood Projection Approaches across Hydry-climatological Diverse River Basins. Water Resource Research

9.1. EXECUTIVE SUMMARY

The challenge of estimating design flood magnitude under climate change has led to the development of multiple approaches to long-term flood projection: stationarity, hydrologic simulation, and informed parameter (comprised of both trend-informed and climate-informed). This is the first study to compare these various approaches and to do so across a set of hydro-climatologically diverse basins, which are located throughout the contiguous United States, using a split-sample test conducted over the historic period. The results provide a quantitative perspective on key long-term flood projection issues, showing that assuming stationarity can lead to large biases, but no nonstationary alternative provides satisfactory performance in the validation period. Additionally, the results show that the climate-informed and hydrologic simulation approaches are more flexible than the trend-informed approach, but, as expected, use of general circulation model simulations as forcing introduces additional bias. Finally, the results show that even though performance seems to be largely specific to each basin and each approach, flood magnitude in the validation period is likely to be overestimated by at least one of the approaches; thus, the highest value in the ensemble could function as a conservative estimate for design. Future research and application efforts could include performing more comprehensive comparison studies and developing improved model structures and projections of climatic forcing data, but our conclusion is that such effort is best spent on using all projection approaches as an ensemble to inform uncertainty-based design paradigms.

9.2. TECHNICICAL APPROACH

There are currently three main approaches to long-term flood projection (François et al., 2019): (1) stationarity, (2) hydrologic simulation, and (3) informed parameter. The stationarity approach maintains that the parameters of the flood distribution are time-invariant; clear guidance for its implementation exists at the federal level in the United States (England et al., 2018). The hydrologic simulation approach uses projected climate to force a hydrologic model and has been employed at spatial scales ranging from small basins to the global scale (e.g., Arnell & Gosling, 2016; Madsen et al., 2014; Mallakpour et al., 2019; Milly et al., 2002). The informed-parameter approach conditions the parameter(s) of the flood distribution on time-varying covariates and can take two forms: the trend-informed approach in which the covariate is time (e.g., Hu et al., 2017; Luke et al., 2017; Šraj et al., 2016), and the emergent climate-informed approach in which the covariate(s) are (projected) large-scale oceanic-atmospheric fields (e.g., Bracken et al., 2018; Delgado et al., 2014; Schlef et al., 2018; Steirou et al., 2019) or synoptic-scale patterns (Knighton et al., 2019).

The literature on flood projection approaches contains many conflicting claims regarding the validity and strengths and weaknesses of each approach. The assumption of stationarity has been maintained based on the following claims: (1) there is no established approach for nonstationary long-term flood projection (i.e., this is an unresolved challenge) (England et al., 2018), (2) it is impossible to statistically distinguish a stationary process with short- and long-term persistence from a nonstationary process driven by anthropogenic impacts on land and climate given our current state of knowledge (Lins & Cohn, 2011), (3) relatedly, stationary Hurst-Kolmorgorov dynamics are both adequate and appropriate for representing hydrological systems (Koutsoyiannis, 2011), and (4) finally, convincing evidence for non-stationarity in the historic record is elusive (Archfield et al., 2016; Hirsch & Ryberg, 2012; Luke et al., 2017; Peterson et al., 2013; Villarini et al., 2009; Wouter et al., 2017).

This study implements long-term flood projection approaches for nine hydro-climatologically diverse river basins that span the contiguous United States (Figure 9-1); specifically, the stationarity, hydrologic simulation, and trend- and climate-informed approaches are implemented. In split-sample tests over the historical period, the performance of each approach is compared during calibration (1951-1980) and validation (1981-2010) periods. The results facilitate a discussion of the strengths and weaknesses of each approach based on directly comparing the modeling results. The results also engender recommendations regarding the appropriate use of each approach.

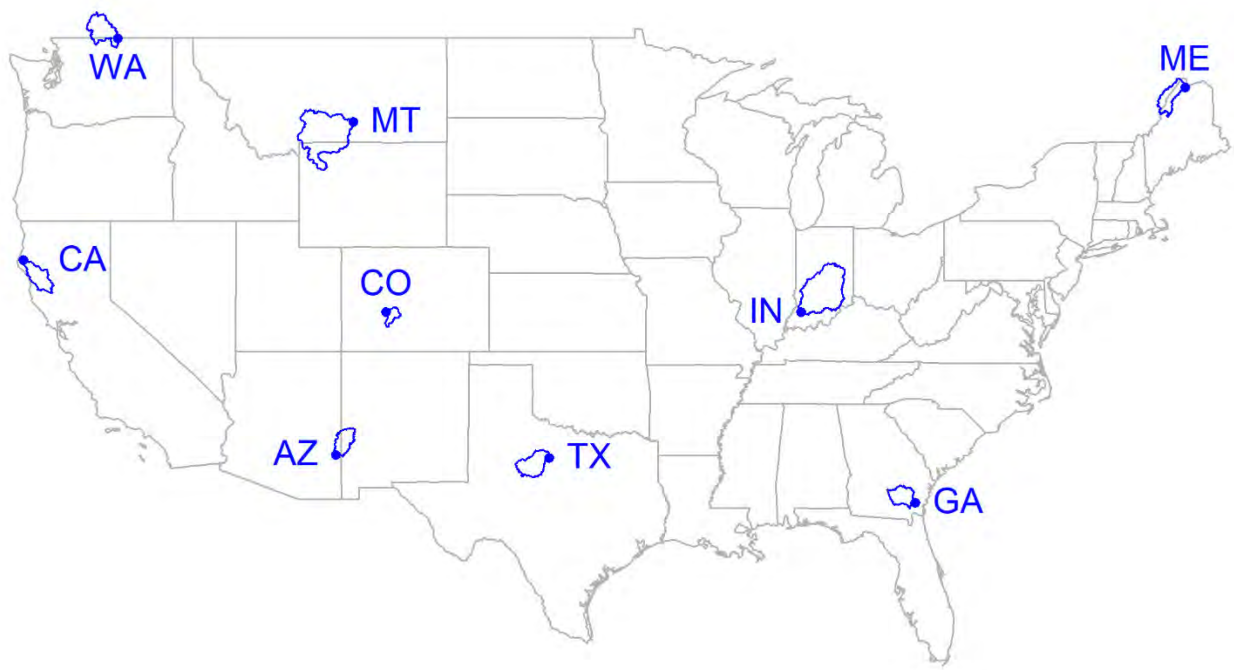


Figure 9-1 Map of gages and associated basins (labeled by abbreviation of the state where the gage is located).

9.3. DATA AND METHODS

9.3.1. CASE STUDIES

Characteristics of the nine case study river basins are summarized in Table 9-1. Throughout, the basins are referenced by the abbreviation of the state that contains the gage as indicated in Table 9-1. The basins were chosen such that (1) each gage is part of the United States Geological Survey (USGS) Hydro-Climatic Data Network (Landwehr & Slack, 1992) indicating unimpaired status at least up to 1988, (2) each gage has complete daily streamflow data from water years 1950 through 2010 (i.e., 10/1/1949-9/30/2010), and (3) the final set spans a variety of climate classifications (Kottek et al., 2006), flood generating processes (Berghuijs et al., 2016), locations (e.g., coastal vs. inland), and altitudes across the contiguous United States.

For each basin, the following were calculated based on the water year over the calibration period: (1) the average of the R-B flashiness index (Baker et al., 2004), (2) annual maximum daily streamflow as the flood timeseries, and (3) the mean flood date and flood seasonality index based on circular statistics (Berghuijs et al., 2016; Villarini, 2016; Ye et al., 2017). Although not used as criteria for choosing basins, the final set spans a variety of changes in the flood duration curve between the calibration and validation periods; in the AZ basin, there is minimal change, in the CO, IN, and ME basins there is an increase of varying magnitudes, in the CA, TX, and WA basins there is a decrease of varying magnitudes, and in the GA and MT basins the shape changes (i.e., small return periods decrease in magnitude while large return periods increase).

9.3.2. Flood distribution and fitting

The log-normal distribution was chosen to represent floods, in part due to the simplicity of interpretation (i.e., the location and scale parameters correspond to the mean and standard deviation of the log of the data); l-moments diagrams (not shown) indicate that the log-normal distribution is usually suitable for the data. The distribution was fit using a Bayesian model, where the prior distribution for the location parameter or covariate coefficients is a normal distribution with zero mean and standard deviation of five, and the prior distribution for the scale parameter is a uniform distribution between zero and 10,000. A Bayesian model was chosen because it naturally estimates parameter uncertainty and hence credible intervals, thus removing any need for bootstrapping. The Bayesian models were fit with JAGS in R (Plummer, 2016; Yu-Sung & Yajima, 2015) using three chains each with 2,000 samples of which 1,000 was discarded as burn-in, resulting in a total of 3,000 samples. For all models, the potential scale reduction factor, also known as Gelman's R, and the effective sample size were well within accepted rules of thumb (less than 1.1 and greater than 300, respectively).

The deviance information criterion (DIC) and statistical significance of model parameters (Table 9-2) can be used to evaluate the model fit of all approaches except hydrologic simulation. These metrics cannot be used for hydrologic simulation because the DIC is calculated relative to simulated, not observed, flood magnitudes, and because no statistical significance is associated with the hydrologic model parameters. Lower DIC values are better, but based on rules of thumb, differences of five or less are negligible. Because all the Bayesian models were fit to the same time period of 30 years with the same number of 3,000 samples, all DIC values for a given basin are directly comparable. The DIC values of the stationary and trend-informed approaches are nearly identical, but larger than the DIC values of the climate-informed approach. For the trend- and climate-informed approaches, where the location parameter is linearly regressed on a covariate, the covariate was considered significant if the coefficient's 95% credible interval does not include zero. The model parameters of the climate-informed and stationary approaches are always statistically significant; however, for the trend-informed approach, $\alpha_{0,c}$ is never significant and $\alpha_{1,c}$ is only significant when a trend is detected, which occurs in five out of the nine basins (see Section 9.3.3 for explanation of $\alpha_{0,c}$ and $\alpha_{1,c}$).

Table 9-1 Characteristics of case study river basins. The state indicates the location of the gage and the abbreviation is used to reference the basin throughout the text. Refer to Kottek et al. (2006) for further explanation of climate. Flood cause is from Berghuijs et al. (2016).

State	**Arizona (AZ)**	**California (CA)**	**Colorado (CO)**	**Georgia (GA)**	**Indiana (IN)**	**Maine (ME)**	**Montana (MT)**	**Texas (TX)**	**Washington (WA)**
Basin									
USGS Gage	09444500	11477000	09119000	02228000	03374000	01010500	06214500	08085500	12442500
Area (sq. mi.)	2766	3113	1061	2790	11125	2680	11807	3988	3550
Gage Elevation (ft.)	3436	36	7629	15	399	590	3081	1164	1138
Gage Latitude	33.05	40.49	38.52	31.22	38.51	47.11	45.8	32.93	48.98
Gage Longitude	-109.3	-124.1	-106.94	-81.87	-87.29	-69.09	-108.47	-99.22	-119.62
Climate									
Classification	BSk	Csb	Dfb	Cfa	Cfa	Dfb	BSk	Cfa	Dfb
Climate	Arid	warm temperate	snow	warm temperate	warm temperate	snow	arid	warm temperate	snow
Precipitation	steppe	dry summer	fully humid	fully humid	fully humid	fully humid	steppe	fully humid	fully humid
Temperature	cold steppe/desert	warm summer	warm summer	hot summer	hot summer	warm summer	cold steppe/desert	hot summer	warm summer
Floods									
Mean Flood	10-07	01-22	05-26	03-26	03-17	04-30	06-13	07-20	05-30
Seasonality	0.4	0.91	0.89	0.3	0.68	0.86	0.99	0.38	0.98
Flashiness	0.33	0.34	0.08	0.09	0.1	0.16	0.07	0.54	0.08
Flood Cause	daily & weekly P	daily & weekly P, P excess	snow+rain	daily P	P excess	snow+rain	daily & weekly P, snow+rain	daily & weekly P	snow+rain

Meaning of flood cause (Berghuijs et al., 2016)

- daily precipitation (P) means that “flooding is caused by the single largest precipitation events”
- weekly P means that “flooding is caused by the single largest series of precipitation events”
- P excess means that “flood is caused by the single largest precipitation excess event” where “precipitation excess is defined as the rainfall excess compared to available soil moisture capacity
- snow+rain means that “flooding is caused by the single largest snowmelt or rain-on-snow event”

Table 9-2 Model parameters and performance. Where applicable, the values are mean (standard deviation) and the mean is bolded if statistically significant. Months are those used for finding the sum of precipitation (e.g., ONDJFM is October-March).

State	Arizona (AZ)	California (CA)	Colorado (CO)	Georgia (GA)	Indiana (IN)	Maine (ME)	Montana (MT)	Texas (TX)	Washington (WA)
Stationary									
DIC	107	78	58	68	42	27	7	95	29
μ_c	**8** (0.27)	**12** (0.16)	**6.5** (0.12)	**9.6** (0.14)	**11** (0.088)	**11** (0.067)	**11** (0.05)	**8.5** (0.22)	**9.7** (0.072)
σ_c	**1.4** (0.2)	**0.88** (0.12)	**0.63** (0.089)	**0.75** (0.1)	**0.48** (0.067)	**0.38** (0.052)	**0.27** (0.038)	**1.2** (0.16)	**0.39** (0.053)
Trend-Informed									
DIC	107	79	58	68	42	25	8	95	31
$\alpha_{0,c}$	-0.3 (5)	0.97 (4.9)	0.082 (4.9)	-0.62 (5)	-0.022 (4.8)	-2 (4.8)	1.4 (4.6)	0.025 (5)	2.3 (4.9)
$\alpha_{1,c}$	0.0043 (0.0025)	**0.0055** (0.0025)	0.0033 (0.0025)	**0.0052** (0.0025)	**0.0056** (0.0025)	**0.0065** (0.0024)	**0.0047** (0.0023)	0.0043 (0.0025)	0.0038 (0.0025)
σ_c	**1.4** (0.2)	**0.89** (0.12)	**0.62** (0.087)	**0.74** (0.1)	**0.48** (0.068)	**0.36** (0.05)	**0.27** (0.037)	**1.2** (0.17)	**0.4** (0.056)
Climate-Informed									
Months	ONDJFM	NDJFMA	NDJFMA	DJFMA	DJFMAM	DJFMAM	DJFMAM	MJJAS	NDJFM
DIC	84	59	46	51	34	11	-13	89	18
$\alpha_{0,c}$	**5.8** (0.43)	**9.7** (0.42)	**5.2** (0.33)	**7.4** (0.44)	**9.2** (0.54)	**9.7** (0.21)	**9.6** (0.2)	**6.7** (0.61)	**9** (0.21)
$\alpha_{1,c}$	**0.01** (0.0017)	**0.0015** (0.00029)	**0.0046** (0.0012)	**0.0045** (0.0009)	**0.0035** (0.001)	**0.0024** (0.00049)	**0.003** (0.00055)	**0.005** (0.0017)	**0.0021** (0.00054)
σ_c	**0.95** (0.13)	**0.62** (0.089)	**0.5** (0.072)	**0.55** (0.078)	**0.41** (0.058)	**0.28** (0.04)	**0.19** (0.027)	**1** (0.14)	**0.32** (0.044)
Hydrologic Simulation									
NSE calibration	0.81 (0.0095)	0.91 (0.0014)	0.79 (0.0076)	0.81 (0.0095)	0.88 (0.027)	0.84 (0.0025)	0.9 (0.013)	0.71 (0.0091)	0.87 (0.0032)
validation	0.62 (0.043)	0.89 (0.004)	0.76 (0.019)	0.82 (0.0079)	0.91 (0.0063)	0.64 (0.015)	0.83 (0.061)	0.38 (0.041)	0.42 (0.036)
DIC calibration	104 (2.5)	61 (1.2)	42 (2.4)	46 (1.6)	41 (2.2)	14 (1.4)	10 (3.6)	91 (2.6)	6 (2.3)
validation	103 (3.1)	37 (1.3)	56 (2.2)	59 (1.5)	33 (1.2)	-7 (2.1)	0 (3.2)	74 (1.9)	7 (3.2)

9.3.3. Flood Projections Approaches

Each flood projection approach was applied separately to each basin. To fit the models for approaches that incorporate climate data (climate-informed and hydrologic simulation), observed gridded daily precipitation and daily minimum and maximum temperature were obtained from Livneh et al. (2013). Those fitted models were also forced with simulation data for the period 1950 through 2010 from 32 GCMs; the GCMs were forced with the representative concentration pathway of 8.5 W/m2 (RCP 8.5) associated with the 5th generation of the Coupled Model Intercomparison Project (CMIP5) of the Intergovernmental Panel on Climate Change (IPCC) (Taylor et al., 2012), which has been bias-corrected and downscaled following the locally constructed analogs approach (Pierce et al., 2014, 2015).

9.3.3.1. STATIONARITY

For the stationary approach, the parameters of the log-normal distribution were fit to the calibration period and assumed to be the same for the validation period (e.g., England et al., 2018); that is,

$$\log Q \sim N(\mu_c, {\sigma_c}^2) \tag{9.1}$$

where Q is the flood magnitude, μ and σ are the location and scale parameters, respectively, of the normal distribution N, and the subscript c indicates calibration.

9.3.3.2. TREND-INFORMED

For the trend-informed approach, the location parameter of the log-normal distribution was linearly regressed on time (e.g., Luke et al., 2017); parameter values were fit to the calibration period and assumed valid for the validation period; that is:

$$\log Q \sim N\left(\alpha_{0,c} + \alpha_{1,c}t, {\sigma_c}^2\right) \qquad 9.2$$

where t is time and a_0 and a_1 are the regression coefficients. Only five of the nine basins have a statistically significant $\alpha_{1,c}$, all of which are positive, indicating an increasing trend.

9.3.3.3. CLIMATE-INFORMED

For the climate-informed approach, the location parameter of the log-normal distribution was linearly regressed on a covariate (e.g., Schlef et al., 2018); parameter values were fit to the calibration period and assumed valid for the validation period; that is:

$$\log Q \sim N\left(\alpha_{0,c} + \alpha_{1,c}x, {\sigma_c}^2\right) \qquad 9.3$$

where x is the covariate. Ideally, the covariate(s) should be large-scale oceanic-atmospheric patterns (Schlef et al., 2018).

Here, the covariate was the sum of monthly precipitation over a given period of months; this choice of covariate was motivated by two reasons: (1) the climatic forcing data is the same as for the hydrologic simulation approach and (2) identification of sufficiently statistically significant covariates based on large-scale oceanic-atmospheric patterns for every basin was prohibitively difficult. For each basin, the period of months was chosen to obtain (1) physical interpretability (e.g., floods in the WA basin are primarily caused by melting of snow accumulated over the winter, so the months of November through March are a logical choice) and (2) highly significant correlation between the flood and covariate timeseries. Monthly precipitation was calculated by aggregating the observed daily precipitation from 1/16th to 1/8th degree grid resolution, summing over all days in the month to obtain monthly precipitation, and then averaging over all the grid cells in the basin. The $\alpha_{1,c}$ values are statistically significant, with positive values, across all case study basins.

9.3.3.4. HYDROLOGIC SIMULATION

The hydrologic simulation approach was implemented with a distributed version of the Soil Moisture Accounting model (SACSMA) coupled with a river routing and snow model (Brown et al., 2016), running on a daily time step at 1/8th degree grid resolution; all basins had good Nash-Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970) values during the calibration period (at least 0.7). Once calibrated, the model was used to simulate the calibration and validation periods; the resulting streamflow timeseries were post-processed to calculate the annual maximum daily streamflow. The log-normal distribution was then fit separately to each period; that is,

$$\log Q_c \sim N(\mu_c, {\sigma_c}^2);\ \log Q_v \sim N(\mu_v, {\sigma_v}^2) \qquad 9.4$$

where the subscript v indicates validation. Flow length and digital elevation model data used to define the hydrologic model were calculated in ArcGIS using digital elevation data from the CGIAR-CSI SRTM 90m Digital Elevation Database v4.1 (Jarvis et al., 2008) following the procedure provided in Wi et al. (2017).

For each basin, 25 sets of model parameters were calibrated using a genetic algorithm for the calibration period, plus a one-year warm-up, by maximizing the NSE. The observed daily precipitation and minimum and maximum temperature used as forcing data, initially at 1/16th degree grid resolution, was aggregated to 1/8th degree; mean temperature was obtained by averaging the minimum and maximum temperature. In each period, because the log-normal distribution was fit separately to the 25 timeseries corresponding to the different model parameter sets, 120 samples were randomly chosen from each set of 3,000 samples so that the total number of samples across all 25 parameter sets was 3,000. The DIC values from the 25 model parameter sets were averaged to obtain a mean and standard deviation.

9.3.4. Model performance metrics

A variety of evaluative techniques were used to assess the performance of each long-term flood projection approach. Given the large amount of generated data, to maintain clarity of presentation, the results often focus on the 100-year flood magnitude, chosen for its high societal importance in part due to the National Flood Insurance Program (FEMA, 2011; NRC, 2000), notwithstanding scientific critiques (Read & Vogel, 2015). As needed, the log-normal distribution fit to the observed data (i.e., the stationary approach) was used as the null hypothesis or control set representing true values.

For evaluating performance in one basin only, two metrics were used. First, flood frequency curves in the calibration and validation periods from each approach were compared to gain a broad view of performance across multiple return periods. Second, the 100-year flood magnitude was used to evaluate the impact of using GCM simulations to force the climate-informed and hydrologic simulation approaches.

For evaluating performance across all basins, in addition to evaluating the impact of using GCM simulations as forcing, three metrics were used: bias, spread as a measure of uncertainty, and change (described below). In the case of forcing by GCM simulations for the climate-informed and hydrologic simulation approaches, the three metrics were calculated separately for each GCM. The median value across GCMs is reported; statistical significance of the GCM median value was not assessed.

To calculate bias, first an unpaired two-sided t-test was performed between a treatment set, consisting of the 3,000 samples of the 100-year flood magnitude from a given approach and period (i.e., calibration or validation), and the control set in the same period. Bias was calculated as the difference between the treatment set mean and control set mean. To enable comparison across basins, the results report percent bias, calculated as the bias divided by the control set mean.

Spread was calculated as the difference between the 2.5th and 97.5th percentiles of the 100-year flood magnitude across the 3,000 samples from a given approach, basin, and period (i.e., calibration or validation). To enable comparison across basins, the percent spread was calculated as the spread divided by the 50th percentile of the 100-year flood estimate. Subsequently, percent spread bias was calculated as the difference between the percent spread of the treatment and control sets, divided by the percent spread of the control set. No statistical significance is associated with the spread.

Change was calculated for both the 2- and 100-year flood magnitudes; the 2-year flood is considered representative of more frequent nuisance flooding. To calculate change, first an unpaired two-sided t-test was performed between the 3,000 samples of the flood magnitude from the calibration and validation periods of a given approach. Change was calculated as the difference between the validation period mean and the calibration period mean. To enable comparison across basins, the results report percent change, calculated as the change divided by the calibration period mean.

9.4. RESULTS AND DISCUSSION

The results first evaluate performance for the WA basin as an example; specifically, the results examine bias in flood frequency curves and examine the impact of using GCM simulations as forcing data for the climate-informed and hydrologic simulation approaches. Subsequently, the results then examine aggregated results for all basins, specifically looking at bias, uncertainty, and change in the 100-year flood magnitude and change in the 2-year flood magnitude. The results also evaluate the uncertainty arising from the ensemble of GCM simulations used as forcing.

9.4.1. EXAMPLE: EVALUATING PERFORMANCE IN THE WA BASIN

As an example, performance for the WA basin is evaluated. Importantly, while the results discussed below are valid for the WA basin, they are not necessarily generalizable to the other basins.

9.4.1.1. BIAS ILLUSTRATED WITH FLOOD FREQUENCY CURVES

A broad view of the performance of each approach is gained by examining the flood frequency curves. In the WA basin, between the calibration and validation period, there is a large decrease in the flood frequency curve (Figure 9-2a). All approaches overestimate flood magnitude in the validation period, such that many of the observed floods are below the modeled 95% credible intervals (Figure 9-2b-d). The 95% credible intervals of the climate-informed and hydrologic simulation approaches are smaller than those obtained when the distribution is fit directly to the observed floods; in contrast, the trend-informed credible intervals are larger.

The reasons for a positive bias in the validation period are different for each approach. The stationary approach is based on the calibration period and thus cannot reflect any changes in the validation period (Figure 9-2a). The trend-informed approach adequately models the calibration period, but, constrained by a positive albeit insignificant trend coefficient ($\alpha_{1,c}$) (Table 9-2), simulates a significance increase in the validation period (Figure 9-2b). The climate-informed approach also adequately models the calibration period but exhibits minimal change in the validation period (Figure 9-2c), indicating that the distribution of the precipitation index is approximately the same in both periods and that other hydrologic processes not in the model (e.g., snowpack dynamics) are affecting the predominately snowmelt or rain-on-snow floods in the basin (Table 9-1). The hydrologic simulation approach is negatively biased in the calibration period and exhibits an increase in the validation period; ironically, the modeled calibration period nearly corresponds to the observed validation period, and vice versa (Figure 9-2d). While a detailed investigation of the causes of this behavior is beyond the scope of this study, some possibilities are that negative bias in the calibration period may be due to insufficient snow accumulation or incorrect timing of snowmelt, while the increase between the two periods may be due to temperature changes that affect the snow-rainfall partitioning of precipitation.

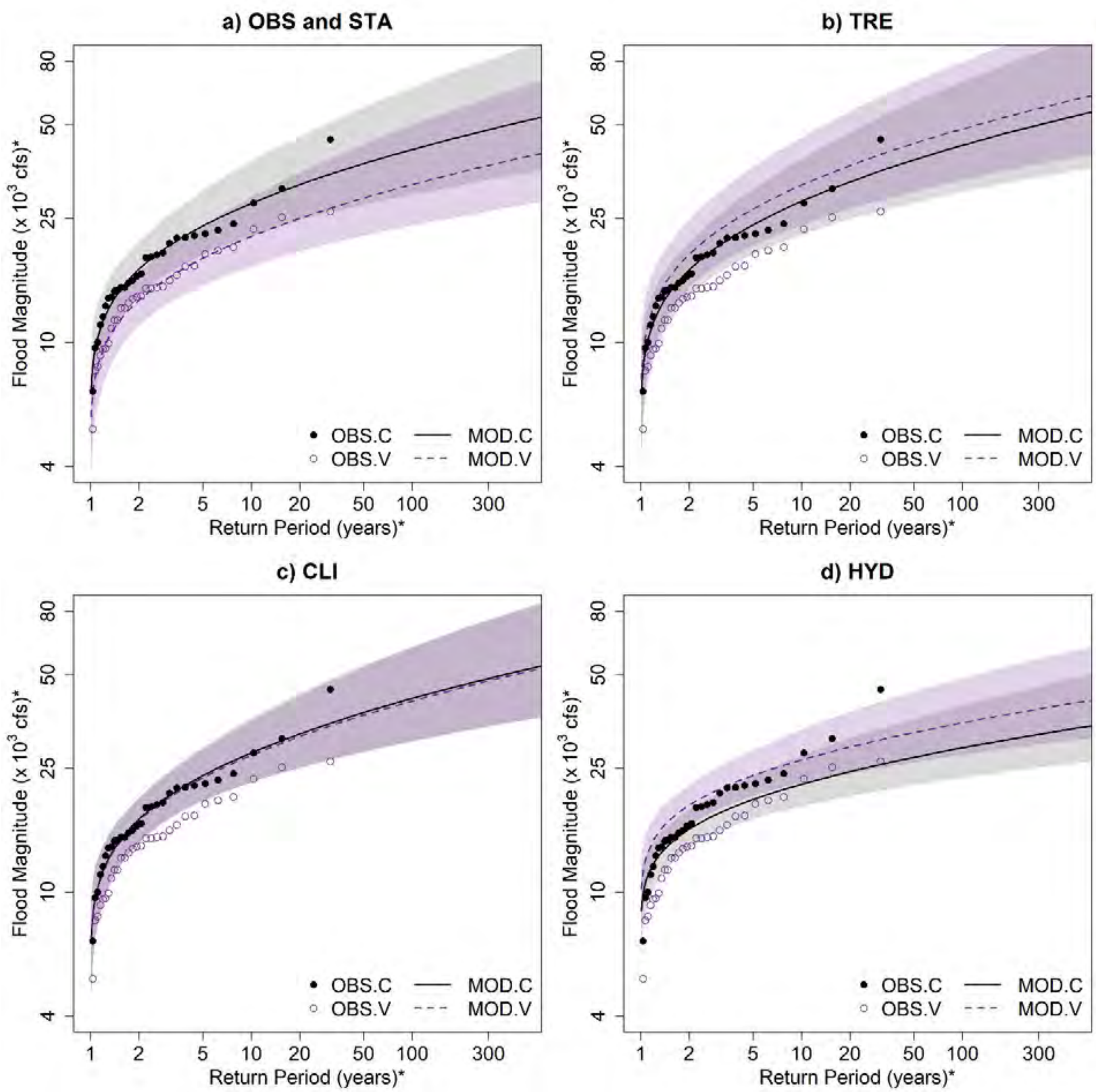


*Figure 9-2 Flood frequency curves for the WA basin for each method. OBS is observed, STA is stationary, TRE is trend-informed, CLI is climate-informed, and HYD is hydrologic simulation. MOD is modeled and the suffixes C and V indicate calibration and validation, respectively. *Axis is on log scale.*

9.4.1.2. IMPACT OF USING GCM SIMULATIONS AS FORCING DATA

The impact of using GCM simulations, rather than observed climate, as forcing for the climate-informed and hydrologic simulation approaches, is evaluated using the 100-year flood. In the WA basin, as expected, additional bias relative to the modeled 100-year flood is introduced when GCM simulations are used to the force the model (Figure 9-3). Although the exact magnitude varies by GCM, the direction of bias relative to the modeled 100-year flood in the calibration and validation periods is nearly the same across all GCMs. The GCM bias relative to the model is negative for the climate-informed approach, indicating that GCM simulation of precipitation over the winter and spring months is too low. The GCM bias relative to the model is positive for the hydrologic simulation approach, which could be due to a variety of reasons not explored here. As a result, for both approaches, the bias relative to the observed value of the 100-year flood is negative in the calibration period and positive in the validation period. Interestingly, the range of values simulated across all GCMs is much larger for the hydrologic simulation approach than for the climate informed approach. Notably, GCM bias can offset model bias and engender inaccurate conceptions of model performance; for example, during the calibration period for the hydrologic simulation approach, forcing the model with GCM simulations produces better estimates of the 100-year flood than forcing the model with observed climate data.

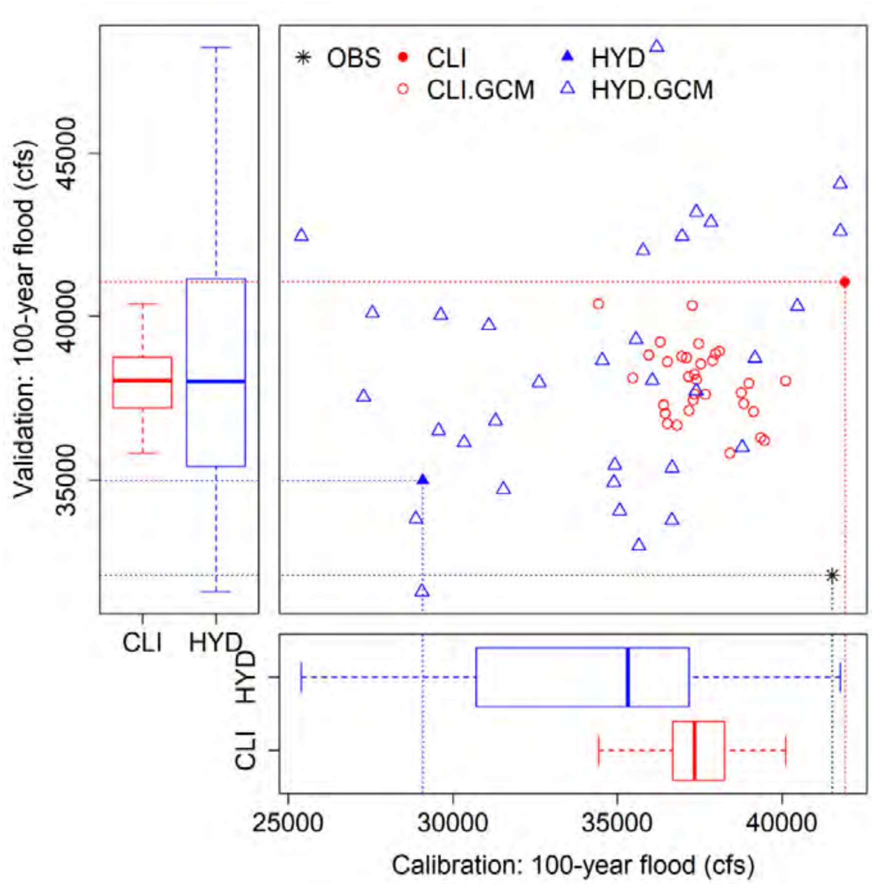


Figure 9-3 The median 100-year flood magnitude for the WA basin in calibration and validation periods from observations (OBS) and from the climate-informed and hydrologic simulation approaches forced with observed climate (CLI and HYD, respectively) and forced with GCM simulations of climate (suffix GCM). The boxplots are of the 32 GCMs.

9.4.2. Evaluating performance across all basins

In aggregated form, results for bias, spread as a measure of uncertainty, and change in the 100-year flood magnitude are presented for all approaches and all basins. Changes in the 2-year flood magnitude are also examined as well as the uncertainty in flood magnitude estimates from the ensemble of GCMs.

9.4.2.1. BIAS AND SPREAD

The percent bias of the 100-year flood estimate during the calibration period enables comparison of the ability of all approaches to simulate flood magnitude (Figure 9-4a). Using the fit to the observed data in the calibration period (i.e., the stationary approach) as the null hypothesis, the hydrologic simulation approach is always negatively biased, likely due to model structure errors, inadequate resolution of the forcing data, and the fact that the model is calibrated to the full timeseries and not only annual peaks. In contrast, the trend- and climate-informed approaches have relatively small but significant biases at only five and three basins, respectively. While the trend-informed bias is a statistical artifact, the bias of the climate-informed approach can be explained physically; the low seasonality, high flashiness, and late mean flood date characteristics of the AZ and TX basins are difficult for the climate-informed model to adequately represent, while the strong influence of soil moisture in the IN basin (e.g., Schlef et al., 2018) is not included in the climate-informed model (Table 9-1). In the calibration period, as expected, GCM forcing data causes additional bias beyond that of the model (Figure 9-4a), sometimes offsetting the model bias (e.g., the CO and MT basins for the hydrologic simulation approach). The direction and magnitude of this additional bias varies by approach (e.g., negative for the climate-informed approach but positive for the hydrologic simulation approach in the MT basin) and by basin (e.g., for the climate-informed approach, positive in the GA basin and negative in the CO basin). Overall however, for a given basin, the bias resulting from forcing with GCM simulations is generally on the same order of magnitude as the bias when observations are used as forcing.

In the validation period, using the fit to the observed data in the validation period as the null hypothesis, all approaches are statistically significantly biased at all or nearly all basins (Figure 9-4b). When forced with observed data, the hydrologic simulation approach remains negatively biased, except for the AZ and WA basins; for other approaches, the direction of bias is basin-dependent. The absolute magnitude of bias is relatively similar across all approaches and basins, except for the TX and CA basins, where a large observed decrease in flood magnitude (discussed subsequently, see Figure 9-6) is not well-simulated by the stationary and trend- and climate-informed approaches. As in the calibration period, when GCM simulations are used to force the climate-informed and hydrologic simulation approaches, additional bias is introduced; the direction and magnitude varies by approach and basin but is generally on the same order of magnitude for each basin. Notably, at least one approach overestimates the 100-year flood magnitude for all basins except GA; this holds true even when physically-based approaches forced with observations are not included.

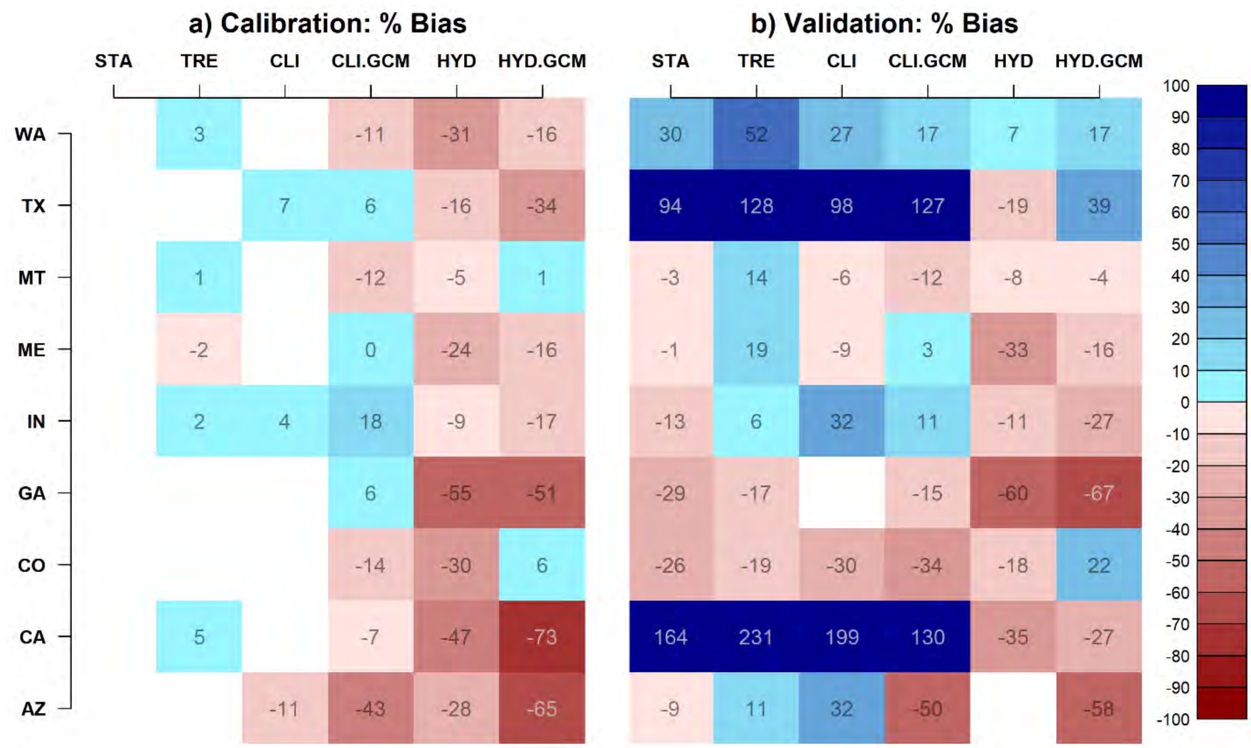


Figure 9-4 Percent bias of the 100-year flood magnitude in (a) calibration and (b) validation periods for each approach (OBS is observed, STA is stationary, TRE is trend-informed, CLI is climate-informed, and HYD is hydrologic simulation) for each basin (indicated by the state abbreviation). White indicates statistically insignificant values. Some values are outside the color range.

In the calibration period, the uncertainty associated with the 100-year flood estimate is reduced relative to the null hypothesis across nearly all approaches and basins (Figure 9-5a). In the validation period, whether the uncertainty associated with the 100-year flood estimate from the various approaches is smaller or larger than the null hypothesis depends on the basin (Figure 9-5b). The uncertainty tends to be larger for the TX, IN, and CA basins regardless of approach, and smaller for the other basins. While the use of GCM simulations as forcing alters the magnitude of percent spread bias, the difference in magnitude is usually not large and infrequently results in a change of sign; one of the few exceptions is the CO basin for the hydrologic simulation approach.

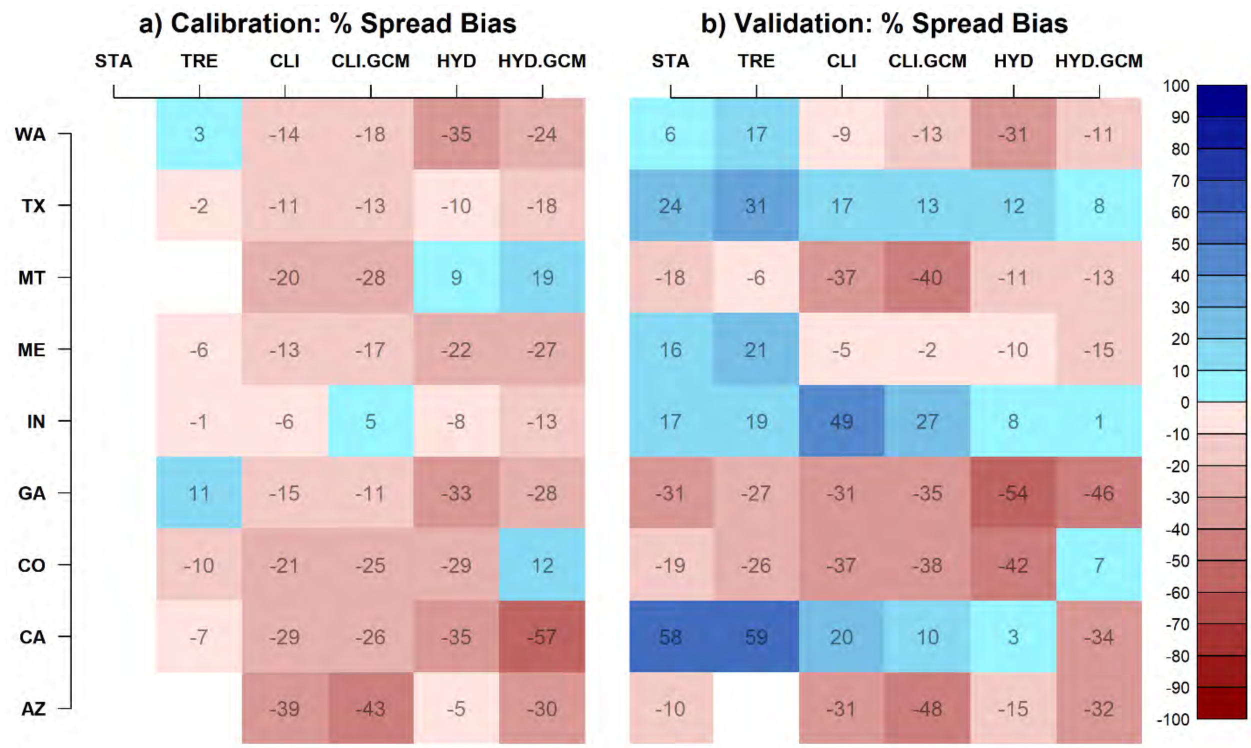


Figure 9-5 Percent spread bias for the (a) calibration and (b) validation periods for each basin and each approach (abbreviations are the same as Figure 9-4). White indicates absolute magnitude less than 0.5%.

9.4.2.2. CHANGE

Based on observations, six basins show a decrease in the 2-year flood magnitude between the calibration and validation periods and three show an increase (Figure 9-6a). Between the 2- and 100-year flood magnitudes, the direction of change switches from negative to positive for three basins (AZ, GA, and MT) (Figure 9-6b). This behavior indicates that the calibration and validation period distributions cross at some point between the 2- and 100-year return periods. By definition, the stationary approach projects no change. The trend-informed approach, constrained by the positive $\alpha_{1,c}$ found during model fitting, projects an increase of similar magnitude across all basins and across return periods. Notably, the statistical formulation of the trend-informed model prohibits any crossing of the distribution between calibration and validation periods.

When forced with observations, the performance of the climate-informed and hydrologic simulation approaches varies by basin and by return period and does not seem to follow a consistent pattern (Figure 9-6). Interestingly, when forced with observations, the hydrologic simulation approach is able to project large decreases in flood magnitude (e.g., the 100-year flood for CA and TX basins), while the climate-informed approach appears limited to decreases of small magnitude. When forced with GCM simulations, several noteworthy behaviors are exhibited by the climate-informed and hydrologic simulation approaches. First, the additional bias from GCM forcing can reverse the projected direction of change; for example, projected change in the 100-year flood magnitude is -50% and 10% in the TX basin for the hydrologic simulation approach forced by observations and by GCM simulations, respectively. Second, the additional bias from GCM forcing usually reduces the absolute magnitude of change; in combination with the first behavior, the result is that most GCM-forced projected changes are positive and of relatively small magnitude. Some exceptions are the negative projected changes in the AZ and CA basins and the large projected increase in the CO basin by the hydrologic simulation approach.

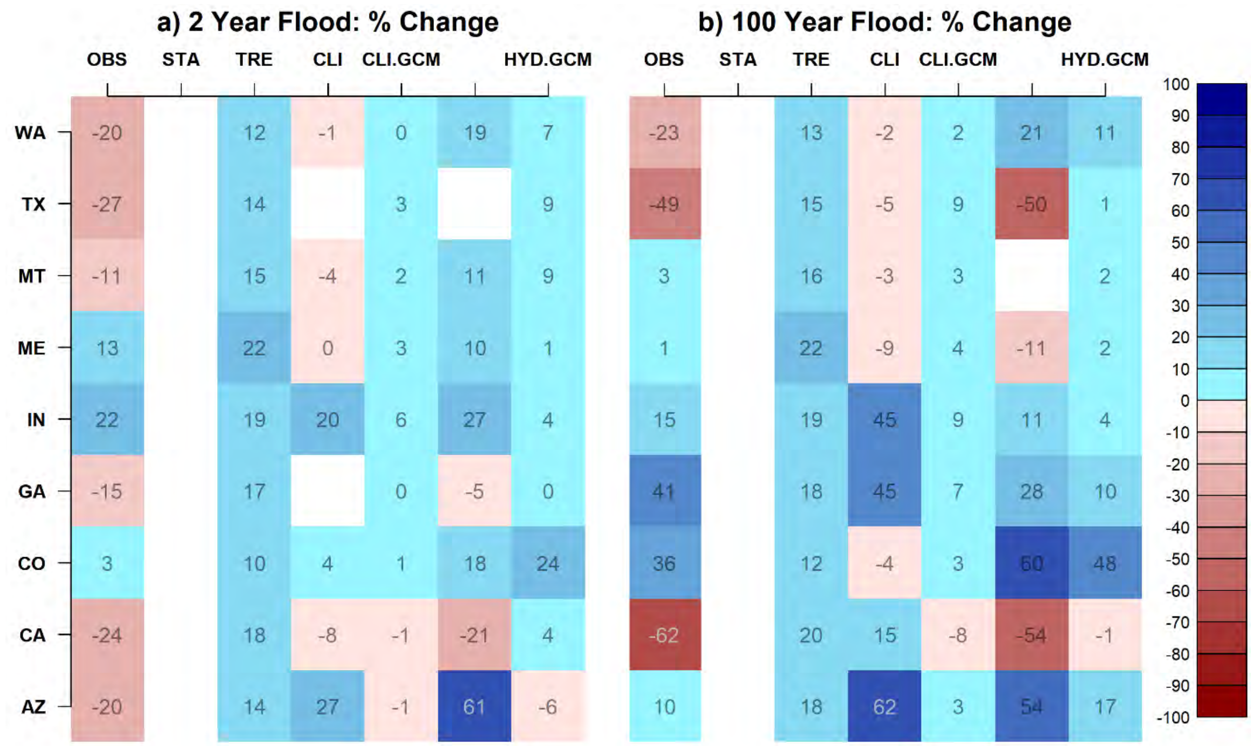


Figure 9-6 Percent change in the (a) 2-year and (b) 100-year flood magnitude between the calibration and validation periods for each approach and each basin (abbreviations are the same as Figure 9-4). White indicates statistically insignificant values.

9.4.2.3. GCM ENSEMBLE UNCERTAINTY

The range of median 100-year flood magnitudes across GCMs for each basin demonstrates the impact of using GCM simulations as forcing data for the climate-informed and hydrologic simulation approaches, as opposed to using observed climate (Figure 9-7). As stated previously, clearly the GCMs introduce additional bias in both the calibration and validation periods, to the extent that in some cases neither the observed nor the modeled value lies within the range resulting from forcing with GCM simulations (the hydrologic simulation approach for AZ and CA in the calibration period). Furthermore, the additional bias sometimes offsets model bias, which can provide a false impression of satisfactory model performance relative to the observed value (e.g., the hydrologic simulation approach for CO and MT in the calibration period). The additional bias introduced by GCMs relative to the model has no consistent direction or magnitude across basins, approaches, or periods (i.e., calibration and validation). Additionally, it appears that neither approach consistently results in more additional bias and neither approach consistently has a wider range across all GCMs.

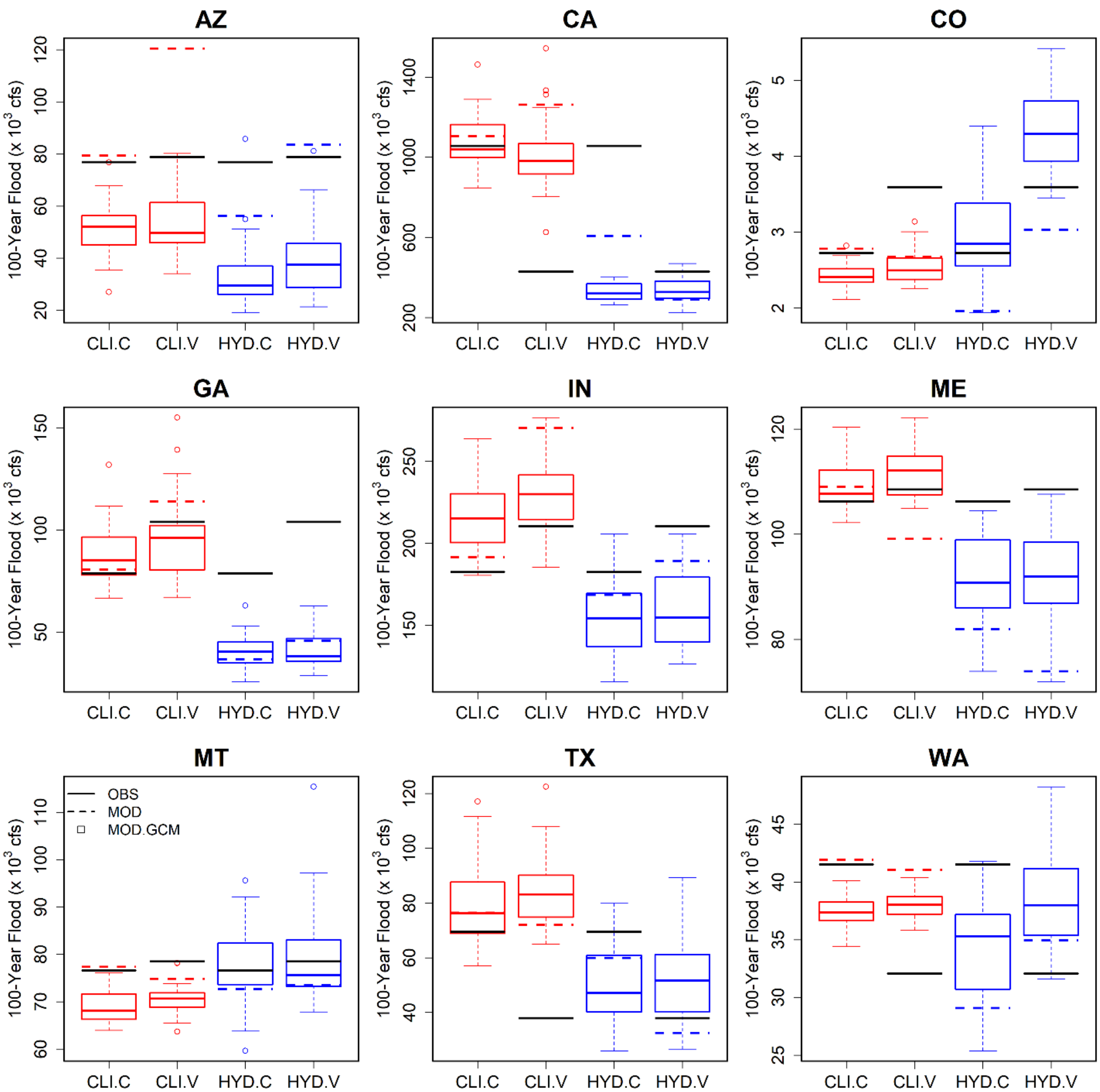


Figure 9-7 Boxplots showing the range of median 100-year flood magnitudes in each basin across 32 GCMs for both the climate-informed (CLI) and hydrologic simulation (HYD) approaches in the calibration (C) and validation (V) periods. Legend is in lower left plot; OBS is observed, MOD is model forced with observed climate, and MOD.GCM is model forced with GCM simulations. Note the varying y-axis scales.

9.4.3. Discussion

The results provide a quantitative perspective on key issues surrounding the challenge of long-term flood projection. Overall, performance seems to be highly basin- and approach-specific; however, some general observations can still be made.

The assumption of stationarity leads to large and statistically significant biases of estimated flood magnitude in the validation period, because all basins in this study exhibit a statistically significant change in flood magnitude between the validation and calibration periods. None of the nonstationary alternatives provide an obvious advantage as a replacement. In other words, in the validation period, none consistently performs better than all others over all metrics and all basins, nor do any consistently result in negligible bias. Notably, however, at least one approach overestimates the 100-year flood for all basins except GA; this holds true even when physically-based approaches forced with observations are excluded.

In the calibration period, nonstationary approaches reduce uncertainty, as quantified by spread (although model structure uncertainty is not explored here; see discussion below). For the trend- and climate-informed approaches, this is expected as the additional parameters in the model explain and hence reduce variability (Koutsoyiannis, 2011). The reduction in variability is less pronounced for the trend-informed approach because the additional parameters are frequently statistically insignificant, leading to larger standard deviations (Table 9-2). In the validation period, however, uncertainty is more basin-specific, rather than approach-specific. Spread resulting from forcing with GCM simulations tends to be of the same sign, albeit different magnitude, as spread from forcing with observed climate. The fact that the trend-informed approach is constrained by its mathematical formulation is clearly illustrated by the results focusing on change in flood magnitude. The positive coefficients constrain change to be positive in all basins and constrain the magnitude of change to be similar across return periods (i.e., the calibration and validation distributions do not cross). The climate-informed and hydrologic simulation approaches do allow for both positive and negative changes in flood magnitude; however, for the climate-informed approach, the magnitude of negative change is small, and for both approaches, negative change is primarily associated with using observations, rather than GCM simulations, as forcing.

As expected, GCMs introduce additional bias, which may engender false confidence in model performance by offsetting model bias. Furthermore, this additional bias may reverse the direction of projected change and often reduces the magnitude of projected change. However, the inconsistency of the additional bias across basins, approaches, and time periods indicates that bias correction techniques would not be effective. The range of values across GCMs can be large and may not contain either the observed value or the modeled value forced by observed climate. Since neither approach has consistently greater additional bias or a consistently larger range across GCMs, by implication aggregated precipitation over time as space (i.e., the covariate of the simplified climate-informed approach employed here; see discussion below), as opposed to grid-based localized precipitation (i.e., the input to the hydrologic simulation approach), does not consistently reduce bias or range of uncertainty across GCMs, perhaps because any benefit gained from improved GCM simulation of aggregated fields is offset by the simplifications inherent in the covariate-based statistical model.

There are several limitations to this study. One limitation is having only nine case study basins. In the original experimental design, nine basins were assumed sufficient to illustrate differences in the performance of each approach across various hydro-climatologic regimes; however, the variability in the results across basins and the range of other basin and flood characteristics (e.g., area, elevation, flood cause, etc.) precluded any such conclusions. While more basins would provide replicates for statistical analysis, applying all long-term flood projection approaches to nine basins is already a significant advancement beyond previous studies.

Another limitation, engendered by data constraints and the split-sample experimental design, is that only 30 years were used for each time period; while more years of data would reduce the bias and uncertainty in estimates of the 100-year flood, use of Bayesian credible intervals does account for sample size and 30 years is still representative of the data limits faced by many water managers (Lanfear & Hirsch, 1999; Lins, 2012). A time period of 30 years also implies that the climate change signal is likely dominated by natural climate variability (Hawkins & Sutton, 2011; Martel et al., 2018); while a larger climate change signal may enable better differentiation between approaches, the poor performance observed over even small climate changes is a cause for caution.

A third limitation, primarily engendered by a desire to maintain clarity in the presentation of results, is the focus on the 100-year flood magnitude, chosen for its societal significance. As illustrated by the flood frequency curve results for the WA basin and the change results for the 2-year flood magnitude, the performance at other return periods may be different.

A fourth limitation is that the climate-informed approach is based on precipitation rather than on large-scale oceanic-atmospheric patterns, which are preferred as covariates (Schlef et al., 2018), although precipitation-based models have been implemented in previous studies (Condon et al., 2015; Šraj et al., 2016). Assuming large-scale covariates can be found, previous experience with the climate-informed approach indicates that the associated model coefficients are likely to be less statistically significant than the coefficient for precipitation, which may offset the gain from more skillful GCM simulations of large-scale patterns relative to localized precipitation.

Finally, the experimental design is limited by only a partial exploration of uncertainty. There are two unexplored sources of uncertainty which are applicable to every approach, and hence should not affect the results: choice of flood distribution and streamflow data measurement. For the physically-based approaches, additional unexplored sources of uncertainty include climate data measurement, natural climate variability (which can be explored through stochastic weather or streamflow generators, e.g., Chen et al., 2015; Steinschneider et al., 2015), forcing scenario for GCMs and downscaling approach. Finally, for all nonstationary approaches, a major source of unexplored uncertainty is model structure, such as hydrologic model type or the form of the statistical regression; in fact, an appropriate mathematical formulation for non-stationarity is still an open research question (François et al., 2019). All these sources of uncertainty, but especially that of model structure, will likely increase the magnitude of spread (Serinaldi & Kilsby, 2015).

9.5. CONCLUSION

This is the first study to compare various approaches to long-term flood projection, using a split-sample test over the historic record, and to do so across a set of hydro-climatologically diverse basins, which are located across the contiguous United States. Based on the results and the ensuing discussion, the findings can be summarized as:

- no approach provides satisfactory performance in the validation period, and there are no consistently apparent patterns of performance across basins, approaches, or time periods, which implies that over- or under-design can occur just as readily with the stationarity approach as with other approaches,
- the climate-informed and hydrologic simulation approaches, which incorporate physical mechanisms, are more flexible than the trend-informed approach, but, as expected, use of GCM simulations as forcing for the climate-informed and hydrologic simulation approaches introduces additional inconsistent bias,
- flood magnitude in the validation period is likely to be overestimated by at least one of the approaches, which implies that the highest value in the ensemble could function as a conservative estimate in design for future periods without observations.

We realize that for anyone invested in the topic of long-term flood projection from either a scientific or engineering perspective, including ourselves, these results demonstrate that this field still requires much development, and we are forced to acknowledge the need for a posture of humility as advocated by Lins & Cohn (2011). However, we are of the perspective that this humility can be a catalyst for further investigation; in other words, where do we go from here?

One (often suggested) path would be to perform a more comprehensive comparison study. If information about the performance of all approaches (including a climate-informed approach based on large-scale oceanic-atmospheric patterns) was available for not just nine but, for example, 1,000 basins across the United States (as was performed by Luke et al., 2017 for the stationarity and trend-informed approaches) or even around the globe, it might be possible to determine the statistical likelihood that one approach will perform better than another and to identify under what hydro-climatological conditions better performance is observed. From our perspective of having implemented a comparison of nine basins, the technical and computational cost of such a comprehensive comparison study is daunting and even prohibitive for an individual researcher or team. Instead, such a study would be well suited to a large-scale international collaborative effort by the hydro-meteorological community that could be modeled after similar existing efforts to compare climate models (e.g., CMIP; Taylor et al., 2012) and downscaling methods (Gutiérrez et al., 2018).

Another path (already being followed) is to improve the model structure and forcing data of the physically-based approaches through improved understanding of the physical systems driving floods. One particularly interesting research avenue is the identification and characterization of synoptic-scale patterns associated with extremes (e.g., Knighton et al., 2019; Schlef et al., 2019). Notably, this path is not applicable to the stationary or trend-informed approaches, which can only, but do not necessarily, improve with longer data records. One challenge is that it is unclear whether attainable improvements will be sufficient to generate results useful to practitioners. A similar idea is to incorporate paleo-data into models to expand the record of observed variability (e.g., Støren & Paasche, 2014).

A third path (borrowed from the climate scientists) would be to treat the suite of approaches as an ensemble that provides some information about the uncertainty associated with not definitively knowing the causal mechanisms or statistical structure of changes in flood magnitude. While the results of this study show that agreement between approaches does not imply accuracy, at least one approach overestimates the 100-year flood for nearly all basins in the validation period, indicating that the ensemble of approaches may be able to provide a conservative (i.e., too high) estimate of future flood magnitude.

We conclude by observing that the results of this study clearly illustrate the need for design paradigms which account for uncertainty (e.g., climate factors, the prudent approach, and robustness-based decision methods, as discussed in François et al., 2019) and can also be used to inform those paradigms. For example, given that under-design should be more of a concern than over-design, although both are damaging to society (Rosner et al., 2014), and given that the results indicate that one projection approach is likely to provide an overestimate, an ensemble of projections from different approaches applied to multiple basins could be used to develop climate factors specific to a region or a hydroclimatic regime or could be used as both motivation and justification for a prudent approach which chooses a different design value than estimated from historic data. Finally, an ensemble of flood projection approaches is simply one more dimension of uncertainty to be included into robustness-based decision methods, such as that illustrated by Spence & Brown (2018). Use of these design paradigms seems to us the only way to both explicitly acknowledge of our lack of knowledge in the face of such complex challenges and to attempt to use what little knowledge we do have to mitigate against damages from devastating floods in the future.

9.6. REFERENCES

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

Arnell, N. W., & Gosling, S. N. (2016). The impacts of climate change on river flood risk at the global scale. Climatic Change, 134(3), 387–401. https://doi.org/10.1007/s10584-014-1084-5

Baker, D. B., Richards, R. P., Loftus, T. T., & Kramer, J. W. (2004). a New Flashiness Index : Characteristics and Applications To Midwestern Rivers and Streams 1, 44883(03095), 503–522. https://doi.org/10.1111/j.1752-1688.2004.tb01046.x

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

Blöschl, G., & Montanari, A. (2010). Climate change impacts - throwing the dice? Hydrological Processes, 24, 374–381.

Bracken, C., Holman, K. D., Rajagopalan, B., & Moradkhani, H. (2018). A Bayesian Hierarchical Approach to Multivariate Nonstationary Hydrologic Frequency Analysis. Water Resources Research, 54(1), 243–255. https://doi.org/10.1002/2017WR020403

Brown, C., Weatherly, J., & Mearns, L. (2016). Decision-Scaling : A Decision Framework for DoD Climate Risk Assessment and Adaptation Planning.

Chen, L., Singh, V. P., Guo, S., Zhou, J., & Zhang, J. (2015). Copula-based method for multisite monthly and daily streamflow simulation. Journal of Hydrology, 528. https://doi.org/https://doi.org/10.1016/j.jhydrol.2015.05.018

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., et al. (2013). Long-term Climate Change: Projections, Commitments and Irreversibility. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 1029–1136). Cambridge, UK, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.024

Condon, L. E., Gangopadhyay, S., & Pruitt, T. (2015). Climate change and non-stationary flood risk for the upper Truckee River basin. Hydrology and Earth System Sciences, 19(1), 159–175. https://doi.org/10.5194/hess-19-159-2015

Delgado, J. M., Merz, B., & Apel, H. (2014). Projecting flood hazard under climate change: An alternative approach to model chains. Natural Hazards and Earth System Sciences, 14(6), 1579–1589. https://doi.org/10.5194/nhess-14-1579-2014

England, J. J. F., Cohn, T. A., Faber, B. A., Stedinger, J. R., Thomas Jr., W. O., Veilleux, A. G., et al. (2018). Guidelines for Determining Flood Flow Frequency - Bulletin 17C. USGS Techniques and Methods 4-B5. https://doi.org/10.3133/tm4B5

FEMA. (2011). Answers to Questions About the NFIP, (March), 70.

François, B., Schlef, K. E., Wi, S., & Brown, C. M. (2019). Design Considerations for Riverine Floods in a Changing Climate - A Review. Journal of Hydrology.

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., et al. (2018). An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. International Journal of Climatology, 1–36. https://doi.org/10.1002/joc.5462

Hawkins, E., & Sutton, R. (2011). The potential to narrow uncertainty in projections of regional precipitation change. Climate Dynamics, 37(1–2), 407–418.

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

Hu, Y., Liang, Z., Chen, X., Liu, Y., Wang, H., Yang, J., et al. (2017). Estimation of design flood using EWT and ENE metrics and uncertainty analysis under non-stationary conditions. Stochastic Environmental Research and Risk Assessment, 31(10), 2617–2626. https://doi.org/10.1007/s00477-017-1404-1

IPCC. (2013). Summary for Policymakers. In T. Stocker, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, et al. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.004

Jain, S., & Lall, U. (2001). Floods in a changing climate: Does the past represent the future? Water Resources Research, 37(12), 3193–3205. https://doi.org/10.1029/2001WR000495

Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe Version 4 available from the CGIAR-CSI SRTM 90m Database. Retrieved from http://srtm.csi.cgiar.org

Knighton, J., Pleiss, G., Carter, E., Lyon, S., Walter, M. T., & Steinschneider, S. (2019). Potential Predictability of Regional Precipitation and Discharge Extremes using Synoptic-Scale Climate Information via Machine Learning: An Evaluation for the Eastern Continental United States. Journal of Hydrometeorology, 20, 883–900. https://doi.org/10.1175/JHM-D-18-0196.1

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift, 15(3), 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Koutsoyiannis, D. (2011). Hurst-Kolmogorov Dynamics and Uncertainty. Journal of the American Water Resources Association, 47(3), 481–495. https://doi.org/10.1111/j.1752-1688.2011.00543.x

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., et al. (2014). Flood risk and climate change: global and regional perspectives. Hydrological Sciences Journal, 59(1), 1–28.

https://doi.org/10.1080/02626667.2013.857411
Kundzewicz, Z. W., Krysanova, V., Dankers, R., Hirabayashi, Y., Kanae, S., Hattermann, F. F., et al. (2017). Differences in flood hazard projections in Europe–their causes and consequences for decision making. Hydrological Sciences Journal, 62(1), 1–14. https://doi.org/10.1080/02626667.2016.1241398
Landwehr, J. M., & Slack, J. R. (1992). Hydro-Climatic Data Network: A U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1874-1988. Retrieved from https://pubs.usgs.gov/wri/wri934076/1st_page.html
Lanfear, K. J., & Hirsch, R. M. (1999). USGS study reveals a decline in long-record streamgages. Eos, 80(50), 605–607. https://doi.org/10.1029/99EO00406
Lins, H. F. (2012). USGS hydro-climatic data network 2009 (HCDN-2009). U.S. Geological Survey Fact Sheet 2012–3047, 2009(April), 4. https://doi.org/http://water.usgs.gov/osw/hcdn-2009/
Lins, H. F., & Cohn, T. A. (2011). Stationarity: Wanted dead or alive? Journal of the American Water Resources Association, 47(3), 475–480. https://doi.org/10.1111/j.1752-1688.2011.00542.x
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
Luke, A., Vrugt, J. A., Aghakouchak, A., Matthew, R., & Sanders, B. F. (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. Water Resources Research, 53, 1–26. https://doi.org/10.1002/2016WR019676
Madsen, H., Lawrence, D., Lang, M., Martinkova, M., & Kjeldsen, T. R. (2014). Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. Journal of Hydrology, 519, 3634–3650.
Mallakpour, I., Aghakouchak, A., & Sadegh, M. (2019). Climate-Induced Changes in the Risk of Hydrological Failure of Major Dams in California. Geophysical Research Letters, 46(4). https://doi.org/10.1029/2018GL081888
Martel, J.-L., Mailhot, A., Brissette, F., & Caya, D. (2018). Role of Natural Climate Variability in the Detection of Anthropogenic Climate Change Signal for Mean and Extreme Precipitation at Local and Regional Scales. Journal of Climate, 31, 4241–4263. https://doi.org/10.1175/JCLI-D-17-0282.1
Merz, B., Aerts, J., Arnbjerg-Nielsen, K., Baldi, M., Becker, A., Bichet, A., et al. (2014). Floods and climate: Emerging perspectives for flood risk assessment and management. Natural Hazards and Earth System Sciences, 14(7), 1921–1942. https://doi.org/10.5194/nhess-14-1921-2014
Milly, P. C. D., Wetherald, R. T., Dunne, K. A., & Delworth, T. L. (2002). Increasing risk of great floods in a changing climate. Nature, 415(6871), 514–517.
Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: whither water management? Science, 319(5863), 573–574. https://doi.org/10.1126/science.1151915
Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., et al. (2015). On Critiques of "Stationarity is Dead: Whither Water Management?" Water Resources Research, 51, 7785–7789. https://doi.org/10.1002/2015WR017408
Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part 1 - a discussion of principles. Journal of Hydrology, 10, 282–290.
NRC. (2000). Risk analysis and uncertainty in flood damage reduction studies. National Academies Press.
Peterson, T. C., Heim, R. R., Hirsch, R., Kaiser, D. P., Brooks, H., Diffenbaugh, N. S., et al. (2013). Monitoring and Understanding Changes in Heat Waves, Cold Waves, Floods, and Droughts in the United States: State of Knowledge. Bulletin of the American Meteorological Society, 94(6), 821–834. https://doi.org/10.1175/BAMS-D-12-00066.1
Pierce, D. W., Cayan, D. R., & Thrasher, B. L. (2014). Statistical Downscaling Using Localized Constructed Analogs (LOCA)*. Journal of Hydrometeorology, 15(6), 2558–2585. https://doi.org/10.1175/JHM-D-14-0082.1
Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., & Hegewisch, K. C. (2015). Improved Bias Correction Techniques for Hydrological Simulations of Climate Change*. Journal of Hydrometeorology, 16(6), 2421–2442. https://doi.org/10.1175/JHM-D-14-0236.1
Plummer, M. (2016). rjags: Bayesian Graphical Models using MCMC. R package version 4-6. Retrieved from https://cran.r-project.org/package=rjags
Read, L., & Vogel, R. M. (2015). Reliability, return periods, and risk under nonstationarity. Water Resources Research, 51(1), 6381–6398. https://doi.org/10.1002/2014WR016259
Rosner, A., Vogel, R., & Kirshen, P. (2014). A risk-based approach to flood management decisions in a nonstationary world. Water Resources Research, 1–15. https://doi.org/10.1002/2013WR014561.Received
Salas, J., Rajagopalan, B., Saito, L., & Brown, C. (2012). Special Section on Climate Change and Water Resources: Climate Nonstationarity and Water Resources Management. Journal of Water Resources Planning and Management, 138(5), 385–388. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000279
Schlef, K. E., François, B., Robertson, A. W., & Brown, C. M. (2018). A General Methodology for Climate Informed Approaches to Long-Term Flood Projection - Illustrated with the Ohio River Basin. Water Resources Research, 54. https://doi.org/10.1029/2018WR023209
Schlef, K. E., Moradkhani, H., & Lall, U. (2019). Atmospheric Circulation Patterns Associated with Extreme United States Floods Identified via Machine Learning. Scientific Reports, 9(1), 7171. https://doi.org/10.1038/s41598-019-43496-w

Serinaldi, F., & Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. Advances in Water Resources, 77, 17–36. https://doi.org/10.1016/j.advwatres.2014.12.013

Spence, C. M., & Brown, C. M. (2018). Decision Analytic Approach to Resolving Divergent Climate Assumptions in Water Resources Planning. Journal of Water Resources Planning and Management, 144(9). https://doi.org/10.1061/(ASCE)WR.1943-5452.0000939.

Šraj, M., Viglione, A., Parajka, J., & Blöschl, G. (2016). The influence of non-stationarity in extreme hydrological events on flood frequency estimation. Journal of Hydrology and Hydromechanics, 64(4), 426–437. https://doi.org/10.1515/johh-2016-0032

Steinschneider, S., Wi, S., & Brown, C. (2015). The integrated effects of climate and hydrologic uncertainty on future flood risk assessments. Hydrological Processes, 29(12), 2823–2839. https://doi.org/10.1002/hyp.10409

Steirou, E., Gerlitz, L., Apel, H., Sun, X., & Merz, B. (2019). Climate influences on flood probabilities across Europe. Hydrology and Earth System Sciences, 23, 1305–1322. https://doi.org/10.5194/hess-23-1305-2019

Støren, E. N., & Paasche, Ø. (2014). Scandinavian floods: From past observations to future trends. Global and Planetary Change, 113, 34–43. https://doi.org/10.1016/j.gloplacha.2013.12.002

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society, 93(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Villarini, G. (2016). On the seasonality of flooding across the continental United States. Advances in Water Resources, 87, 80–91. https://doi.org/10.1016/j.advwatres.2015.11.009

Villarini, G., Serinaldi, F., Smith, J. A., & Krajewski, W. F. (2009). On the stationarity of annual flood peaks in the continental United States during the 20th century. Water Resources Research, 45(8), 1–17. https://doi.org/10.1029/2008WR007645

Wi, S., Ray, P., Demaria, E. M. C., Steinschneider, S., & Brown, C. (2017). A user-friendly software package for VIC hydrologic model development. Environmental Modelling and Software, 98, 35–53. https://doi.org/10.1016/j.envsoft.2017.09.006

Wouter, B., Emma, A., Joshua, L., Ralph, T., & Ross, W. (2017). Recent changes in extreme floods across multiple continents. Environmental Research Letters. Retrieved from http://iopscience.iop.org/10.1088/1748-9326/aa8847

Ye, S., Li, H.-Y., Leung, L. R., Guo, J., Ran, Q., Demissie, Y., & Sivapalan, M. (2017). Understanding Flood Seasonality and Its Temporal Shifts within the Contiguous United States. Journal of Hydrometeorology, 18(7), 1997–2009. https://doi.org/10.1175/JHM-D-16-0207.1

Yu-Sung, S., & Yajima, M. (2015). R2jags: Using R to Run "JAGS". R package version 0.5-7. Retrieved from https://cran.r-project.org/packages=R2jags

Zhang, Q., Gu, X., Singh, V. P., Xiao, M., & Chen, X. (2015). Evaluation of flood frequency under non-stationarity resulting from climate indices and reservoir indices in the East River basin, China. Journal of Hydrology, 527(August), 565–575. https://doi.org/10.1016/j.jhydrol.2015.05.029

10. CLUSTERING APPROACHES FOR PREDICTING CLIMATE EXTREMES

10.1. EXECUTIVE SUMMARY

The Ohio River Basin and the overall Mississippi River Basin has a long history of large-scale regional flooding. This study explores long and short-term horizons in predictability of climate extremes. The short-term horizon methods involve use of newly developed mathematical tools based on dynamical systems theory. The results from this approach show a distinct large-scale organization leading up to the regional scale flooding events i.e. climatic extremes. The long-term horizon methods involve use of clustering and wavelet analysis to study the spatiotemporal variability. The method results clearly illustrates the low frequency and decadal variability modes acting in the region. These two methods provide two distinct pieces of information with crucial implications in the management of water and crucial infrastructure systems in the region.

10.2. TECHNICICAL APPROACH

Short Term Dynamical System Methods – Almost all climate systems are dynamical systems evolving chaotically over time. Recently academic literature has developed mathematical tools for identifying the persistence and complexity of extreme values in dynamical systems (Faranda, Messori, & Yiou, 2017; Messori, Caballero, & Faranda, 2017). These tools have been extended to analyze bivariate systems to learn about how rainfall extremes are clustered in space and time, and the patterns that cause them. This can be used to forecast future short-range rainfall patterns and to understand physical mechanisms. The approach used here establishes connections based on the local properties of the dynamical systems and climatic extremes.

Long Term Spatiotemporal Methods - The section above deals with time horizons of a few days leading up to storm events, whereas long-term decadal spatiotemporal predictability is important for infrastructure and economic planning. Here we propose combined wavelet analysis along with clustering methods. Using specific chronological operations of Principal Component Analysis, Hierarchical Clustering and Wavelet Analysis, different aspects of the time-frequency structure and the spatial dependence structure of the streamflow field is uncovered in the region. This is followed by the analysis of 10-year flood events and periods/ regimes associated with it. This helps to incorporate aspects of spatiotemporal risks faced in the Ohio River Basin.

10.3. DATA AND METHODS

With a focus on the Ohio River Basin, we analyze the predictability of climatic extremes over different time horizons. There is utility for predictions across every time horizon, with short term predictability crucial for real time disaster management and saving lives, whereas long term predictability is crucial for infrastructure and economic planning. The sections below are divided keeping into account the approached needed to handle each time horizon.

10.3.1. SHORT TERM DYNAMICAL SYSTEM APPROACH

Dynamical Systems at their simplest are systems which evolve or depend on time in one form or the other. Higher order dynamical systems tend to exhibit sensitive dependence to initial conditions, a phenomenon called chaos (Lorenz, 1980). The dynamical system can be characterized by its state space which has the dimension equal to the total number of variables. In this state space the individual states evolve by following a particular trajectory and combined all these trajectories make up the dynamical system state space. Most climatic systems are chaotic dynamical systems.

The dynamical systems approach to predictability uses the instantaneous system dimension(d) and inverse of persistence(θ) as the two quantities which describe the evolution of the states in the state space. The computation and interpretation of these two quantities is based on previous work focused on the Northern Atlantic region (Faranda et al., 2017; Messori et al., 2017). A distance matrix is computed between all the points in the phase space, which in this case is a time series. The values over a high threshold q (98th percentile) are then fit to a Generalized Pareto Distribution (GPD). The instantaneous dimension(d) is the inverse of the standard deviation of the GPD.

$$d(\zeta) = \frac{1}{\sigma(\zeta)} \qquad (10.1)$$

Here, ζ is a point in the phase space. The instantaneous inverse of persistence(θ) is computed using the Suveges likelihood estimator (Süveges, 2007). It is interpreted as inverse of mean residence time in the sphere corresponding to the threshold and is such that:

$$P\big(g\big(x(t)\big) > q\big) \simeq exp\left[-\theta\left(\frac{x-\mu(\zeta)}{\sigma(\zeta)}\right)\right], \qquad (10.2)$$

where, x(t) are all other points in the phase space and ζ is the current point of interest. μ and σ are the parameters of the GPD. The instantaneous dimension and inverse of persistence at a given moment (ζ) describe its current state and corresponds to the information from the dynamical system.

The approach was extended from a univariate field to a bivariate field to better capture the dynamics of the overall system. The data were acquired from the ECMWF's ERA Interim reanalysis data product (Dee et al., 2011). Daily data ranging from 1979-2017 and extrapolated on a 2.5o x 2.5o grid were utilized for this study (Note: - The downloaded dataset had 6 hourly timesteps and grid size was 0.75o x 0.75o). The atmospheric fields used for the bivariate analysis were the 500 hPa geopotential height & total column water (precipitable water). The data spanned the entire region of the contiguous United States from 70 W to 122.5 W and 25N to 50N.

Climatic Precipitation Extremes - The above sub-section outlines a method to compute the properties of the dynamical system thereby also identifying dynamical extremes. Here we outline a method to identify climate extremes which correspond to floods and extreme precipitation events in the basin. The climatic extremes were computed based on a gridded regional precipitation product, and days which had 15% or more grid cells exceeding their 99th percentile of daily precipitation were classified as Regional Extreme Precipitation (REP) Days (Farnham, Doss-Gollin, & Lall, 2018). The data for the gridded precipitation was taken from NOAA's Climate Prediction Centre (CPC).

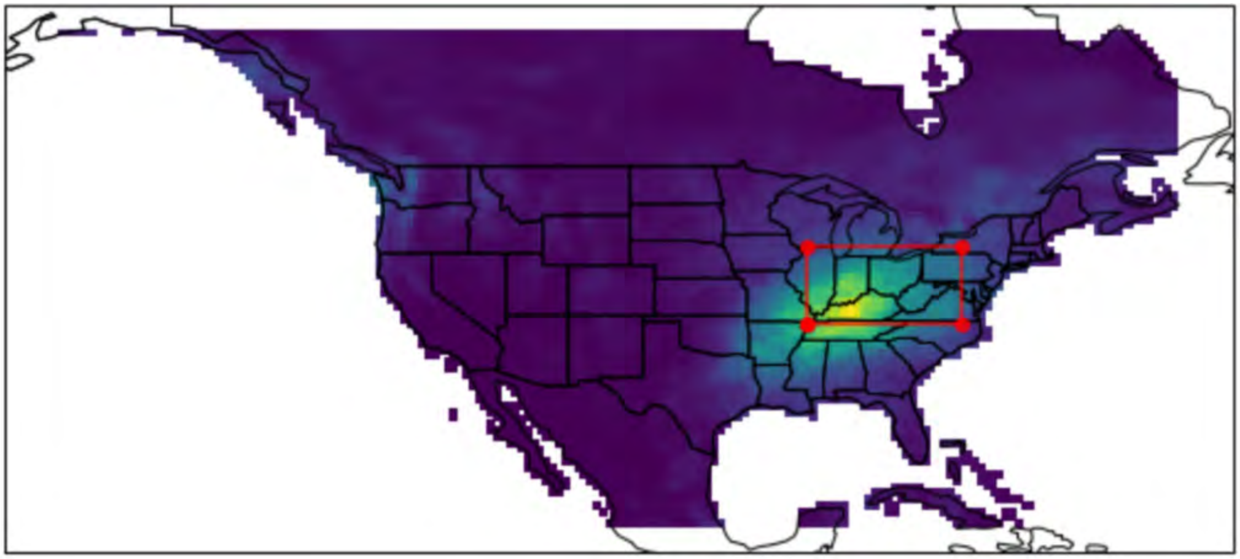


Figure 10-1 Rainfall in North America during a Regional Extreme Precipitation (REP) Days. The red box denotes the area over which the grid cells for the computations were included.

10.3.2. Long Term Spatiotemporal Approach

The estimation of climate risk is currently done by means of point risk estimation methods. This traditional method of risk estimation ignores the clustered spatiotemporal aspect of risk, the dominant modes of global and regional climate variability and decadal variations in climate. This risk estimation becomes important for critical infrastructure systems like flood protection dams which are present in the Ohio River Basin. This kind of infrastructure is also prone to catastrophic failure leading to cascading effects to further downstream systems. Here relying on the streamflow characteristics in the region, along with the major climate indices we look at the broader regions of predictability and risk clustering in the region.

Implicit Clustering - The streamflow gauge information is transformed from a daily time series to its annual maximum contemporary. The first step in the analysis consists of computation of wavelet analysis at each gauge, helping uncover the time-frequency structure (Torrence & Compo, 1998). The wavelet analysis sheds light on the periodicity of floods and the years during which these periodicities were dominant. This is followed by hierarchical clustering using the output from the wavelets i.e. location-based time-frequency domain. Here the number of clusters are defined by the Silhouette Analysis (Rousseeuw, 1987). For each of the clusters, the 1st PC for the streamflow field is computed by means of Principal Component Analysis which is followed by another wavelet analysis on the 1st PC of each cluster, to uncover the time-frequency structure of the entire cluster. Since the 1st PCs are a time series we also compute the wavelet coherence of this PC with known climate indices and their corresponding interactions. For the climate indices, the JFM mean climatology is considered.

Explicit Clustering – Similar to the approach described above, but instead of dividing them into clusters, first a Principal Component Analysis is carried out on the data to extract the PC which explained the most variance. And for each PC a wavelet analysis is carried out, uncovering the time-frequency structure of the entire domain.

The chief difference between both these methods is that the former method allows for explicit spatial clustering whereas the latter allows implicit clustering which can be seen via the distribution of loadings of the PCs. This method allows us to get an understanding of the time – frequency structure in the maximum annual flow. For the 1st PC across the domain and all subsequent PC's the wavelet coherence structure with climate indices are also computed, which help in the study of the time-frequency dependence structure between the two and point towards possible teleconnections.

Flood Return Period Exceedance Modelling – Floods are characterized by return periods, where a 'y' year flood corresponds to a 1/'y' chance of occurrence. Tying this to the flood occurrences across the region, the 10-year site specific return period is used as the metric of interest. These return periods were computed individually for each site by using the Maximum Likelihood Estimation (MLE) method. The R package extRemes was used for this purpose. Once these return periods were computed, the annual site exceedances were counted and aggregated for the entire domain, giving the total count of the exceedances in the entire domain. This exceedance time series across the whole domain was modelled using Poisson Regression with the traditional climate indices – ENSO, PDO, NAO and their interactions as the predictors. Based on the regression output the 95% confidence intervals were computed, which serves as a measure of the upper limit and exceedances far beyond them point towards clustering of events across the river basin.

The Streamflow data were compiled by USGS and the R package dataretrival was used as an interface (Hirsch & De Cicco, 2015). The climate indices used in this study were ENSO (Nino 3.4), PDO and NAO. The ENSO data was taken from Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST), PDO was taken from the Joint Institute for the Study of the Atmosphere and Ocean(JISAO) and NAO from (Jones, Jonsson, & Wheeler, 1997). All the data for the climate indices was extracted from the KNMI Climate Explorer, an online tool to access climate data. The data for the dams was taken from the National Inventory of Dams (NID) and the R package 'dams' was the interactive interface tool.

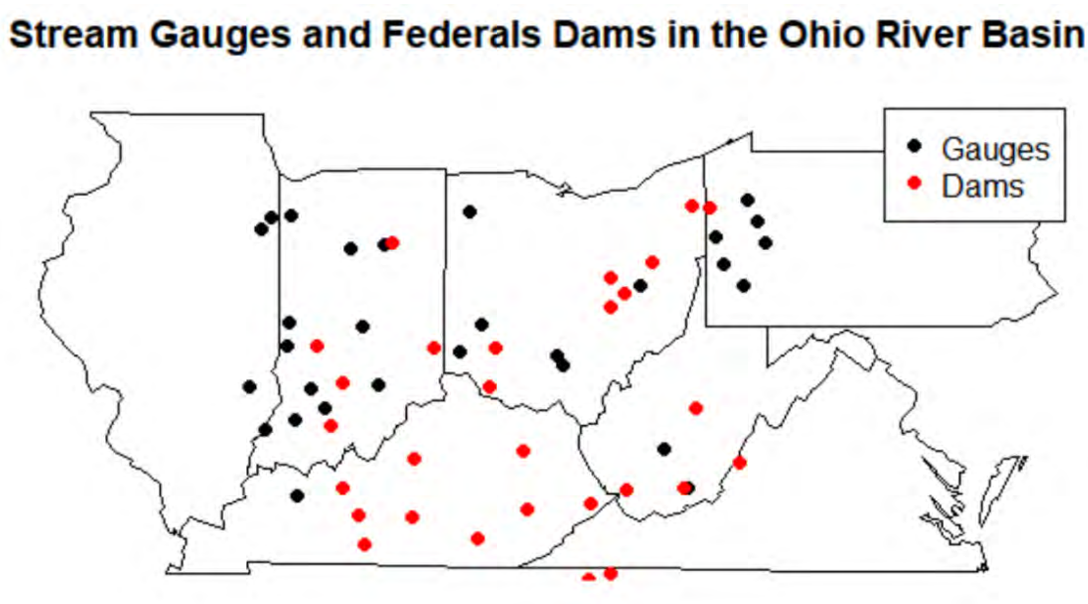


Figure 10-2 Locations of stream gauges with continuous 81 years data and flood control dams in the Ohio River Basin. Source: - Location and Dam characteristics were obtained from the National Inventory of Dams (NID) and were selected based on their storage, primary use and if they were federally owned.

10.4. RESULTS AND DISCUSSION

10.4.1. SHORT TERM DYNAMICAL SYSTEM APPROACH

The 500 hPa Geopotential Height and the Total Column Water were the two fields used in the computation of the daily system properties using the methods described above. Further for the Ohio River Basin, based on the methodology elaborated above, the REP days were also computed.

The first part of the analysis consisted of clustering the data based on their instantaneous dynamical properties, d and θ. Since the total variables were large, a principal component analysis was carried out and the first three PC's which explain about 80% variance were selected. Based on these PCs, clustering was carried out using K-Means clustering. Figure 10-3 shows the composites for the 9 clusters. Shaded colors denote the geopotential height anomalies whereas contour lines denote total column water anomalies, with solid lines representing positive anomalies and negative lines denoting negative anomalies. Cluster 9 corresponds most to what would be expected of a heavy rainfall or REP day over the Ohio River Basin. It is also the most organized cluster among all of them and represents characteristics which would be expected of such a systems i.e. low dimensionality and low persistence.

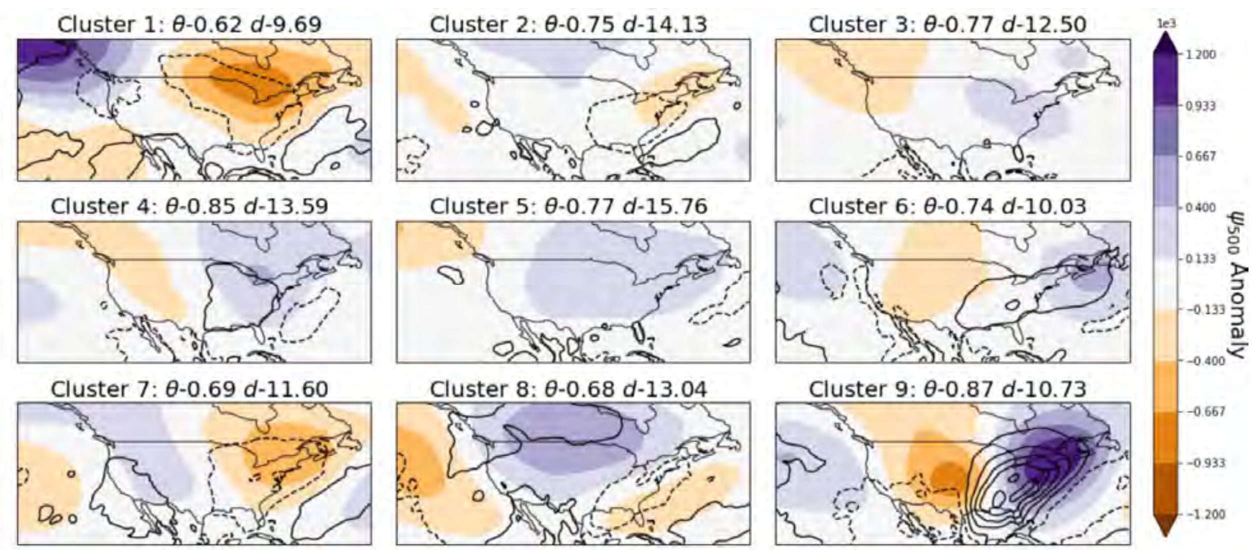


Figure 10-3 Composites of the individual clusters computed on the instantaneous system dynamical properties. At the top of each cluster are the mean instantaneous dimension(d) and inverse of persistence(θ) for each cluster. Shades denote the 500 hPa geopotential height anomalies and contour lines denote column water anomalies.

Another crucial aspect of this mathematical theory is the connection between climatic extremes and dynamical extremes. The prime utility of this relationship between the two is to enhance predictability of climatic extremes, with conditional knowledge of the current dynamical systems extremes. To check for this, the REP days are used as a metric of climatic extremes and evolution of the system in the vicinity of the REP days is analyzed. The vicinity of the event – REP Days is Day -5 to Day +3 after the event and composites across all the REP Days are shown in Figure 10-4.

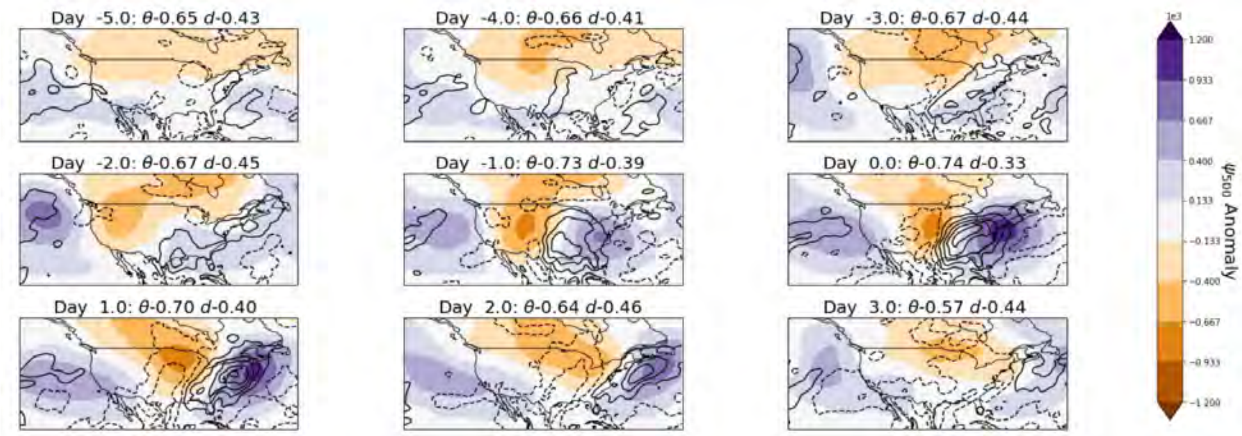


Figure 10-4 Composites for Day -5 to +3 in the vicinity of REP days during 1979-2017. Shaded contours denote anomalies in the geopotential height field and contour lines denote anomalies in the total column water field. The values associated with d and θ at the top are their corresponding percentiles.

There is a certain amount of organization which happens in the system as the storm (REP) approaches, which can be seen in the d and θ quantiles associated with each day. The general trend can be characterized by a decreasing instantaneous dimension along with decreasing instantaneous persistence (Note: - θ is the inverse of instantaneous persistence). Similar patterns are seen in a lot of other studies based on the Ohio River Basin. Regional extreme precipitation days are characterized by presence of low local dimensions and low persistence, meaning that these are organized and fleeting events. Further an important point to note is the similarity between the composites on Day 0 and Day 1 and Cluster 9 in Figure 10-3 and Figure 10-4. For events leading to a REP Day, the organization viewed by the lower dimensionality and persistence, is during the storm event and dissipates soon after it.

10.4.2. Long Term Spatiotemporal Approach

The daily streamflow data from the streamflow gauges is used as the primary field of interest for the study of long-term spatiotemporal variability and predictability. Since the primary interest lies in the spatiotemporal nature of streamflow a compromise is made between the number of stations and the time availability of the data (Since fewer stations go back 80 years or more). One subset of the analysis included 29 stations but covered 81 years (1937-2017) and the other subset consisted of 49 stations but covered only 39 years (1979-2017). Figure 10-2 shows locations of all stream gauges (black dots) which have data for 81 years (1937-2017). The implicit and explicit clustering methods were applied to both sub-datasets separately.

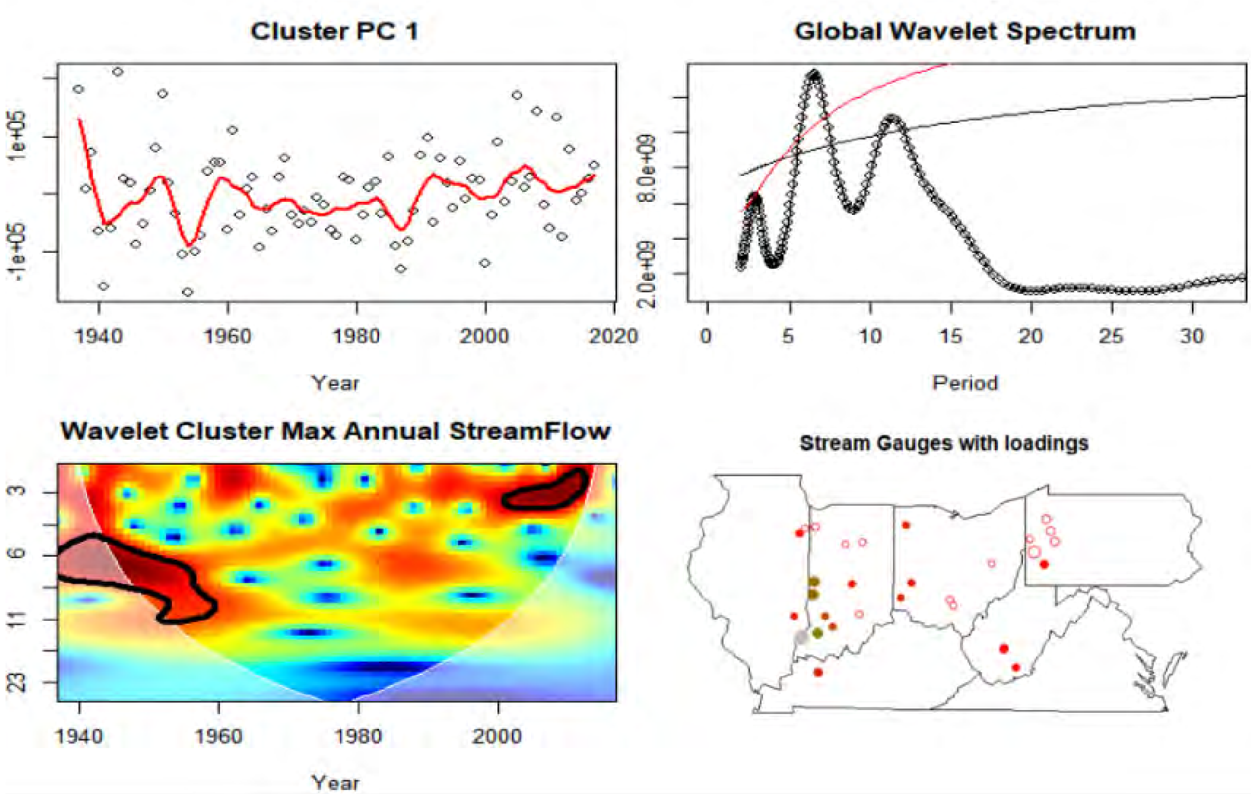


Figure 10-5 Wavelet Analysis on the 1st Principal Component of the spatial cluster for the 81-year dataset. TOP-LEFT – 1st PC of cluster with a loess line. TOP-RIGHT – Global Wavelet Spectrum of the PC. BOTTOM-LEFT – Power Spectrum of the Wavelet. BOTTOM-RIGHT – Spatial Distribution of the Loadings (Size indicates their drainage area.

The analysis shown in Figure 10-5 and Figure 10-6 are the results of the explicit clustering methodology. Both, 81- and 39-years datasets are characterized by sharp peaks at 6-7 years in their global wavelet spectrums (Figure 10-5 and Figure 10-6 respectively), with the black and red lines corresponding to the 95% confidence intervals for white and red noise respectively. The wavelet spectrum for the 81 years dataset also shows a peak for the 11-12 years dataset, though this does not exceed the 95% confidence limit corresponding to red noise but exceeds the white noise threshold. Between these two analyses, the results point to a periodicity of 6-7 years and a near decadal period.

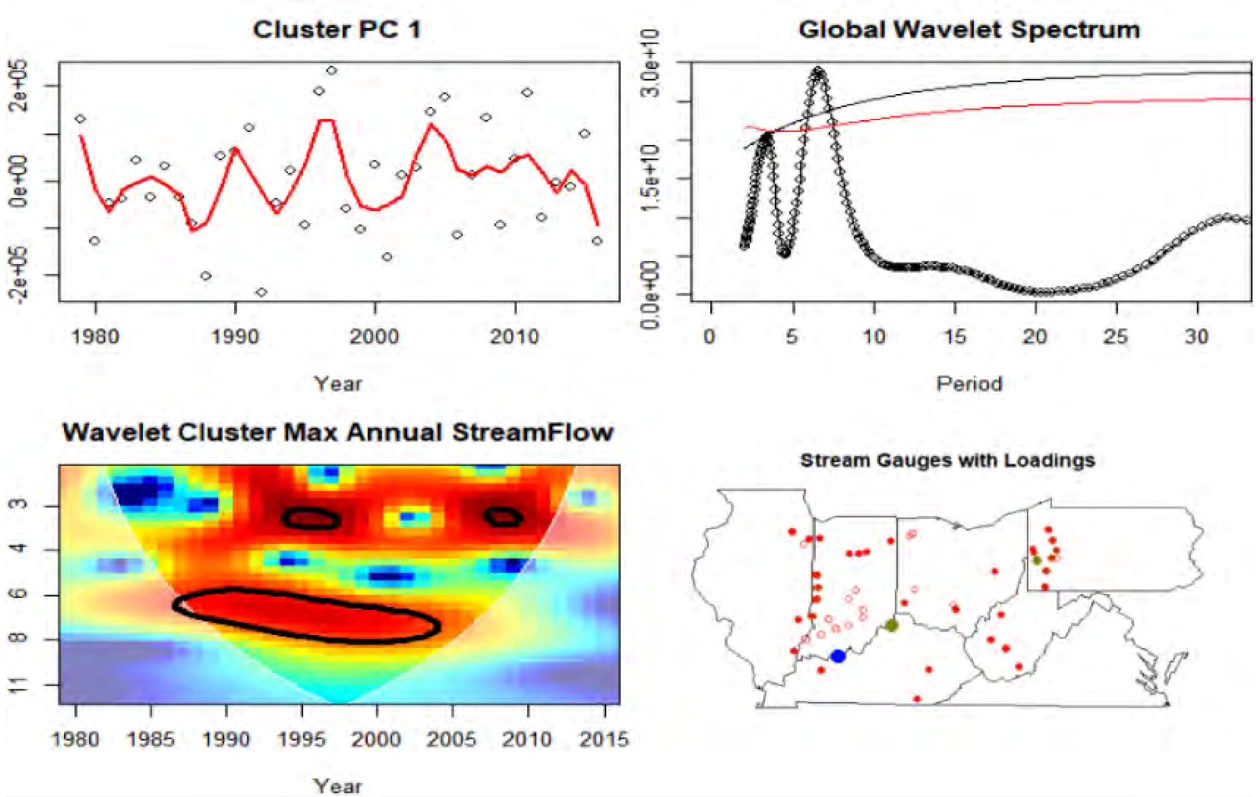


Figure 10-6 Wavelet Analysis on the 1st Principal Component of the spatial cluster but for 39 years of data. TOP-LEFT – 1st PC of cluster with a loess line. TOP-RIGHT – Global Wavelet Spectrum of the PC. BOTTOM-LEFT – Power Spectrum of the Wavelet. BOTTOM-RIGHT – Spatial Distribution of the Loadings (Size indicates their drainage area.

The implicit clustering methodology was applied to the 81 years dataset and the 1st Principal Component of the entire field was extracted. The similar 6-7-year peak is identified in this case too. Further the time-frequency wavelet power structure is also similar for the implicit and explicit methods increasing confidence in these methods. For the PC's of each cluster, the wavelet coherences with individual climate indices and their interactions were also computed and are shown in Figure 10-8. Across NAO, PDO and ENSO we see broad regions where the coherence is statistically significant as computed by Torrence and Compo et al. The raw correlations between the 1st Principal Component of the entire domain with ENSO, NAO, the 1st PC of the dominant cluster along with streamflow exceedances were computed and are attached in Table 10-1.

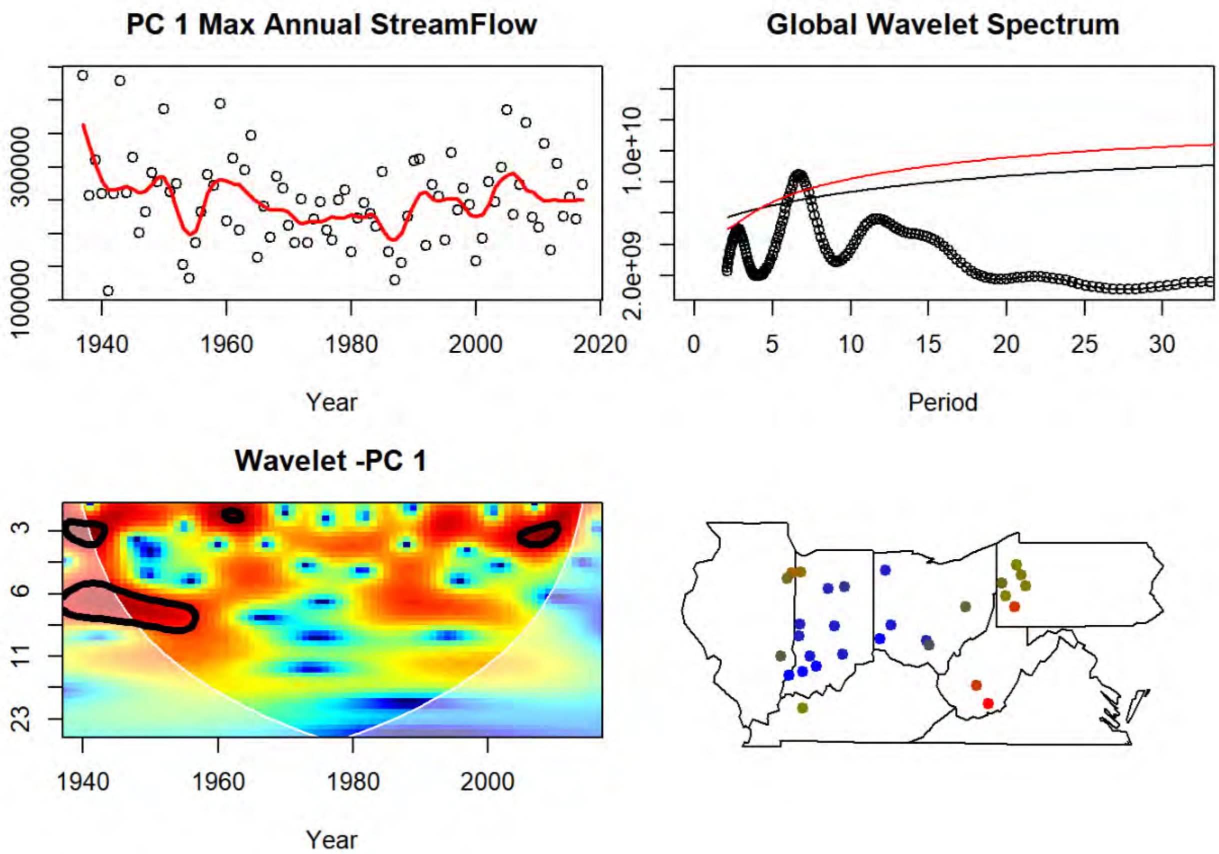


Figure 10-7 Implicit Clustering Wavelet Analysis on the 1st Principal Component of the spatial cluster but for 81 years of data. TOP-LEFT – 1st PC of cluster with a loess line. TOP-RIGHT – Global Wavelet Spectrum of the PC. BOTTOM-LEFT – Power Spectrum of the Wavelet. BOTTOM-RIGHT – Spatial Distribution of the Loadings (Size indicates their drainage area.

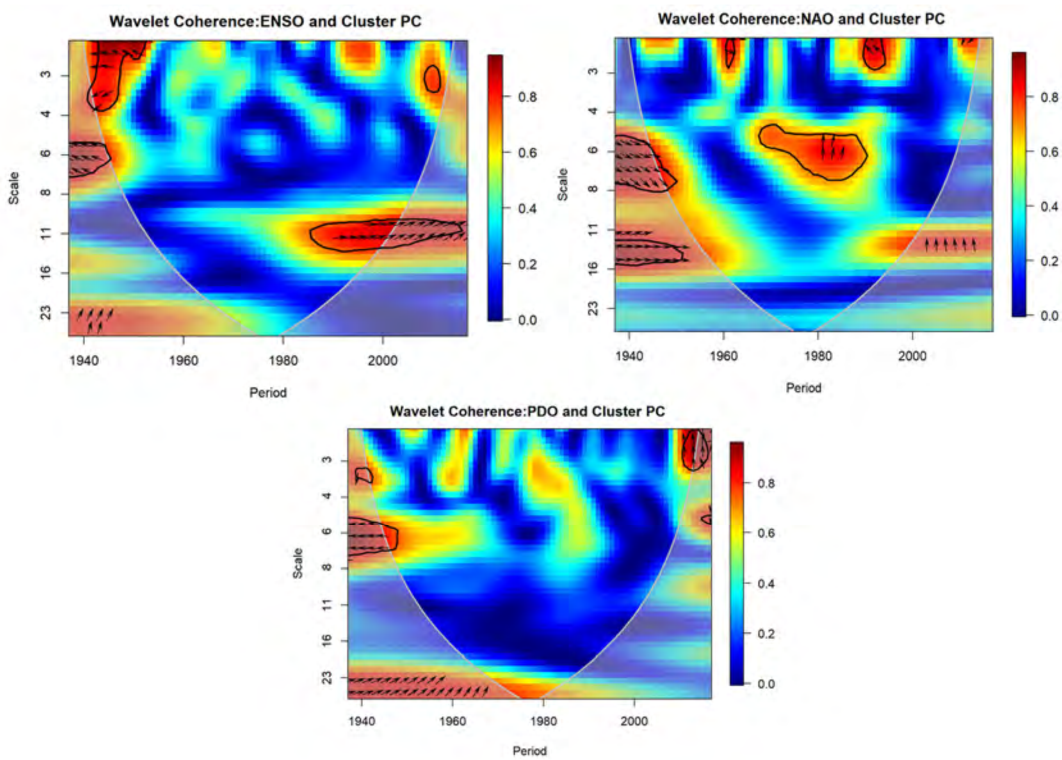


Figure 10-8 Wavelet Coherences of the 1st PC of the entire domain (implicit clustering) with ENSO, PDO and NAO. Regions enclosed in black are regions with 95% confidence and the arrows point to the phase relations between the two.

Long – Term Clustering/Regimes in the Streamflow Domain in the Ohio River Basin - For each site, using the MLE approach a 10-year return period was computed. The R package 'extRemes' was used to compute these site-specific return periods, using which the total count exceedances time series was computed. The annual streamflow exceedance plot in Figure 10-9 shows up the presence of two significant regimes in streamflow patterns in the Ohio River Basin. The period from the beginning of the dataset (1937) to mid-1960's is characterized by sharp peaks in streamflow exceedances. This is followed by a period of low activity which extends till the mid 2000's, following which the increased activity as was seen earlier returns. The green line corresponds to the 95th percentile of a Poisson Distribution fit to the data and the exceeding large number of peaks above this threshold point to presence of clustered spatiotemporal risk during certain periods of time in the river basin. [Note: - The sharp peak at the beginning of the dataset corresponds to the 1937 flood in the region which is the second largest flood event (largest being the floods of 1927)].

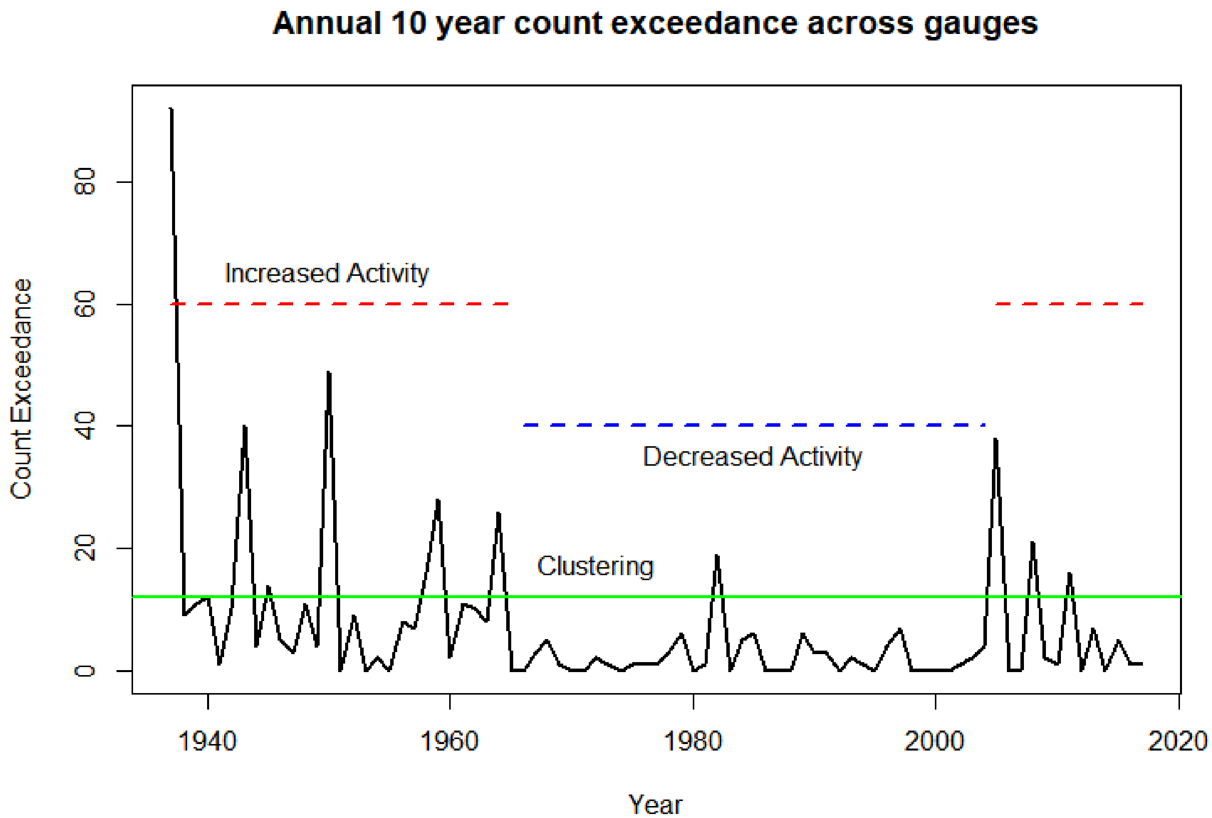


Figure 10-9 Annual Streamflow exceedances crossing the 10-yr return period at the gauges. The red-dotted lines show a period of increased activity whereas the blue dotted lines show periods of decreased activity. The green solid line is the 95% confidence interval for the Poisson Regression fit to the data.

The traditional climate indices capture the dominant modes of climate variability throughout the globe. These modes affect distant local climate through means of teleconnections. Correlations were utilized to examine these relations between the 1st PC of the entire streamflow domain (implicit method) and the DJFM mean climatology of ENSO, NAO along with the 1st PC of the dominant cluster and annual exceedance time series, and results are shown in Table 10-1. Here we see significant correlations of the 1st PC across the entire domain with ENSO and NAO two dominant modes of climate variability affecting the Northern Atlantic Region. The relationship between ENSO, rather the negative correlation is seen in various other studies focused on the region too (Nakamura, Lall, Kushnir, Robertson, & Seager, 2013). A high correlation is also present with the North Atlantic Oscillation, pointing to possible teleconnections or influences of these global models of variability in this local region.

Table 10-1 Correlations of ENSO – Nino 3.4, NAO, 1st PC of cluster based on 81 years of data and annual exceedances with the 1st PC of ann. Max flow across the entire domain for the 81 years

Variable	*Correlation with 1st PC across entire domain – Implicit Clustering Method*
ENSO	-0.24
NAO	0.3
1st PC of Cluster (Explicit Method - Figure 10-6)	0.91
Count Exceedances	0.72

10.5. CONCLUSION

The Ohio River Basin provides an excellent opportunity to advance our understanding of floods, streamflow characteristics and extreme precipitation activity in the region which could be extrapolated to other river basins. Since this region is scattered with aging dams along with other critical infrastructure systems, analysis of the structure and nature of climate extremes in the Ohio River Basin and the greater Mississippi region is of paramount importance.

The analysis of the dynamical extremes in conjunction with its climatic extremes led to valuable insights about the inherent clustering in the system along with the evolution of the system leading up to an extreme precipitation event. This short-term predictability horizon is characterized by a fair amount of organization, as measured by the instantaneous system dimension(d) and instantaneous inverse of persistence(θ), during the build-up towards the event which vanishes rapidly once the event has occurred.

Spatiotemporal wavelet-clustering on the streamflow field allows for the examination of the decadal and semi-decadal variations in streamflow and helps find broad power spectrums at the 6-7 years and 11-12 years band. Major climate indices like ENSO, PDO and NAO along with their interactions are shown to influence the streamflow across the Ohio River Basin capturing a significant mode of variability in this region. Further this also led to identification of time periods corresponding to increased activity and long periods corresponding to diminished activity. Knowledge and predictability of these periods would prove to be valuable in the maintenance of water management infrastructure systems like dams and levees in the Ohio River Basin.

10.6. REFERENCES

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., … Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137(656), 553–597. https://doi.org/10.1002/qj.828

Faranda, D., Messori, G., & Yiou, P. (2017). Dynamical proxies of North Atlantic predictability and extremes. Scientific Reports, 7, 41278. https://doi.org/10.1038/srep41278

Farnham, D. J., Doss-Gollin, J., & Lall, U. (2018). Regional Extreme Precipitation Events: Robust Inference From Credibly Simulated GCM Variables. Water Resources Research, 54(6), 3809–3824. https://doi.org/10.1002/2017WR021318

Jones, P. D., Jonsson, T., & Wheeler, D. (1997). Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. International Journal of Climatology, 17(13), 1433–1450. https://doi.org/10.1002/(SICI)1097-0088(19971115)17:13<1433::AID-JOC203>3.0.CO;2-P

Lorenz, E. N. (1980). Attractor Sets and Quasi-Geostrophic Equilibrium. Journal of the Atmospheric Sciences, 37(8), 1685–1699. https://doi.org/10.1175/1520-0469(1980)037<1685:ASAQGE>2.0.CO;2

Messori, G., Caballero, R., & Faranda, D. (2017). A dynamical systems approach to studying midlatitude weather extremes. Geophysical Research Letters, 44(7), 3346–3354. https://doi.org/10.1002/2017GL072879

Nakamura, J., Lall, U., Kushnir, Y., Robertson, A. W., & Seager, R. (2013). Dynamical Structure of Extreme Floods in the U.S. Midwest and the United Kingdom. Journal of Hydrometeorology, 14(2), 485–504. https://doi.org/10.1175/JHM-D-12-059.1

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Süveges, M. (2007). Likelihood estimation of the extremal index. Extremes, 10(1), 41–55. https://doi.org/10.1007/s10687-007-0034-2

Torrence, C., & Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. Bulletin of the American Meteorological Society, 79(1), 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

11. DECISION-ANALYTIC FRAMEWORK FOR ENGINEERING DESIGN UNDER NON-STATIONARY RISK: CASE STUDY

This chapter is published as a thesis chapter of the PhD thesis manuscript of Dr. Katherine Schlef:

Schlef, K., 2018. Flood risk assessment, management and perceptions in a changing world (PhD thesis). University of Massachusetts, Amherst, United-States. https://scholarworks.umass.edu/dissertations_2/1273/

11.1. EXECUTIVE SUMMARY

Risk-based flood management approaches are challenged by the possibility of non-stationarity in flow magnitudes due to, among other factors, climate change. The key contribution of this work is to advance the field of risk-based flood management by demonstrating how to integrate climate informed projections of flood magnitudes into the decision scaling framework, which can be used to determine optimal design values under uncertainty. Furthermore, this work compares the results of the climate informed method and the decision scaling framework to results from traditional risk analysis based on model chains. This is accomplished for the flood-prone city of Louisville, Kentucky, located on the Ohio River in the Midwest U.S. In the case study, the key decision is the return period of the design flood for the concrete floodwall and earthen levee along the Ohio River channel through the heart of the city. The best design is determined by minimizing the expected total costs, which is the sum of the expected flood damages and the cost of the levee. The results indicate that the best design varies both by the approach, whether traditional or decision-scaling, and by the GCM used to force the models. The conclusion discusses some of the benefits and limitations of both risk-based flood management approaches and highlights areas of future research.

11.2. TECHNICICAL APPROACH

The methodology consists of two main components: developing nonstationary flood projections and performing risk-based analyses. Nonstationary flood projections are developed following the traditional model chain method and following the climate informed method. As extensively discussed in Chapter 2 and 3, the climate informed method capitalizes on the fact that GCMs more skillfully simulate large-scale climate patterns compared to local-scale precipitation and uses a statistical model rather than a hydrologic model to estimate future flood magnitude and frequency of occurrence. Projections following the traditional model chain method were developed by performing flood frequency analysis on the output of a calibrated hydrologic model forced with downscaled projections of precipitation and temperature from GCMs. Projections following the climate informed method are based heavily on the process described in Chapter 2, where flood frequency analysis is performed on the output of a statistical model forced by projections of large-scale ocean-atmospheric patterns from GCMs. For each method, the same GCMs were used to facilitate comparisons and the projections were combined using Bayesian model averaging.

Risk-based analysis was performed following traditional methods and following the decision scaling methodology. Unlike traditional methods, which are limited by the scenarios chosen to force the analysis, decision scaling calculates system response to a wide array of stressors, identifies ex-post scenarios, and only then uses projections to assess probability of occurrence (Brown et al., 2012). Risk is quantified as expected total cost, defined as the sum of the levee cost (calculated as a function of levee height) and expected damages (calculated as the integral of flood probability and modeled damages). The traditional risk-based analysis was implemented using the flood probabilities obtained from the model chain method. The decision scaling analysis was implemented using flood probabilities obtained by forcing the climate informed statistical model with stochastic realizations of the large-scale patterns altered by systematically applied linear trends.

11.3. DATA AND METHODS

11.3.1. CASE STUDY

Louisville, Kentucky, which is located on the Ohio River and has a population of over 600,000 as of 2016 (US Census Bureau, 2018), has experienced a number of devastating floods. The largest recorded flood occurred in 1937, causing damages estimated at approximately 250 million USD (over 4 billion USD in 2016 dollars) (National Weather Service, 2018b). The 1937 flood, and a subsequent major flood in 1945, motivated investment in flood risk management infrastructure (Louisville/Jefferson County Metropolitan Sewer District, 2018a, 2018b). Despite this investment, floods continue to occur; for example, a flood in 1997 caused 200 million USD (nearly 300 million USD in 2016 dollars) in damages to the city, and a flood in 2009 caused 45 million USD (over 50 million USD in 2016 dollars) in damages to the state (National Weather Service, 2018a, 2018b).

The city's flood risk management infrastructure consists of a major concrete floodwall and earthen levee along the Ohio River main channel as well as pumping stations and smaller levees throughout the city (Louisville/Jefferson County Metropolitan Sewer District, 2018a, 2018b). The main levee system is nearly 26 miles long and was built to withstand a flood crest three feet higher than that observed in 1937 (Louisville/Jefferson County Metropolitan Sewer District, 2018a, 2018b); as recorded in the National Levee Database, it was built to the 500 year flood with three feet of freeboard (USACE, 2018) (based on fitting the log Pearson type 3 distribution using l-moments to the annual maximum series streamflow at the Louisville USGS gage for the full record from 1928 to 2017, the 1937 flood crest of 111,000 cfs has a return period of just under 300 years). However, like much of the infrastructure across the U.S., it is aging; the most recent inspection labeled it as "minimally acceptable" (USACE, 2018). Thus, the Louisville levee system is facing many of the same investment questions that are being asked for flood risk management structures across the U.S.

11.3.2. OBSERVED FLOOD EVENTS

Observed daily streamflow data was obtained from USGS gage 03294500, which is located on the Ohio River at Louisville. The gage has a drainage area of 91,170 square miles, has elevation 373.18 feet above NGVD29, and is located at latitude 38o16'49" and longitude 85o47'57". The gage is considered impaired according to the Hydro-Climatic Data Network (Landwehr & Slack, 1992), due to a system of locks and dams upstream. The impact of impairment on flood peaks was investigated by comparison to naturalized data, aggregated from an hourly to daily time step, obtained from the U.S. Army Corps of Engineers (USACE) over the period 2004 through 2015. Surprisingly, annual maximum series (AMS) flood events in the USACE data were higher than those in the USGS gage data for only five out of the 11 years, and the highest flow over the whole time period is recorded by the USGS gage data. As another indication of relative impairment, a reservoir index was calculated following López & Francés (2013). Accounting for all man-made water bodies on the Ohio River main-stem above Louisville, the maximum reservoir index is 0.034, which is much smaller than the threshold value of 0.25 cited by López & Francés (2013) as indicating significant impairment, likely because the capacities of the man-made water bodies are much smaller than the mean annual flow of the river. Both the comparison to the USACE naturalized flow and the calculation of the reservoir index indicate a lack of significant impairment, especially in regards to flood peaks, and thus the USGS data was used without adjustment.

The work on flood events in the northwest region of the Ohio River Basin described in Chapter 4, which is bordered by Louisville, has shown that January through April (JFMA) AMS flood events are mechanistically linked to winter large-scale climate processes. For this reason, the remainder of this work will focus on JFMA AMS flood events. As with AMS flood events, the JFMA AMS flood events are only minimally impacted by upstream impairment. Furthermore, there is no significant trend in JFMA AMS flood events based on the Mann-Kendall trend test. Here and throughout the remainder of the chapter, significance is reported at the 95% level unless noted otherwise.

Realizing that the full AMS is more useful for management decisions than JFMA AMS, we note that nearly 80% of AMS flood events occur in JFMA. Furthermore, a preliminary analysis (not shown), indicates that the model developed for climate informed projections of JMFA AMS (discussed below) is still statistically significant (although less strongly so) when applied to the full AMS. This likely occurs due to the high percentage of AMS events in JFMA. Furthermore, for those AMS events which occur outside JFMA, more than 80% occur during either December or May and are thus likely to be somewhat influenced by winter climate patterns. The caveat is that the climate informed model derives its credibility from the demonstrated mechanistic link between winter flood events and winter climate processes; applying the model without modification to the full AMS reduces the strength of this credibility.

11.3.3. Traditional Model Chain Flood Projections

The model chain method was implemented by forcing a hydrologic model with GCM projections of precipitation and temperature. The hydrologic model is a distributed version of the Soil Moisture Accounting model (SAC-SMA) coupled with a river routing model as described in Brown et al. (2016). The model was implemented on a daily time step at 1/8th degree grid resolution, with three hydrologic response units (i.e., within each hydrologic response unit, the parameter values are the same for each grid cell). Observed daily gridded 1/16th degree precipitation and average temperature were obtained and aggregated to 1/8th degree (Livneh et al., 2013). Model parameters were calibrated using a genetic algorithm over the period 1970 through 1995 inclusive of a five year warm-up period by maximizing the Nash-Sutcliffe Efficiency (Nash & Sutcliffe, 1970), which yielded 0.88. Model performance in the full time period from 1950 through 2010 is also good although the model over-estimates the upper quantiles; the NSE is 0.86 and the JFMA AMS streamflow as a function of return period is shown in Figure 11-1.

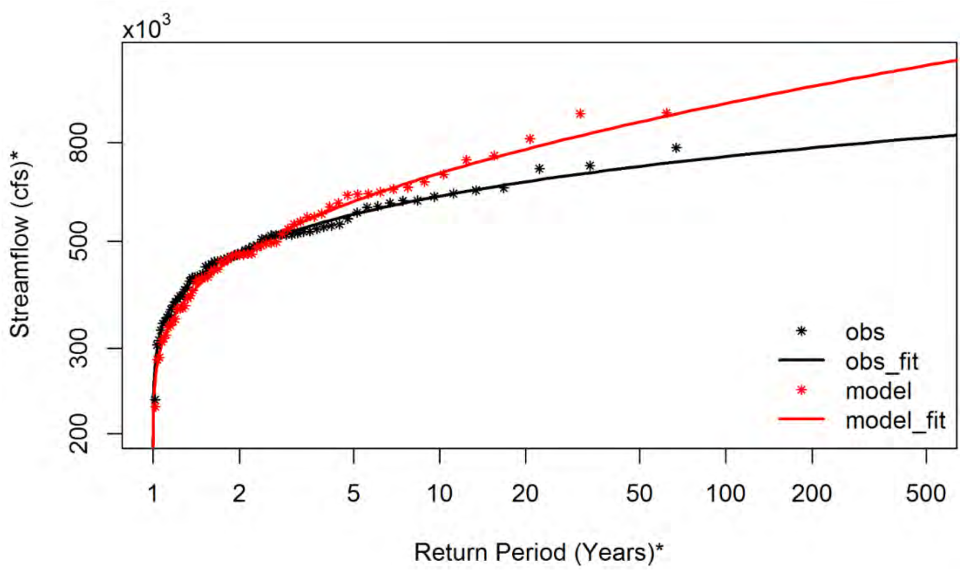


*Figure 11-1: Performance of the hydrologic model (model) relative to observations (obs). "fit" refers to the log Pearson type 3 (LP3) distribution fit to the data. * indicates the axis is on log-scale.*

To create streamflow projections, historical (1950 through 2005) and projected (2006 through 2099) spatially downscaled and bias corrected data from 10 GCMs in the fifth generation of GCM experiments (CMIP5) directed by the Intergovernmental Panel on Climate Change (Van Vuuren et al., 2011) for two representative concentration pathways (RCPs) (4.5 and 8.5) was obtained (Bracken, 2016; Pierce et al., 2014, 2015); the method used for downscaling and bias correction is the localized constructed analog method. The 10 GCMs are CanESM2, CCSM4, CNRM-CM5, CSIRO-Mk3.6.0, GFDL-CM3, GISS-E2-H, HadGEM2-AO, IPSL-CM5A-MR, MPI-ESM-LR, and NorESM1-M. The GCMs were chosen based on availability of the predictor variables used in both the model chain method and the climate informed method. The ensemble member is r1i1p1, except for CCSM4, which uses r6i1p1, and GISS-E2-H, which uses a combination of r6i1p1, r6i1p3, and r2i1p1, due to data availability constraints (Bracken, 2016). The GCM historical and projected data was then used to force the calibrated hydrologic model. Flood events were estimated by fitting the log Pearson type 3 (LP3) distribution to the JFMA AMS of modeled streamflow in 61 year increments ending on every decade from 2010 through 2099 (because 2100 is not available, the last increment has only 60 years). Confidence intervals for the fitted distribution were obtained by sampling with replacement from the time series to create 3000 alternative time series and fitting the LP3 to each. The flood projections from the multiple GCMs were combined using Bayesian model averaging (Hoeting et al., 1999), as described in Raftery et al. (2005). Briefly, Bayesian model averaging is a method for calculating a probabilistic weighted average of multiple models. The weights reflect model performance, which is assessed by linearly regressing observations onto projections of the variable of interest. Here, the observations (projections) are the flood events calculated from the hydrologic model forced with observed (GCM historical) climate. Furthermore, because GCM and observed climate variability are not temporally aligned, the flood events are sorted before linear regression is performed. Thus, the performance of each GCM is assessed by comparing the empirical distribution function of the model output when forced with observed climate to that of the model output when forced with GCM historical climate. The weights ranged from 0.04 to 0.22, with the lowest assigned to CanESM2 and the highest assigned to MPI-ESM-LR.

11.3.4. Climate Informed Flood Projections

The climate informed flood projections are closely based on the methods and results described in Chapter 2 and 4. Gridded monthly climate data (sea surface temperatures – Rayner, 2003; geopotential heights at the 500 mbar pressure level – Kalnay et al., 1996; soil moisture – Fan & van den Dool, 2004) was converted to annual time series by taking the maximum value within December through February (DJF) or within JFMA. Correlation maps between the annual climate data and JFMA AMS flood events show significant relationships to DJF sea surface temperatures in the eastern tropical Pacific, DJF geopotential heights at the 500 mbar level over central Canada and the North Pacific, and JFMA soil moisture over the basin (Figure 11-2). From the correlation maps, the following predictors were developed: sst_{ETP}^{DJF} is the DJF sea surface temperatures averaged over the eastern tropical Pacific region (15S – 5S, 145W – 85W), hgt_{CC-NP}^{DJF} is the difference in DJF geopotential heights at the 500 mbar level averaged over central Canada (38N – 46N, 97W – 90W) and averaged over the North Pacific (46N – 52N, 150W – 140W), and $soil_{basin}^{JFMA}$ is the JFMA soil moisture averaged over the center of the basin (37N – 40N, 86W – 81W). The predictors are standardized and the resulting correlations are given in Table 11-1 Correlations between the standardized predictors and the standardized log of JFMA AMS, denoted X. *, **, and *** indicates that the p-value lies between 0.05 and 0.01, between 0.01 and 0.001, and is less than 0.001, respectively..

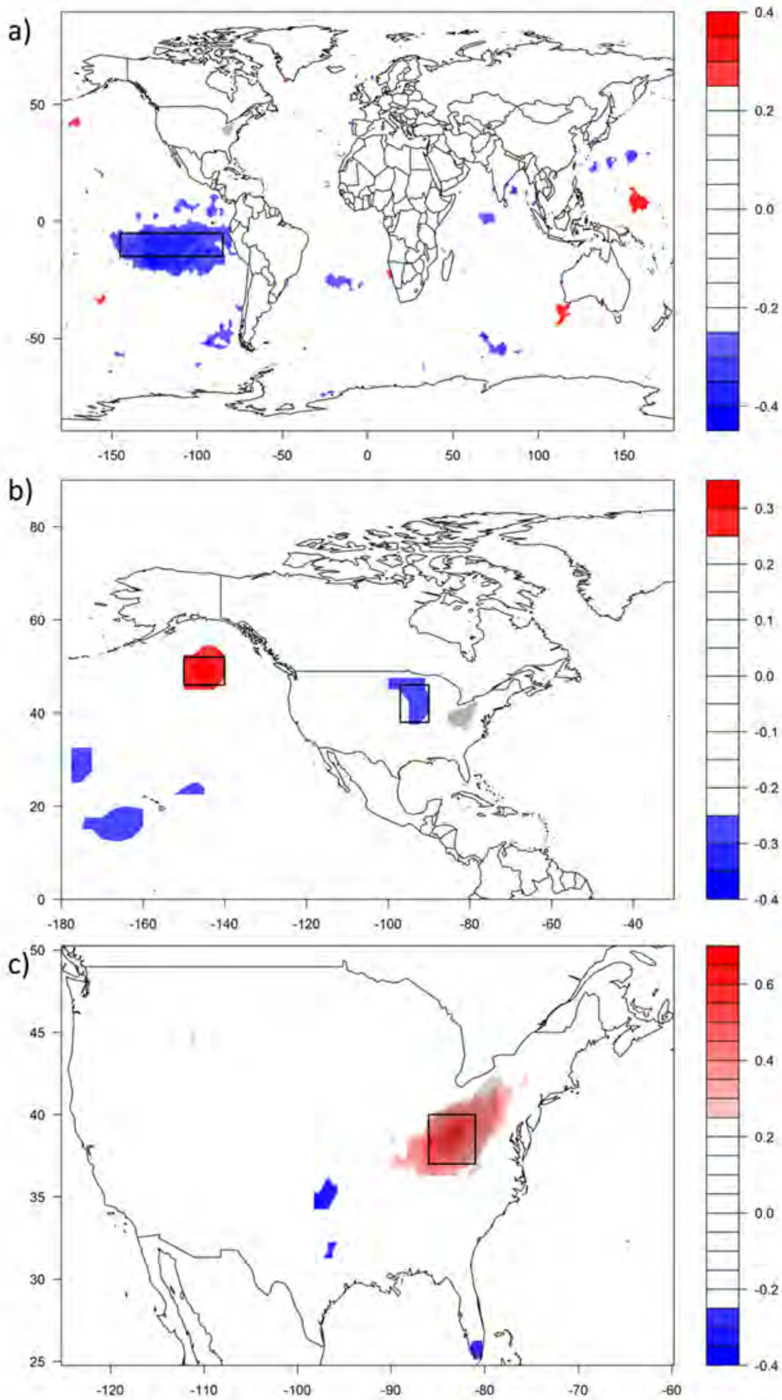


Figure 11-2: Correlation maps of the standardized logarithm of flood events to climate variables. (a) DJF sea surface temperatures (the region in the eastern tropical Pacific is outlined by a rectangle), (b) DJF geopotential heights at the 500 mbar level (the central Canada and the North Pacific regions are outlined by rectangles), (c) JFMA soil moistures (the region over the basin is outlined by a rectangle and the Louisville gage is represented by a point. The scale indicates the magnitude of the correlation (white areas are not significant), the basin is shaded grey, and the x- and y-axis labels are longitude and latitude, respectively.

*Table 11-1 Correlations between the standardized predictors and the standardized log of JFMA AMS, denoted X. *, **, and *** indicates that the p-value lies between 0.05 and 0.01, between 0.01 and 0.001, and is less than 0.001, respectively.*

	$\boldsymbol{sst_{ETP}^{DJF}}$	$\boldsymbol{hgt_{CC-NP}^{DJF}}$	$\boldsymbol{soil_{basin}^{JFMA}}$
$\boldsymbol{X}$	-0.378**	-0.376**	0.611***
$\boldsymbol{sst_{ETP}^{DJF}}$	1	0.403***	-0.270*
$\boldsymbol{hgt_{CC-NP}^{DJF}}$		1	-0.349**

Given the multiple predictors identified, multiple models were developed (Table 11-2: Model form and associated parameters and performance.). The models were fit over the time period 1950 through 2015 by JAGS in R (Plummer, 2016; Yu-Sung & Yajima, 2015) using three model chains each having 2000 samples with 1000 samples discarded as burn-in. Sufficiently vague priors were placed on the variances (a uniform distribution from zero to 10) and on the coefficients (a normal distribution with mean zero and variance 25). For all models, both the potential scale reduction factor, also known as Gelman's R, and the effective sample size were well within accepted rules of thumb (less than 1.1 and greater than 300, respectively). Predictors are deemed to be significant if the 95% credible interval of the coefficient does not include zero. Model performance is judged by two statistics. The first is the coefficient of determination, R^2, between the simulated and observed; higher is better. The second is the deviance information criterion (DIC) which accounts for parameter uncertainty and is appropriate even when the prior is non-informative or improper; lower is better (Spiegelhalter et al., 2002; Sun et al., 2014).

Table 11-2: Model form and associated parameters and performance. $N()$ indicates the normal distribution. Values are given as the mean (standard deviation).

Model	Model Equation	$\boldsymbol{\alpha_1}$	$\boldsymbol{\alpha_2}$	$\boldsymbol{\alpha_3}$	$\boldsymbol{\beta}$	$\boldsymbol{\sigma}$	$\boldsymbol{R^2}$	DIC
all3	$Q_{AMS}^{JFMA} \sim N\left(\alpha_1 sst_{ETP}^{DJF} + \alpha_2 hgt_{CC-NP}^{DJF} + \alpha_3 soil_{basin}^{JFMA} + \beta, \sigma^2\right)$	-0.19 (0.11)	-0.12 (0.11)	0.52 (0.10)	0.00 (0.10)	0.79 (0.07)	0.18 (0.08)	159
soil&hgt	$Q_{AMS}^{JFMA} \sim N\left(\alpha_2 hgt_{CC-NP}^{DJF} + \alpha_3 soil_{basin}^{JFMA} + \beta, \sigma^2\right)$	-	-0.19 (0.10)	0.55 (0.11)	0.00 (0.10)	0.80 (0.07)	0.16 (0.08)	160.5
soil&sst	$Q_{AMS}^{JFMA} \sim N\left(\alpha_1 sst_{ETP}^{DJF} + \alpha_3 soil_{basin}^{JFMA} + \beta, \sigma^2\right)$	-0.23 (0.10)	-	0.55 (0.10)	0.00 (0.10)	0.79 (0.07)	0.18 (0.08)	158.5
soil	$Q_{AMS}^{JFMA} \sim N\left(\alpha_3 soil_{basin}^{JFMA} + \beta, \sigma^2\right)$	-	-	0.61 (0.10)	0.00 (0.10)	0.82 (0.08)	0.14 (0.07)	161.9

As in Chapter 4, the sign of coefficients of the fitted models match what is expected from the correlation maps and the literature and the intercept is essentially zero for all models, as expected. R^2 and DIC are inversely related, and the variance decreases as model performance improves. Following logic similar to that in Chapter 4, the model with sst_{ETP}^{DJF} and $soil_{basin}^{JFMA}$ as predictors (soil&sst) is the best and is used in all subsequent analysis. Based on the Shapiro-Wilk test for normality, the residuals of the soil&sst model are normal for more than 95% of the 3000 model runs. Simulated data can be obtained by sampling from the model to stochastically generate a time series, de-standardizing, and taking the exponent. Quantiles of interested are developed by using l-moments to fit the simulated data to a LP3 distribution. When compared to observations, the model does a good job of fitting the data (Figure 11-3).

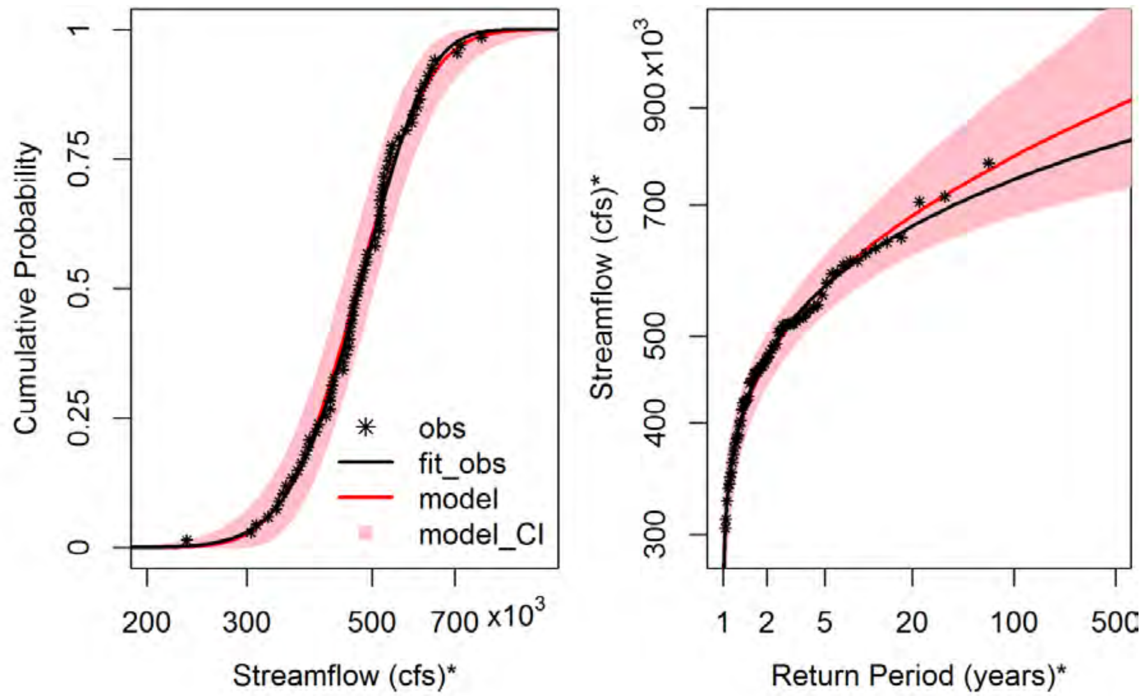


*Figure 11-3: The performance of the climate informed model. The empirical cumulative distribution function based on the Weibull plotting position of the observed data (obs) and the LP3 fit to the observed data (fit_obs) and to the model forced with observed climate (model) with associated credible intervals (model_CI). * indicates the axis is log-scale.*

Unlike the model in Chapter 4 which includes predictors based on geopotential height, this model includes a predictor based on sea surface temperature. From a climate science perspective, sea surface temperature is a largely thermodynamic variable and can be expected to increase under global warming. Consequently, flood events, which are negatively correlated to sea surface temperatures, can be expected to decrease absent other regulating mechanisms. In the model, soil moisture does provide some regulation, however, a revised model was tested that would better account for both the dynamics and thermodynamics of climate change by replacing the sst_{ETP}^{DJF} predictor with the Southern Oscillation Index (SOI), which is another measure of the ENSO phenomenon and is based on sea-level pressure anomalies in the tropical Pacific. The SOI data, obtained from (NCAR, 2018), was processed in the same manner as sst_{ETP}^{DJF} to obtain a SOI^{DJF} predictor. However, despite the highly significant correlation between sst_{ETP}^{DJF} and SOI^{DJF}, the revised model had poor performance and was not used for subsequent analysis.

To make projections of future flood events, projections of the predictors are obtained from GCM simulations and used to force the statistical model. Specifically, monthly gridded historical runs from 1950 through 2005 and projections from 2006 through 2100 of sea surface temperature and soil moisture are obtained from the same 10 CMIP5 GCMs used for the model chain projections for RCP 4.5 and 8.5 and the r1i1p1 ensemble member (except that GFDL-CM3 did not have data available for RCP 4.5). Flood events are estimated by fitting the LP3 in 61-year increments ending on every decade from 2010 through 2100. Simulations from each GCM are combined using Bayesian model averaging following the same procedure as described for the model chain method. The weights ranged from 0.094 to 0.103 with the lowest assigned to IPSL-CM5A-MR and the highest assigned to NorESM1-M.

11.3.5. Traditional Risk-Based Analysis

Traditional risk-based analysis consists of optimizing a risk-based metric across a range of probable scenarios. Here, we chose to minimize the expected total cost (similar to (Qi, 2017; Qi and Liu, 2018; Rehan and Hall, 2016) , T_k, associated with a levee built to withstand a flood of return period k,

$$T_k = C_k + ED_k \qquad 11.1$$

where C_k is the cost of the levee and ED_k is the expected damages,

$$ED_k = \int_0^\infty P(q) D_k(q) dq \qquad 11.2$$

where $P(q)$ is the probability and $D_k(q)$ is the damages associated with a flood of magnitude q. The flood probability is given by the traditional model chain projections described previously, while calculation of flood damages and levee cost is described below. The calculation of ED_k was accomplished by numerical integration for the first and last moving window of the projections (1950 through 2010 and 2040 through 2099, respectively).

Flood damages were determined from a HAZUS model. HAZUS is a program developed by the Federal Emergency Management Agency and has been applied to a variety of questions concerning flood damage estimation; some examples include the cities of Atlanta, Georgia (Ferguson and Ashley, 2017) Cairo, Illinois (Luke et al., 2015), and Cedar Rapids, Iowa (Tate et al., 2016), the regions of the Middle Mississippi River (Remo et al., 2012) and the Sacramento – San Joaquin Delta in California (Burton and Cutter, 2008), the states of Illinois (Remo et al., 2016) and Pennsylvania (State of Pennsylvania, 2013) and the country of Canada (Nastev and Todorov, 2013). In brief, given a flow volume, HAZUS simulates flooded area elevation and extent using a digital elevation model and flow routing, links that data to census data regarding the type and location of infrastructure, and calculates building loss damages from elevation-cost functions specific to each infrastructure type. HAZUS also estimates indirect damages; that is, “dislocations in economic sectors no sustaining direct damage” (Scawthorn et al., 2006a). However, indirect damages are not reported in this study due to the high uncertainty associated with their estimation. Flood risk management options (e.g., levees/floodwalls, dams, and early warning systems) can also be incorporated into a HAZUS model. HAZUS has different levels of simulation complexity; here, a level 1 analysis (the simplest) was used due to the increased data requirements associated with levels 2 and 3. For a full description of flood damage simulation in HAZUS, see (Scawthorn et al., 2006a, 2006b).

Despite the relative simplicity of a level 1 analysis compared to levels 2 and 3, there is still a number of modeling choices required to successfully define and run a HAZUS model. The study region was chosen to be Jefferson County, Kentucky, which includes the city of Louisville, with an area of 900 km^2 (350 square miles). Topographic data was obtained from the USGS's National Elevation Database. The Manning's roughness coefficient was set to the default value of 0.160. Based on a sensitivity analysis, the drainage threshold was chosen to be 225 square miles, corresponding to the smallest area (rounded up to the nearest 5 square miles) for which only the Ohio River is delineated. This choice of drainage area excludes direct modeling of flooding on small tributaries; however, this simplification was deemed appropriate given that only the levee along the main channel is analyzed and not the system of pumps and smaller levees spread throughout the city. The magnitudes of the 2, 5, 10, 25, 50, 100, 200, and 500 year floods used to define the flood event distribution in HAZUS were calculated from the quantiles of the LP3 fitted by maximum likelihood estimation to the JFMA AMS data at the Louisville gage from 1950 through 2015 (the fitted values are 14.4, 0.036, and 37.6 for the location, shape, and scale parameters respectively). The location of the current levee in Louisville was added to the model using data obtained from the USACE's National Levee Database. In HAZUS, the protection level provided by a levee is not specified by its height, but rather by choosing the flood return period for which it protects (within an allowable range of 5 to 500 years). For this analysis, the return periods for the levee protection level were chosen to be 5, 10, 25, 50, 100, and 500; the case of no levee was also modeled.

A continuous damage function is needed for calculation of ED_k, but is computationally expensive. Instead, we assumed that the case of no levee represents an upper limit to possible damages (Figure 11-4 and Table 11-3). We note that the high damages caused by the two-year flood in the absence of a levee likely occur because the city has experienced significant development after the completion of the levee which relies on the levee's protection. To determine the functional form of damages in the presence of a levee, we performed a preliminary analysis using the levee built for the 100 year flood (Figure 11-4). The preliminary analysis showed that the damages are linear up to the 100-year flood. Immediately after the 100 year flood, the damages jump up and follow the magnitude of the damages associated with no levee. Intuitively this makes sense; once the flood is greater than 100 years, all the formerly protected areas are now inundated. Based on these results, strategic combinations of levee return period and flood volumes were chosen to minimize computational expense while still fully characterizing the system (Table 11-3). For any levee, damages from floods below its protection level are assumed to follow the lowest simulated value, while damages from floods above its protection level are assumed to follow the case of no levee. A continuous function is created by assuming a linear piece-wise regression as a function of streamflow between points. For the lower tail of the distribution, damages are assumed to go to zero at the flood with return period 1.01 years, and for the upper tail of the distribution, damages are assumed to increase to 17 billion USD at the flood with return period 3000 years. Damages remain capped at 17 billion USD for all greater floods. Since the return period associated with no levee cannot be calculated, the expected total cost for the levee with return period 1.01 was calculated by linear interpolation between the expected total cost of the case of no levee and the two year levee.

The levee cost, C_k, was estimated using a function modified from (Al-Futaisi and Stedinger, 1999)

$$C_k = a{h_k}^b \qquad 11.3$$

where h_k is the average height of the levee, a is a scaling parameter, and the exponent b ranges from 2 to 3.5 (here, values of 2.65, 2.75, and 2.85 were used). Because levees in HAZUS are specified by return period rather than height, the average height associated with each levee was determined by running the model without the levee, averaging the modeled height of the water at 40 randomly picked locations along the levee, and adding three feet to represent freeboard. The value of a was estimated using the following approximations, given a lack of more precise data on levee cost. Recalling that the current Louisville levee was designed to the 500 year flood plus three feet of freeboard (USACE, 2018), then its height in HAZUS is approximately 22.7 feet, which is the average height associated with a 500 year protection level including three feet of freeboard. The cost of the 26 mile long levee (USACE, 2018) is approximated to range between

100 to 120 million USD per mile (in increments of 10 million USD); this ratio is roughly estimated from the 14.5 billion USD used to repair and upgrade New Orleans flood protection infrastructure, which includes 133 miles of levees encircling the city, after hurricane Katrina (Llanos, 2015). The nine different possible cost parameter combinations (three values of *b* by three values of cost per mile) were used in all future analysis.

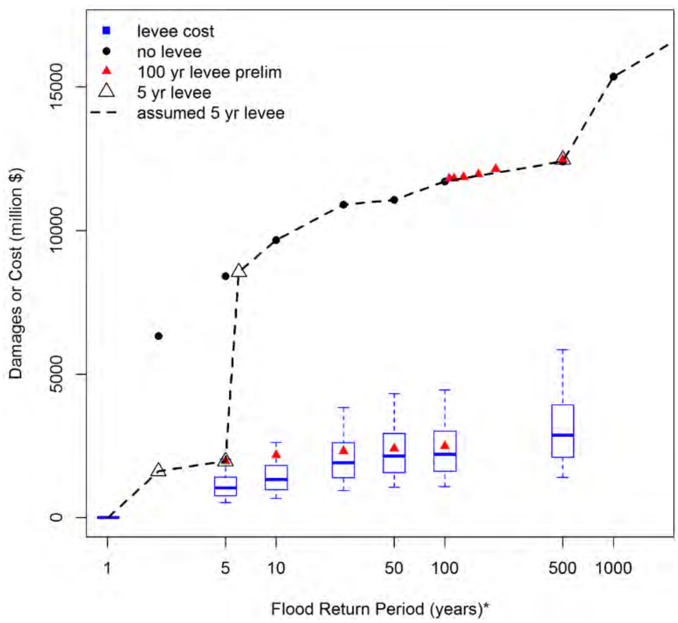


Figure 11-4: Damages and costs associated with levees. The levee cost ("levee cost") as well as damages with no levee ("no levee"), for the preliminary analysis with the 100 year levee ("100 yr levee prelim"), the 5 year levee ("5 yr levee"), and the assumed damage function for the 5 year levee ("assumed 5 yr levee"). The shape of the assumed damage function is similar across all levee protection levels but is not shown for clarity.

11.3.6. Decision Scaling Risk-Based Analysis

Like traditional risk-based analysis, decision scaling risk-based analysis also seeks to minimize a risk-based metric and often, though not investigated here, to apply robustness-based approaches (e.g., Spence and Brown, 2016); the key difference compared to traditional analyses is that decision scaling centers around a system vulnerability analysis. Thus, while cost and damages are assessed in the same way using the functions described previously, the flood probabilities do not come from the traditional model chain flood projections, but are systematically and stochastically generated. Only after the system vulnerability analysis is complete are projections superimposed on the results.

Previous decision scaling studies of floods have demonstrated two approaches to generating floods. (Poff et al., 2015) and (Steinschneider et al., 2015) obtain time series of temperature and precipitation from a stochastic weather generator, apply systematic additive or multiplicative changes to those time series, force a hydrologic model with the perturbed stochastic time series, and then calculated floods from the hydrologic model output. Alternatively, Spence and Brown (2016) apply systematically chosen linear trends to the location parameter of the log-normal distribution. With the climate informed model, there is now a third option in which new flood probabilities are generated from perturbations in the predictors.

*Table 11-3: HAZUS data inputs and outputs. The units are as follows: flood return period (FRP) and protection level (PL) (years), flood volume (FV) (1000 cfs), damages (million USD), average height including freeboard (AH) (feet). Grey indicates the value is assumed. *The 1000 year flood magnitude is not a HAZUS input.*

FRP		2	5	6	10	11	25	27	50	55	100	110	500	1000
FV		475	568	583	620	625	676	681	714	719	748	753	819	846*
PL	AH	Damages												
0	0	6,325	8,411		9,673		10,906		11,063		11,703		12,414	15,366
5	15.7	1,617	1,965	8,552	9,673		10,906		11,063		11,703		12,472	15,366
10	17.2	1,617	1,965		2,173	9,730	10,906		11,063		11,703		12,472	15,366
25	19.6	1,617	1,966		2,173		2,315	10,990	11,063		11,703		12,472	15,366
50	20.4	1,617	1,966		2,173		2,315		2,405	11,090	11,703		12,472	15,366
100	20.7	1,617	1,965		2,173		2,315		2,405		2,488	11,807	12,472	15,366
500	22.7	1,617	1964		2,173		2,315		2,405		2,488		2,488	15,366

Here, perturbations in the predictors are accomplished by bootstrap sampling of the historic record of sst_{ETP}^{DJF} and $soil_{basin}^{JFMA}$ based on sequences from a lag 1 Markov chain built to reproduce the states of sst_{ETP}^{DJF}. The Markov chain operates on an annual time step and has three discrete states, representing El Nino, Neutral, and La Nina conditions. The Markov chain is specified by

$$\pi_j^{t+1} = \sum_{i=1}^{S} p_{ij}\pi_i^t \quad \forall\, j = 1 \dots S \qquad 11.4$$

where p_{ij} is the probability of transitioning from state i to j, π_i^t is the unconditional probability of state i in time period t, and S is the total number of states. The chain is constrained such that the sum of the unconditional probabilities equals one ($\sum_{i=1}^{S} \pi_i = 1$) and the sum of the transition probabilities from a given state to any other state equals one ($\sum_{j=1}^{S} p_{ij} = 1 \;\forall\, i = 1 \dots S$).

To calculate the unconditional and transition probabilities, monthly sea surface temperatures averaged over the sst_{ETP}^{DJF} region were obtained from (Rayner, 2003) for the years 1870 through 2015. Monthly anomalies were calculated using a 31-year moving window ending on the year of interest. For example, the February 1900 monthly anomaly is the February 1900 monthly value minus the mean of all February values from 1870 through 1900. The monthly anomalies were smoothed using a three-month moving average, resulting in a dataset from February 1990 to November 2015. A monthly state time series was developed by identifying El Nino (La Nina) months as those for which the smoothed anomaly is ≥ 0.3°C (≤ -0.3°C) for at least six consecutive months; all other months were designated as Neutral. Subsequently, an annual state time series, based on a July to June year, was developed from the monthly time series by identifying as El Nino (La Nina) those years for which as least five months were designated El Nino (La Nina); all other years were designated as Neutral. This process is similar in form to that used by the National Weather Service’s Climate Prediction Center (NOAA, 2015). From the annual state time series, the unconditional probabilities are calculated as the number of years in a given state divided by the total number of years and the transition probabilities are calculated as the number of times in which a given initial state is followed by another given state divided by the number of years in the initial state (Table 11-4). The resulting unconditional probabilities are similar to those reported by (Trenberth, 1997), in which ENSO state is calculated with slightly different thresholds using the Nino3.4 region.

Table 11-4: Unconditional and transition probabilities of the Markov chain. The transition probabilities are from the state in the row to the state in the column. EL is El Nino, NU is Neutral, and LA is La Nina.

				EL	NU	LA
EL	Unconditional	0.27	Transition	0.32	0.42	0.26
NU		0.45		0.215	0.57	0.215
LA		0.28		0.28	0.31	0.41

Stochastic realizations of annual states are generated by sampling the state of the first year according to the unconditional probabilities, and then iteratively sampling the state of each successive year according to the transition probabilities associated with the current state. The realizations are 150 years long to match the length of the model chain results from GCM historical runs and projections. In total, 500 realizations are generated; to reduce computational expense, only the 10 whose unconditional probabilities are closest to observed are retained for subsequent analysis. The realizations are then used to perform bootstrap sampling of years in the historic record with replacement (e.g., if the ENSO state is El Nino for a given year, then one of the years designated as El Nino is randomly sampled). Time series of sst_{ETP}^{DJF} and $soil_{basin}^{JFMA}$ are created by drawing the data associated with each bootstrapped year.

To create the stress test, systematic linear trends are added to the stochastic realizations of sst_{ETP}^{DJF} and $soil_{basin}^{JFMA}$. The trends are created such that the total change over the length of the realization ranges from zero to six in increments of two for sst_{ETP}^{DJF} and from -1 to one in increments of one for $soil_{basin}^{JFMA}$. These ranges nearly encompass the range of change projected by the GCMs (Figure 11-5a). In total, there are four sst_{ETP}^{DJF} scenarios by four $soil_{basin}^{JFMA}$ scenarios by 10 realizations for a total of 160 scenarios used to force the climate informed model. The climate informed model generates 3000 samples, which, when combined with 7 possible levee return periods and 9 possible cost function parameter sets, is highly computationally expensive, especially for numerical integration. Thus, 51 of the 3000 samples which span the sample space are retained for subsequent analysis (Figure 11-5b). The expected damages and expected total cost are calculated for the last 60 years of the time series, which matches the last moving window used for the traditional risk analysis. Finally, the expected value of the expected total cost over the GCM projections for each levee design is calculated by bilinear numerical integration of a bicubic approximation of T_k and a bivariate normal distribution fitted to the GCM projections for both RCP 4.5 and 8.5.

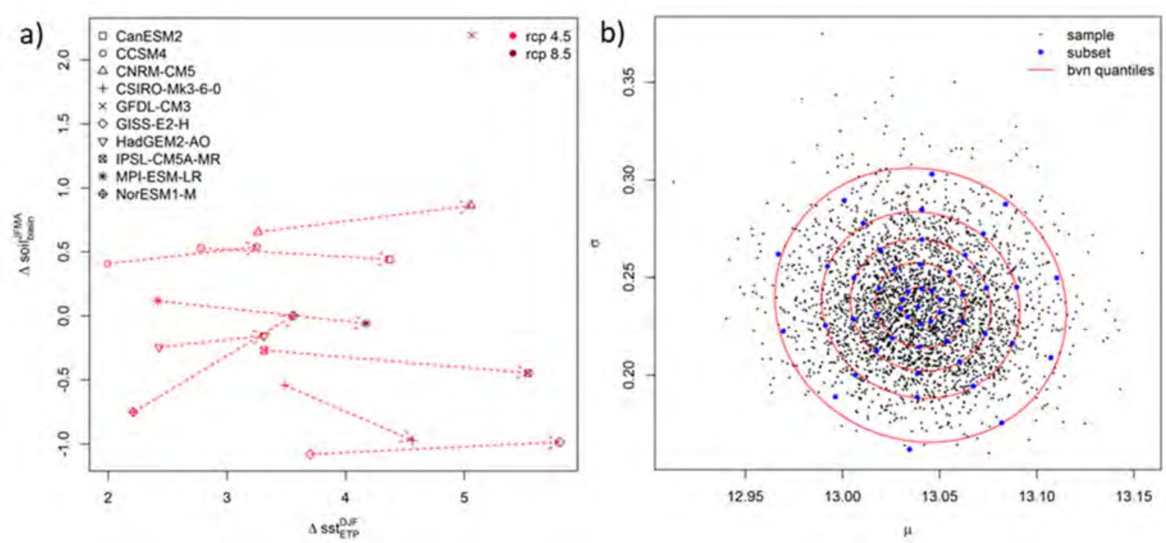


Figure 11-5: System vulnerability analysis information. (a) Changes in the climate predictors projected by the GCMs (the arrows indicate the change from RCP 4.5 to RCP 8.5 and the axes are unit-less because the values are standardized). (b) Subsets of the samples based on quantiles of a bivariate normal distribution (bvn) fitted to the mean and standard deviation of the fitted LP3 for each sample.

11.4. RESULTS AND DISCUSSION

The results are comprised of two parts. The first is the GCM projections of all climate variables used as drivers for the models (precipitation, temperature, sst_{ETP}^{DJF} , and $soil_{basin}^{JFMA}$) and the resulting flood projections from both the model chain and climate informed approach. The second is the expected total cost results from traditional risk analysis and decision scaling and a comparison of the decision-relevant information from both methods.

11.4.1. Projections

Projections of the climate variables are shown in Figure 11-6. GCM simulation of sst_{ETP}^{DJF} performs well over the historic period except for underestimation of the high extremes. Future sst_{ETP}^{DJF} is projected to increase, which is expected because temperature-based variables are increasing due to global warming; the greatest increase is associated with RCP 8.5, which is the more extreme scenario. GCM simulation of $soil_{basin}^{JFMA}$ also performs relatively well over the historic period except for under- (over-) estimation of the high (low) extremes and the unusual behavior of IPSL-CM5A-MR. Future $soil_{basin}^{JFMA}$ may increase or decrease depending on the GCM, with no consistent difference in magnitude of change between RCP 4.5 and 8.5. Notably, GFDL-CM3 projects an exceptionally high increase under RCP 8.5. GCM simulation of extreme precipitation, defined as any daily JFMA data above the 98th percentile, exhibits nearly consistent overestimation over the historic period except HadGEM-AO which consistently underestimates. Future extreme precipitation is projected to increase, with no consistent difference in magnitude of change between RCP 4.5 and 8.5. GCM simulation of temperature exhibits very little bias over the historic period, although the comparison to the other predictors is not direct because the temperature quantiles are calculated from the full daily data. As expected with global warming, temperatures are projected to increase, with the greatest increase associated with RCP 8.5.

Flood projections from both the model chain and climate informed methods for select GCMs and the Bayesian model average are shown in Figure 11-7. GCM performance can be assessed by comparing the model forced with observed climate, hereafter the “observed model”, to the model forced with GCM historic climate, hereafter the “GCM historic model”. All GCMs perform satisfactorily using the climate informed method (i.e., the GCM historic model closely follows the observed). However, GCM performance varies widely using the model chain method; while the NorESM1-M historic model closely follows the observed model, both the GISS-E2-H and GFDL-CM3 historic models greatly underestimate the upper return periods. Even though this underestimation results in a closer alignment to the observed data, this does not indicate improved performance, but rather that the GCMs are introducing additional error on top of that contributed by the hydrologic model. For both the model chain and climate informed method, the Bayesian model average of the GCM historic models slightly underestimates the observed model.

The direction and magnitude of change projected by the GCMs can be assessed by comparing the GCM historic model to the model forced with GCM future climate from RCP 4.5 and 8.5. Furthermore, the likely causes of the projected changes can be determined from the projected changes in the predictors shown in Figure 11-6. For the model chain method, flood events are projected to increase in the future by both individual GCMs and the Bayesian model average, likely due to the projected increase in extreme precipitation. Furthermore, the magnitude of the increase is greater for RCP 8.5, the more extreme scenario, than for RCP 4.5; GFDL-CM3 is an exception likely because the projected increase in extreme precipitation is greater for RCP 4.5 than for RCP 8.5. For the climate informed method, flood events are projected to decrease in the future by both individual GCMs and the Bayesian model average, likely because the large projected increases in sea surface temperature are either not completely offset by projected increases in soil moisture, or for some GCMs, are even accentuated by projected decreases in soil moisture. Here again the exception is GFDL-CM3, where the exceptionally large projected increase in soil moisture offsets the projected increase in sea surface temperature.

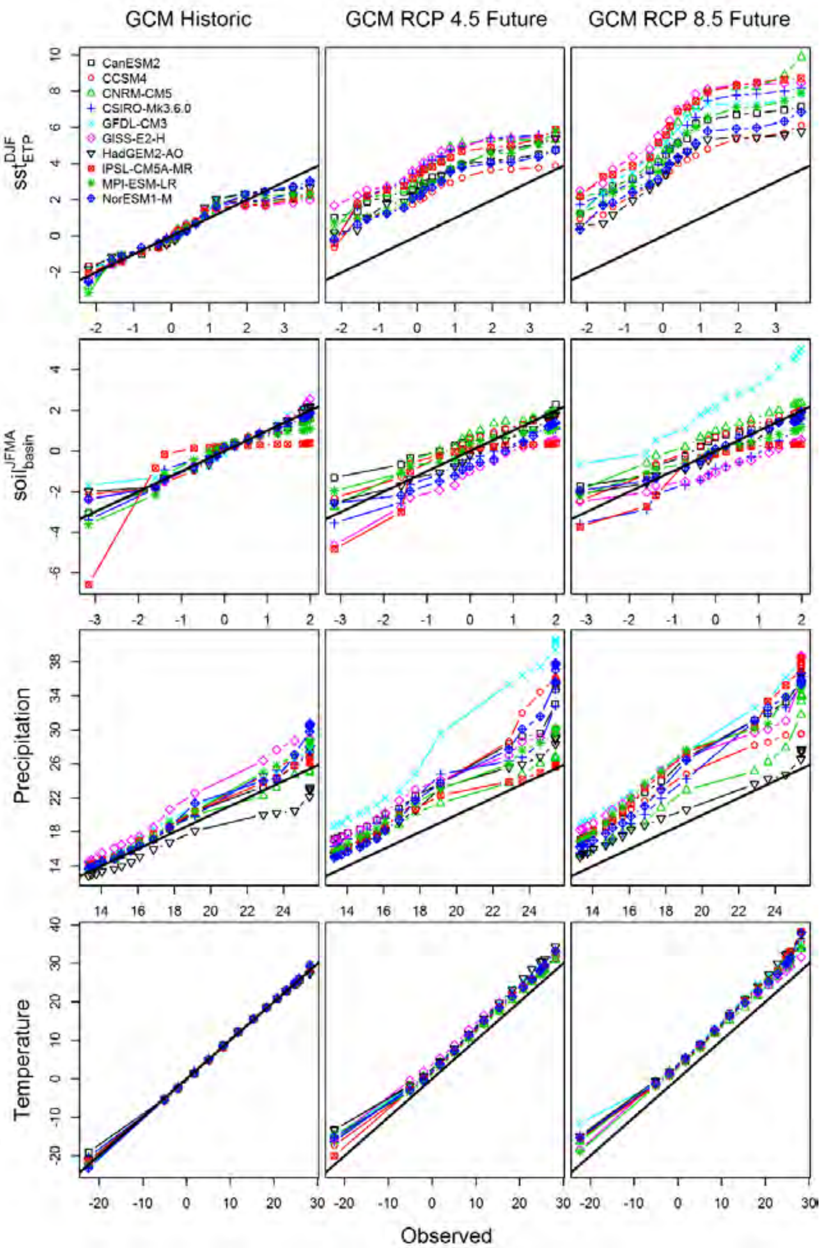


Figure 11-6: GCM performance and projections of climate variables. Selected quantiles of the standardized annual sst_{ETP}^{DJF} and $soil_{basin}^{JFMA}$ (unit-less) and daily precipitation (mm) and temperature (°C) of observations (1950 through 2010) versus GCM historic (1950 through 2005) and future (2040 through 2100 for sst_{ETP}^{DJF}, and $soil_{basin}^{JFMA}$ and through 2099 for precipitation and temperature) values for both RCP 4.5 and RCP 8.5. The precipitation quantiles are from JFMA data above the 98th percentile.

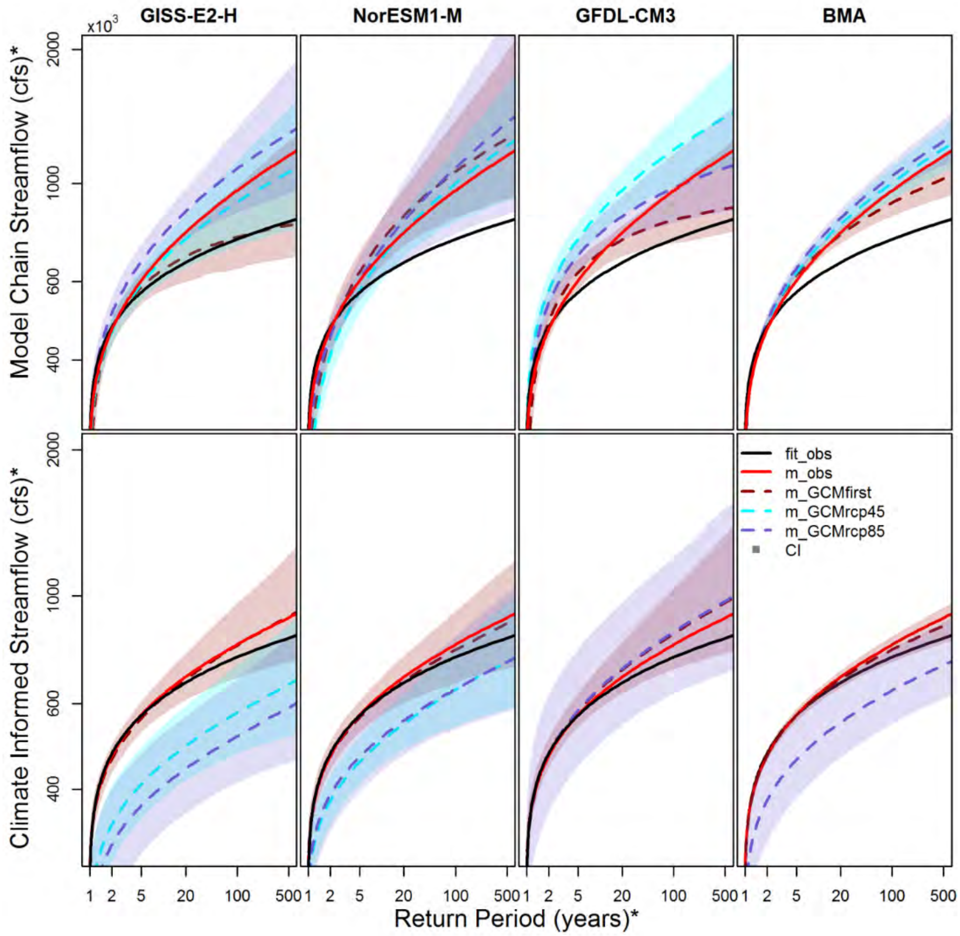


Figure 11-7: Flood projections from both the model chain and climate informed methods for select GCMs (GISS-E2-H, NorESM1-M, and GFDL-CM3) and the Bayesian model average (BMA). "fit_obs" is the LP3 fit to the observed data, "CI" is the confidence or credible intervals (where the colors correspond to the model), and "m_obs", "m_GCMfirst", "m_GCMrcp45", and "m_GCMrcp85" are median of the LP3 fit to the outputs of the model forced with observed, the model forced with the historic time period from GCMs, and the model forced with the future time period from GCMs for RCP 4.5 and RCP 8.5, respectively. The time periods are the same as those in Figure 11-6.

11.4.2. Traditional Risk Analysis and Decision Scaling

Expected total cost calculated using the observed flood event probability is shown in Figure 11-8. The full confidence interval has been partitioned into the confidence interval arising from the 9 possible cost function parameter sets (associated with the LP3 fit to observed) and the 3000 samples of possible LP3 fits (associated with the mean cost function). The medians largely overlap and are relatively flat between the 10 year and 100 year levees, although the 100 year levee does minimize expected total cost and would thus be declared the best design. The partitioned confidence intervals indicate that levee cost primarily drives uncertainty at higher return periods. This likely occurs because the upper tail of the flood distribution contributes little probability mass and the uncertainty in the levee cost function is more pronounced at higher return periods compared to lower return periods. Sampling uncertainty only contributes at lower return periods, likely because the bulk of the flood distribution is at lower return periods. The confidence intervals displayed in all subsequent results show both uncertainties.

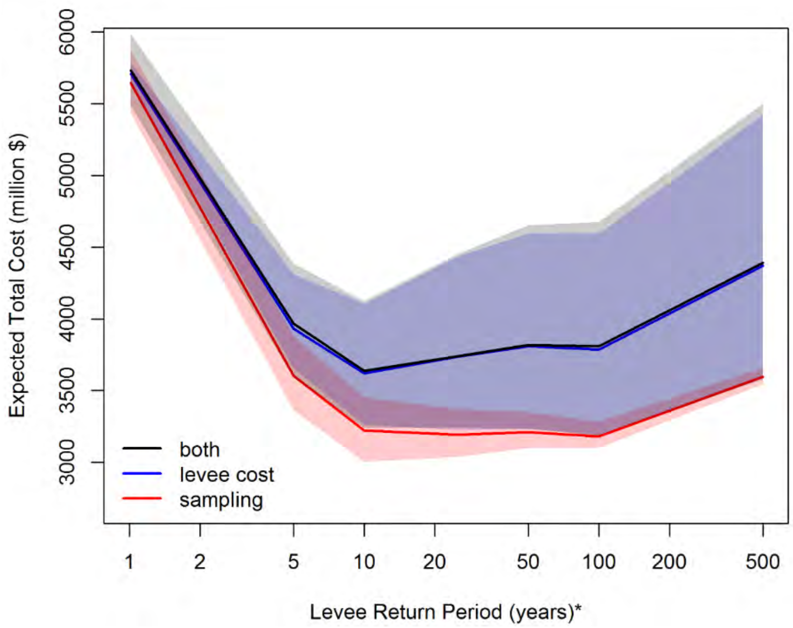


*Figure 11-8: Expected total cost using the observed data, divided into uncertainty arising from levee cost, sampling, and both. The solid lines indicate the median, the shaded areas indicate the range between the 25th and 75th quantiles. * indicates the axis is log-scale. The x-axis return period is based on the historic record.*

The expected total cost from traditional risk analysis using the model chain results is shown in Figure 11-9 for the same GCMs as Figure 11-7 (the Bayesian model average is discussed below). The model forced with observed precipitation and temperature is different from the observed data due to the hydrologic model error discussed previously. As with the flood projections, the expected total cost and the best return period for design varies widely among GCMs. The expected total cost associated with GISS-E2-H aligns with expectations based on the flood projections shown in Figure 11-7. There is a clear increase from the GCM historic to GCM RCP 4.5 and then GCM RCP 8.5 in both the flood projections and the total expected cost. As the flood distribution increases, the expected total cost associated with levees built for lower return periods increases, such that the best design increases from a return period of 100 years to 500 years in GCM RCP 8.5. Additionally, the alignment seen in the flood projections is maintained in the expected total cost; GCM historic aligns with observations and GCM RCP 4.5 aligns with the model forced by observations. However, for NorESM1-M, the small but apparent projected increases in flood magnitude do not translate to increases, but rather decreases, in total expected cost. Furthermore, the alignment between GCM RCP 4.5 and the model forced with observations seen in the flood projections is not maintained in the expected total cost. This can be explained by the very low bias in the lower flood quantiles, particularly observed in GCM RCP 4.5, which translates to a lower expected total cost even though the upper flood quantiles are high. The results associated with GFDL-CM3 exhibit a mix of the characteristics of GISS-E2-H and NorESM1-M.

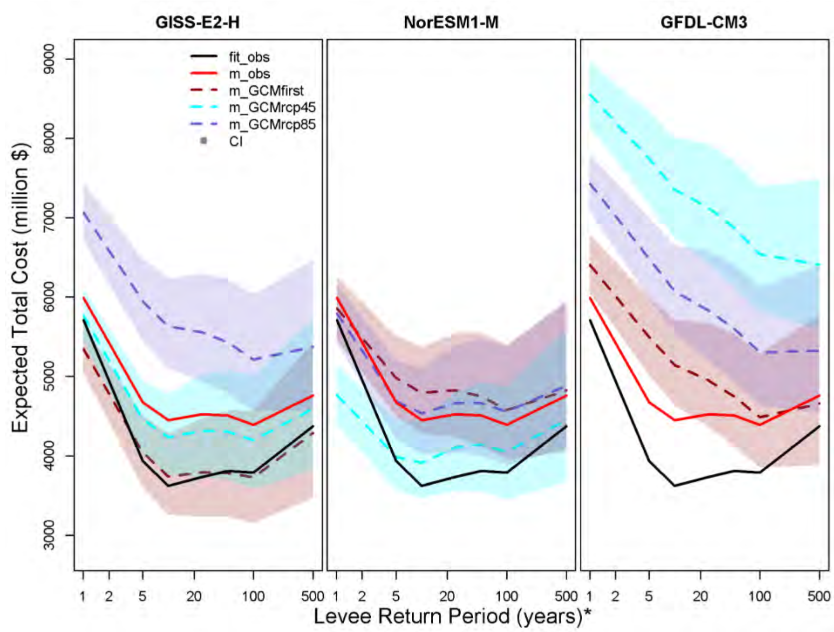


*Figure 11-9: Expected total cost from traditional risk analysis using the model chain. * indicates the axis is log-scale. The x-axis return period is based on the historic record. The meaning of the legend is the same as Figure 11-8.*

The expected total cost results from decision-scaling (Figure 11-10) take a much different form that those from the traditional risk analysis using the model chain. This is primarily due to the vulnerability analysis, which introduces more variables (in particular, four sst_{ETP}^{DJF} scenarios by four $soil_{basin}^{JFMA}$ scenarios by 10 realizations) and only includes GCM projections (which are sst_{ETP}^{DJF} and $soil_{basin}^{JFMA}$, not precipitation and temperature) after the vulnerability analysis is complete. The vulnerability analysis results are shown for the 100 year levee in Figure 11-10. The relationship of expected total cost to the predictors matches the correlations between flood events and the predictors; expected total cost increases with decreasing sst_{ETP}^{DJF} and increasing $soil_{basin}^{JFMA}$. The diagonal angle of the contours, rather than horizontal or vertical alignment, indicates that neither predictor dominates, but both affect expected total cost. The GCM projections fall below the contour line of expected total cost associated with the no-change scenario, indicating expected total cost may decrease in the future. The exception is soil moisture associated with GFDL-CM3 for RCP 8.5, which results in an elongated bivariate normal distribution for RCP 8.5. The standard deviation of expected total cost across the 10 realizations and the 51 samples is relatively small compared to the magnitude of the median (ranging from 50 to 250 million USD) and is positively correlated to the median values (i.e., the contours follow the same pattern) (not shown).

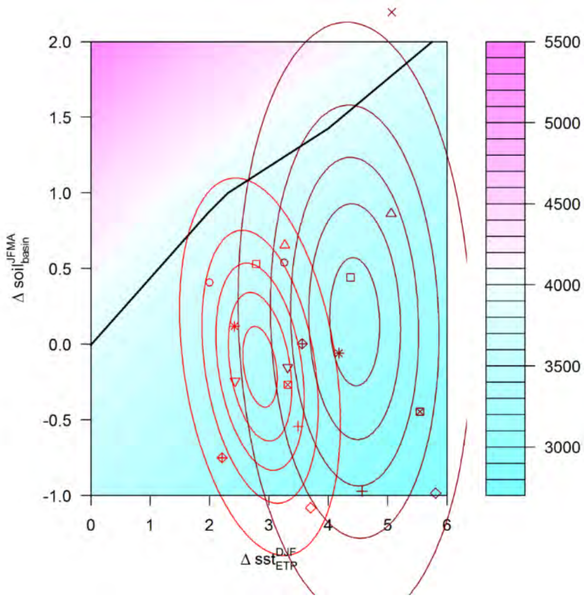


Figure 11-10: The median vulnerability analysis results for the 100 year levee across the 10 realizations and the 51 samples. The color scale indicates the expected total cost in million USD, the x- and y-axis are the change in the indicated predictor, the solid line is the contour of expected total cost associated with the no-change scenario, the points indicate GCM projections of the predictors (symbols have same meaning as in Figure 11-5a), and the two ellipses are bivariate normal distributions fit to the GCM 4.5 and 8.5 projections at the 10%, 30%, 50%, 70% and 90% quantiles.

Figure 11-11 shows a comparison between the decision relevant results of the traditional risk analysis and those of decision scaling for RCP 4.5. The traditional risk analysis results are the Bayesian model average of expected total cost across the 10 GCMs while the decision scaling results are the integral of the distribution of GCM projections with the response surface of expected total cost for each levee size. The results show that the biases in model performance observed in the flood distributions propagate into the decision relevant results. Specifically, the small (large) overestimation of the climate informed model when forced with observations (hydrologic model when forced with GCM historic data) results in a small (large) overestimation of expected total cost. Consequently, the optimal levee design size, defined as the design size which minimizes the median expected total cost, is the 10 year flood when calculated based on the observations and the climate informed model forced with observations, but is the 100 year flood for the traditional risk analysis using GCM historic data.

The direction of projected change in flood distributions also propagates into the decision relevant results. Specifically, the model chain method projection of an increase in the flood distribution due to increases in extreme precipitation causes a corresponding increase in the expected total cost, but not enough to shift the optimal levee size to a higher return period. Conversely, because the GCMs generally project warmer sea surface temperatures but decreasing soil moisture, which causes a decrease in flood events, the expected total cost from decision scaling over the region of likely changes as indicated by the GCMs is lower than the expected total cost of the climate informed model when forced with observations. As a result, the optimal levee design size decreases from the 10 to 5 year flood.

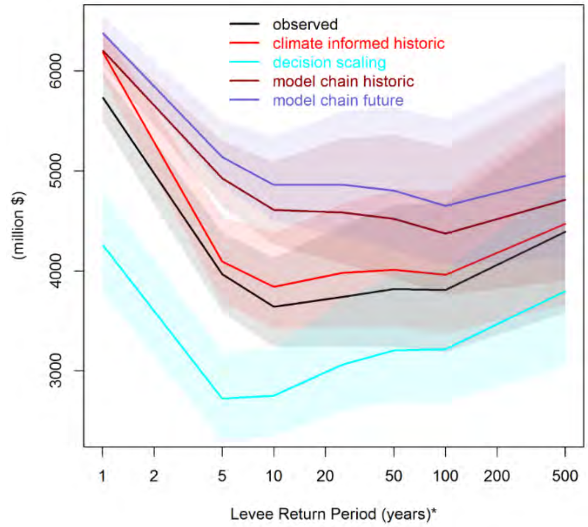


*Figure 11-11: A comparison of the traditional risk analysis and decision scaling results. "Observed" is the expected total cost from the observed data, "climate informed historic" is the expected total cost of the climate informed model forced with observed climate data, "decision scaling" is the result obtained by numerical integration of the response surface of expected total cost for each levee size with the bivariate normal distribution fit to the GCM RCP 4.5 projections and "model chain historic/future" is the Bayesian model average of the model chain results forced with historic and RCP 4.5 future data from the 10 GCMs. The shaded areas indicate the range between the 25th and 75th quantiles, * indicates the axis is log-scale, and the x-axis return period is based on the historic record.*

11.5. CONCLUSION

This study created flood projections using both the traditional model chain method and the climate informed method for Louisville, Kentucky. It subsequently compared the results of risk-based analyses of the design flood for a levee using both a traditional analysis forced by the model chain scenarios and a decision scaling analysis forced by imposed systematic variations in stochastic realizations of the large-scale climate variables. Thus the contributions of this work are two-fold: the integration of climate informed flood projections into decision scaling and a direct comparison of the model chain approach to the climate informed method and to decision scaling.

The analysis showed that the decision relevant results of the traditional risk analysis, in which the flood distribution and total expected costs increase between the historic and future period, are very different from those of decision scaling, which shows a decrease. This difference can be traced to the projected changes in predictors, since the levee cost and damage functions are the same for both methods. Given that the predictors in both methods come from the same GCMs, one possible explanation for the difference is the inability of GCMs to maintain teleconnections between large-scale ocean-atmospheric patterns and localized precipitation and temperature (Lee and Black, 2013; Polade et al., 2013; Sheffield et al., 2013). Another possible explanation is that the climate informed model, which captures the thermodynamic response of sea surface temperatures to global warming, is missing a feedback mechanism, such as the atmospheric response as represented in geopotential heights, which would capture the dynamics of climate change. An illustration of the importance of accounting for both dynamic and thermodynamic impacts of climate change, in a downscaling application, is given in (Greene et al., 2011).

Choosing between the two methods should be based on considerations of both methodology and model credibility. In terms of methodology, as has been convincingly argued elsewhere (Brown et al., 2012;

Spence and Brown, 2016), in comparison to the traditional method which is scenario-led, one strength of decision scaling is its exploration of system response which facilitates evaluation of the robustness of design options. Knowing the system response is valuable information apart from any projection of future changes (e.g., if system performance is satisfactory across all plausible changes in driving forces, then projections and an assessment of their credibility is not necessary). In terms of model credibility, that is, the ability of the model to accurately and precisely represent the important physical processes, Chapter 4 argued that the climate informed approach to flood projection is expected to be more credible than a model chain approach because GCM simulation of large-scale ocean-atmospheric patterns on a seasonal basis is less biased than simulation of daily localized extreme precipitation. Thus, apart from specific case study results, abstract consideration of methodology and model credibility results in a preference for choosing climate informed decision scaling over the model chain.

Specifically for the case study of Louisville, while some bias is observed in the climate informed model predictors as simulated by GCMs over the historic period, large biases are observed in GCM simulation of extreme precipitation over the historic period. The inability of GCMs to reproduce teleconnections for the model chain method and the possibility of a missing feedback mechanism in the climate informed method has already been discussed. Additionally, the hydrologic model was found to be more biased than the climate informed model. When the models are forced by GCM historical climate, performance varied more widely for the model chain method than for the climate informed method. However, it should be noted that the climate informed model explains only a small portion of the variance in flood events, likely because the contributing area for Louisville includes portions of the Ohio River Basin where tropical Pacific sea surface temperatures and geopotential height patterns similar to the Pacific North American pattern are not strong explanatory variables (Figure 4-7). Thus, for the specific case study results, consideration of model credibility still indicates a preference for climate informed decision scaling but is tempered by some caveats regarding the credibility of the climate informed model.

There are several avenues of future research which build off this study. The first is improvement of the climate informed model by including more predictors based on a better understanding of the processes driving flood events in the Louisville catchment. One starting point would be to investigate the influence of snow (see Appendix A) or look for large-scale factors which contribute to precipitation in excess of soil moisture holding capacity (Berghuijs et al., 2016). Another starting point would be a detailed analysis into the GCM processes to identify whether teleconnections are maintained, and if not, where biases are introduced. Such an analysis would help explain the observed projections (e.g., why for some GCMs daily extreme precipitation is projected to increase while seasonal soil moisture is projected to decrease), will yield further insight into model credibility, and may also provide increased insight into the driving mechanisms of floods in the region, which could be used to improve the climate informed model. A second avenue of future research is an exploration of uncertainty. In particular, does the elimination of a hydrologic model reduce the uncertainty in the climate informed approach compared to the model chain approach? Additionally, the levee cost function was found to contribute a large portion of uncertainty due to lack of data; better data would reduce this uncertainty. Furthermore, HAZUS is known to have large uncertainties (Tate et al., 2015) which were not accounted for in this analysis. As in (Schlef et al., 2018), an analysis of variance could be used for uncertainty attribution. Finally, a third avenue of future research is increasing the accessibility of these methods both in terms of the scientific knowledge required to develop the models (e.g., a study like (Berghuijs et al., 2016) which catalogs major large-scale driving forces of floods across the U.S. based on literature review and correlation analysis, as is indicated in Chapter 4) and software platforms that facilitate model development and result visualization (e.g., a web-based application similar to (Whateley et al., 2015) tailored to flood events). Increased accessibility of these methods would allow them to be more widely used by decision makers for flood risk management.

11.6. REFERENCES

Al-Futaisi, A., Stedinger, J., 1999. Hydrologic and Economic Uncertainties and Flood-Risk Project Design. J. Water Resour. Plan. Manag. 125, 314–324.

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

Bracken, C. (2016). Downscaled CMIP3 and CMIP5 Climate Projections - Addendum: Release of Downscaled CMIP5 Clmate Projections (LOCA) and Comparison with Preceding Information.

Brown, C.M., Steinschneider, S., Wi, S., Weatherly, J., Case, M., Hayden, T., Koster, A., Mearns, L.O., Bukovsky, M.S., McCrary, R., 2016. Climate Risk Assessment: Technical Guidance Manual for DoD Installations and Built Environment (Technical Report No. AD1029408). University of Massachusetts Amherst United States.

Brown, C., Ghile, Y., Laverty, M., Li, K., 2012. Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector. Water Resour. Res. 48, W09537. https://doi.org/doi:10.1029/2011WR011212

Burton, C., & Cutter, S. L. (2008). Levee Failures and Social Vulnerability in the Sacramento-San Joaquin Delta Area, California. Natural Hazards Review, 9(3), 136–149. https://doi.org/10.1061/(ASCE)1527-6988(2008)9:3(136)

Fan, Y., & van den Dool, H. (2004). Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present. Journal of Geophysical Research D: Atmospheres, 109(10), 1–8. https://doi.org/10.1029/2003JD004345

Ferguson, A. P., & Ashley, W. S. (2017). Spatiotemporal analysis of residential flood exposure in the Atlanta, Georgia metropolitan area. Natural Hazards, 87(2), 989–1016. https://doi.org/10.1007/s11069-017-2806-6

Greene, A.M., Robertson, A.W., Smyth, P., Triglia, S., 2011. Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model. Q. J. R. Meteorol. Soc. 137, 347–359. https://doi.org/10.1002/qj.788

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. Statistical Science, 14(4), 382–417. https://doi.org/10.2307/2676803

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., … Woollen, J. (1996). The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society, 77(3), 437–471.

Landwehr, J. M., & Slack, J. R. (1992). Hydro-Climatic Data Network: A U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1874-1988. Retrieved from https://pubs.usgs.gov/wri/wri934076/1st_page.html

Lee, Y.Y., Black, R.X., 2013. Boreal winter low-frequency variability in CMIP5 models. J. Geophys. Res. Atmospheres 118, 6891–6904. https://doi.org/10.1002/jgrd.50493

Livneh, B., Rosenberg, E.A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K.M., Maurer, E.P., Lettenmaier, D.P., 2013. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. J. Clim. 26, 9384–9392.

Llanos, M., 2015. Will New Orleans' $14.5 Billion Walls Stand Up to the Next Big Storm? [WWW Document]. NBC News. URL https://www.nbcnews.com/storyline/hurricane-katrina-anniversary/new-orleans-14-5-billion-walls-n415816

López, J., Francés, F., 2013. Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. Hydrol. Earth Syst. Sci. 17, 3189–3203. https://doi.org/10.5194/hess-17-3189-2013

Louisville/Jefferson County Metropolitan Sewer District. (2018a). Flood Protection. Retrieved January 1, 2018, from http://www.louisvillemsd.org/what-we-do/flood-protection

Louisville/Jefferson County Metropolitan Sewer District. (2018b). Upgrade Ohio River Flood Protection. Retrieved January 1, 2018, from http://www.louisvillemsd.org/upgrade-ohio-river-flood-protection

Luke, A., Kaplan, B., Neal, J., Lant, J., Sanders, B., Bates, P., & Alsdorf, D. (2015). Hydraulic modeling of the 2011 New Madrid Floodway activation: a case study on floodway activation controls. Natural Hazards, 77(3), 1863–1887. https://doi.org/10.1007/s11069-015-1680-3

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nastev, M., & Todorov, N. (2013). Hazus: A standardized methodology for flood risk assessment in Canada. Canadian Water Resources Journal, 38(3), 223–231.

National Weather Service. (2018a). Hydrologic Information Center - NWS Annual Flood Loss Summary Reports To U.S. Army Corps of Engineers. Retrieved January 1, 2018, from http://www.nws.noaa.gov/hic/summaries/index.shtml

National Weather Service. (2018b). Top Ten Flood Events. Retrieved January 1, 2018, from https://www.weather.gov/lmk/toptenfloodevents

NCAR. (2018). CGD's Climate Analysis Section: Southern Oscillation Index (SOI). Retrieved March 4, 2018, from http://www.cgd.ucar.edu/cas/catalog/climind/soi.html

NOAA, 2015. Cold & Warm Episodes by Season.

Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., & Hegewisch, K. C. (2015). Improved Bias Correction Techniques for Hydrological Simulations of Climate Change*. Journal of Hydrometeorology, 16(6), 2421–2442. https://doi.org/10.1175/JHM-D-14-0236.1

Pierce, D.W., Cayan, D.R., Thrasher, B.L., 2014. Statistical downscaling using localized constructed analogs (LOCA). J. Hydrometeorol. 15, 2558–2585.

Plummer, M. (2016). rjags: Bayesian Graphical Models using MCMC. R package version 4-6. Retrieved from https://cran.r-project.org/package=rjags
Poff, N.L., Brown, C., Grantham, T.E., Matthews, J.H., Palmer, M.A., Spence, C.M., Wilby, R.L., Haasnoot, M., Mendoza, G.F., Dominique, K.C., Baeza, A., 2015. Sustainable water management under future uncertainty with eco-engineering decision scaling. Nat. Clim. Change. https://doi.org/10.1038/nclimate2765
Polade, S.D., Gershunov, A., Cayan, D.R., Dettinger, M.D., Pierce, D.W., 2013. Natural climate variability and teleconnections to precipitation over the Pacific-North American region in CMIP3 and CMIP5 models. Geophys. Res. Lett. 40, 2296–2301. https://doi.org/10.1002/grl.50491
Qi, W. (2017). A non-stationary cost-benefit analysis approach for extreme flood estimation to explore the nexus of "Risk, Cost and Non-stationarity." Journal of Hydrology, 554, 128–136. https://doi.org/10.1016/j.jhydrol.2017.09.009
Qi, W., & Liu, J. (2018). A non-stationary cost-benefit based bivariate extreme flood estimation approach. Journal of Hydrology, 557, 589–599. https://doi.org/10.1016/j.jhydrol.2017.12.045
Raftery, A., Gneiting, T., Balabdaoi, F., & Polakowski, M. (2005). Using Bayesian Model Averaging to Calibrate Forecast Ensembles. Monthly Weather Review, 1155–1174.
Rayner, N. A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. Journal of Geophysical Research, 108(D14), 4407. https://doi.org/10.1029/2002JD002670
Rehan, B. M., & Hall, J. W. (2016). Uncertainty and sensitivity analysis of flood risk management decisions based on stationary and nonstationary model choices, 20003. https://doi.org/10.1051/e3sconf/20160720003
Remo, J. W. F., Carlson, M., & Pinter, N. (2012). Hydraulic and flood-loss modeling of levee, floodplain, and river management strategies, Middle Mississippi River, USA. Natural Hazards, 61(2), 551–575. https://doi.org/10.1007/s11069-011-9938-x
Remo, J. W. F., Pinter, N., & Mahgoub, M. (2016). Assessing Illinois's flood vulnerability using Hazus-MH. Natural Hazards, 81(1), 265–287. https://doi.org/10.1007/s11069-015-2077-z
Scawthorn, C., Flores, P., Blais, N., Seligson, H., Tate, E., Chang, S., … Lawrence, M. (2006a). HAZUS-MH Flood Loss Estimation Methodology. I: Overview and Flood Hazard Characterization. Natural Hazards Review, 7(2), 72–81. https://doi.org/10.1061/(ASCE)1527-6988(2006)7:2(72)
Scawthorn, C., Flores, P., Blais, N., Seligson, H., Tate, E., Chang, S., … Lawrence, M. (2006b). HAZUS-MH Flood Loss Estimation Methodology. II. Damage and Loss Assessment. Natural Hazards Review, 7(2), 72–81. https://doi.org/10.1061/(ASCE)1527-6988(2006)7:2(72)
Schlef, K.E., Steinschneider, S., Brown, C.M., 2018. Spatiotemporal Impacts of Climate and Demand on Water Supply in the Apalachicola-Chattahoochee-Flint Basin. J. Water Resour. Plan. Manag. 144, 1–12. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000865.
Sheffield, J., Barrett, A.P., Colle, B., Fernando, D.N., Fu, R., Geil, K.L., Hu, Q., Kinter, J., Kumar, S., Langenbrunner, B., Lombardo, K., Long, L.N., Maloney, E., Mariotti, A., Meyerson, J.E., Mo, K.C., Neelin, J.D., Nigam, S., Pan, Z., Ren, T., Ruiz-Barradas, A., Serra, Y.L., Seth, A., Thibeault, J.M., Stroeve, J.C., Yang, Z., Yin, L., 2013. North American Climate in CMIP5 experiments. Part I: Evaluation of historical simulations of continental and regional climatology. J. Clim. 26, 9209–9245. https://doi.org/10.1175/JCLI-D-12-00592.1
Spiegelhalter, D. J., Best, N., Carlin, B. P., & Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. Journal of the Royal Statistical Society. Series B, 64(4), 583–639.
State of Pennsylvania. (2013). Pennsylvania 2013 Standard State All-Hazard Mitigation Plan.
Steinschneider, S., McCrary, R., Wi, S., Mulligan, K., Mearns, L.O., Brown, C., 2015. Expanded Decision-Scaling Framework to Select Robust Long-Term Water-System Plans under Hydroclimatic Uncertainties. J. Water Resour. Plan. Manag. 141. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000536.
Sun, X., Thyer, M., Renard, B., & Lang, M. (2014). A general regional frequency analysis framework for quantifying local-scale climate effects: A case study of ENSO effects on Southeast Queensland rainfall. Journal of Hydrology, 512, 53–68. https://doi.org/10.1016/j.jhydrol.2014.02.025
Tate, E., Strong, A., Kraus, T., & Xiong, H. (2016). Flood recovery and property acquisition in Cedar Rapids, Iowa. Natural Hazards, 80(3), 2055–2079. https://doi.org/10.1007/s11069-015-2060-8
Trenberth, K.E., 1997. The Definition of El Nino. Bull. Am. Meteorol. Soc. 78, 2771–2777. https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2
USACE, 2018. National Levee Database [WWW Document]. URL http://nld.usace.army.mil/egis/f?p=471:1:0::NO
US Census Bureau. (2018). QuickFacts Louisville/Jefferson County (balance), Kentucky. Retrieved January 1, 2018, from https://www.census.gov/quickfacts/fact/table/louisvillejeffersoncountybalancekentucky/PST045216
van Vuuren, D.P., Edmonds, J.A., Kainuma, M., Riahi, K., Weyant, J., 2011. A special issue on the RCPs. Clim. Change 109, 1–4. https://doi.org/10.1007/s10584-011-0157-y
Whateley, S., Walker, J.D., Brown, C., 2015. A web-based screening model for climate risk to water supply systems in the northeastern United States. Environ. Model. Softw. 73, 64–75. https://doi.org/10.1016/j.envsoft.2015.08.001
Yu-Sung, S., & Yajima, M. (2015). R2jags: Using R to Run "JAGS". R package version 0.5-7. Retrieved from https://cran.r-project.org/packages=R2jags

12. ROBUST ADAPTATION TO MULTI-SCALE CLIMATE VARIABILITY

A modified version of this chapter is published in Earth's Future*:*

Doss-Gollin, J., Farnham, D.J., Steinschneider, S., Lall, U., 2019. Robust Adaptation to Multi-Scale Climate Variability. Earths Future. [*https://doi.org/10.1029/2019EF001154*](https://doi.org/10.1029/2019EF001154)

Supplementary material is available at:
https://agupubs.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1029%2F2019EF001154&file=eft2568-sup-0001-2019EF001154-S1.pdf

12.1. EXECUTIVE SUMMARY

The assessment and implementation of structural or financial instruments for climate risk mitigation requires projections of future climate risk over the operational life of each proposed instrument. A point often neglected in the climate adaptation literature is that the physical sources of predictability differ between projects with long and short planning periods: while historical and paleo climate records emphasize modes of variability at interannual to multidecadal time scales, anthropogenic climate change is expected to alter their occurrence at longer time scales. We present a set of stylized experiments to assess the uncertainties and biases involved in estimating future climate risk over a finite future period, given a limited observational record. These experiments consider both quasi-periodic and secular change for the underlying risk, as well as statistical models for estimating this risk from an N-year historical record. The uncertainty of IPCC-like future scenarios is considered through an equivalent sample size N. The relative importance of estimating the short- or long-term risk extremes depends on the investment life M. Shorter design lives are preferred for situations where inter-annual to decadal variability can be successfully identified and predicted, suggesting the importance of sequential investment strategies for adaptation.

12.2. TECHNICAL APPROACH

We explore how the temporal structure of climate risk and how the uncertainty associated with its estimation influences the answers to questions about whether to invest in permanent or flexible solutions. To do this we begin with three specific observations about climate risk which. We then carry out a set of stylized experiments to probe their implications.

12.2.1. OBSERVATIONS

12.2.1.1. PLANNING DECISIONS ARE MADE WITH FINITE HORIZONS

Public or private sector investments in climate adaptation require not only the design of each potential structural instrument, but also selecting between instruments with vastly different operational planning periods. This project planning period, which we define as being M years, describes the nominal economic or physical lifespan of the structure or contract. Typical planning periods may vary from M=1 year or less for a financial contract to M=100 years or longer for a structural instrument, as illustrated in Table 12-1. The planning period can also be interpreted as the finite period over which (CBA) is conducted when assessing the project.

Table 12-1 Six real-world risk mitigation instruments and the associated project planning period (M).

Location	Description	M	Reference
Iowa River	Purchase options for inundation of downstream agricultural lands to allow higher release flows from the flood control reservoir	1	Spence and Brown (2016)
New York City	Catastrophe bond for protection against storm surge caused by named storms and earthquakes	3	
County of Santa Barbara, California	Emergency improvements to portions of the Santa Maria Levee to reduce risk of levee failure	5	USACE (2007)
Iowa River	Raise levees by six feet	30	Spence and Brown (2016)
Dallas, TX	Evacuation of Rockefeller Boulevard	50	USACE (2014)
Central California	Tulare Lake storage and floodwater protection project	100	GEI Consultants, Inc. (2017)

Typical climate risk management policies do not use a single risk mitigation instrument, but rather build a portfolio of several instruments. Each has its own operational period, which may or may not match the planning horizon of the portfolio as a whole. This means that even if the portfolio has a long planning period, i.e. if long-term plans are a priority, this goal may be best accomplished through a series of flexible and adaptive instruments with short individual planning periods. For example, the optimal policy for New York City to manage uncertain hurricane risk in the 21st century might potentially be to keep areas devastated by hurricane Sandy zoned for low-impact development for the next 10 years. This would reduce future risk over all climate scenarios while postponing major investments until large uncertainties as to the magnitude of future sea level rise are resolved. The costs and benefits of each individual instrument will be assessed over its individual, finite planning period, but decisions about the portfolio structure are evaluated over the longer planning horizon.

The availability of precise climate information in the near future may significantly alter the choice between a large, long-duration instrument and a sequence of smaller, short duration instruments that can be executed quickly. For example, if above-average climate risk is projected over the next few years, a more costly project might be justified. However, in the plausible case of a long construction period for the large, permanent instrument, a financial risk mitigation instrument might be needed in the immediate term to cover potential losses before the large project is completed. Conversely, if the near-term risk is projected to be low, then deferral of the large, potentially expensive instrument may be warranted. These cases highlight how the precision of short- and long-term climate risk projections plays directly into climate adaptation.

12.2.1.2. CLIMATE VARIES ON MANY SCALES

Climate risk is governed by a variety of physical processes which occur on scales ranging from local and transient to global and permanent. Of these processes, anthropogenic climate change (ACC) has received the most attention in the climate adaptation literature and its influence on some river floods, droughts, hurricanes, urban flooding, and many other climate hazards has been the subject of substantial investigation (e.g., Coumou and Rahmstorf 2012; Milly et al. 2008; O'Gorman and Schneider 2009; Trenberth et al. 2003). Human activities can also affect climate risk through modification of local land or river systems (see Merz et al. 2014), and through changes in exposure to extremes (Di Baldassarre et al. 2018; Jongman, Ward, and Aerts 2012). In combination, these effects highlight that the past may not be an adequate representation of future climate risk (termed "nonstationarity" by Milly et al. 2008).

Secular change is not the only mechanism which can cause historical records to provide a biased view of future risk. The Hurst phenomenon is a well known mathematical relationship which describes the long memory of processes found in in geophysics, physics, biology, medicine, traffic, network dynamics, and finance (O'Connell et al. 2016). The extensive observations of such behavior in hydrologic and climatic time series emphasize the need to consider such processes as underlying any discussion of climate change or nonstationarity (Koutsoyiannis 2003; Markonis and Koutsoyiannis 2013; Palmer 1993). The Hurst phenomenon has also been connected to low frequency quasi-periodic phenomenon, especially where fractal scaling is expected. For example, wavelet methods have been used to estimate the Hurst exponent (Simonsen, Hansen, and Nes 1998; Chamoli, Ram Bansal, and Dimri 2007), and to design simulation algorithms that reproduce self-similarity, long range dependence and quasi-periodic regimes (Kwon, Lall, and Khalil 2007; Bullmore et al. 2001; Geweke and Porter-Hudak 1983; Feng, Willemain, and Shang 2005). The Hurst phenomenon also provides a link between catchment hydrology and global climate dynamics (Blöschl and Montanari 2010; Montanari 2003). The Hurst exponent is directly related to the fractal dimension of a process, and there is a rich multi-disciplinary literature as to the process level and statistical justification of long memory and fractal processes in hydrology (Mandelbrot 1985; Mandelbrot and Wallis 1969; Beran 1994). These processes have also been used to describe multi-scale dynamics of the climate (Lovejoy and Schertzer 2012; Shaun Lovejoy and Schertzer 2013; S. Lovejoy 2013; Selvam 2017), including enso (Maruyama 2018; Živković and Rypdal 2013) and the pdo (Mantua et al. 1997).

External forcing from structured climate signals ("teleconnections"; Ångström 1935) and catchment dynamics are both useful in explaining the Low-Frequency Variability (LFV) observed in natural hydroclimate time series. We illustrate such LFV in Figure 12-1, which shows a 500 year drought reconstruction from the lbda (Cook et al. 2010), a 100 year record of annual maximum streamflowflow on the American River at Folsom, and the global wavelet power spectrum for both (Torrence and Compo 1998; Roesch and Schmidbauer 2016). Peaks for the American River time series are apparent at 2.3 and 15 years and in the lbda time series at approximately 8, 20, and 64 years. This is illustrated by the blue line in Figure 12-1(b), which shows a 20 year moving average of the lbda time series. A detailed analysis of these time series is beyond the scope of this paper, but we note that the high amplitude and long time periods of the quasi-periodic oscillations they exhibit are consistent with analyses of LFV in other hydroclimate systems (Kiem, Franks, and Kuczera 2002; Swierczynski et al. 2012; Woollings et al. 2014; Hodgkins et al. 2017). The key implication is that the observations, (Jain and Lall 2001), trends (Bhattacharya, Gupta, and Waymire 1983), and frequencies (Newman et al. 2016) observed in the past are often poor predictors of future behavior.

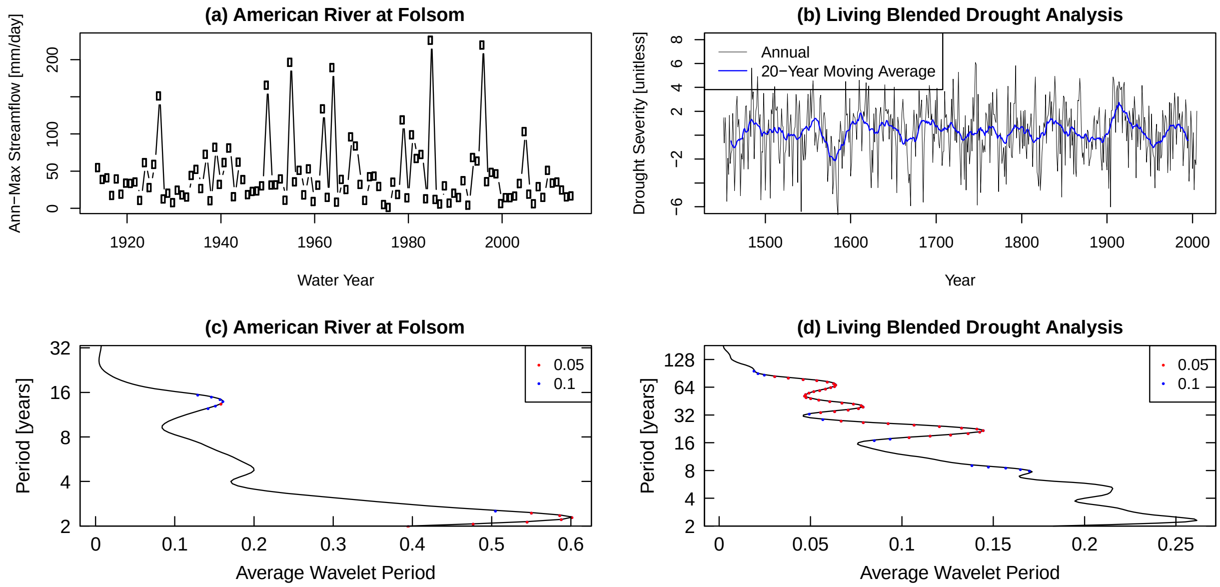


Figure 12-1 Hydroclimate time series vary on many time scales. (a) A 500 year reconstruction of summer rainfall over Arizona from the lbda. Lower values indicate more severe drought. A 20-year running mean is also shown in blue. (b) A 100 year record of annual-maximum streamflow for the American River at Folsom. Daily streamflow values were divided by the catchment area to yield a normalized flow in units of . (c) The global wavelet power spectrum of the lbda time series (a). Blue (red) dots indicate frequencies which are significant at $\alpha = 0.10(0.05)$ *compared to white noise. (d) Global wavelet power spectrum, like (c), for the American River data.*

12.2.1.3. THE DOMINANT PROCESSES DEPEND ON THE PLANNING PERIOD

Evaluating a particular risk mitigation instrument involves projecting climate risk over the M-year planning period. Consequently, the physical mechanisms which impart predictability on the system differ between projects with long and short planning periods. As illustrated in Figure 12-2 (a), the lifetime risk of a permanent structure with a 100-year planning period depends on the magnitude and extent of future human activities, with very large associated uncertainty. Even in the idealized and unrealistic case of a perfect climate model, these uncertainties will be large. By contrast, this perfect climate model may usefully inform estimates of climate hazard over a three-year insurance contract with much less associated uncertainty.

Of course, scientists are not equipped with perfect models. Since different physical processes control climate risk at different timescales, successful integration of climate projections into decision frameworks depends on identifying, and subsequently predicting, these processes. A key question is whether the limited information in an N-year observational record permits the identification and projection of cyclical climate variability and secular change, and what the resulting bias and uncertainty portend for risk mitigation instruments with a planning period ranging from a few years to several decades. As shown in Figure 12-2 (b), the combination of LFV, stochastic variability, and secular change in a limited record can lead to large uncertainty in estimated future risk. Although Figure 12-2 focuses on physical processes, similar conclusions would also be valid for the socioeconomic processes which drive exposure to floods and other hydroclimate hazards.

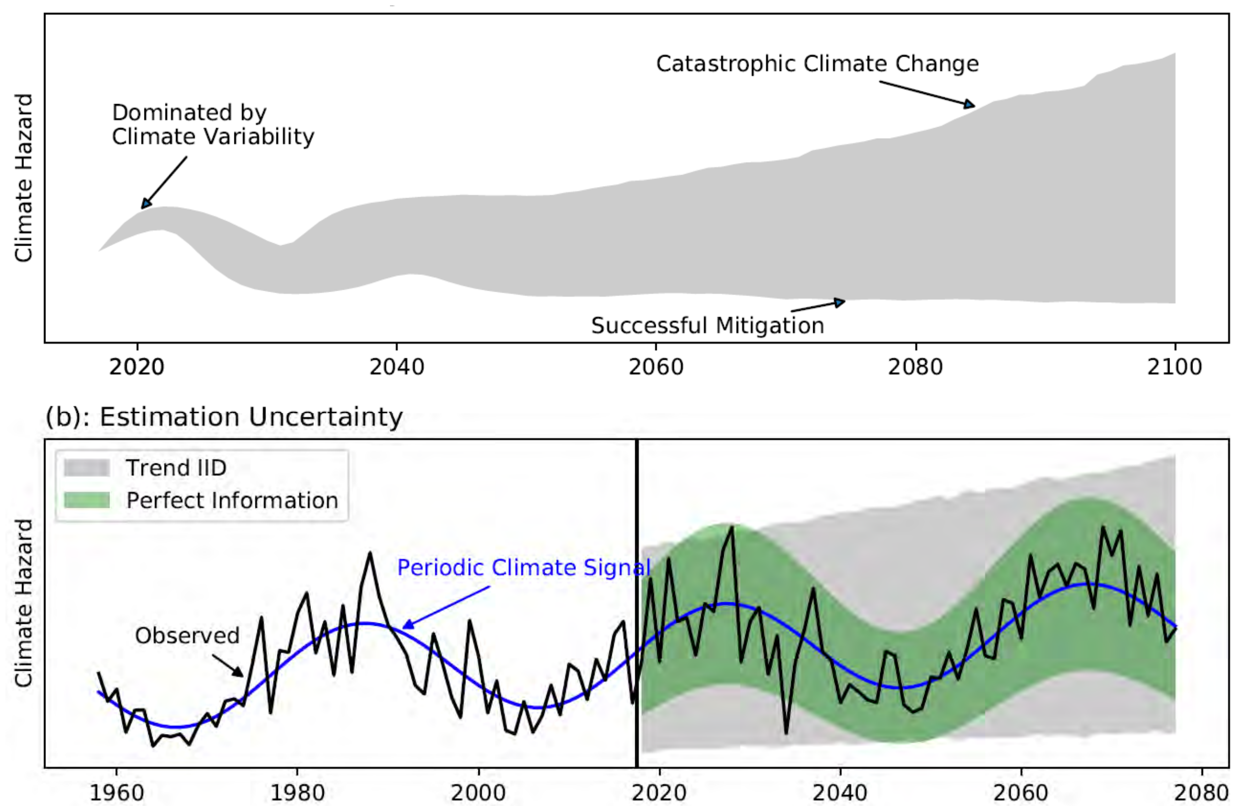


Figure 12-2 A stylized illustration of (a) irreducible and (b) estimation uncertainty. (a): Irreducible uncertainty cannot be resolved with better models or data and is dominated in the short term by chaotic behavior of the climate, and in the long term by the uncertainty in future acc. (b): Informational uncertainty limits the potential to identify different climate signals. The blue line shows an idealized climate signal and the black line shows observations, which are scattered stochastically around the signal line. The green shading shows the true range within which observations will occur 95% of the time, while the gray shading the 95% confidence interval as estimated with a linear trend model.

12.3. DATA AND METHODS

We consider a set of stylized experiments to assess how well one can identify and predict risk associated with cyclical and secular climate signals for the M-year planning period and the probability of over- or under-design of a climate adaptation strategy based on these projections. We consider different temporal structures for the underlying risk which encompass quasi-periodic, regime-like, and secular change, as well as simple statistical models for estimating this risk from an N-year historical record. The relative importance of estimating the short- or long-term risk associated with these extremes depends on the design life M, but the potential to understand and predict these different types of variability depends on the informational uncertainty in the N-year historical record. Though we illustrate our findings with a simple flood risk example, the conclusions drawn apply to other hydroclimate hazards, and in particular those typically characterized through a time series of annual maxima or minima.

We consider three scenarios for climate risk, which we define by the structure of the underlying climate signal: secular change only; LFV only; and LFV plus secular change. For each scenario, and for its identification from the N year length historical data, the bias and variance of the estimated flood risk over the M year design life relative to the "true model" are computed. We repeat the simulations J=1000 times for each combination of experiment parameters to obtain estimates of the expected bias and variance for each scenario given M and N.

We caution the reader that the models for sampling climate risk and for statistically projecting future risk were chosen for their intuitive interpretation, rather than their general validity (see Held 2005 for a thoughtful discussion of the value of simple models). We do not, in general, endorse these models for practical use but instead argue that the conclusions drawn from these simple models may be straightforwardly applied to more complex and realistic models.

12.3.1. SAMPLING CLIMATE RISK

The first step is to sample climate risk by generating synthetic streamflow sequences. To do this, we model annual-maximum flood peaks with a log-normal distribution, conditional on a location parameter which varies in time:

$$log\, Q\,(t) \sim \mathcal{N}\big(\mu(t), \sigma(t)\big). \qquad 12.1$$

We further assume a constant coefficient of variation of the log streamflow,

$$\sigma(t) = \xi\mu(t) \qquad 12.2$$

and apply a lower threshold on the standard deviation

$$\sigma(t) \geq \sigma_{\min} > 0. \qquad 12.3$$

This formulation describes all scenarios for future climate considered in this paper within a single equation. To add climate variability to the system, the only component which needs to change is the dependence of μ(t) on time, which we parameterize as

$$\sigma(t) \geq \sigma_{\min} > 0. \qquad 12.4$$

where $x(t)$ represents a climate time series which itself exhibits LFV but not secular change. This parameterization is analagous to the “climate-informed” approach described in several studies for estimating climate risk (Delgado, Merz, and Apel 2014; Merz et al. 2014; Farnham, Doss-Gollin, and Lall 2018). Following this model when $\beta \neq 0$ there will be LFV, and when $\gamma \neq 0$ there will be secular change. The values of all parameters used for sampling climate risk are listed in the online supporting information for each of the three scenarios considered.

We represent the climate state variable $x(t)$ through an index for enso, which has been shown to impact flood risk around the world (Ropelewski and Halpert 1987; Ward et al. 2014) and has characteristic variability on timescales of 3 to 7 years (Sarachik and Cane 2009) as well as a “staircase” of lower-frequency scales (Jin, Neelin, and Ghil 1994). We model enso variability by taking a 20000 year integration of the Cane-Zebiak model (Zebiak and Cane 1987) to produce a monthly NINO3 index (Ramesh et al. 2016). To create an annual time series, we average the October-December values of the NINO3 index for each year. Supplemental figure S1 shows a wavelet spectrum and time series plot of the resulting annual time series. In the online supplemental information we consider an alternative parameterization of $\mu(t)$, which considers a Markovian state transition rather than an explicit enso model, and note a general agreement of results.

12.3.2. PROJECTING CLIMATE RISK OVER THE FUTURE M YEARS

Once a synthetic streamflow sequence has been generated, we evaluate the identifiability and predictability of the dominant climate modes by fitting the sequence to statistical models and creating probabilistic projections of the future. We use three well-studied statistical methods for future flood risk, each of which parameterizes time in a different way. One is purely stationary, another captures LFV, and the third captures secular change. We choose these models for their interpretability and simplicity, rather than because of a belief that they are generally valid. For each synthetic flood sequence to be analyzed, the first N years are treated as observations. Once a statistical model is fit to these observations, then K=1000 sequences of future annual-maximum streamflow over the future M-year record are generated from the fitted model using Monte Carlo simulation.

In the first case we fit a stationary model to the observed flood record, following classical assumptions of iid sequences. In this model annual-maximum streamflow are taken to follow a log-normal distribution with constant mean and variance. We refer to this model as “LN2 Stationary.” The parameters of the model are fit in a Bayesian framework to fully represent the posterior uncertainty, using the stan probabilistic computing package (Carpenter et al. 2017) with weakly informative priors (Gelman, Simpson, and Betancourt 2017; Simpson et al. 2017).

Next, we modify this stationary model to incorporate secular change. Many studies have done this by regressing certain parameters of the model on time (see Salas, Obeysekera, and Vogel 2018 for a comprehensive review). We consider an extension of the stationary log-normal model by adding a time trend on the scale parameter and maintaining a constant coefficient of variation. We refer to this model as “LN2 Linear Trend.” This model gives a lower bound on total informational uncertainty because it correctly represents the trend’s known form, whereas in real-world analyses the form of the trend is unknown.

Finally, we explicitly model LFV using a Hidden Markov Model (HMM). An HMM is a latent variable model in which the system being modeled is assumed to follow a Markov process with unobserved (i.e. hidden) states $S(t)$ (Rabiner and Juang 1986). The (unobserved) states evolve following a first-order Markov process, and the observed variable (e.g. streamflow) depends only on the underlying state. HMMs have been widely used for modeling streamflow sequences (Bracken, Rajagopalan, and Woodhouse 2016) and enso (Rojo Hernandez, Lall, and Mesa 2017). We fit streamflow sequences using a HMM with two states. The model is fit using the Baum-Welch algorithm, assuming that the data follow a log-normal distribution that is conditional only on the unobserved state variables. This algorithm simultaneously estimates the transition matrix of the Markov process and the conditional parameters of each distribution. For simplicity, we fit only a two-state HMM to each sequence. Future floods are then estimated by simulating future states from the estimated transition matrix and then drawing $Q(t)$ conditional on the simulated state.

12.3.3. EVALUATING FITTING MODELS

Both estimation bias and estimation uncertainty affect the utility of a climate risk projection. An instrument whose design was based on projections with overestimated variance or positive bias will be over-designed, either causing the risk manager to avoid the investment, given its higher cost, or will lead to unnecessary diversion of funds from other instruments. Similarly, an instrument designed based on underestimated variance or negative bias may be under-designed, and thus fail to protect the public.

We evaluate both the estimation bias and estimation uncertainty. For a given choice of M, N, and generating model, we compare the synthetic streamflow sequence’s N-year “historical record” and the $K = 1000$ posterior simulations of future flows. The quantity $\hat{p}_T$, the estimated expected number of floods per year, is taken by calculating, for each of the K posterior simulations, the number of exceedances of the flood design threshold, then dividing by M to get exceedances per year. We then compute the variance of these K estimates. We further calculate the bias of $\hat{p}_T$ by averaging it across the K samples and comparing this to the number of times the M-year "future period" of the synthetic streamflow sequence exceeds the flood design threshold. Since the "observed" number of flood exceedances from the generating model is inherently noisy for an M-year period, we average the bias and variance across $J = 1000$ different streamflow sequences to compute expected values of both. These sequences are generated with the same underlying parameters, but the specific synthetic NINO3 sequence (or set of Markov states) may differ between the J sequences.

12.3.4. Experiment Deign

Figure 12-3 describes the experimental design. We assess estimation bias and variance for three scenarios of future climate. First, we consider an idealized scenario where only secular change is present in the system and LFV is fully damped ("secular change only"). Next, we consider the "pre-industrial" case where there is no secular change but LFV modulates climate risk in time ("LFV only"). Finally, we consider a more realistic (though still idealized) case with both LFV and secular change ("LFV plus secular change"). Model parameters for each scenario are given in the supplemental methods section of the online supporting materials.

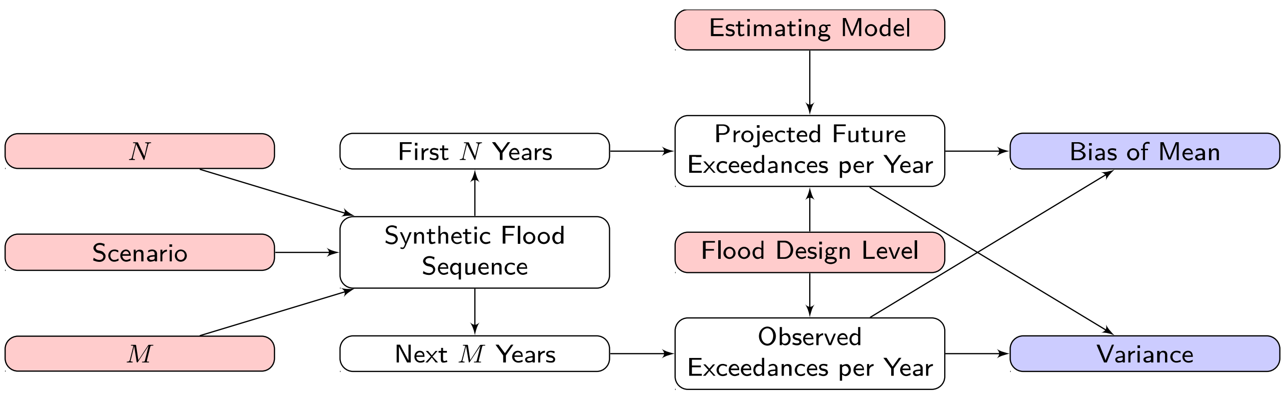


Figure 12-3 Flow chart describing experiment design. Parameters are shown in red. N denotes the informational uncertainty (length of historical record) and M the amount of extrapolation (project design life). Calculated quantities are shown in white. Quantities used for analysis are shown in blue.

Computation was carried out in the python programming language, making particular use of the matplotlib, numpy, pandas, pomegranate, scipy, and xarray libraries for scientific computing (Hunter 2007; Walt, Colbert, and Varoquaux 2011; McKinney 2010; Schreiber 2017; Jones, Oliphant, and Peterson 2001; Hoyer and Hamman 2017). Wavelet analysis was conducted using the WaveletComp package (Roesch and Schmidbauer 2016) in the R programming language. Bayesian models were written in the stan probabilistic programming language (Carpenter et al. 2017) using the No U-Turn Sampler (Hoffman and Gelman 2011; Betancourt 2017). The codes used to generate the figures and text of this paper are available at https://github.com/jdossgollin/2018-robust-adaptation-cyclical-risk.

12.4. RESULTS AND DISCUSSION

These three scenarios for future climate considered are illustrated belowfig:example-fit, which shows a single synthetic streamflow sequence generated with $N = 50$ and $M = 100$. We also show projected future climate risk with each of the three estimating models. This figure highlights that even where projections of average streamflow are unbiased, if the spread is too large then projection of the threshold exceedance probability may be too large. In the remainder of this section we present a more systematic analysis of each of these three cases.

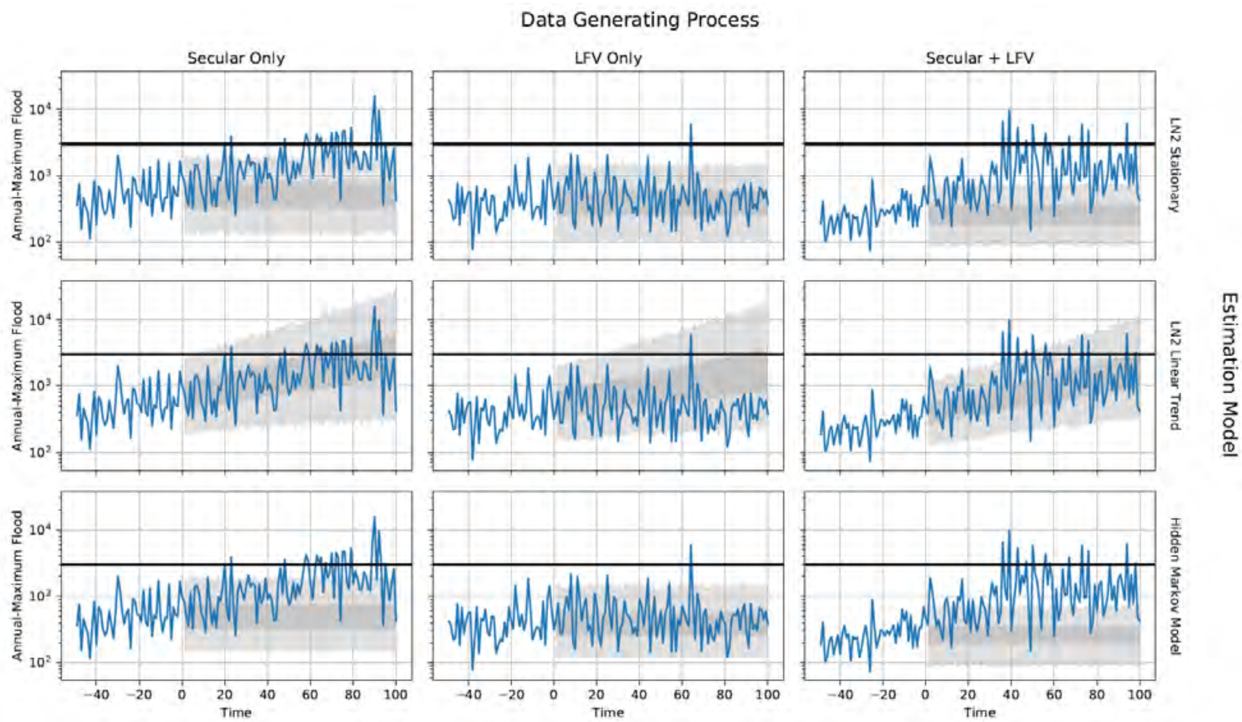


Figure 12-4 An illustration of the estimation procedure. A single streamflow sequence with $N = 50$ and $M = 100$ is shown for each of the three cases (secular only, LFV only, and secular plus LFV) considered. The blue line shows the observed sequence. The gray shading indicates the 50% and 95% confidence intervals using each of the three fitting methods discussed (rows). The horizontal black line indicates the flood threshold.

12.4.1. Secular Change only

In the idealized case where only secular change exists, accurate climate predictions need to either use a long record to identify and model this trend, or to ignore the trend and predict only a few years ahead. This is shown in Figure 12-5, which depicts the estimation bias and variance for each of the three estimation models for many combinations of M and N.

The log-normal trend model tends to over-estimate risk (positive bias), except when N is large, because the model gives a non-zero probability to the trend being larger than it actually is. The variance of these estimates is also large. This again highlights the difficulty of fitting complex models for estimating risk when informational uncertainty is large. By contrast, the stationary log-normal model and HMM, which do not account for secular change, show relatively low variance of their estimates and exhibit low bias for short M. As $N \to \infty$, these (mis-specified) models can only represent the trend by setting the scale parameter very large, leading to high estimation variance and (as $M \to \infty$) also a large bias. This principle has prompted some to consider only the most recent years of the data, deliberately shortening N (i.e., Müller et al. 2014). However, these results also highlight that the increase in variance as N is reduced may quickly outpace the utility of any bias reductions.

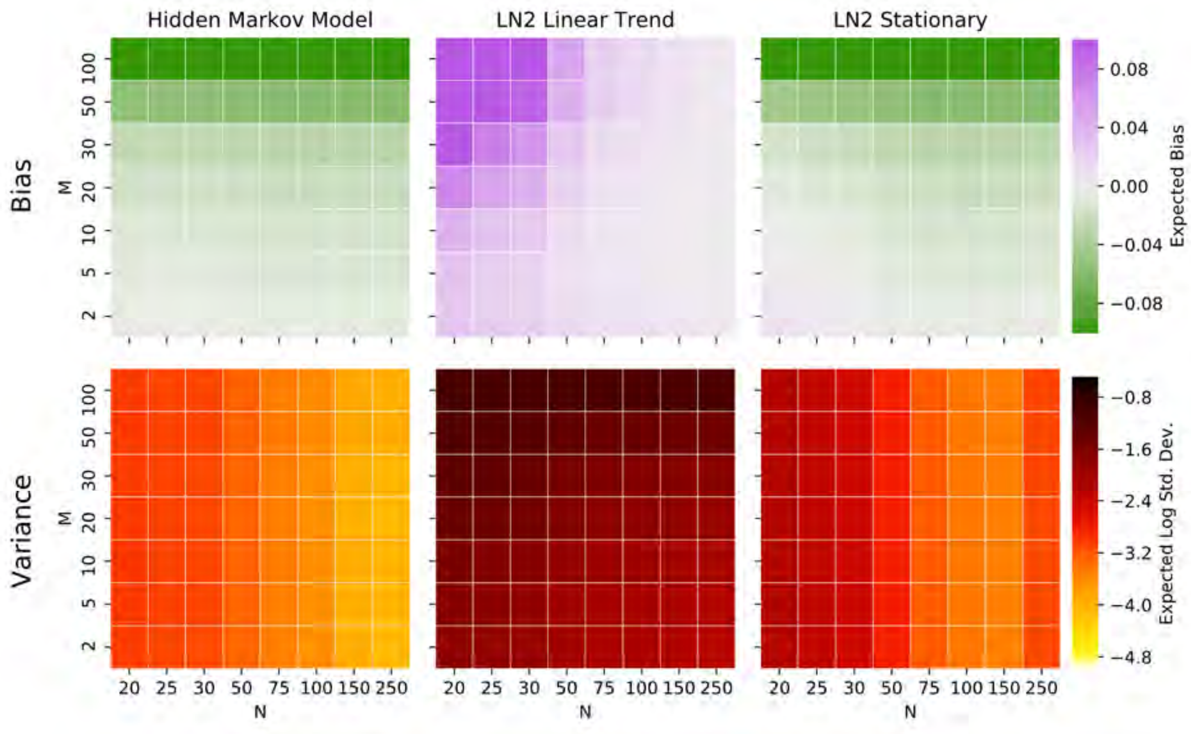


Figure 12-5 Expected estimation bias and variance for sequences generated with secular change only (no LFV). Sequences were fit to each of three statistical models (columns) for different N and M (x and y axis, respectively). Top row shows estimation bias and bottom row shows log standard deviation of estimates. Note the uneven spacing of the x and y axes.

If the analyst could know *a priori* that secular change is present in a time series, and if M is long, then the use of a complex model which represents the processes causing this change is required. Here the log-normal linear trend model has the advantage of being correctly specified (both the generating and fitting processes assume a log-normal distribution conditional on a linear time trend), which is generally not the case in the real world (Montanari and Koutsoyiannis 2014; Serinaldi and Kilsby 2015). As a result, in real-world settings longer N may be required to identify trends whose exact form is not known. Alternatively, if M is small then it may be reasonable to use a stationary estimate, since the bias will be small and the variance substantially lower.

12.4.2. Low-Frequency Variability Only

We next turn to the idealized case where LFV is present but there is no secular change in the system. Figure 12-6 highlights that identification of nonexistent trends from limited data may lead to gross over-estimation of true risk through an increase in the variance of the estimated risk. As expected, the stationary log-normal model performs well overall, with low bias and low variance. The HMM actually out-performs the stationary model, with slightly lower variance than the stationary model, because it better captures the multimodal distribution that emerges from dependence on the enso index, which exhibits several regimes (see supplemental figure S1). By contrast, the linear trend model performs poorly for low N and high M because a positive probability is assigned to the existence of a positive trend.

Of particular relevance to analysis of real-world data sets is the ratio of the project planning period M to the characteristic periods of variability of the LFV. If this period is much larger than M, then a stationary assumption may provide reasonable estimates, and fewer observations may be required (shorter N). As shown in supplemental figure S1, the enso time series is most active in the band. In the real world, however, many hydroclimate time series vary at multidecadal and longer frequencies. In this case, as illustrated in 2, the characteristic periods may be as large or larger than M, particularly if multidecadal modes such as the pdo or amo are involved, and the LFV must therefore be estimated explicitly. This in turn requires a longer observational record N in order to identify and predict these different signals.

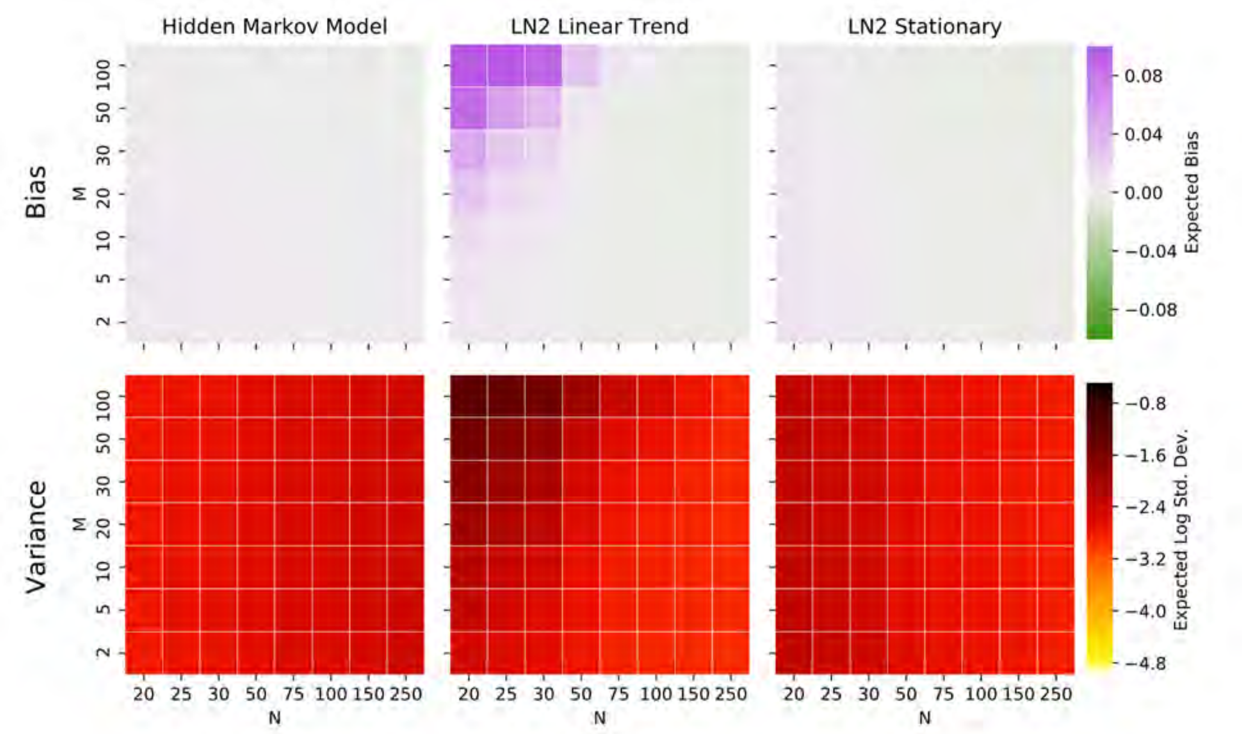


Figure 12-6 As previous figure but for sequences generated with zero secular change and strong LFV.

12.4.3. Low-Frequency Variability and Secular Change

In the final and most realistic case, where both LFV and secular change are present, stationary models perform well for short M while for long M the trend must be identified from a long record and modeled explicitly.

Consistent with the conceptual illustration of Figure 12-2, the results of Figure 12-7 highlight that the relative importance of secular change and LFV depends on M. When M is long, climate risk is dominated by secular change and it becomes essential to model this risk explicitly with a more complex model (i.e. , the linear trend model). Alternatively, when M is short, LFV dominates and the increased variance associated with estimating a trend is not worth the modest reduction in bias. As before, when the informational uncertainty is large (small N), the identifiability and predictability of the trend are limited.

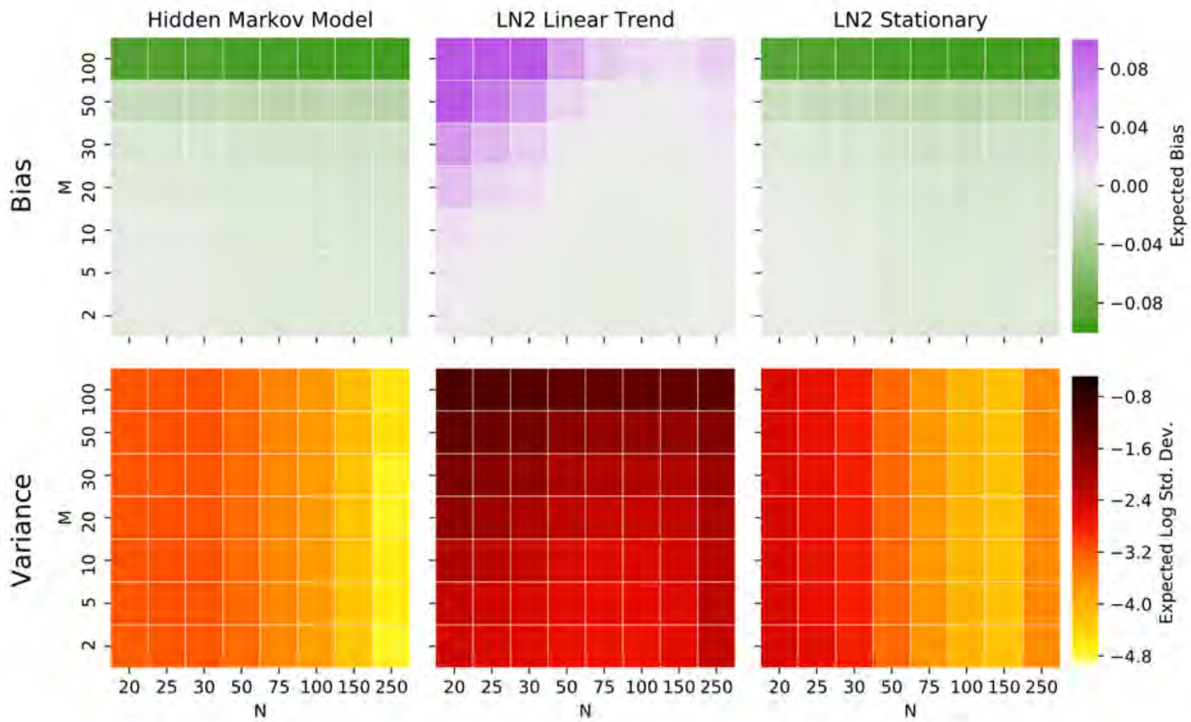


Figure 12-7 As previous figures but for sequences generated with both LFV and secular change.

12.4.4. Discussion

Evaluating and implementing investments for climate risk mitigation involves making projections of climate risk, which generally exhibits both LFV and secular trends, over the M-year project life of the instrument. The success of this prediction will depend on the identifiability of different signals from limited information, the time scales of LFV relative to the project life of the instrument, and the degree of intrinsic uncertainty in the system. In this paper we took a synthetic data approach to explore the implications of varying M and N in stylized scenarios that represent important features of real-world hydroclimate systems.

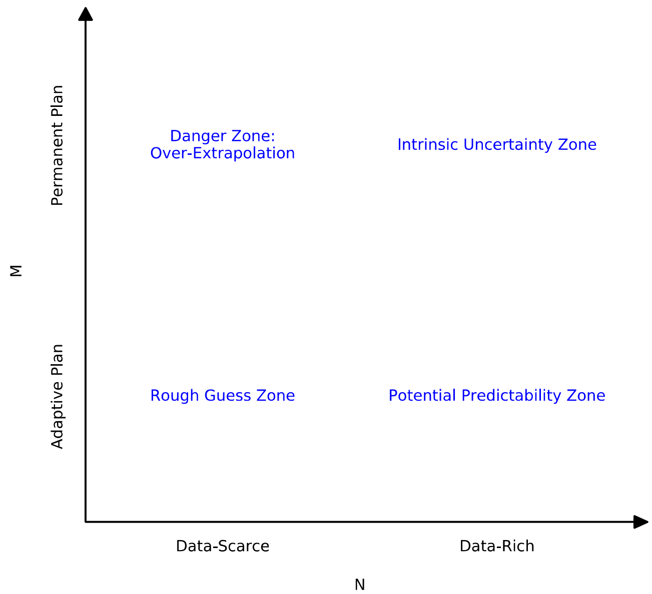


Figure 12-8 The importance of predicting different signals, and the identifiability and predictability of the signals, depends on the degree of informational uncertainty (N) and the project planning period (M).

Figure 12-5 and Figure 12-7 show that for projects where M is sufficiently short, intrinsic uncertainty is low and cyclical climate variability is dominant over the project planning period (Jain and Lall 2001; Hodgkins et al. 2017). However, one's ability to identify and predict this variability depends on having a model of sufficient complexity to represent the processes that cause LFV, and the data to fit the model. In this case, the project may be in the "potential predictability zone" of Figure 12-8. If sufficient information is not available, however, then simple models which represent fewer processes may be preferred (the "rough guess zone").

For projects with longer M, our results highlight the importance of identifying and predicting secular change. As illustrated schematically in Figure 12-2, large uncertainties (e.g., as to future CO_2 concentrations and local climate impacts) lead to large intrinsic uncertainty in projections of future climate risk. As the physical mechanisms cascade from global (e.g., global mean surface temperature) to regional (e.g., storm track position; Barnes and Screen 2015) and local (e.g., annual-maximum streamflows) scales, informational uncertainties also compound and increase (Dittes et al. 2017). With sufficient information (large N), this informational uncertainty may be reduced, but this data cannot address intrinsic uncertainty and this zone is thus named the "intrinsic uncertainty zone". Finally, if N is limited then there will be strong potential for misleading estimates and over-extrapolation (i.e. a "danger zone" for planning).

These findings were derived conceptually and through idealized computational experiments for simulating and predicting climate risk, but the principles are applicable to more complex, physically based methods. For example, flood frequency analysis may join observations across time and space (Lima et al. 2016; Merz and Blöschl 2008) or apply model chains based on general circulation models and hydrologic models (see Merz et al. 2014) to increase N. We suggest that the sample size N defined in our experiments may be straightforwardly interpreted as a measure of the total informational uncertainty in the analysis; as N increases, informational uncertainty decreases.

Similarly, real-world climate adaptation plans will typically include multiple instruments which may be placed in different locations and times in a sequential fashion. Even if the planning period of a portfolio is long, the individual instruments within the portfolio may have short planning periods. Since our results show that the bias and variance of climate risk projections tend to increase with M, the total bias and variance associated with sequencing 20 consecutive M=5 year projects will be less than that associated with making a single M=100 year project. This effect will be compounded by the fact that if the first M=5 year project is based on estimates with informational uncertainty N, the second will have $N+5$, the third $N+10$, and so on.

The climate adaptation decisions which our analysis can inform are typically framed as economic cost-benefit analyses which discount future cash flows at some annual rate (Sodastrom, Sokolove, and Fairfax 1999; Powers 2003). The application of a positive discount rate, mandated for many public sector projects in the United States (Powers 2003), further emphasizes the importance of predicting near-term risk. Projects with long planning periods must therefore overcome future discounting, the potential for large bias or variance, and that all estimates are made with informational uncertainty N. By contrast, the informational uncertainties for a sequence of short-term instruments are $N, N+M, N+2M, \ldots$, potentially yielding improved identifiability and predictability of relevant climate signals.

12.5. CONCLUSION

In this paper we considered how the temporal structure of the climate affects the potential for successful prediction over a finite M-year future period. We began with three premises, or observations, about the nature of climate risk: that different climate risk mitigation instruments have different planned lifespans; that climate risk varies on many scales; and that the processes which dominate this risk over the planning period depend on the planning period itself. Although the simulations presented here are neatly divided into secular change, LFV only, and LFV plus secular change, real-world hydroclimate time series exhibit LFV on many timescales and several sources of (not necessarily linear) secular change, adding further informational and intrinsic uncertainties.

Depending on the specific climate mechanisms that impact a particular site, and the predictability thereof, the cost and risk associated with a sequence of short-term adaptation projects may be lower than with building a single, permanent structure to prepare for a worst-case scenario far into the future. For most large actors, a portfolio of both large M and small M projects will likely be necessary, none of which precludes the need for mitigation of global and local climate change and the development or the execution of vulnerability reduction strategies.

12.6. REFERENCES

Ångström, Anders. 1935. "Teleconnections of Climatic Changes in Present Time." Geografiska Annaler 17 (3-4): 242–58. https://doi.org/10.1080/20014422.1935.11880600.

Barnes, Elizabeth A, and James A Screen. 2015. "The Impact of Arctic Warming on the Midlatitude Jet-Stream: Can It? Has It? Will It?" Wiley Interdisciplinary Reviews-Climate Change 6 (3): 277–86. https://doi.org/10.1002/wcc.337.

Beran, Jan. 1994. Statistics for Long-Memory Processes. New York: Chapman & Hall.

Betancourt, Michael. 2017. "A Conceptual Introduction to Hamiltonian Monte Carlo." arXiv.org. http://arxiv.org/abs/1701.02434v2.

Bhattacharya, R. N., Vijay K. Gupta, and Ed Waymire. 1983. "The Hurst Effect Under Trends." Journal of Applied Probability 20 (3): 649–62. https://doi.org/10.2307/3213900.

Blöschl, Günter, and Alberto Montanari. 2010. "Climate Change Impacts—Throwing the Dice?" Hydrological Processes 24 (3): 374–81. https://doi.org/10.1002/hyp.7574.

Borgomeo, Edoardo, Mohammad Mortazavi-Naeini, Jim W Hall, and Benoit P Guillod. 2018. "Risk, Robustness and Water Resources Planning Under Uncertainty." Earth's Future 6 (3): 468–87. https://doi.org/10.1002/2017EF000730.

Bracken, C, B Rajagopalan, and C Woodhouse. 2016. "A Bayesian Hierarchical Nonhomogeneous Hidden Markov Model for

Multisite Streamflow Reconstructions." Water Resources Research 52 (10): 7837–50. https://doi.org/10.1002/2016WR018887.
Brown, Casey M. 2010. "The End of Reliability." Journal of Water Resources Planning and Management 136 (2): 143–45. https://doi.org/10.1061/(ASCE)WR.1943-5452.65.
Bullmore, Ed, Chris Long, John Suckling, Jalal Fadili, Gemma Calvert, Fernando Zelaya, T. Adrian Carpenter, and Mick Brammer. 2001. "Colored Noise and Computational Inference in Neurophysiological (fMRI) Time Series Analysis: Resampling Methods in Time and Wavelet Domains." Human Brain Mapping 12 (2): 61–78. https://doi.org/10.1002/1097-0193(200102)12:2<61::AID-HBM1004>3.0.CO;2-W.
Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." Journal of Statistical Software 76 (1): 1–29. https://doi.org/10.18637/jss.v076.i01.
Chamoli, Ashutosh, Abhey Ram Bansal, and V. P. Dimri. 2007. "Wavelet and Rescaled Range Approach for the Hurst Coefficient for Short and Long Time Series." Computers & Geosciences 33 (1): 83–93. https://doi.org/10.1016/j.cageo.2006.05.008.
City of New York. 2013. "A Stronger, More Resilient New York." New York.
Cook, Edward R, Richard Seager, Richard R Heim Jr, Russell S Vose, Celine Herweijer, and Connie Woodhouse. 2010. "Megadroughts in North America: Placing IPCC Projections of Hydroclimatic Change in a Long-Term Palaeoclimate Context." Journal of Quaternary Science 25 (1): 48–61. https://doi.org/10.1002/jqs.1303.
Coumou, Dim, and Stefan Rahmstorf. 2012. "A Decade of Weather Extremes." Nature Climate Change 10 (7): P100013–496. https://doi.org/10.1038/nclimate1452.
Delgado, J M, B Merz, and Heiko Apel. 2014. "Projecting Flood Hazard Under Climate Change: An Alternative Approach to Model Chains." Natural Hazards and Earth System Science 14 (6): 1579–89. https://doi.org/10.5194/nhess-14-1579-2014.
Di Baldassarre, Giuliano, Heidi Kreibich, Sergiy Vorogushyn, Jeroen Aerts, Karsten Arnbjerg-Nielsen, Marlies Barendrecht, Paul Bates, et al. 2018. "Hess Opinions: An Interdisciplinary Research Agenda to Explore the Unintended Consequences of Structural Flood Protection." Hydrology and Earth System Sciences 22 (11): 5629–37. https://doi.org/https://doi.org/10.5194/hess-22-5629-2018.
Dittes, Beatrice, Olga Špačková, Lukas Schoppa, and Daniel Straub. 2017. "Climate Uncertainty in Flood Protection Planning." Hydrological and Earth System Science, 1–28. https://doi.org/10.5194/hess-2017-576.
Dugan, Patrick J., Chris Barlow, Angelo A. Agostinho, Eric Baran, Glenn F. Cada, Daqing Chen, Ian G. Cowx, et al. 2010. "Fish Migration, Dams, and Loss of Ecosystem Services in the Mekong Basin." AMBIO 39 (4): 344–48. https://doi.org/10.1007/s13280-010-0036-1.
Farnham, David J, James Doss-Gollin, and Upmanu Lall. 2018. "Regional Extreme Precipitation Events: Robust Inference from Credibly Simulated GCM Variables." Water Resources Research. https://doi.org/10.1002/2017wr021318.
Feng, Huaguang, Thomas R. Willemain, and Nong Shang. 2005. "Wavelet-Based Bootstrap for Time Series Analysis." Communications in Statistics - Simulation and Computation 34 (2): 393–413. https://doi.org/10.1081/SAC-200055722.
GEI Consultants, Inc. 2017. "Tulare Lake Storage and Floodwater Protection Feasibility Report."
Gelman, Andrew, Daniel Simpson, and Michael Betancourt. 2017. "The Prior Can Often Only Be Understood in the Context of the Likelihood." Entropy 19 (10): 555. https://doi.org/10.3390/e19100555.
Geweke, John, and Susan Porter-Hudak. 1983. "The Estimation and Application of Long Memory Time Series Models." Journal of Time Series Analysis 4 (4): 221–38. https://doi.org/10.1111/j.1467-9892.1983.tb00371.x.
Haasnoot, Marjolijn, Jan H Kwakkel, Warren E Walker, and Judith ter Maat. 2013. "Dynamic Adaptive Policy Pathways: A Method for Crafting Robust Decisions for a Deeply Uncertain World." Global Environmental Change 23 (2): 485–98. https://doi.org/10.1016/j.gloenvcha.2012.12.006.
Held, Isaac M. 2005. "The Gap Between Simulation and Understanding in Climate Modeling." Bulletin of the American Meteorological Society 86 (11): 1609–14. https://doi.org/10.1175/BAMS-86-11-1609.
Hodgkins, Glenn A, Paul H Whitfield, Donald H Burn, Jamie Hannaford, Benjamin Renard, Kerstin Stahl, Anne K Fleig, et al. 2017. "Climate-Driven Variability in the Occurrence of Major Floods Across North America and Europe." Journal of Hydrology 552: 704–17. https://doi.org/10.1016/j.jhydrol.2017.07.027.
Hoffman, Matthew D, and Andrew Gelman. 2011. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." arXiv.org. http://arxiv.org/abs/1111.4246v1.
Hoyer, Stephan, and Joe Hamman. 2017. "Xarray: N-D Labeled Arrays and Datasets in Python." Journal of Open Research Software 5 (1): 304. https://doi.org/10.5334/jors.148.
Hunter, John D. 2007. "Matplotlib: A 2D Graphics Environment." Computing in Science & Engineering 9 (3): 90–95. https://doi.org/10.1109/MCSE.2007.55.
IPCC. 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation.
Jain, Shaleen, and Upmanu Lall. 2001. "Floods in a Changing Climate: Does the Past Represent the Future?" Water Resources Research 37 (12): 3193–3205. https://doi.org/10.1029/2001WR000495.
Jin, F F, J D Neelin, and M Ghil. 1994. "El-Nino on the Devils Staircase - Annual Subharmonic Steps to Chaos." Science 264 (5155): 70–72.
Jones, Eric, Travis Oliphant, and Pearu Peterson. 2001. "SciPy: Open Source Scientific Tools for Python."
Jongman, Brenden, Philip J Ward, and Jeroen C J H Aerts. 2012. "Global Exposure to River and Coastal Flooding: Long Term Trends and Changes." Global Environmental Change 22 (4): 823–35. https://doi.org/10.1016/j.gloenvcha.2012.07.004.

Kiem, Anthony S, Stewart W Franks, and George Kuczera. 2002. “Multi-Decadal Variability of Flood Risk.” Geophysical Research Letters 30 (2): 3. https://doi.org/10.1029/2002GL015992.

Koutsoyiannis, Demetris. 2003. “Climate Change, the Hurst Phenomenon, and Hydrological Statistics.” Hydrological Sciences Journal 48 (1): 3–24. https://doi.org/10.1623/hysj.48.1.3.43481.

Kwon, Hyun-Han, Upmanu Lall, and Abedalrazq F Khalil. 2007. “Stochastic Simulation Model for Nonstationary Time Series Using an Autoregressive Wavelet Decomposition: Applications to Rainfall and Temperature.” Water Resources Research 43 (5): 287. https://doi.org/10.1029/2006WR005258.

Lempert, Robert J., and Myles T. Collins. 2007. “Managing the Risk of Uncertain Threshold Responses: Comparison of Robust, Optimum, and Precautionary Approaches.” Risk Analysis 27 (4): 1009–26. https://doi.org/10.1111/j.1539-6924.2007.00940.x.

Lima, Carlos H R, Upmanu Lall, Tara Troy, and Naresh Devineni. 2016. “A Hierarchical Bayesian GEV Model for Improving Local and Regional Flood Quantile Estimates.” Journal of Hydrology 541: 816–23. https://doi.org/10.1016/j.jhydrol.2016.07.042.

Lovejoy, S. 2013. “What Is Climate?” EOS, Transactions American Geophysical Union 94 (1): 1–2. https://doi.org/10.1002/2013EO010001.

Lovejoy, Shaun, and Daniel Schertzer, eds. 2013. The Weather and Climate Emergent Laws and Multifractal Cascades /. Cambridge : Cambridge University Press,

Lovejoy, S., and D. Schertzer. 2012. “Low-Frequency Weather and the Emergence of the Climate.” In Extreme Events and Natural Hazards: The Complexity Perspective, 231–54. American Geophysical Union (AGU). https://doi.org/10.1029/2011GM001087.

Mandelbrot, Benoit B. 1985. “Self-Affine Fractals and Fractal Dimension.” Physica Scripta 32 (4): 257–60. https://doi.org/10.1088/0031-8949/32/4/001.

Mandelbrot, Benoit B., and James R. Wallis. 1969. “Robustness of the Rescaled Range R/S in the Measurement of Noncyclic Long Run Statistical Dependence.” Water Resources Research 5 (5): 967–88. https://doi.org/10.1029/WR005i005p00967.

Mantua, N J, S R Hare, Y Zhang, J M Wallace, and R C Francis. 1997. “A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production.” Bulletin of the American Meteorological Society 78 (6): 1069–79. https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

Markonis, Yannis, and Demetris Koutsoyiannis. 2013. “Climatic Variability over Time Scales Spanning Nine Orders of Magnitude: Connecting Milankovitch Cycles with Hurst–Kolmogorov Dynamics.” Surveys in Geophysics 34 (2): 181–207. https://doi.org/10.1007/s10712-012-9208-9.

Maruyama, Fumio. 2018. “The Relation Among the Solar Activity, the Total Ozone, QBO, NAO, and ENSO by Wavelet-Based Multifractal Analysis.” Journal of Applied Mathematics and Physics 06: 1301. https://doi.org/10.4236/jamp.2018.66109.

McKinney, W. 2010. “Data Structures for Statistical Computing in Python.” In Proceedings of the 9th Python in Science Ldots.

Merz, Bruno, Jeroen C J H Aerts, Karsten Arnbjerg-Nielsen, M Baldi, A Becker, A Bichet, Günter Blöschl, et al. 2014. “Floods and Climate: Emerging Perspectives for Flood Risk Assessment and Management.” Natural Hazards and Earth System Science 14 (7): 1921–42. https://doi.org/10.5194/nhess-14-1921-2014.

Merz, Ralf, and Günter Blöschl. 2008. “Flood Frequency Hydrology: 1. Temporal, Spatial, and Causal Expansion of Information.” Water Resources Research 44 (8): 3. https://doi.org/10.1029/2007WR006744.

Milly, P C D, Julio Betancourt, M Falkenmark, R M Hirsch, Z W Kundzewicz, D P Lettenmaier, and R J Stouffer. 2008. “Stationarity Is Dead: Whither Water Management?” Science 319 (5863): 573–74. https://doi.org/10.1126/science.1151915.

Montanari, Alberto. 2003. “Long-Range Dependence in Hydrology.” Theory and Applications of Long-Range Dependence, 461–72.

Montanari, Alberto, and Demetris Koutsoyiannis. 2014. “Modeling and Mitigating Natural Hazards: Stationarity Is Immortal!” Water Resources Research 50 (12): 9748–56. https://doi.org/10.1002/2014WR016092.

Müller, Björn, Martin Wild, Anton Driesse, and Klaus Behrens. 2014. “Rethinking Solar Resource Assessments in the Context of Global Dimming and Brightening.” Solar Energy 99: 272–82. https://doi.org/10.1016/j.solener.2013.11.013.

Newman, Matthew, Michael A. Alexander, Toby R. Ault, Kim M. Cobb, Clara Deser, Emanuele Di Lorenzo, Nathan J. Mantua, et al. 2016. “The Pacific Decadal Oscillation, Revisited.” Journal of Climate 29 (12): 4399–4427. https://doi.org/10.1175/JCLI-D-15-0508.1.

O’Connell, P. E., D. Koutsoyiannis, H. F. Lins, Y. Markonis, A. Montanari, and T. Cohn. 2016. “The Scientific Legacy of Harold Edwin Hurst (1880–1978).” Hydrological Sciences Journal 61 (9): 1571–90. https://doi.org/10.1080/02626667.2015.1125998.

O’Gorman, Paul A, and Tapio Schneider. 2009. “The Physical Basis for Increases in Precipitation Extremes in Simulations of 21st-Century Climate Change.” Proceedings of the National Academy of Sciences of the United States of America 106 (35): 14773–7. https://doi.org/10.1073/pnas.0907610106.

Palmer, T N. 1993. “A Nonlinear Dynamical Perspective on Climate Change.” Weather 48 (10): 314–26. https://doi.org/10.1002/j.1477-8696.1993.tb05802.x.

Papakonstantinou, Vangelis, Thomas Maier, Rashad R Kaldany, Andre Belelieu, and Michael Drexler. 2016. “Risk Mitigation Instruments in Infrastructure.” World Economic Forum.

Peduzzi, P, B Chatenoux, H Dao, A De Bono, C Herold, J Kossin, F Mouton, and O Nordbeck. 2012. “Global Trends in Tropical Cyclone Risk.” Nature Climate Change 2 (4): 289–94. https://doi.org/10.1038/nclimate1410.

Powers, Kyna. 2003. “Benefit-Cost Analysis and the Discount Rate for the Corps of Engineers’ Water Resource Projects: Theory and Practice.” RL31976. Washington, DC: Congressional Research Service.

Rabiner, L, and B Juang. 1986. “An Introduction to Hidden Markov Models.” IEEE ASSP Magazine 3 (1): 4–16. https://doi.org/10.1109/MASSP.1986.1165342.

Ramesh, Nandini, Mark A Cane, Richard Seager, and Dong Eun Lee. 2016. “Predictability and Prediction of Persistent Cool States of the Tropical Pacific Ocean.” Climate Dynamics 49 (7-8): 2291–2307. https://doi.org/10.1007/s00382-016-3446-3.

Roesch, A, and Harald Schmidbauer. 2016. WaveletComp: Computational Wavelet Analysis. R Package Version 1.1.

Rojo Hernandez, J. D., U. Lall, and O. J. Mesa. 2017. “ENSO Dynamics and Trends, an Alternate View.” AGU Fall Meeting Abstracts 41.

Ropelewski, C F, and M S Halpert. 1987. “Global and Regional Scale Precipitation Patterns Associated with the El-Nino Southern Oscillation.” Monthly Weather Review 115 (8): 1606–26. https://doi.org/10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2.

Salas, J D, J Obeysekera, and R M Vogel. 2018. “Techniques for Assessing Water Infrastructure for Nonstationary Extreme Events: A Review.” Hydrological Sciences Journal 63 (3): 325–52. https://doi.org/10.1080/02626667.2018.1426858.

Sarachik, Edward S, and Mark A Cane. 2009. The El Niño–Southern Oscillation Phenomenon. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511817496.

Schreiber, Jacob. 2017. “Pomegranate: Fast and Flexible Probabilistic Modeling in Python.” Journal of Machine Learning Research 18 (1): 5992–7.

Selvam, A. M. 2017. “Universal Inverse Power-Law Distribution for Fractal Fluctuations in Dynamical Systems: Applications for Predictability of Inter-Annual Variability of Indian and USA Region Rainfall.” Pure and Applied Geophysics 174 (1): 413–26. https://doi.org/10.1007/s00024-016-1394-9.

Serinaldi, Francesco, and Chris G Kilsby. 2015. “Stationarity Is Undead: Uncertainty Dominates the Distribution of Extremes.” Advances in Water Resources 77: 17–36. https://doi.org/10.1016/j.advwatres.2014.12.013.

Shaw, T A, M Baldwin, Elizabeth A Barnes, Rodrigo Caballero, C I Garfinkel, Y T Hwang, C Li, et al. 2016. “Storm Track Processes and the Opposing Influences of Climate Change.” Nature Geoscience 9 (9): 656–64. https://doi.org/10.1038/ngeo2783.

Simonsen, Ingve, Alex Hansen, and Olav Magnar Nes. 1998. “Determination of the Hurst Exponent by Use of Wavelet Transforms.” Physical Review E 58 (3): 2779–87. https://doi.org/10.1103/PhysRevE.58.2779.

Simpson, Daniel, Håvard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sørbye. 2017. “Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors.” Statistical Science 32 (1): 1–28. https://doi.org/10.1214/16-STS576.

Sodastrom, Elizabeth, Jennifer Sokolove, and Sally K. Fairfax. 1999. “Federal Reserved Water Rights Applied to School Trust Lands Land and Water Division: Article.” Land and Water Law Review 34: 1–38.

Spence, Caitlin M, and Casey M Brown. 2016. “Nonstationary Decision Model for Flood Risk Decision Scaling.” Water Resources Research 52 (11): 8650–67. https://doi.org/10.1002/2016WR018981.

Swierczynski, Tina, Achim Brauer, Stefan Lauterbach, Celia Martín-Puertas, Peter Dulski, Ulrich von Grafenstein, and Christian Rohr. 2012. “A 1600 Yr Seasonally Resolved Record of Decadal-Scale Flood Variability from the Austrian Pre-Alps.” Geology 40 (11): 1047–50. https://doi.org/10.1130/G33493.1.

Torrence, Christopher, and Gilbert P Compo. 1998. “A Practical Guide to Wavelet Analysis.” Bulletin of the American Meteorological Society 79 (1): 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2.

Trenberth, Kevin E, Aiguo Dai, Roy M Rasmussen, and David B Parsons. 2003. “The Changing Character of Precipitation.” Bulletin of the American Meteorological Society 84 (9): 1205–17. https://doi.org/10.1175/BAMS-84-9-1205.

USACE. 2007. “Flood Damage Analysis Santa Maria Levee Project.” Los Angeles, CA: United States Army Corps of Engineers Los Angeles District.

———. 2014. “Dallas Floodway, Dallas TX.” Fort Worth, TX: United States Army Corps of Engineers, Fort Worth District.

Walker, Warren, Marjolijn Haasnoot, and Jan Kwakkel. 2013. “Adapt or Perish: A Review of Planning Approaches for Adaptation Under Deep Uncertainty.” Sustainability 5 (3): 955–79. https://doi.org/10.3390/su5030955.

Walt, Stéfan van der, S Chris Colbert, and Gaël Varoquaux. 2011. “The NumPy Array: A Structure for Efficient Numerical Computation.” Computing in Science & Engineering 13 (2): 22–30. https://doi.org/10.1109/MCSE.2011.37.

Ward, P J, B Jongman, M Kummu, Michael D Dettinger, F C Sperna Weiland, and H C Winsemius. 2014. “Strong Influence of El Nino Southern Oscillation on Flood Risk Around the World.” Proceedings of the National Academy of Sciences of the United States of America 111 (44): 15659–64. https://doi.org/10.1073/pnas.1409822111.

Woollings, T, C Franzke, D L R Hodson, B Dong, Elizabeth A Barnes, C C Raible, and J G Pinto. 2014. “Contrasting Interannual and Multidecadal NAO Variability.” Climate Dynamics 45 (1-2): 539–56. https://doi.org/10.1007/s00382-014-2237-y.

Zebiak, Stephen E, and Mark A Cane. 1987. “A Model El Niño-Southern Oscillation.” Monthly Weather Review 115 (10): 2262–78. https://doi.org/10.1175/1520-0493(1987)115<2262:AMENO>2.0.CO;2.

Živković, T., and K. Rypdal. 2013. “ENSO Dynamics: Low-Dimensional-Chaotic or Stochastic?” Journal of Geophysical Research: Atmospheres 118 (5): 2161–8. https://doi.org/10.1002/jgrd.50190.

13. APPENDIX

13.1. PUBLICATIONS

13.1.1. Published or Accepted

Doss-Gollin, J., Farnham, D.J., Steinschneider, S., Lall, U., 2019. Robust Adaptation to Multi-Scale Climate Variability. Earths Future. https://doi.org/10.1029/2019EF001154

François, B., Schlef, K.E., Wi, S., Brown, C.M., 2019. Design considerations for riverine floods in a changing climate – A review. J. Hydrol. 574, 557–573. https://doi.org/10.1016/j.jhydrol.2019.04.068

Schlef, K. E., B. François, A. Robertson, and C. Brown (2018). A General Methodology for Climate Informed Approaches to Long-Term Flood Projection – Illustrated with the Ohio River Basin. Water Resources Research 54, https://doi.org/10.1029/2018WR023209

Schlef, K. E. (2018). Flood Risk Assessment, Management, and Perceptions in a Changing World. PhD thesis, University of Massachusetts Amherst, Department of Civil and Environmental Engineering, Amherst, MA (chpt. 1 and 2). Available at: https://scholarworks.umass.edu/dissertations_2/1273/

David J Farnham, James Doss-Gollin, and Upmanu Lall. (2018) "Regional Extreme Precipitation Events: Robust Inference From Credibly Simulated GCM Variables". Water Resources Research .doi:10.1002/2017wr021318.

McCrary, R. R., S. McGinnis, and L. O. Mearns, (2017): Evaluation of Snow Water Equivalent in NARCCAP Simulations, Including Measures of Observational Uncertainty. *J. Hydromet*, **18**, 28.

13.1.2. Submitted or Under Review

Schlef, K.E., François, B., Brown, C., **Submitted**. Comparing Flood Projection Approaches across Hydry-climatological Diverse River Basins. Water Resource Research

McCrary, R. R and L. O. Mearns, **Accepted Pending Revisions**: Quantifying and Diagnosing Sources of Uncertainty in Mid-Century Changes in North American Snowpack from NARCCAP. *J. Hydromet*,.

13.1.3. In preparation

McCrary, R. R, M. Bukovsky, L. O. Mearns, 2019: Future Changes in snow over the Missouri River Basin, J. Hydromet, In Preparation.

Bukovsky, M.S., R.R. McCrary, S. McGinnis, and L. Mearns, 2019: The future of cool-season precipitation in the East-South-Central U.S. in regional climate models. In preparation.

13.2. PRESENTATION IN CONFERENCES AND WORKSHOP

Yash Amonkar, Upmanu Lall. 2019 "Spatiotemporal Clustered Risk of Flooding in the Ohio River Basin and American Midwest". In: Workshop on Correlated Extremes - Initiative on Extreme Weather and Climate. New York, NY, May, 29, 2019.

James Doss-Gollin, 2019, Evaluating staged investments in critical infrastructure for climate adaptation, Interdisciplinary Ph.D. Workshop in Sustainable Development 2019, Columbia University, New York,NY, talk.

McCrary, R.R., J. Jacobs, and L.O.Mearns, 2019: Future Changes in Mean and Extreme Daily Snowfall over the United States. 76th Eastern Snow Conference, Fairlee, Vermont, June 2019.

James Doss-Gollin, 2018 Robust Adaptation to Multi-Scale Climate Variability,The Nexus of Climate Data, Insurance, and Adaptive Capacity, Asheville, NC, poster.

Schlef, K., B. François, S. Wi, and C. Brown (2018). A Comparison of Approaches to Long-Term Flood Projection, 2018 CUAHSI Biennial Colloquium, Poster.

McCrary, R.R., M. Bukovsky, and L.O.Mearns, 2018: Uncertainty in future changes in snowpack and rain-on-snow events in the U.S> northern Great Plains using high-resolution climate models. 75th Eastern Snow Conference, College Park, Maryland, June 2018.

James Doss-Gollin, David J Farnham, Scott Steinschneider, and Upmanu Lall. “RobustAdaptation to Multi-Scale Climate Variability”. In:American Geophsyical Union Fall Meeting.Washington, DC, Dec. 14, 2018.doi:10.13140/RG.2.2.28447.20649.

James Doss-Gollin, 2017, Regional Intense Precipitation: Inferences From GCM Atmospheric Circulation Fields,Modeling Research in the Cloud, NCAR, Boulder, Colorado, poster.

James Doss-Gollin, 2017, Statistical-Dynamical Analysis of Climate Projections for Flood Infrastructure Design,Interdisciplinary Ph.D. Workshop in Sustainable Development 2017, Columbia University, NewYork, NY, talk.

James Doss-Gollin, David J Farnham, and Upmanu Lall. “Designing and Operating In-frastructure for Nonstationary Flood Risk Management”. In:American Geophsyical UnionFall Meeting. New Orleans, LA, Dec. 2017.doi:10.13140/RG.2.2.16110.46403.

McGinnis, Seth, Melissa Bukovsky, and Linda Mearns. “From Descriptive to Explanatory: Using Metrics to Identify Candidate Phenomena for Process Evaluation in NA-CORDEX and NARCCAP.” Fall Meeting, American Geophysical Union, New Orleans LA, 12 December 2017. (Talk)

Davide Faranda, Gabriele Messori,James Doss-Gollin, David J Farnham, Upmanu Lall,and Pascal Yiou. “Dynamics and Thermodynamics of Weather Extremes: A Dynamical Systems Approach”. In:American Geophsyical Union Fall Meeting. New Orleans, LA, Dec.2017.

Schlef, K., A. Robertson, and C. Brown (2017). Projections of Flood Risk using Credible Climate Signals in the Ohio River Basin, Abstract H21L-04, 2017 Fall Meeting AGU, Oral.

François, B., Wi, S., Brown, C. (2017) Using deep learning to assess future flood magnitude and frequency in the semi-arid and snowmelt-dominated Missouri River headwater catchments. Abstract 274511, AGU conference, New Orleans, December 2017

Bukovsky, M.S., R.R. McCrary, T.S. Rendfrey, A.S. Schroeder, and L. Mearns, 2017: Causes of Cool-Season Precipitation Bias in the East-South-Central U.S. Poster. *2017 AGU Fall Meeting*, New

Orleans, LA. A23D-2396.

McCrary, R.R., L.O Mearns, 2017: Uncertainty in the Future of Seasonal Snowpack over North America, *2017 AGU Fall Meeting*, New Orleans, LA. GC24D-06.

Kravits, J., and Schlef, K. (2017). A Comparison of Modeling Choices when Sizing Levees in a Changing Climate: A Louisville Kentucky Case Study, 2017 AGU Undergraduate Virtual Poster Showcase - Fall 2017 (http://abstractsearch.agu.org/vps/2017/8/99.html)

Schlef, K., C. Spence, and C. Brown (2016). Modeling non-stationary flood magnitude and frequency in West Africa using a hierarchical Bayesian framework conditioned on large-scale atmospheric processes, Abstract NH53E-06, 2016 Fall Meeting AGU, Oral.

James Doss-Gollin, David J Farnham, and Upmanu Lall. “Global-Local Interactions Mod-ulate Tropical Moisture Exports to the Ohio River Basin”. In:American Geophsyical UnionFall Meeting. San Francisco, CA, 2016.doi:10.13140/RG.2.2.36009.19044.

David J Farnham,James Doss-Gollin, and Upmanu Lall. “Space-Time Characteristics andStatistical Predictability of Extreme Daily Precipitation Events in the Ohio River Basin”. In:American Geophsyical Union Fall Meeting. San Francisco, CA, Dec. 2016.

Caitlin M Spence, Casey Brown, and James Doss-Gollin. “Exploiting Synoptic-Scale Climate Processes to Develop Nonstationary, Probabilistic Flood Hazard Projections”. In: American Geophysical Union Fall Meeting. San Francisco, CA. 2016

James Doss-Gollin, 2016, Understanding the Physical Drivers of Extreme Rainfall for Flood Prediction, Oxford Water Network, Oxford University, talk