Using Climate Change Projections to Increase the Resilience of Stormwater Infrastructure Designs Under Uncertainty

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Department of Civil and Environmental Engineering

Lauren M. Cook

B.S. in Civil and Environmental Engineering, University of Maryland
M.S. in Industrial Engineering, Institute Français du Petrole (IFP School)

Carnegie Mellon University
Pittsburgh, PA 15213

December 2018

ii

For Clifford, my most annoying and favorite companion

Acknowledgements

Dissertation Committee

Constantine Samaras (Chair)

Assistant Professor

Civil and Environmental Engineering

Carnegie Mellon University

Jeanne M. VanBriesen (Co-advisor)

Duquesne Light Company Professor

Civil and Environmental Engineering & Engineering and Public Policy

Carnegie Mellon University

Jared Cohon

President Emeritus and University Professor

Civil and Environmental Engineering & Engineering and Public Policy

Carnegie Mellon University

Mitchell Small

H. John Heinz Professor

Civil and Environmental Engineering & Engineering and Public Policy

Carnegie Mellon University

Kathleen White

Senior Lead for Global and Climate Change

Institute for Water Resources

U.S. Army Corps of Engineers

Abstract

Climate change is expected to increase the intensity of rainfall events, posing a major threat to stormwater infrastructure systems. To ensure that these systems continue to be resilient and reliable under changing conditions, traditional engineering design methods must be updated to incorporate changing rainfall patterns. Climate models can be used to gain insight into these changes; however, the path from climate projections to design decisions of future stormwater structures is unclear. The objective of this research was to determine how to use climate change projections during the stormwater design process to increase the resilience of stormwater infrastructure under uncertain, future conditions.

To advance this objective, a general framework was developed for the use of climate model data in engineering analyses. The framework consists of five main steps: define historical data requirements, select the appropriate climate model data source, bound uncertainty, convert model output to the format for the engineering analysis, and interpret results for an engineering audience. The framework was applied to updating intensity duration frequency (IDF) curves, an important aspect of the stormwater design process.

Findings from the updating of IDF curves suggest that the range of potential increases in extreme precipitation are large and different modeling choices can alter the size and level of protection of infrastructure designs. As a result, the second recommended revision to the design process is to assess performance of infrastructure over time as the climate changes. Continuous simulation can be used as a tool to test the performance of designs once they have been conceptualized or built. Findings suggest that annual measures of rainfall could also be used to anticipate performance degradation and necessary adaptation actions.

ACRONYMS

AMS   Annual Maximum Series

ATF   AMS Transfer Function

BCSD   Bias-Corrected Spatial Disaggregation method

CDF   Cumulative Distribution Function

CF   Change Factor

CI   Confidence Interval

DDF   Depth-Duration-Frequency

ECP   Scripps model

ESM   Earth System Model

FEMA   Federal Emergency Management Agency

GCM   General Circulation Model

GEV   Generalized Extreme Value

HIRHAM   Hadley model

IDF   Intensity-Duration-Frequency

KDDM   Kernel Density Distribution Mapping

MACA   Multivariate Constructed Analogs

MM5   Iowa State model

NARCCAP   North American Regional Climate Change Assessment Program

NA-CORDEX North American Coordinated Regional Downscaling Experiment

NOAA National Oceanic and Atmospheric Administration

PCSWMM PC Stormwater Management Model

RCM Regional Climate Model

RCP Representative Concentration Pathway

RG Rain Garden

SRES Special Report on Emissions Scenarios

USGS U.S. Geological Survey

WRFP Weather Research Forecasting Model

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Motivation

Urban water systems of the 21st century face many challenges, including aging infrastructure, increased urbanization, and lack of support for long term investments (Larsen et al. 2016). These issues are amplified by climate change, which has the potential to affect infrastructure systems in multiple ways, including changes in extreme temperatures, sea levels, droughts, streamflows, frequency and severity of rain storms, and flooding (Walsh, J.D., Wuebbles, K, and Kossin, J 2014; IPCC 2014; Kilgore et al. 2016). Increases in the intensity of rainfall events (amount of rain per hour) are a major threat to infrastructure systems, especially stormwater infrastructure systems and the transportations systems they protect. Higher rainfall intensities lead to more severe storms, with expected increases in damages and fatalities related to residential, street, and flash flooding (Willems et al. 2013; Arnbjerg-Nielsen et al. 2013; Mailhot and Duchesne 2009), as well as the potential for an increase in the volume of sewer overflows (Goorè Bi et al. 2015; Semadeni-Davies et al. 2008).

Increases in the severity and frequency of extreme storms are already occurring throughout the United States (U.S.) (Karl and Knight 1998; Groisman et al. 2005). Severity and frequency of storms are described by precipitation measures, called indices, which represent different quantity and frequency characteristics. Reviewing data for the contiguous United States from 1948 to the present, Groisman et al (2012) calculated the number of "very heavy" rainfall events (days with rain above 76.2 mm) and extreme precipitation events (storms with more than 154.9 mm of rain); they detected a significant increase in these events over the past 30-years, compared to the period from 1948 – 1978 (Groisman, Knight, and Karl 2012). Kunkel et al. (2013) used a different measure to estimate changes in extreme rainfall, namely, the 20-year daily rainfall event, which is the amount of daily rainfall that has a 1 in 20 chance of being exceeded in any year. They found that from 1948 to 2010, about 75% of all weather stations across the U.S. experienced increases in the daily, 20-year rainfall event (Kunkel et

al. 2012). In comparison, Degaetano (2009) evaluated the parameters of the extreme value distribution, which are used to calculate the probability of any extreme rainfall event (like the 20-year storm) (Coles 2001; Katz 2013; DeGaetano 2009; Arnbjerg-Nielsen et al. 2013; Wi et al. 2015)). He found a significant upward trend in the distribution from 1950 to 2007 across numerous rain gauges in the Northeastern, Midwestern and Northwestern U.S, confirming that extreme rainfall in the U.S. has been changing significantly over time.

In addition to changes in extreme rainfall, flooding and damages are occurring more often, leading to an increase in the number of disasters declared by the Federal Emergency Management Agency (FEMA). From 1960 to 2008, there were approximately 165 disaster declarations per year related to severe storms. Over the past decade, this annual number has nearly quadrupled. There have been approximately 620 severe storm declarations per year since 2008. These numbers do not include disaster declarations relating to flooding and hurricanes; however, those numbers have also increased considerably. Figure 1.1 presents the number of disasters declared by FEMA since 1960 by type: flood (blue), hurricane (green), and severe storms (red). It is clear from the figure that the total number of disaster declarations is increasing over time, as well as the proportion of declarations related to severe storms. Disasters declared in 2017 and 2018 are not shown in the figure due to lack of updated data. However, the number of declarations related to Hurricane Harvey (in Texas, 2017), Hurricane Irma (in Florida, 2017), Hurricane Maria (in Puerto Rico, 2017), and Tropical Storm Florence (in North and South Carolina, 2018) are expected to be considerable.

Extreme rainfall storms are increasing because as global temperatures rise (as a result of increases in greenhouse gas emissions to the atmosphere and other feedbacks), water vapor also increases, which leads to more water available for precipitation (Soden et al. 2005; Allan and Soden 2008; Allen and Ingram 2002; Karl and Trenberth 2003). This phenomenon could lead to changes in the number of storms, the amount of rain per storm, and/or the total rainfall in a given period (e.g., annual rainfall). In some regions, increases in heavy and very heavy rainfall events have occurred along with a decrease in rain events (and an increase in
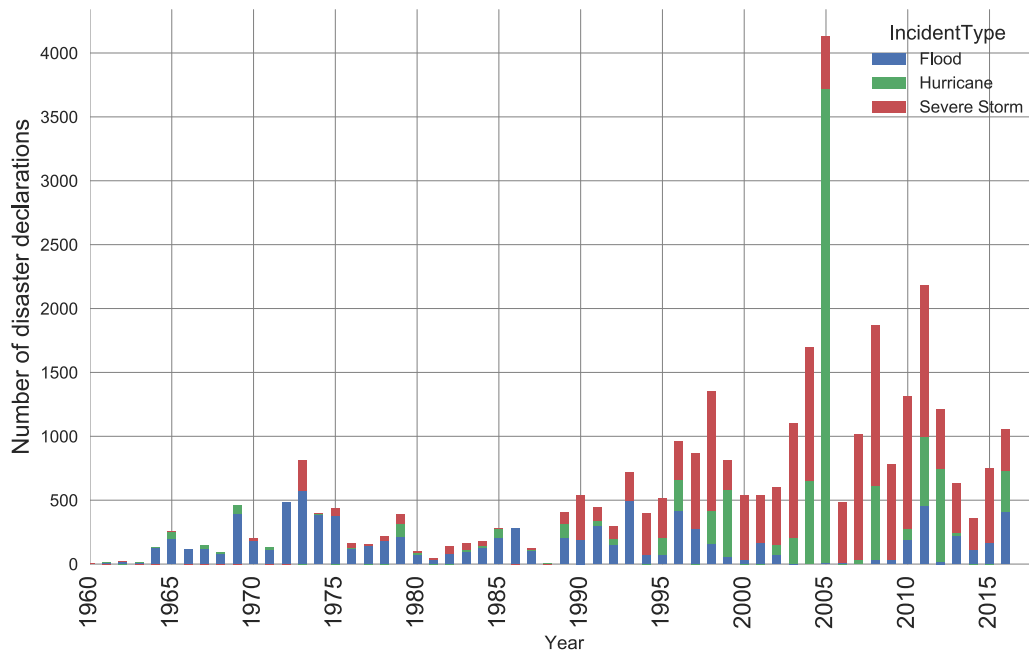
Figure 1.1: Number of disasters declared by the U.S. Federal Emergency Management Agency (FEMA) since 1960. Three types of disasters are shown: flood (blue), hurricanes (green) and severe storms (red). Data was retrieved from fema.gov; however FEMA and the Federal Government cannot vouch for the data or analyses derived from these data after the data have been retrieved from the Agency's website.

the number of consecutive dry days). This can lead to increased extremes with no change, or a decrease, in total annual rainfall (Groisman et al. 2005).

Downscaled climate models, which create simulations of the past and future climate on a small spatial and temporal scale, predict disproportionate increases in days with heavy rainfall across North America; the number of days with rainfall greater than the 95th percentile of all rain days is expected to increase by nearly 30% (Easterling et al. 2017). The daily, 20-year return period storm (or the storm that occurs, on average, every 20-years) is expected to increase by 13 to 22% under the high greenhouse gas emissions scenario by late century (Easterling et al. 2017). The largest increases are expected in shorter duration events (less than a day)(Westra et al. 2014; Kuo, Gan, and Gizaw 2015). Hourly extreme precipitation events are expected to increase significantly – by as much as 400% in North America (Prein et al. 2016).

1.2 Design and Function of Stormwater Systems

A major concern for civil and environmental engineers is how these changes in rainfall are affecting the function of stormwater systems and will affect them in the future. Stormwater systems are designed to carry or store a specific intensity of rainfall — the rate of precipitation depth per unit time (e.g., mm/hr). Further, when these design conditions are exceeded, many stormwater systems are designed to overflow into the environment, in a controlled failure that can cause inundation (e.g., street flooding) and sewer discharges to waterways. The intensity of rainfall used in design of these systems is called the "design storm;" it is a precipitation "event" that is characterized by its probability of occurrence (return period) and its duration. For example, a stormwater pipe could be designed to convey the "10 year-1 hour storm," indicating its capacity is selected to convey the amount of water that would reach it during one hour from a storm with a volume of precipitation that is likely to occur only once in 10 years. In general, stormwater structures are sized to convey rare rainfall events with a low probability of occurrence (i.e., less than a 50% chance of occurring in any year). However, the specific design return period is selected by stakeholders based on the acceptable risk level for a design to fail (causing inundation or overflow), as well as the available investment funding.

Rainfall intensities and their associated probability of occurrence are standardized in practice in the form of intensity-duration-frequency (IDF) curves (CSA Standards 2012; Bonnin et al. 2006; PennDOT 2011). These curves are typically created from statistical analysis of long records of observed rainfall data. The use of IDF curves created in this way means that current stormwater infrastructure has been designed using rainfall information that is based on assumptions that future rainfall will be similar to past rainfall (i.e., the assumption of a stationary climate). Changes in rainfall due to climate change are expected to decrease the performance of existing stormwater infrastructure in regions where the structures will expe-

rience extreme rainfall conditions that are outside the bounds of their historical design parameters. Furthermore, the performance of new stormwater installations will degrade over time, as the climate continues to become more extreme. To improve future performance and resilience of stormwater infrastructure, the design process must be updated to account for these changes, and infrastructure may need to adapt over time (Milly et al. 2008; Barros 2006; Karsten Arnbjerg-Nielsen 2011; K. Arnbjerg-Nielsen et al. 2013; Rootzèn and Katz 2013; Infield, Abunnasr, and Ryan 2018).

Figure 1.2 presents the steps involved in the current stormwater design process that may require modification. The first panel of the figure presents design inputs. In addition to standardized precipitation information in the form of IDF curves, the current stormwater design process also requires watershed characteristics (e.g., land use and drainage area of the site (McCuen 2005)). The combination of precipitation and watershed inputs are then used to calculate the maximum flow rate that a particular site would receive without stormwater infrastructure in place (i.e., the peak runoff discharge), as well as the amount of time it would take for this peak to occur (i.e., the time to peak). This is shown in the second panel. In highly urbanized watersheds, the time to peak can occur very quickly since impervious surfaces do not allow for infiltration or storage, and water moves quickly to the outlet (lowest point) of the watershed.

Conventional stormwater infrastructure, sometimes called "grey" infrastructure to distinguish it from infrastructure containing plants ("green") was built in urbanized watersheds to convey runoff quickly off-site and reduce localized flooding. These structures (e.g. inlets and storm sewers) are sized based on the rate of runoff discharge that needs to be conveyed off-site (top image of panel 3 in Figure 1.2). To avoid back-ups and overflows of the collection system, storage structures, like detention basins, or control structures, like weirs and orifices, can be built on-site so runoff is gradually released into the sewer system. This process decreases the peak discharge rate, and the risk that sewers will back-up and flood roadways and nearby buildings. Instead of temporary storage and release back to the sewer, stormwa-
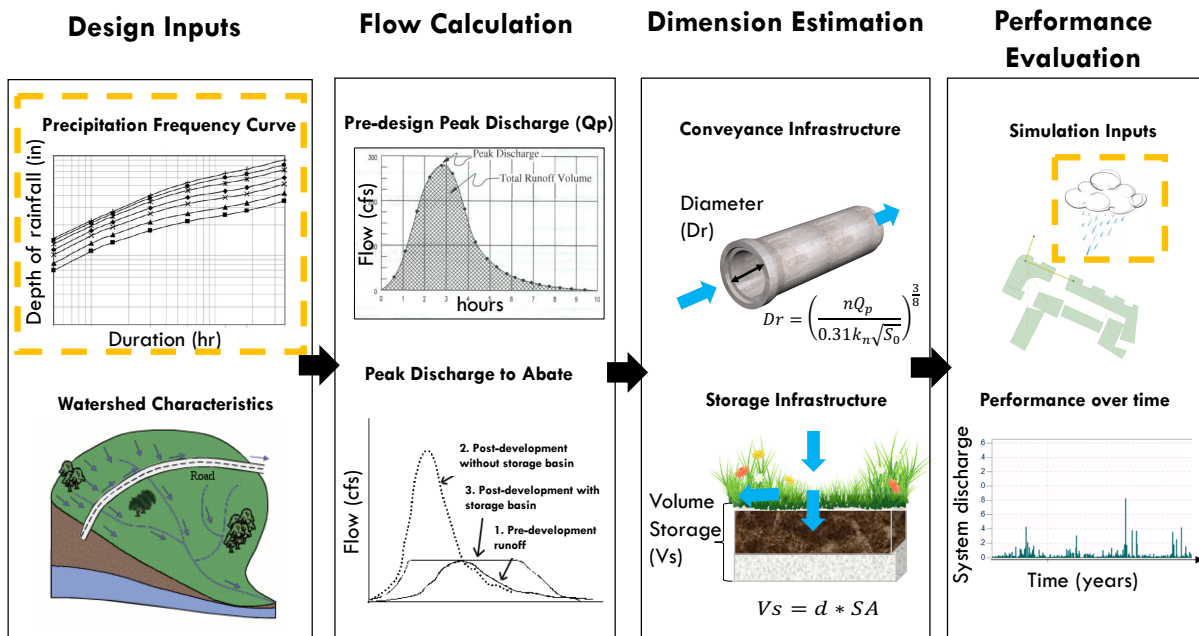
Figure 1.2: Overview of the suggested revised stormwater design process. Yellow, dashed boxes represent places where climate model output can be used to update design

ter infrastructure can also be designed to allow water to infiltrate on-site. Green stormwater infrastructure, which includes bio-retention basins (or rain gardens), bio-swales, permeable pavers and green roofs, is intended to mimic or restore natural hydrologic processes within the built environment (U.S. EPA 2016). The dimensions of infrastructure designed to store or infiltrate water on-site are determined based on the total volume of runoff under the discharge curve (see panel 2 in Figure 1.2). Thus the depth of a rain garden is determined by the available surface area (SA) and the volume of discharge (V_s) that would need to enter the basin during the design storm (bottom image of panel 3 in Figure 1.2).

1.3 Adapting Stormwater Systems to Incorporate Climate Change Trends

Many communities are interested in increasing the resilience of infrastructure to a changing climate, which may require the use of robust and flexible designs that incorporate many possible future states (Lempert 2010; Gregersen and Arnbjerg-Nielsen 2012; Lempert and Schlesinger 2001; Lempert and Groves 2010). Designing for robustness means minimizing regret against a range of possible futures, in order to ensure good performance throughout the lifetime as conditions change (Lempert, Popper, and Bankes 2010). This could mean potentially over-designing early on, which may be desirable for replacement of grey infrastructure systems. These systems require substantial capital investments and are expected to last for 50-years or more. Alternately, designing for flexibility involves replacing, adapting, or adding installations as time goes on (Adger, Arnell, and Tompkins 2005; Spiller et al. 2015). Green infrastructure, which consists primarily of soil, rocks, and plants, could be more easily expanded or added over time. In order for infrastructure systems to be resilient or adaptive as climate changes, current hydrologic design methods must be revised to include robust and adaptive design techniques.

Strategies that are informed by engineering and climate science are needed to ensure all communities have the information they will need to make good choices in rehabilitation and protection of their infrastructure. Tools like high resolution climate models can be used to provide insight into potential future conditions (Musau, Sang, and Gathenya 2013; Fowler, Blenkinsop, and Tebaldi 2007; Cooney 2012); however, different models rarely agree on the amount of future change. Different models can even disagree on the direction (e.g. more or less rainfall expected in a region, and not just the magnitude), a phenomenon classified as "deep uncertainty" in model results (Allen et al. 2000; Walker, Lempert, and Kwakkel 2013; Kirtman et al. 2013). Furthermore, climate model output is seldom in a format (e.g. at the

appropriate spatial or temporal resolution) that is useful to engineers and planners (Goorè Bi et al. 2017).

Translational work between the climate and engineering communities is needed in order to adequately utilize climate model outputs and update stormwater decision making techniques to incorporate uncertainty. In its absence, most engineering standards have not been updated to account for uncertainty in long-term climate conditions. Government agencies have differing ad hoc approaches and information availability; and there remains a large gap in the fundamental research to inform the practitioners who are making decisions for long-lived infrastructure (Olsen et al. 2015; Larkin et al. 2015; Lee and Ellingwood 2017; GAO (General Accountability Office) 2015; Linkov et al. 2014; Viner and Howarth 2014; Bocchini et al. 2014; Kennedy and Corfee-Morlot 2013; Francis and Bekera 2014; Wilbanks, Fernandez, and Allen 2015; Lempert 2013; Groves and Lempert 2007).

Updating the standardized precipitation information used by stormwater engineers is one way that the design process can incorporate expected increases in extreme rainfall events. Intensity-duration-frequency curves can be updated using climate change projections; however, different climate models can lead to different prediction results (Sarr et al. 2015; Schmidli et al. 2007; M'Po et al. 2016). Furthermore, there are many methods available for updating, and a consensus has not yet been reached on the preferred method. Inconsistencies and uncertainty in IDF curves could lead to variation in stormwater designs that may change the level of protection afforded by the infrastructure. It is possible that safety-margins built into the design process will outweigh these uncertainties, leading to no change in dimensions of the design. However, it is also possible that increases to design storms informed by climate models are so high that systems designed with these values are cost and space prohibitive. These uncertainties need to be investigated before updated IDF curves can be used with confidence to inform design.

In addition to updating design to consider possible future climate conditions, engineers must also evaluate how climate change will alter as-built infrastructure performance. It may

no longer be sufficient to assume that the point estimate used in design will result in infrastructure that performs well in an uncertain future, particularly if that future includes more extreme events. Simulating rainfall/runoff interactions during the design process can help to determine if the proposed structure will meet desired performance goals or metrics (see Figure 1.2 panel 4). Hydrologic simulation models (e.g. EPA SWMM (USEPA 2015)) simulate physical system response over time (Durrans et al. 1999) and can be used in conjunction with observed rainfall data or simulated rainfall data produced with downscaled climate models. After designs are built, it may also be possible to use simulation methods to determine how or when systems are expected to fail and how adaptation planning should be undertaken. Performance metrics and rainfall measures could be tracked over time, and adaptation triggered when thresholds are surpassed.

1.4   Research Objectives

The goal of this dissertation is to demonstrate how to use available information about climate change to increase the resilience and performance of urban drainage designs under uncertain future conditions. The stormwater design process can be updated in two main ways to improve the resilience of the designed infrastructure: first, at beginning of the design process, by selecting an updated value of expected precipitation from an updated regional IDF curve, and second, towards the end of the design process, by evaluating designs against a range of futures using specific performance metrics. Output from climate models is recommended for use in both of these strategies; however, this requires methods to present the climate data at the right temporal and spatial resolution and in a format familiar to the design engineer. The objectives of this thesis are to:

1. Develop a general framework for incorporating climate model output into engineering analyses and apply this framework to the updating of precipitation frequency curves

(a) What process should be followed in order to choose and utilize climate data in engineering applications?

(b) How can this process be applied to the updating of IDF curves?

2. Evaluate the effects of modeling choices on updated precipitation frequency curves and grey stormwater infrastructure

(a) Is one spatial adjustment method preferred over another?

(b) How does the spatial resolution of the climate model affect design dimensions?

(c) Does uncertainty in IDF curves outweigh design safety factors or vice versa?

3. Assess how rainfall measures, also called indices, and performance metrics can be used to evaluate performance and adaptation of green infrastructure systems under climate change

(a) Can rainfall indices be used to track performance?

(b) If so, how can this process be used to predict when to adapt?

1.5   Structure of Dissertation

This dissertation is made up of five chapters, including an introduction, three research papers, and a conclusion. Chapter 1, the introduction (and current chapter) provides motivation for the research topic, and background information. Chapter 2, which has already been published in 2017 in the Journal of Infrastructure Systems[1], presents a framework for the use of climate model output in engineering analyses and applies this framework to the updating of precipitation frequency curves. Chapter 3 investigates how modeling choices made during the updating of precipitation frequency curves can alter design storm values and the dimen-

---

[1] Cook, L., Anderson, C.J., and Samaras, C. (2017) A Framework for Incorporating Downscaled Climate Output into Existing Engineering Methods: Application to Precipitation Frequency Curves. Journal of Infrastructure Systems. 23 (4). doi:10.1061/(ASCE)IS.1943-555X.0000382

sions of stormwater conveyance pipes. Chapter 4 presents a method that uses continuous hydrologic simulation and rainfall indices to inform how and when adaptation should occur in green infrastructure systems. Chapter 5 summarizes major findings of the dissertation and suggests potential future work.

CHAPTER 2   A FRAMEWORK FOR INCORPORATING DOWNSCALED CLIMATE MODEL OUTPUT INTO EXISTING ENGINEERING METHODS: APPLICATION TO PRECIPITATION FREQUENCY CURVES [1]

Abstract

To improve the resiliency of designs, particularly for long-lived infrastructure, current engineering practice must be updated to incorporate a range of future climate conditions that are likely to be different from the past. However, a considerable mismatch exists between climate model outputs and the data inputs needed for engineering designs. The present work provides a framework for incorporating climate trends into design standards and applications, including selecting the appropriate climate model source based on the intended application, understanding model performance and uncertainties, addressing differences in temporal and spatial scales, and interpreting results for engineering design. The framework is illustrated through an application to depth-duration-frequency curves, which are commonly used in stormwater design. A change factor method is used to update the curves used in a case study of Pittsburgh, PA. Extreme precipitation depth is expected to increase in the future for Pittsburgh for all return periods and durations examined, requiring revised standards and designs. Doubling the return period and using historical, stationary values may enable adequate design for short duration storms; however, this method is shown to be insufficient to enable protective designs for larger duration storms.

Keywords

2.1 Introduction

Record-breaking rainfall has triggered more than 20 severe flood events in parts of Texas, Oklahoma, Louisiana, Arkansas, Missouri, Iowa, Florida, North Carolina, and South Carolina in 2015 and 2016. These events have led to the closure of two airports, flooding of more than 200 homes, numerous evacuations, cars stalled in high water requiring rescue, and deadly flash flooding.High water also led to spillway activation to protect New Orleans, as well as structural failure of more than 100 roads and retaining walls (Erdman 2016). Existing infrastructure systems were inadequate to deal with these events, which occurred outside of historical experience frequency. Design standards rely on historical observations and the assumption that climate is stationary (i.e., climate will not change over time). However, recent events and numerous simulations of future climate conditions indicate that the past is no longer a reliable indicator of the conditions under which infrastructure will have to perform in the future (Walsh et al. 2014; Milly et al. 2008).

Climate change has the potential to affect infrastructure systems in multiple ways, including: (i) changes in average and/or extreme temperatures; (ii) variations in frequencies, intensities, and duration of precipitation causing extreme rainfall and flooding in some regions; (iii) changes in storm tracks and severe weather; (iv) an increase in sea levels and the risk of storm surge; and (v) a decrease of water availability in some areas (Walsh et al. 2014; IPCC 2014; Kilgore et al. 2016).

Recently, increased attention has been directed to infrastructure reliability (the ability of systems to remain functional during a disaster) and resiliency (the ability to resist, absorb, and adapt to disruptions) (Faturechi and Miller-Hooks 2014). To ensure reliable and resilient infrastructure, engineering design standards must account for anticipated future conditions (Milly et al. 2008; Olsen 2015; Mailhot and Duchesne 2009; Moss et al. 2013; A. Barros and Evans 1997). These standards are set by organizations such as the American Society of Civil

Engineers (ASCE and ASCE 2013), agencies such as the Federal Highway Administration (Kilgore et al. 2016) or the National Oceanic and Atmospheric Administration (NOAA) (Bonnin et al. 2006), or by collaborations among organizations (e.g., 10 States Standards (Wastewater Committee of the Great Lakes - Upper Mississippi River 2014)).

One of the most advanced tools available to decision makers seeking to increase reliability and resilience of infrastructure is the use of high-resolution, or "downscaled," climate models. Compared to general circulation models (GCMs) that simulate global climate systems, these downscaled models provide insight into localized conditions by generating finer-scale (4 – 50 km), future projections of air temperature, precipitation, evapotranspiration, wind speed, and other factors that affect regional patterns (Musau et al. 2013). Most models agree on the direction of temperature change; however, for precipitation there are variations in trend and magnitude across models and geographic regions, leading to large uncertainty in results. For precipitation data particularly there is often a mismatch in the spatial and temporal resolution of the downscaled climate model and the micro scale (e.g., < 1 km) of inputs needed for engineering design standards and applications. Furthermore, the use of climate models introduces uncertainties and complicates data extraction and preparation requirements, compared to the current use of recorded historical data. A clear path from climate model predictions to development of updated design standards is needed.

Despite these challenges, by building on historical observations, scientists have successfully used global and downscaled climate models to inform higher spatial and temporal resolution precipitation trends for engineering applications. Weather generators, which can be adapted to different anticipated changes in climate, have been used to simulate synthetic, rainfall time series at the station (point) scale at monthly, daily and hourly time steps (Wilks and Wilby 1999; Kilsby et al. 2007; Willems et al. 2013). Quantile mapping has been used to apply expected changes to the empirical distribution of observed rainfall events at the temporal and spatial resolution of the observations (Laflamme et al. 2016; Boé et al. 2007; Gudmundsson et al. 2012; Wood et al. 2004). Numerous studies utilize a "delta" or "change

factor" technique, which applies the expected absolute (delta) or relative (ratio) change between current and future gridded projections to historical rainfall data (Wilks and Wilby 1999; Boé et al. 2007; Wood et al. 2004; K. Arnbjerg-Nielsen et al. 2013; Forsee and Ahmad 2011). These climate-informed local-scale models have also been used to update intensity-duration-frequency curves used in design of infrastructure affected by rainfall (Chandra et al. 2015; Cheng and AghaKouchak 2014; Forsee and Ahmad 2011; Zhu 2012; Kuo et al. 2015; Hassanzadeh et al. 2013; Mirhosseini et al. 2013).

Applications of these climate-informed methods can provide important insights; however, many reported studies provide insufficient detail regarding the importance and difficulty of obtaining a reliable historical record; selecting and extracting the appropriate climate output source; accounting for reliability and uncertainty in climate modeling; and incorporating findings into infrastructure planning and design. In the absence of a consensus on methods to update design standards to account for climate change, many stakeholders avoid the use of climate model output. Further, there is the potential for misuse through simplified choices, such as using output from a single climate model instead of an ensemble (or group) of models, or failing to account for model reliability and uncertainty in the interpretation of results. Given the widespread use of infrastructure design standards and the potential consequences to the public if they are improperly applied (including failure due to under-design or misallocation of taxpayer dollars due to overdesign), it is critical that the most advanced and appropriate methods are used to update standards and that the challenges and limitations associated with this updating are well understood by those who will apply these techniques.

With this problem in mind, a five-step framework is proposed that can guide the revision of design standards, as well as engineering practice, through the use of publicly available, downscaled climate model outputs of future precipitation. By applying the framework, engineers will be able to define relevant aspects of the historical method that need to be updated; select the relevant climate data sources and extract output; manage model reliability and bound uncertainty; adjust for spatial and temporal resolution, and apply results to engi-

neering design under climate non-stationarity. In the present work, as a demonstration, the framework is applied to a common input to stormwater design: depth-duration-frequency (DDF) curves.

2.2   Framework Steps

The steps of the framework for updating engineering design standards are: (0) Define the existing design standard (or application) that is based on or incorporates precipitation data; (1) Understand the historical basis and data requirements used to develop the existing standard (or application) and retrieve data; (2) Access appropriate climate model output based on requirements for the existing standard (or application); (3) Account for climate model uncertainty and reliability; (4) Adjust existing method to incorporate expected future trends; and (5) Interpret results and incorporate changes into design practice. A flow chart of these steps is presented in Figure 2.1. Solid arrows display the suggested sequence of the steps from 0 to 5; dashed arrows represent the flow of information or data from step 1 to step 4.

2.2.1   Step 0. Define the existing design standard (or application) that is based on or incorporates precipitation data

Standards for engineering design have been developed for a variety of engineering applications that are expected to be affected by a non-stationary climate, including: water supply management, water quality regulations, flood forecasting, stormwater management, and wastewater collection and treatment. These and many other applications rely on different types of estimates of expected precipitation for a region. For example, floodplain delineation and stormwater management rely on duration-specific estimates of rainfall depth from intensity-
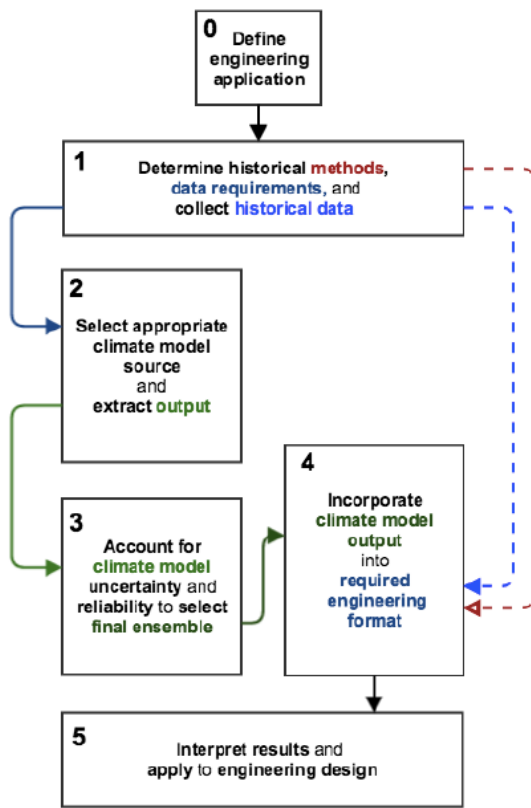
Figure 2.1: Flowchart of the framework for incorporating downscaled climate data into existing engineering applications.

duration-frequency (IDF) curves; whereas wastewater collection and treatment system design requires a peaking factor, usually relating to maximum daily or monthly rainfall. Some applications may require a time series for precipitation (a sequence of data points collected over a time period, usually provided at evenly spaced time intervals). These different specific precipitation data determine the type of modifications that will be needed, and thus, defining the ways that rainfall data are used in the current standard is the initial step to updating.

2.2.2 Step 1. Understand historical basis and data requirements used for existing standard (or application) and retrieve data

In this step, the engineer defines the nature of the data on which the design standard was based and obtains the historical data to enable re-creation of the supporting calculations underlying the current standard. Definition of the data includes the length of record, as well as the temporal and spatial resolution of the data required as input to re-create the components of the method. These specifications are important as they dictate the source of climate output that will be needed. Downscaled climate model outputs are only publically available at specific spatial and temporal resolutions (refer to step 2), which may not be consistent with the resolution of the inputs required for the engineering application or standard. Model outputs are also only provided for specific historical and future dates, and the length of this simulation period may not be equivalent to the length of rainfall record (e.g., 50 - 100 years) utilized in some methods to inform standards. Table 2.1 provides information on spatial and temporal resolutions required for analysis of several types of engineering applications and design standards. Large-scale optimization models used for reservoir or drought management use monthly or seasonal data, while stream flow and water quality simulations require daily or hourly data. Continuous hydrologic simulation models (e.g., EPA SWMM (USEPA 2015)) use data at a sub-hourly time step and 1 km spatial resolution (Wood et al. 2000; Wilks and Wilby 1999), while IDF curves use multi-decadal time series of observed rainfall at individual geographic locations (point measurements) for durations ranging from 5 minutes to 72 hours (CSA 2012; Bonnin et al. 2006).

In support of these different data needs, historical precipitation data, collected through rain gauges, can be obtained at the point or grid scale. Many regional airports and local stormwater agencies collect rain gauge data at specific locations (points), at hourly intervals

Table 2.1: Select design standard applications and the requirements of precipitation data inputs (format, temporal and spatial)

Engineering application	Example Analysis	Format	Data Requirements Spatial	Temporal
Water supply management	Reservoir routing, simulation and optimization (e.g. STELLA)	Timeseries	Basin scale (e.g. $1000\,km^2$) to point scale ($<1km^2$)	Seasonal to daily
Water quality monitoring	Water Quality Analysis Simulation (e.g. streamflow simulation)	Timeseries	Riverine scale (e.g. $10\,km^2$)	Seasonal to hourly
Flood forecasting	Continuous hydrologic simulation	Timeseries	City scale (e.g. $1\,km^2$ or less)	Sub-hourly
Flood plain delineation	Event based simulation (e.g., HECRAS)	IDF Curve	City scale (e.g. $1\,km^2$ or less)	Hourly to sub-hourly
Stormwater management planning	Peak discharge estimation (e.g. Rational method, TR-55)	IDF Curve	City scale (e.g. $1\,km^2$ or less)	Daily to sub-hourly
Wastewater collection and treatment planning	Wet weather flow estimation (e.g. Peaking factor)	Monthly maximum	Basin scale (e.g. $1000\,km^2$) to point scale ($<1km^2$)	Monthly

or less. Airport records tend to be longest (50 to 100 years); however, local agency data may be available at higher resolution (sub-hourly, multiple gauge sites) for a shorter time period. Barros (2006) suggested that the usefulness of a rain gauge network is dependent on the density of gauges, the number of years of data, the type of rain gauge, and frequency of data collection (Barros 2006). NOAA National Centers for Environmental Information (NOAA 2016) provides rainfall data at the point scale, often at hourly intervals.

Gridded rainfall data are also available using two methods. The first approach, which produces rainfall grids through interpolation of point measurements, is based entirely on the assumption of spatial correlation of rainfall point measurements. Thus, the accuracy is dependent on the spatial density of rain gauges and how terrain influences the correlation of precipitation measurements. The second method, called data assimilation, uses weather models to infer spatial correlation and temporal evolution of rainfall, and then when combined with point measurements, adjusts model-predicted rainfall toward observed values. By systematically merging numerous observations (available at different resolutions from rain gauges, satellites, or radar) with weather model output, data assimilation creates gridded precipitation data that is uniform and consistent with simulated weather conditions. These assimilation data sets are called "re-analysis" data. The quality of gridded data, especially relating to precipitation and extremes, is variable by location and time period, due to the changing combination of observation density and quality as well as model bias (Dee et al. 2011; Dee et al. 2016; Bosilovich et al. 2008; Sun and Barros 2014). Reanalysis data are publicly available for multiple temporal and spatial resolutions for the North American domain (see Table 2.2). Additional information can be found at the University Corporation for Atmospheric Research (UCAR) Climate Data Guide website (Dee et al. 2016) or at the reanalysis site maintained by the University of Colorado at Boulder (Reanalysis.org 2016).

Table 2.2: Temporal and spatial characteristics of select sources of re-analyzed observations for North American domains

Data Source	Origin	Spatial scale	Time-step	Period of record	Website
NCEP North American Regional Reanalysis (NARR)	NOAA Earth System Research Laboratory	32-km	3-hour, daily, monthly	1979 to 2015	https://www. ncdc.noaa.gov /data-access
Observation-reanalysis hybrid dataset	Princeton University	1, 1/2, & 1/4 degree (110, 55, 25 km)	3-hour, daily, monthly	1949 to 2010	http:// hydrology. princeton.edu /data.pgf.php
ERA-Interim Global Atmospheric Reanalysis	ECMWF	80-km	daily, monthly	1979 to present	http:// apps.ecmwf. int/datasets
Parameter-elevation Regressions on Independent Slopes Model (PRISM)	Oregon State University (USDA, NOAA)	4-km	daily	1895 to present	http://www. prism. oregonstate.edu
Variable Infiltration Capacity (VIC) Hydrologic Model	University of Washington	1/8 degree (12 km)	daily	1950 to 2000	http://www. esrl.noaa.gov

2.2.3   Step 2. Access appropriate climate model output based on requirements for the existing standard (or application)

Open source downscaled model output is becoming increasingly prevalent and diverse as data sources continue to emerge and climate models continue to evolve. These outputs are created using "downscaling models" that reduce the coarse, spatial-resolution of the global atmospheric general circulation models (GCMs) that are used as input (Cooney 2012; Di Luca et al. 2015; McGuffie and Henderson-Sellers 2001). Sources of downscaled output provide higher spatial-resolution (4–50 km) outputs than GCMs (75–250 km); however, characteristics of the output differ as a result of three main factors: (1) choices made by climate scientists in the downscaling process, including the downscaling method and the number of general circulation models (GCMs) and emissions scenarios used; (2) consequences of the computational power and data storage that were available to the climate modelers, affecting length of simulation, and temporal and spatial resolution, which are aspects most relevant to engineers; and (3) decisions made by the climate modelers to store and allow access to the data. The next step in the framework provides context and information to allow an engineer to select and extract downscaled data, most suitable to the engineering application, from the myriad of available sources.

Figure 2.2 presents a comparison of characteristics of six publicly available sources of downscaled climate model output for North America, which include:

1. the North American Regional Climate Change Assessment Program (NARCCAP) [narccap.ucar.edu] (Mearns et al. 2008);

2. the North American Coordinated Regional Climate Downscaling Experiment (NA-CORDEX)

[na-cordex.org] (Mearns, et al. 2013);

3. the USGS regional climate model (RegCM3) [regclim.coas.oregonstate.edu] (Hostetler et al. 2011);

4. Eighth degree-CONUS Statistical Asynchronous Regional Regression Daily Downscaled Climate Projections (ARRM) [cida.usgs.gov] (Stoner et al. 2013);

5. The "Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections" from the Bureau of Reclamation (and other collaborators) [gdo-dcp.ucllnl.org] (Brekke et al. 2013); and (vi) Multivariate Adapted Constructed Analogs (MACA) [maca.northwestknowledge.net] (Abatzoglou and Brown 2012).

As discussed above, there are three categories of differences in model output provided from these sources, shown in Figure 2.2, as the first group (model simulation choices), the second group (model output attributes), and the third group (extraction and access features). The figure may be approached from top to bottom, to select a downscaled model source based on desired characteristics of the output; or from left to right, in order to understand the available data characteristics for a particular source. A climate modeler begins the downscaling process from far left column (selection of the global climate model group) then advances to the far right (providing access to output). An engineer often begins the process of using climate model output at the far right (attempting to access the output) and subsequently makes choices from right to left.

The process of selecting an appropriate model source should begin with the characteristics of the model output that are most relevant to the engineering application (referred to above as the second group of characteristics in Figure 2.2), such as the time step, or size of the grid cell. For many water resources applications, the most limiting of these characteristics is the temporal resolution. Daily data is available from all sources; however, there are only two sources of sub-daily data (3-hour increment) available: NARCCAP and NA-CORDEX. These sources are able to provide a finer temporal resolution because they use a technique called

Source of output	Model Simulation Choices				Model Output Attributes				Extraction and Access Features			
	Global climate model group	Emission scenarios	Nb. model simulations (does not include same model multiple runs)	Downscaling method	Spatial resolution	Historical period	Future period	Time step	Units	Date format	File format	Data access
(legend)	CMIP3 / CMIP5	SRES / RCP		Dynamical / Statistical	4 km / 6 km / 12 km / 15 km / 25 km / 50 km	1950 / 1958 / 1965 / 1973 / 1980 / 1988 / 1995 / 2000	2000 / 2015 / 2030 / 2045 / 2060 / 2075 / 2090 / 2100	3 hr / daily / monthly / decadal	flux (kg/m2/s) / depth (mm/day)	calendar / fixed	csv / netCDF	user interface / server download
NARCCAP	CMIP3	A2	12	Dynamical	50 km	1970–2000	2040–2070	3 hr	flux	360 dpy	netCDF	server download
NA-CORDEX	CMIP5	4.5, 8.5 (RCP)	12	Dynamical	25 km, 50 km	1950–2000	2099	3 hr, daily, monthly	flux	360 dpy	netCDF	server download
RegCM3	CMIP3	A2	3	Dynamical	15 km	1960–2000	2020	daily, monthly, decadal	flux	365 dpy	netCDF	user interface
ARRM	CMIP3	B1, A1, A1B, A2	16	Statistical	12 km	1960–2000	2099	daily	depth	calendar	csv	
Reclamation	CMIP3, CMIP5	B1, 2.6, 4.5, 6.0, A1B, 8.5, A2	20+	Statistical	12 km	1950–2000	2000–2099	daily, monthly	flux, depth	calendar	csv, netCDF	user interface
MACA	CMIP5	4.5, 8.5 (RCP)	20+	Statistical	4 km, 6 km	1950–2000	2099	daily	flux, depth	calendar	csv, netCDF	user interface

Figure 2.2: Comparison of characteristics of six publicly available sources of downscaled climate model output for North America. Headings in orange are choices made by climate scientists in the downscaling process; headings in yellow are characteristics of the simulations most relevant to engineering applications; and headings in green relate to manipulation of model output. A climate scientist approaches the figure from left to right; whereas an engineer reads from right to left.

"dynamical downscaling," which uses Regional Climate Models (RCMs) to provide physical characterization of weather processes occurring on small scale and contributing to precipitation (Anderson et al. 2003; Xu 1999; Musau et al. 2013). These models require large amounts of computational power and thus only limited scenarios (e.g., emissions) can be computed. For example, a 30 year, 50 km resolution simulation of the WRF regional model on a super-computer (with 240 processors) lasts for over 2 days; and a 150 year, 25 km simulation lasts for 90 days (Mearns et al. 2013). Prior to the release of NA-CORDEX in 2017, the requirement for a sub-daily time step restricted the user to NARCCAP data, a pioneering project in 2006 that compiled consistent output from numerous regional climate modelers across the globe.

Modeling scenarios were limited to a single emission scenario (SRES A2), a relatively low resolution (50 km), and short simulation periods (30 years); meaning the end-user did not have flexibility to select different characteristics. NA-CORDEX will provide longer simulation periods (1950 - 2100), a higher spatial resolution (25-km) and two emissions scenarios (RCP 4.5 and 8.5), providing more flexibility. When the engineering application is not limited to a sub-daily time step, additional resources are available at the daily level through sources that utilize "empirical downscaling" techniques, which rely on existing statistical relationships between large-scale climate systems and local weather patterns (Abatzoglou and Brown 2012; Khan et al. 2006; Murphy 1999; Chen et al. 2013). These techniques are less computationally intense than RCMs, and can provide higher resolution output (4 - 12 km) for long simulation periods (1950 - 2100), and multiple emissions scenarios and global climate models (Cooney 2012).

At the daily level, the user now has more flexibility to select a downscaled data source based on several data characteristics, including others that are relevant to engineering applications, like spatial resolution; as well as those in the first group in Figure 2.2, which are a result of choices made by climate scientists in the downscaling process, like sources of global models, emissions scenarios, or downscaling technique. The Coupled Model Intercomparison Project (CMIP) was established in 1995 as a standard experimental protocol to compare

outputs from general circulation models (Maurer et al. 2007). The sixth phase of the project (CMIP6) is currently in progress (Eyring et al. 2016); however, the downscaled output discussed in this manuscript originated from GCMs from the third (CMIP3) and/or fifth (CMIP5) phases of the project (Taylor et al. 2011). The GCMs used in CMIP3 are early generation models and many evaluations have been completed; whereas the GCMs used in CMIP5 are more experimental, and have a shorter record of development and evaluation. Comparison between CMIP3 and CMIP5 indicates minor differences in future, projection results (Reichler and Kim 2008), and a large majority of comparisons do not address engineering-specific metrics (Wuebbles et al. 2014). The Bureau of Reclamation dataset is the only source to provide output from both CMIP3 and CMIP5 models. A compelling reason to prefer one over the other for precipitation analysis has not been presented, which means engineers have flexibility in choosing from either data set or a combination thereof, depending on which is best suited for their application. One may wish to select a downscaled data source based on the downscaling technique (dynamical or empirical); however, there is no consensus relating to which technique is superior, since both have advantages and disadvantages (Prudhomme and Davies 2009; Fowler et al. 2007). Similar to the decision to choose between CMIP3 and 5, a single reason to prefer one downscaling technique to the other does not exist, and engineers should select a source suitable for the engineering problem rather than accessibility of a particular data archive.

Emission scenarios estimate the potential concentration of greenhouse gases (GHG) in the atmosphere, based on pathways of socio-economic, technological, and political factors. CMIP3 global models use emissions scenarios from the Special Report on Emissions Scenarios (SRES) (Nakicenvoic et al. 2000) created for the IPCC 3rd Assessment report (Houghton et al. 2001); whereas, CMIP5 global models use emissions in the form of Representative Concentration Pathways (RCPs) (van Vuuren et al. 2011) created for IPCC AR5 (IPCC 2012). When it is available to choose between multiple emissions trajectories, the authors recommend analyzing at least two scenarios when possible: (a) an upper bound that will provide the most

conservative estimate of future conditions for use in engineering practice, such as SRES A2 (projecting 2.0 - 5.1 °C of warming by 2100) or RCP 8.5 (5-6 °C by 2100), and (b) a lower bound that is aligned to targets of the Paris agreements (Framework Convention on Climate Change 2015), similar to SRES A1B and RCP 4.5. Choosing between emissions scenarios is most relevant for infrastructure of long lifetimes (40 years or more), since many impacts across emissions scenarios generally diverge after the middle of the century (Collins et al. 2013). Irrespective of the scenarios and model sources that are ultimately chosen by the engineer, it is important to document assumptions and make them available to those interpreting the findings.

Once a downscaled data source has been selected based on the desired characteristics (from Figure 2.2), the user should extract the output at the spatial grid closest to the geographical location of interest, for a historical simulation period (usually one for which the user has historical data from framework step 1) and a future simulation period (for dates and length required for the engineering application). The user should obtain output from several different climate-modeling organizations in order to create an "ensemble" of downscaled model output. The initial ensemble should include all model simulations available for the selected emissions scenario(s), in order to accurately assess model performance and uncertainty, which is discussed in step 3 of the framework.

To access the output, some sources provide a user interface, including the USGS, Bureau of Reclamation, and MACA. The USGS (RegCM3) and Bureau of Reclamation websites allow for selection of multiple grid cells and provide spatially averaged time series at different time steps. NARCCAP, NA-CORDEX, and the ARRM sources do not have the guidance of a user interface and data must be extracted as individual files from a server. These files are usually available in netCDF (network Common DataForm) format, and require software packages (available in Excel, MATLAB, R, python) and the x-y coordinates for the geographical location (from the netCDF data matrix) to obtain the time series of data. Precipitation stored in netCDF files is often provided as instantaneous flux values (in units of kg/m^2s), which is con-

verted to precipitation depth over a time period by dividing by the density of water (1,000 kg/m^2) and multiplying by the number of seconds in the time step.

2.2.4 Step 3. Account for climate model uncertainty and reliability

Step 3 of the framework relates to analyzing performance and uncertainty of the ensemble of models extracted from the downscaled data source (selected in step 2). Simulations from downscaled climate models are susceptible to large uncertainties, meaning models may not agree on the magnitude or direction of future change in rainfall. Uncertainty can be introduced in the downscaling process, as well as through the three sources of uncertainty inherited from the global climate models: (1) scenario uncertainty of future GHG emissions, (2) natural (internal) climate variability (initial conditions), and (3) inter-model discrepancies (modeling assumptions) (Kirtman et al. 2013).

The relative contributions to uncertainty from each source vary depending on the region and simulation year. For precipitation, internal variability contributes the most uncertainty in early years of the 21st century, whereas inter-model uncertainty makes up the largest majority after 2040 (Kirtman et al. 2013; Hawkins and Sutton 2011). Scenario uncertainty is managed through the use of multiple emissions scenarios, described in step 2. Internal conditions uncertainty is bounded by producing several simulations using different initial criteria (Knutti et al. 2009; Musau et al. 2013). Most sources of climate output, however, only provide output from a single simulation (or an average of multiple simulations) for each climate model, with the exception being the Bureau of Reclamation dataset. Finally, scientists recommend the use of multiple models, or a "model ensemble," in order to avoid misleading conclusions from inter-model uncertainty (Barsugli et al. 2013). Uncertainty introduced through the downscaling methods can be similar in magnitude to inter-model uncertainty (Chen et al. 2011), and

managed in the same way (using multiple models).

The number of models to include in the final ensemble will depend on the approach used to manage uncertainty. The various techniques include: the extremes approach, the ensemble approach, and the validation approach (Musau et al. 2013). The "extremes approach" examines the full range of future scenarios by using the full ensemble extracted in step 2. There are several drawbacks to using the extremes approach, including the time and effort expended to consider each model individually; and the stronger argument, which contends that some models may be less reliable than others at simulating the climate, and that considering the full range of models may produce an unrealistic representation of the future (Fowler et al. 2007). Those in favor of this argument suggest accounting for the reliability of the climate models when considering the ensemble.

Climate models considered by scientists to be "more reliable" include those that are well documented, are well established (with many years to make improvements), and that produce stable results. Models may be considered less reliable if they produce output that is "biased" with respect to the observed metric(s). Bias (also referred to as model systematic error) is defined, in this context, as the average deviation between the observed value (or empirical statistic) and the values or statistics obtained from the historical climate model simulations. The deviation may be larger than zero for numerous reasons, including the assumptions and simplifications made in the modeling equations (of the global, regional, and/or statistical models).

Bias can be overcome by changing internal modeling assumptions (although this often shifts bias in another direction) or through bias correction techniques applied to the simulated rainfall output. Bias is assessed through the comparison of observations to "hind-cast" model simulations (i.e., simulation of historical conditions) at the spatial and temporal resolution of the climate model (Gleckler et al. 2008). Poor performance with respect to historical conditions can be used to identify unrealistic models; however, adequate performance in hind-casting does not guarantee the accuracy or reliability of future predictions, . For re-

gional climate models, these hind-cast runs are driven by historical reanalysis data, which is different from simulating the past using simulated atmospheric conditions from GCM outputs. Hind-cast runs must be obtained separately from the simulations of the past that use GCMs.

Nearly all models exhibit some instances of bias; however, the magnitude varies depending on the metric, season, and models examined (Hall 2014). For example, NARCCAP regional climate models, which have not been bias-corrected, were able to estimate mean annual precipitation with relative precision (exhibiting low bias); however, extreme precipitation statistics (e.g., annual maxima, 20-year return period) were often overestimated. The Weather Research Forecasting Model (WRFP), from NARCCAP, exhibited especially high bias for extremes, as the percentage error of the average maxima precipitation and the 20-year return value was greater than 90% for nearly all seasons and US regions (Wehner 2013). A study that examined the 3-hour precipitation totals of five regional climate model predecessors to the NARCCAP models (excluding WRFP) found that the Iowa State model (MM5) and the Scripps model (ECP) outperformed the Hadley model (HIRHAM), the Regional Climate Model (RegCM2), and the Canadian model (CRCM) (Anderson et al. 2003).

Statistical downscaling techniques usually account and correct for bias in the downscaling process, thus it is more important to assess reliability of regional models. Outputs from the ARRM statistical downscaling method, which have been bias-corrected and cross-validated, were found to show improved accuracy at generating extremes and to be efficient and generalizable across regions (Stoner et al. 2013). For precipitation (and other variables), the Multivariate Constructed Analogs (MACA) method has been found to outperform the Bias-Corrected Spatial Disaggregation method (BCSD), used to create the Bureau of Reclamation dataset, due to the ability to jointly downscale certain variables (Abatzoglou and Brown 2012).

The remaining approaches to uncertainty can account for the relative reliability of climate model output. The "ensemble approach" uses a weighted average of the model results to develop a probability distribution of the range. There is no consensus on the minimum and

maximum number of models to consider in the distribution and more research is needed in this area (Mote et al. 2011). Yet, one study found that the relative skill of the ensemble average to reproduce historical conditions converged after 6 or more models were included (Pierce et al. 2009). Models in the ensemble can be weighted equally or based on criteria of reliability (of the regional model or global model). Nonetheless, Wehner (2013) found that averaging the ensemble using complicated weighting schemes was not more effective than simply removing the unreliable or poorly performing models.

Removing unrealistic models from the ensemble refers to the final approach to managing uncertainty, known as "the validation approach" or "culling the ensemble" (Charles et al. 1999; Flato et al. 2013; Mote et al. 2011). The ensemble statistics are calculated solely from models that are considered to be more reliable, based on performance criteria. Some studies have demonstrated that ranking the models based on performance leads to a difference in predictions (Gleckler et al., while others have shown that the differences due to model culling is slight (Mote et al. 2011). One study found that results for a precipitation metric were nearly indistinguishable between the average of the 11 best performing GCMs and 11 randomly selected GCMs, from the CMIP3 ensemble (Knutti et al. 2009). For engineers, however, culling may be an appropriate avenue in order to reduce ensemble size. While all uncertainty approaches provide utility, it must be highlighted that any estimation of uncertainty from a range of climate models will never provide perfect insight into the full spectrum of possible futures (Mote et al. 2011). The engineer should decide which approach is best suited to their application then clearly state all assumptions.

2.2.5   Step 4. Adjust existing method to incorporate expected future trends

After the ensemble of desired climate models has been selected, the precipitation data forecasted by these models for the relevant future time frame must be incorporated into the existing method for developing the design standard (or application). Since climate model output is provided at a "gridded" resolution (4 km or higher), this step will often require adjustment of the model output to an even finer spatial resolution (e.g., < 1 km2). Model output may also need adjustment temporally, to obtain a smaller time step. These adjustments are accomplished using additional downscaling or disaggregation techniques (Durrans et al. 1999) that depend upon the required format of the precipitation data needed to update the specific design standard. In order to adequately account for the range of uncertainty from the selected climate models, these further downscaling techniques should be applied individually to each model output before averaging and should not be applied to the average of the outputs, since this method filters data variation (Wehner 2013).

As discussed in step 1, sometimes the engineering application explicitly requires high-resolution time series (e.g., at the station scale, or at intervals smaller than 3 hourly), for analyses like streamflow simulation or flood forecasting. In this case, further statistical downscaling techniques must be applied at this point to the already downscaled model output that was selected in steps 2 and 3. Statistical downscaling methods include applying transfer functions, weather generators, weather typing, or quantile-mapping to the gridded, downscaled model output (Wood et al. 2004). Weather generators use empirical relationships calculated from observations to simulate synthetic time series for rainfall data (Andréasson et al. 2004; Chen et al. 2015). Weather typing, or resampling, involves relating the weather patterns of the larger scale climate model to observed patterns in the local area (Prudhomme et al. 2002; Onof and Arnbjerg-Nielsen 2009). Quantile mapping, also used for bias-correction, matches the empir-

ical quantiles of re-gridded historical data to those of the historical climate simulation, then adjusts the future climate simulation based on the difference between the historical data and simulation (Boé et al. 2007; Laflamme et al. 2016; Gudmundsson et al. 2012).

If the engineering design application does not require a high-resolution time series and is instead based on a statistical analysis of the observed data (which is the case for IDF curves), then it may not be necessary to modify the underlying time series using complex downscaling techniques. As an alternative, it is possible to incorporate the future trends from the gridded, climate model output, directly into the statistical methods used historically. Two methods have been employed to adjust statistical metrics: the change factor approach and the bias-correction approach. In the former, an observed statistic (usually at the point scale) is adjusted to a future date using a change factor that is calculated from the original, downscaled model output. Change factors are typically calculated as the percentage difference between historical and future model output, or as a ratio between the two values (Forsee and Ahmad 2011; Zhu 2012). In the bias-correction technique, the future, gridded value from the downscaled climate model output is modified based on the difference between the observed statistics (point scale) and past model simulation statistics (grid scale) or hind-cast simulation statistics (grid-scale) (Arnbjerg-Nielsen et al. 2013; Boé et al. 2007; Chen et al. 2015; Wilks and Wilby 1999; Wood et al. 2000). Quantile mapping may be employed as a bias-correction technique (Boé et al. 2007). Detailed methods to accomplish this are described in the example section for depth-duration-frequency curves.

2.2.6 Step 5. Interpret results and incorporate changes into design practice

At this point in the framework, the user should have been able to incorporate trends from an ensemble of downscaled climate model outputs into an existing design standard or engi-

neering application, producing a range of resulting scenarios. To make use of the range of results that incorporate future climate scenarios, current engineering practice must evolve to incorporate principles relating to uncertainty and risk. Uncertainty may be addressed using exhaustive or simplified approaches that build on the climate-model results. Robust Decision Making (RDM) is a technique based on principles of minimizing regret and achieving acceptable thresholds. RDM involves testing future designs against the full plausible range of futures obtained from step 4 (Lempert et al. 2006; Hallegatte 2009; Lempert 2013; Espinet et al. 2015). Though RDM is growing in popularity to address the challenge of uncertainty associated with climate change; these methods can be computationally intensive and may not be appropriate for all applications. Other approaches to addressing uncertainty include defining an acceptable risk level in order to select a design value or strategy from a range of possibilities (Karsten Arnbjerg-Nielsen 2011; Hallegatte 2009; Hallegatte 2014). When applying either method, decision-makers should favor strategies that are adaptable, have low regret, or are reversible (Hallegatte 2009; Olsen 2015).

In addition to uncertainty, engineering designs will also need to better incorporate principles of non-stationarity. This means addressing the fact that the infrastructure system is subject to one or more shifts in exogenous factors (climate, land-use, demand patterns) over the course of the operating lifetime (Kilgore et al. 2016). Best practices may include: (i) testing for non-stationary trends (before assuming them), (ii) explicitly defining the final year in the future that the structure is designed to operate to, with adequate performance, and (iii) defining what adequate performance means for each structure. The Mann-Kendal (MK) test can be used to detect a non-stationary trend in an underlying distribution (Cheng and AghaKouchak 2014; Katz 2013; DeGaetano 2009; Kilgore et al. 2016). If detected, non-stationarity can be addressed by expressing one or more parameters or variables as a function of time (Katz 2013); however, such factors must be calibrated and verified.

2.3   Application of Framework: Depth-Duration-Frequency Curves

The next section illustrates the framework as applied to rainfall duration frequency curves, using Pittsburgh, PA as a case study. The application focuses on updating the depth-duration-frequency (DDF) curves, which are a form of IDF curves that present rainfall as a depth (inches or mm), rather than an intensity (inches or mm per unit time). Each step describes the decisions made in order to update the curves to reflect future trends and uncertainties, following the framework.

2.3.1   Step 0. Define the existing design standard that is based on or incorporates precipitation data

Depth-duration-frequency curves provide estimates the depth of rainfall that characterizes the potential for extreme storms to occur in a particular region. Storms are differentiated based on their duration and frequency, or probability, of occurrence. Duration refers to the length of time that precipitation occurs, and is selected by the engineer based on the length of the design storm (or time of concentration) used to calculate stormwater runoff for a specific method (e.g. the rational method, TR-55). Frequency of occurrence is described as either: (i) an exceedance probability, which is the probability that an event of specific duration and depth will be exceeded in one time period (often 1 year), or (ii) a return period, or recurrence interval, which is the inverse of the exceedance probability, defined as the average length of time between events of the same depth and duration (McCuen 2005).

When the time period is equal to one year, the rainfall depth expected for a storm of 24-hour duration and 25-year return period is equivalent to the depth of precipitation over 24 hours that has a 4% chance of being exceeded in any year. The return period is selected by

stakeholders based on the acceptable risk level for a design to fail or be inundated. Frequency curves for use in design standards are created regionally in the U.S. by NOAA, available from the NOAA Atlas 14, which consists of a compilation of precipitation frequency estimates for all U.S. states and political entities (Bonnin et al. 2006; PennDOT 2011). The north and southeast regions of the continental US have been recently updated (2015 and 2013, respectively); however, many western regions (e.g., Montana, Washington, Oregon) have not been updated since 1973 (Hydrometeorological Design Studies Center and NOAA's National Weather Service 2016). It is important to recognize, however, that significant challenges exist with these curves due in large part to the spatially sparse observed data used to cluster regions with similar characteristics of extreme rainfall (Barros 2006).

2.3.2   Step 1. Understand historical basis and data requirements used to develop the existing standard and retrieve data

IDF curves have historically been created based on the underlying distribution of extreme events that occur in long time series (50 to 100 years) of observed rainfall. This process is applied at different durations of rainfall (5 minutes to 72 hours) by aggregating the data to the appropriate interval before analysis. Two methods are used to extract the extreme events, also known as block maxima or tails, including: Annual Maximum Series (AMS), where the maximum event for each duration storm is extracted for each year of record, or Partial Duration Series (PDS), where all values are taken above a threshold (Kilgore et al. 2016; Bonnin et al. 2006; CSA 2012). The PDS method, also known as Peaks Over Threshold (POT) is able to account for multiple extremes that may occur in a single year and is useful for short periods of record; however, thresholds may be difficult to select, and events within a year may not be hydro-meteorologically independent (Beguería 2005).

AMS data points are often fit to a Generalized Extreme Value (GEV) distribution, described by location, μ, scale, θ, and shape, ξ, parameters (Visser and Petersen 2012; Bonnin et al. 2006; Coles 2001; CSA 2012). The shape parameter (which can be greater than, less than, or equal to zero) determines the form of the distribution (e.g., Gumbel (Type I), Frechet (Type II) or Weibull (Type III)) (Coles 2001). GEV parameters may be estimated using maximum likelihood techniques (Katz 2013; CSA 2012). When using the AMS method, the rainfall depth for a given duration and return period, i.e., the recurrence interval depth (z_p), is found by relating the GEV parameters to the probability, as presented in Equation 2.1 and 2.2.

$$z_p = \mu - \frac{\theta}{\xi}\left[1 - y_p^{-\xi}\right], \text{ for } \xi \neq 0 \tag{2.1}$$

$$\mu - \sigma \, log(y_p), \text{ for } \xi = 0 \tag{2.2}$$

where $y_p = -\log(1 - p)$, p is the probability of exceedence in any year, and μ, θ, and ξ are the location, scale, and shape parameters of the GEV distribution, respectively (Coles 2001).

2.3.3   Step 2. Access appropriate climate model output based on requirements for the existing standard

IDF curves are calculated for short duration (5 minutes to 12 hours) as well as long duration (24 to 72 hours) events. The statistically downscaled datasets (e.g., Bureau of Reclamation, ARRM, MACA) are suitable for long durations at the daily level or higher. However, the dynamically downscaled datasets (e.g. NARCCAP and NA-CORDEX) are more appropriate for this analysis, as they allow calculation of curves at shorter durations (e.g., the 3-hourly interval and greater). Sub-hourly durations would require additional temporal disaggregation techniques not undertaken in this demonstration.

NA-CORDEX output is recommended for use over NARCCAP, if available, since some mod-

els are available at finer temporal and spatial resolution (e.g., hourly, 25 km) and a longer simulation period is available (1950 – 2100). However, at the time of this study, CORDEX outputs were not yet available; thus, NARCCAP outputs were used. NARCCAP precipitation projections are produced using a single emissions scenario (SRES A2) at a 3-hour time step and spatial resolution of 50 km. Since only a single emissions scenario is available, this analysis does not account for scenario uncertainty; however, the A2 scenario is at the upper end of SRES scenarios and represents a conservative estimate of the future. Precipitation output were extracted for the single grid cell with the centroid nearest to the Pittsburgh International Airport (40.49° N, 80.24° W). Grid point maps are available to relate the geographical location of the grid cell to the associated (x,y) coordinates in the NetCDF data matrix. NARCCAP data were extracted for all 6 NCEP driven runs and 11 available RCM-GCM combinations for historical (1970 – 2000) and future (2040 – 2070) simulation periods (see Tabel 2.3). Output were constructed as a complete time series for a single grid cell after individual 5-yr netCDF files over the entire North American domain were downloaded then concatenated using nco toolkit (Zender et al. 2016).

2.3.4   Step 3.   Account for climate model uncertainty and reliability

The reliability of an ensemble of regional climate models can be assessed by comparing "hind casts" of the regional models to historical observations. The hind casts are output from the RCMs after they have been driven by historical reanalysis data (instead of a GCM). For NARCCAP, the re-analysis data are from NCEP NARR (see Table 2.2) and reanalysis driven outputs are available on a 50-km resolution, 3-hour time step, for the time period from 1979 – 2006. In this study, the reliability of the NARCCAP RCMs was assessed by comparing the empirical distributions of the reanalysis outputs of the six regional climate models to those of obser-

vations obtained from the local stormwater authority in Pittsburgh (3 Rivers Wet Weather, 2015). Before comparison, the observed data, recorded on a 15-minute interval at 33 rain gauges throughout Allegheny County (area of 1,930 km2), was first scaled to the resolution of the reanalysis output (3-hour, 50-km). To do so, the observations were aggregated to a 3-hour interval, and then gauges that are geographically located within the 50-km grid cell of the re-analyses were averaged. The 3-hour exceedence probability, which represents the likelihood that a rainfall event of a specific volume will occur in a 3-hour period, was selected as the metric of comparison to represent the empirical distribution of both precipitation time series. The exceedence probability for each rainfall depth above zero (in the reanalysis and adjusted-observed time series) was calculated using a Weibull distribution, commonly used in precipitation analyses. Exceedence probabilities from the scaled-observations were plotted against those from the reanalyses.

Uncertainty was bounded using the validation approach, which uses a performance or reliability analysis to select (or "cull") models to include in the final ensemble. Three NARCCAP RCMs were selected, or culled, based on the visual proximity of the reanalysis exceedence curve to the adjusted-observed curve (see Figure 2.3). The five RCM-GCM simulations available from the three validated RCMs were used in the subsequent analyses (see Table 2.3).

2.3.5   Step 4. Adjust existing method to incorporate expected future trends

After the performance assessment, data from downscaled models may be integrated into future IDF curves. Future trends may be incorporated in one of several steps taken to obtain the IDF values, including to: the underlying time series of the data record, the extreme value series (AMS or PDS), the GEV distribution, or directly to the return level intensities calculated from the distribution. The first approach involves complex statistical downscaling techniques
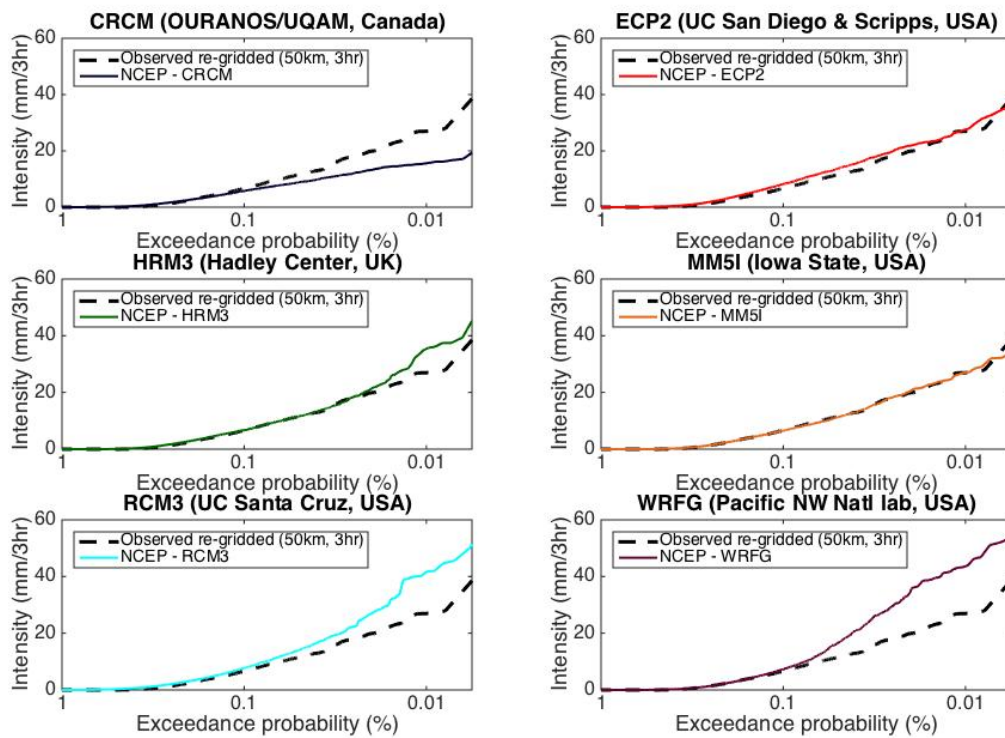
Figure 2.3: Comparison of 3-hr exceedance probabilities from 6 NCEP driven RCM runs in the NARCCAP ensemble (solid line) to observations re-gridded to 3-hr, 50km resolution (dashed line) for a grid cell in the Pittsburgh region (1979 – 2014)

Table 2.3: Regional Climate Models and Associated GCM Drivers Composing the NARCCAP Source of Downscaled Model Output; ECPC, HRM3 and MM5I were Selected After Reliability Analysis (Step 3))

Model	Full Name	Modeling Group	GCM Driver
CRCM	Canadian Regional Climate Model	OURANOS / UQAM	CGCM3 (Third Generation Coupled Global Climate Model)
ECPC (ECP2)	Experimental Climate Prediction Center Regional Spectral Model	University of California-San Diego & Scripps Institute of Oceanography	GFDL (Geophysical Fluid Dynamics Laboratory GCM)
HRM3	Hadley Regional Model 3 / Providing Regional Climates for Impact Studies	Hadley Centre	HADCM3 (Hadley Centre Coupled Model, version 3) GFDL (Geophysical Fluid Dynamics Laboratory GCM)
MM5I	MM5 – PSU / NCAR mesoscale model	Iowa State University	CCSM (Community Climate System Model) HADCM3 (Hadley Centre Coupled Model, version 3)
RCM3	Regional Climate Model version 3	UC Santa Cruz	GFDL (Geophysical Fluid Dynamics Laboratory GCM) CGCM3 (Third Generation Coupled Global Climate Model)
WRFP (WRFG)	Weather Research & Forecasting Model	Pacific Northwest National Lab	CCSM (Community Climate System Model) CGCM3 (Third Generation Coupled Global Climate Model)

to obtain the appropriate temporal and spatial resolution of the time series. However, it has been hypothesized that if the engineer is only concerned with designing for extremes, it may be more manageable to avoid downscaling to a continuous time series, and instead adjust empirical quantiles through mapping functions (Hassanzadeh et al. 2013). A simple method that has been introduced in the engineering literature involves directly adjusting historical rainfall depths at the point scale, for a given return period and duration, based on the expected change from historical to future conditions at the grid scale (Zhu et al. 2012; Forsee and Ahmad 2011). Areal reduction factors have been employed to adjust the station scale rainfall, as reported by Zhu et al (2012), and summarized here in Equation 2.3:

$$I_F^{(s)}(T,d) = I_H^{(s)}(T,d) \frac{I_F^{(g)}(T,d)}{I_H^{(g)}(T,d)} \tag{2.3}$$

where I denotes the intensity for a given return period (T) and duration (d), at the station scale (s) or grid scale (g), for future (F) or historical (H) time periods.

In this analysis, climate signals are incorporated into regional IDF curves using areal reduction factors applied to historical depths at the station scale. This process has three stages: (1) historical DDF curves were recreated for the historical period available from the climate models (1970 – 2000) using airport station data (obtained from NOAA National Centers for Environmental Information); (2) change factors, or areal reduction factors, were calculated from IDF curves estimated from historical and future RCM gridded outputs; and (3) the change factors were applied to update historical curves. Steps (1) and (2) utilize the same method for creating DDF curves, but on the native resolution of each data set. For the historical (1970 – 2000) and future (2040 – 2070) periods, return period depths values are calculated for the 3-, 6-, 12-, 24-, 48-, and 72-hour durations and the 2-, 5-, 10-, 25-, 50-, and 100-year return period. The moving window approach is applied to sum the underlying time series to the appropriate duration to obtain the annual maximum series. The AMS of each duration are fit to a GEV distribution using the method of moments and recurrence interval depths were calculated using

Equation 1 for each 30 year period (1970 – 2000) and (2040 – 2070). GEV distributions are fit independently for the airport station data and each RCM. Change factors were calculated separately for each model as the ratio between the future and historical gridded recurrence interval depths, and are applied to historical depths using Equation 2.

This simplified method is used solely for demonstration purposes of this framework. The method may be appropriate for understanding potential future trends in IDF relationships; however, it is not a reliable alternative to more rigorous methods that alter the extreme value series or the GEV distribution parameters (Mailhot and Duchesne 2009; DeGaetano 2009; Cheng and AghaKouchak 2014; Shahabul Alam and Elshorbagy 2015). In the near future, NA-CORDEX will be available for a continuous time period (1950 – 2100) and could be used to inform a general trend in the future GEV distribution, notably the location parameter.

2.3.6 Step 5. Interpret results and incorporate changes into design practice

Non-stationary conditions imply that the return period of an event will change with time. Mailhot and Duchesne (2009) state that design criteria under non-stationary conditions should explicitly consider (1) the expected lifetime of the structure, (2) that the probability of exceeding the design capacity and risk threshold will change over time, and (3) a statistical model that describes the expected evolution of intense rainfall over time 11/4/16 10:43 PM. The latter comes from the previous steps outlined in this framework and will include bounds of uncertainty represented as a range of plausible values for a given return period and duration. It is the responsibility of regulating agencies to provide guidance on which design value to choose within the range. Traditionally, design criteria have focused on selecting values as close to the expected value as possible, i.e., the mean of the range, assuming a normal distribution. One study suggested that design levels should be selected as the higher-than-median-percentile of

the design criteria in question (Karsten Arnbjerg-Nielsen 2011). Some argue that it is not possible to characterize the distribution and confidence intervals of the uncertainty, and an appropriate value cannot be selected independent of the decision being made (Hallegatte 2014). The authors of this study propose that in addition to presenting uncertainty bounds, agencies could provide two suggested values (possibly the mean and 75th quantile of the range) that could be used in low and high-risk situations.

To address considerations (1) and (2) regarding infrastructure lifetime and changing risk level, Mailhot and Duchesne (2009) propose that the design engineer will need to establish two criteria: (1) the critical return period, which is the return period for which the structure is made to withstand, and (2) the reference year, which is the year in the future to which the critical return period is associated. If the reference year is equal to the design life of the infrastructure, exceeding design capacity would not be accepted and the structure would be overdesigned for the full lifetime. If the reference year is equal to the year of conception, the structure would be under designed for the full lifetime.

Mailhot and Duchesne also suggest that more severe guidelines are needed for infrastructure with long expected lifetimes (e.g., higher critical return periods and longer reference years), since these structures could experience extreme shifts in climate towards the end of life, at which they are most vulnerable to failure do to age and degradation of materials. Furthermore, where uncertainty in projections is especially high, designers may choose to select a shorter reference year to allow for adaptations to be implemented once conditions become more apparent. The 2009 study and the present authors stress the importance of implementing recurring performance evaluations of the drainage system in order to expose evolving system vulnerabilities. Adaptation strategies over time will be required to maintain an acceptable service level.

2.4 Results and Discussion of Framework Application

As framework steps 0 through 2 were detailed previously, this section focuses on results from framework steps 3 through 5. Figure 2.3 presents results from step 3, the performance analysis of NARCCAP ensemble, which compared the exceedance probabilities of the six NCEP-driven RCMs (1979 – 2006) to those of the re-gridded observations for the Pittsburgh region (2004-2014). The black, dashed line represents the 3-hour exceedance probability for the aggregated observations and the solid line represents the 3-hour exceedance of the NCEP driven RCM. Proximity of the solid curve to the dashed curve represents similarity in the underlying empirical distributions and thus a higher skill of the RCM to represent historical statistics. These results show that for southwestern Pennsylvania, the HRM3, MM5I, and ECP2 models perform best in comparison to the other models, since these curves more closely agree with the dashed line. The Canadian model (CRCM) underestimates precipitation volume after the 1% exceedance probability, whereas the RCM3 and WRFG models overestimate the 3-hour precipitation depth after the 0.5% probability. The HRM3, MM5I, and ECP2 RCMs were selected for use in the subsequent analyses. Future research should examine quantitative metrics for objective selection of climate models based on reliability.

Figure 2.4 presents results for the second stage of step 4 of the applied framework, which involves developing the change factors from the gridded climate data in order to adjust existing IDF curves to incorporate future trends. Change factors (CFs), i.e., the ratio of the future rainfall depth to the historical depth, are presented for return periods of 2-, 5-, 10-, 25-, 50-, and 100-years as separate sub-plots, and durations of 3-, 6-, 12-, 24-, 48-, and 72-hours within each return period plot. For each duration, the change factor range represents the spectrum of CFs from each of the five selected RCM-GCM models. Change factors greater than 1.0 represent an increase in the rainfall depth in the future; less than 1.0 denotes a decrease. The

47

range of 5 models is presented as a box plot, where the bar in the box represents the median of the models, the top and bottom of the box plot represent the 25th and 75th quartiles, and the whiskers extend to the 90th quantiles. Values outside these ranges are represented as plus signs.
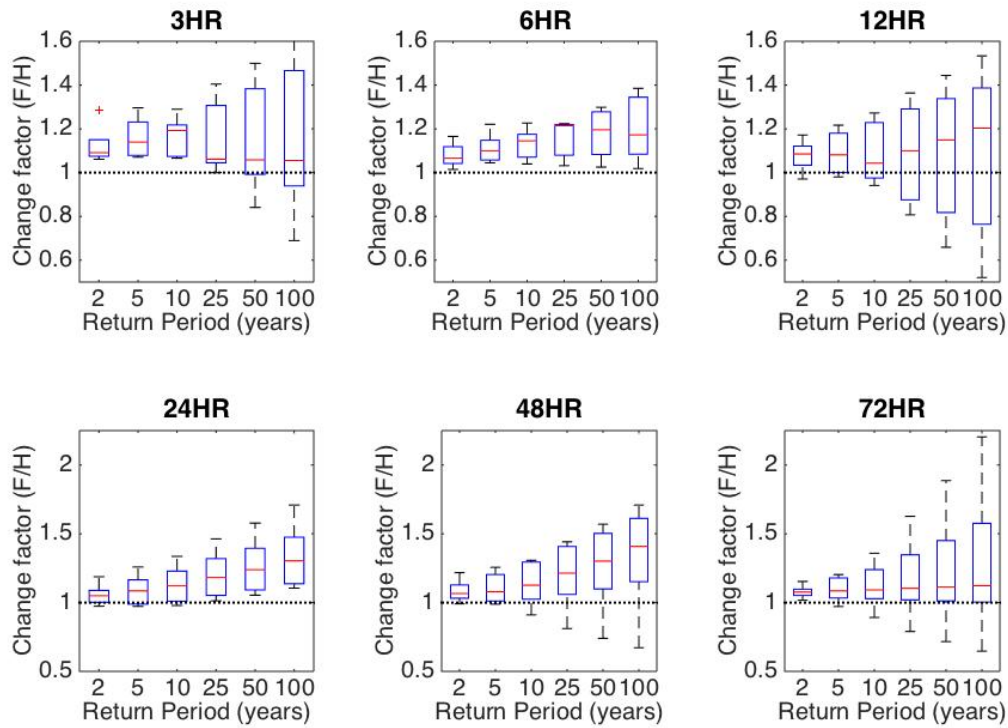
Figure 2.4: Range of change factors relating future (2040 – 2070) and historical (1970 – 2000) gridded rainfall depths calculated for each duration and return period for each model in the culled NARCCAP ensemble (n = 5) and a single grid cell in Pittsburgh. Change factors greater than 1 represent an increase in rainfall depth in the future for a given duration and return period. The range is presented as a box plot, where the red bar in the box represents the median of the models, the top and bottom of the box plot represent the 25th and 75th quartiles, and the whiskers extend to the 90th quantiles. Values outside these ranges are represented as plus signs.

The median change factor for each duration and return period is larger than 1.0, which suggests that the depth of extreme precipitation is expected to increase in the future for Pittsburgh. With the exception of the 3- and 72-hour durations, the median change factor tends to increase as the return period increases. This is also the case for the 75th and 90th quantile

change factor for all durations. This finding implies that the larger recurrence interval storms (e.g., 25-, 50-, 100-year), for the same duration, may increase in severity at a sharper rate than the more frequently occurring storms (e.g. 2-year). It is also interesting to note that for the 2-, 5-, and 10-year return periods, the median change factor of the 3-hour duration storm is the largest of all durations. This is in line with other studies that found only short duration storms are shown to have consistently higher intensities in the future (Kuo et al. 2015; Cheng and AghaKouchak 2014); however, it also suggests further analysis of relative change is needed to produce understanding of whether the result has a clear physical interpretation and is expected to be reliably predicted across downscaling procedures and regions.

It may be possible to interpret change factors as a potential "climate safety factor" that could be applied to existing, stationary, depth-duration-frequency values. Based on these findings, a safety factor of 1.3 would encompass the majority of model uncertainty for depths of smaller return periods (e.g. 2 to 10 years); however, a factor of 1.3 is no longer valid when uncertainty magnifies as the return period increases to 25 years and larger. Change factors for extreme precipitation will vary depending on the duration and return period of the event, as well as the climate model, region, and future year analyzed; thus, additional studies are needed to determine appropriate climate safety factors by region, duration, and return period. As an alternative to applying a safety factor to existing curves, the authors recommend using values from updated, non-stationary, depth- or intensity-duration-frequency curves.

Figure 2.5 presents the range of rainfall depths expected for the future period (2040 – 2070) based on the change factor method, for the previously listed durations and return periods. Change factors (reported in Figure 2.4) as less than 1.0 were converted to 1.0 for this analysis based on recommendations from the Canadian Standards Association, which state that beneficial aspects of climate change that allow for a reduction in design capacity should be neglected due to the inherent risks and costs that could arise from under-design (CSA 2012). When all models agree on findings suggesting change factors less than 1.0, this assumption should be reconsidered. To exemplify how rainfall depth changes with respect to the proba-

bility of occurrence, results are portrayed for a specific duration, as a function of the return period. Uncertainty among the five models is represented as the grey region on the plot. The median of these models is shown as the thin, solid line, and the 75th quantile is the thin, dashed line. The historical curve (1970 – 2000) is shown as the thick, dark line.
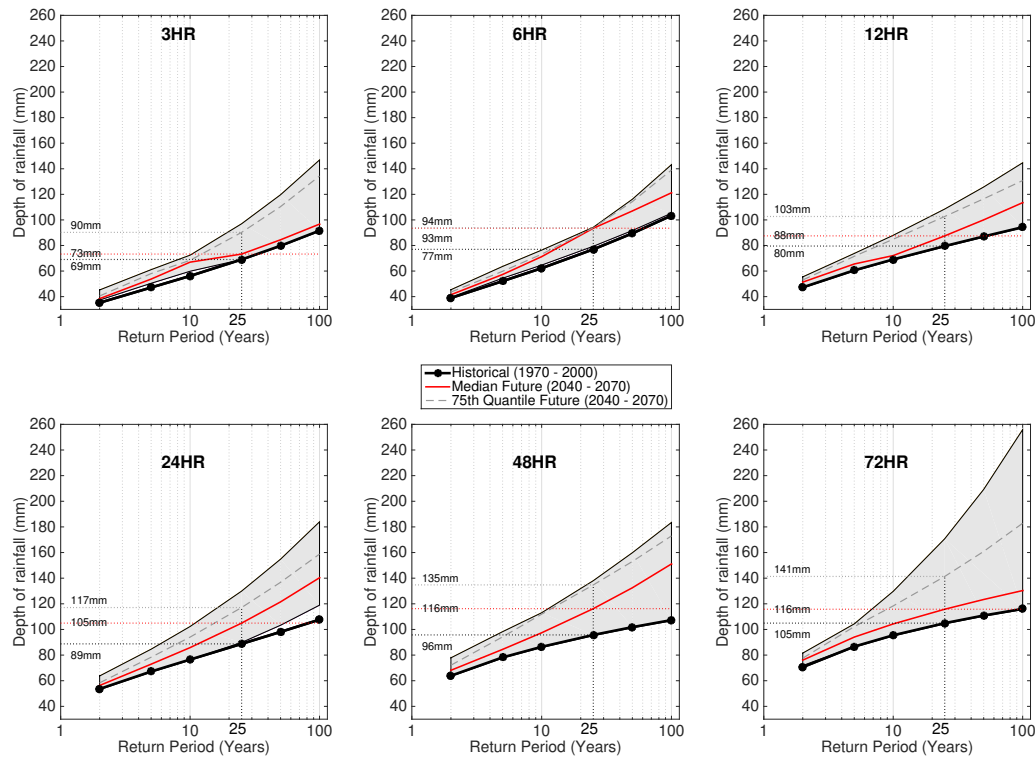
Figure 2.5: Updated IDF curves in Pittsburgh for the future period (2040 – 2070) using change factor method based on the culled NARCCAP ensemble (n = 6). The bolded, black line represents the historical period (1970 – 2000). The uncertainty range of the truncated values is represented as the shaded grey area; the median of the range is shown as the solid, red line; the 75th quantile shown as the thin, grey, dashed line. The historical values (1970 – 2000), calculated using airport data, are shown as the thick, black, dashed line. Future depths for the 25-year return period are highlighted for the mean and 75th quantile of the range.

Uncertainty of future projections tends to increase as return period increases. This phenomenon may be due to model variability of very extreme events; however, it is likely also a result of extrapolation of the GEV distribution to recurrence intervals larger than the underlying 30-year time series. One possible approach to overcome this limitation is to generate

multiple simulations with the same model. Looking specifically at the 25-year return period, increases with respect to the historical depth are inconsistent across durations and do not increase monotonically as duration lengthens. The median, future, depth is equivalent to a 6%, 21%, and 10% increase for the 3-, 6-, and 12-hour durations, respectively, and an 18%, 21%, and a 10% increase for the 24-, 48-, and 72-hour durations, respectively. The 75th quantile depth ranges from a 21% to 41% increase from the historical depth, bounded by the 6-hour and 48-hour durations, respectively.

The future, median, 25-year depth can be extended horizontally right until it intersects with the historical curve. This reflects the historical return period that would have been needed to ensure the 25-year return period performance in the future. For the 3-hour and 6-hour durations, this reflects the 35-year and 60-year return period, respectively; for 12-hour and 24-hour, it is about the 50-year and 85-year depths, and equal to or greater than the 100-year return period for durations 48 hours and larger. This finding indicates that merely doubling the return period (e.g. 25 year to 50 year) and using historical values may be appropriate for shorter duration storms (12-hours and less); however, this simplified method becomes inapplicable for larger duration storms. The historical 25-year depth (bottom, horizontal, dotted line) can also be extended left until it intersects the future, median curve (thin, solid curve). The intersection suggests that designing for depths with respect to a stationary 25-year storm would only provide protection from 7-to 12-year return period storms by 2070, for all durations.

These findings may be applied to the selection of the 25-year, 24-hour duration storm for use as input to the TR-55 method, commonly used in storm water design for calculation of peak discharge. To do so, the authors assume the following: (i) that the updated curves represent the state of the art, (ii) the design structure is located on an arterial road of low traffic volume, and (iii) the reference year, the year after which infrastructure performance is not guaranteed, is 50 years. Assuming that the current year (2016) is equal to the year of conception of the project, the associated calendar year needed to describe the expected rain-

fall depth in the reference year is 2066, which falls within the future period evaluated in this analysis. The storm water structure represents a situation of low risk (due to placement on a low-volume arterial); thus, the authors recommend selection of the median depth of 106 mm for use as input to the TR-55 method.

2.5   Conclusions

Despite the availability of downscaled projections of precipitation, there are considerable limitations to using these as inputs for engineering design applications. These include complex data extraction requirements, uncertain and biased model output, mismatch in temporal and spatial resolution of the gridded climate models and the desired engineering information, as well as the lack of a path to inform future design. This study presented a framework to overcome these limitations that may be used as a guide for agencies and engineers that wish to update current design standards to incorporate future, non-stationary trends.

After defining the design standard to update and how this standard uses precipitation data, the framework begins by understanding the existing methods and precipitation data requirements that have been used to create the current standard. This includes understanding the length of record, and the temporal and spatial resolution of the data required as input to re-create the components of the method. The second step involves selecting and extracting the most appropriate downscaled climate model data source, based on the findings in step 1. A data source may be preferred over another due to available temporal resolution closest to that of the historical data, a higher spatial resolution, the length of the simulation period, and finally the ease of extraction. The third step involves managing the performance and uncertainty of the downscaled climate models. For regional climate models, bias correction and performance evaluation is the responsibility of the user; whereas statistically downscaled data sets are usually corrected for bias and error during the downscaling process. Uncertainty is managed using one of three techniques: the extremes (max/min), ensemble, or validation ap-

proach. The fourth step relates to adjusting model output to the temporal or spatial resolution required for the engineering application, which can be completed using statistical downscaling techniques, bias-correction, or the change factor method. The final step discusses how to incorporate results into engineering practice by accounting for uncertainty ranges, risk levels, and non-stationarity.

The general framework was applied to the updating of depth-duration-frequency curves, a common input to stormwater design, that provides the expected depth of rainfall for given storm duration and recurrence interval. Historical curves were recreated by fitting a GEV distribution to the AMS series of 30 years of historical rainfall data obtained from the gauge station at the Pittsburgh International Airport. Since precipitation inputs are required on a sub-daily time-step for IDF curves, climate output was obtained from NARCCAP, the only source currently available that provides output at a less than daily time step. These dynamically downscaled regional climate models were assessed for performance by comparing NCEP driven runs to re-gridded observations, and 5 RCM-GCM combinations were selected to complete the remaining analyses. Future curves were estimated by applying change factors from the gridded climate model output to the historical curves. A value may be selected from the updated curve based on the lifetime of the structure, critical return period, and risk level.

The median change factor for each duration and return period is larger than 1.0, which suggests that the depth of extreme precipitation is expected to increase in the future for Pittsburgh. Results imply that designing for a rainfall depth equivalent to the future (2040-2070), median 25-year depth is comparable to designing for the stationary (1970 – 2000) 50 to 100+ year depths, depending on the storm duration. If instead the designer selected a 25-year depth from the stationary curve, this would be equivalent to the 7-to 10-year return period storm depths of the future, median value for various durations.

2.6 Recommendations and Future Work

Based on these results and the suggestions of the framework presented, and the ASCE initial guidance for adapting infrastructure and practice to a changing climate (Olsen 2015), the following information should be considered by engineers working with climate output for resiliency applications:

- Match intended engineering application with the appropriate climate model source,

- Different climate model sources require various amounts of effort for data extraction and preparation,

- Climate models have various levels of skill at representing historical mean and extreme statistical metrics and engineers need to understand the major issues and uncertainties involved,

- Create an ensemble and be transparent about assumptions,

- Test robustness of designs to extremes and alternative scenarios,

- Discuss tradeoffs and uncertainties in risk, resiliency, performance, and costs with stakeholders,

- Design for low-regret, adaptability, and robustness, and revisit designs when new information is available.

Because of climate change and stakeholder desires for enhanced resiliency, engineers will need to be familiar with choosing and incorporating climate change projections into planning and design. However, for engineering practitioners constrained by time and resources, it may not be feasible to expend the effort required for the detailed analyses described here. There is a need for collaboration across agencies and the research communities to serve as ad-hoc or standing boundary organizations to translate climate projections into relevant engineering information. Duties of these translational organizations may include providing rigorous stan-

dards for interpretation of climate data, understanding the utility of increasing the number of models considered in an ensemble, development of a single, simplified user interface that accesses all downscaled data sources, and tools that automatically post-process data based on rigorous standards.

2.7 Acknowledgements

CHAPTER 3   THE EFFECT OF MODELING CHOICES ON UPDATED PRECIPITATION FREQUENCY CURVES AND STORMWATER INFRASTRUCTURE[1]

[1]Cook, L., McGinnis, S.., and Samaras, C. (2018) The effect of modeling choices on updated precipitation frequency curves and stormwater infrastructure. In preparation.

Abstract

Precipitation-frequency curves are commonly used in the design of stormwater systems to represent values of extreme rainfall. Many studies and regions have updated these curves to reflect trends in future rainfall from climate model projections. Less often explored in these studies are the model uncertainties such as those related to the spatial resolution of the climate model and the spatial adjustment (downscaling) method. This study investigates these uncertainties by considering climate-corrected depth-duration-frequency curves using sub-hourly regional climate model projections from NA-CORDEX. Results are calculated for 6 U.S. cities (Birmingham, AL; Boston, MA; Boulder, CO; Pittsburgh, PA; Phoenix, AZ; and Seattle, WA) for the period from 2040 to 2099. Shorter duration storms (less than 6 hours) show a larger percent change in the future than longer duration storms for all cities. Out of three correction methods evaluated, the change factor method is preferred due to its consistent results and simple interpretation. The lower spatial resolution (50-km) climate model ensemble generally provides a more extreme estimate of precipitation in the future and wider uncertainty range than the 25-km ensemble. These uncertainties can lead to different choices in pipe size for stormwater designs.

Keywords: Depth-duration-frequency curves; stormwater design; regional climate models; extreme value distribution; climate change

3.1   Introduction

Hydrologic analyses for the design of green and grey stormwater infrastructure require localized precipitation information. This typically takes the form of an "event," which is defined as a certain amount of rain (called a depth or intensity) that takes place over the duration of a storm (Durrans et al. 1999). These values are obtained from regional precipitation-frequency curves, which can be in the form of intensity-duration-frequency (IDF) or depth-duration-frequency (DDF) curves (Bonnin et al. 2006; PennDOT 2011; CSA Standards 2012). IDF curves express the rainfall intensity as a function of storm duration and probability of occurrence, whereas DDF curves express the rainfall depth with respect to duration and probability. The probability is often expressed as a return period, or the inverse probability that rainfall of a specific return level will be exceeded in a year (McCuen 2005). Precipitation frequency curves have been traditionally created based on the statistical distribution of extreme events that occur in the historical rainfall record at a specific weather station (Hershfield 1961).

The use of historical data in hydrologic design is a widely accepted approach; however, as atmospheric water vapor rises along with global temperatures (Allen and Ingram 2002; Soden et al. 2005; Allan and Soden 2008), increases in extreme rainfall are expected (Karl and Trenberth 2003). There is a growing consensus that new approaches to infrastructure design, especially infrastructure with intended long-term use, are required to incorporate projected future changes (Barros and Evans 1997; Milly et al. 2008; Mailhot and Duchesne 2009; Rosenberg et al. 2010; Arnbjerg-Nielsen et al. 2013; Lopez-Cantu and Samaras 2018). In response, several studies and regions have begun to update precipitation frequency curves using trends from climate models in order to capture expected changes and improve resiliency of designed infrastructure, e.g., (Zhu et al. 2012b; Mirhosseini et al. 2013; Hassanzadeh et al. 2013; Kuo et al. 2015; Simonovic et al. 2016; DeGaetano and Castellano 2017; Cook et al. 2017). These studies have analyzed precipitation patterns in the Northeastern U.S. (DeGaetano and Castel-

lano 2017; Cook et al. 2017), Southeast Canada (Srivastav et al. 2014; Kuo et al. 2015; Ahmed and Tsanis 2016), and Northern England (Fadhel et al. 2017). For these regions, the 50-year, 24-hour storm is projected to increase by approximately 25% (for the average of the climate model ensemble). This means that the amount of rainfall during a 24 hour period in an extreme storm (one expected to occur only once in 50 years) is expected to increase by 25%. Few studies expect extreme precipitation to decrease in the future, and then by no more than 10% (DeGaetano and Castellano 2017; Fadhel et al. 2017).

Predictions of future extreme precipitation are subject to large uncertainties, meaning that the results could change depending on the methods and climate simulations used in modeling. Uncertainty is introduced when the system is modeled due to: (1) characteristics of the climate model (Allen et al. 2000; Deser et al. 2012); (2) the greenhouse gas emissions scenario selected; and (3) the downscaling, or spatial correction method. A recent study suggested that global climate models (GCMs) contribute the largest uncertainty in IDF curves, followed by the representative concentration pathway (RCP) emissions scenarios, and then the downscaling method (Shahabul Alam and Elshorbagy 2015). However, a different study (Sarr et al. 2015) found that the choice of downscaling method had a larger effect than the selection of the climate model for storms of 10-year return periods or larger. Additional uncertainty can also be introduced from the spatial resolution of the climate model. Mendoza et al (2016) found that variation introduced due to different spatial resolutions of the climate model surpasses variation from the choice of climate model alone (Mendoza et al. 2016). In addition to differences resulting from the spatial correction technique, DeGaetano and Castellano (2017) compared three downscaling techniques and found that a change factor technique employed with GCMs produced the largest percent change, while the use of regional climate models led to the largest station to station variability (DeGaetano and Castellano 2017). Their comparison also demonstrated that differences in the direction of change could result from using different spatial correction techniques, even when the same climate model was used.

While the prior work has identified the importance of evaluating uncertainties other than

those introduced by the climate model ensemble, most studies have not considered how to address multiple combinations of uncertain factors, nor have they put the introduced uncertainty into context for stormwater infrastructure design. This paper investigates how different choices in climate-corrected DDF curve creation, including the choice of climate model spatial resolution and the method used for spatial adjustment, alter projections of 1-hour to 24-hour rainfall depth and the uncertainty associated with these projections. We then demonstrate how the selection of design storm from this uncertain range affects stormwater infrastructure design dimensions in 6 U.S. cities.

3.2   Data and Approach

This study assesses the uncertainty introduced by the choices made when developing updated, or climate corrected, depth-duration-frequency (DDF) curves using sub-daily climate model output, and how these choices and their related uncertainty can affect stormwater infrastructure designs. In this study, DDF curves are estimated using the annual maximum series (AMS) from observed rainfall and regional climate model simulations, for six locations in the United States (U.S.). The DDF curves in this analysis are created for five durations (1-, 3-, 6-, 12-, and 24-hour) and six return periods (2-, 5-, 10-, 25-, 50-, and 100-years).

Section 3.2.1 presents the data used to create the climate-corrected DDF curves, including observed weather station data and the climate model simulations. Section 3.2.2 describes general methods used in this study to create historically-informed DDF curves. Section 3.2.3 describes the methods used to develop updated DDF curves in the present work, including the three correction techniques and the selection of a future time period for simulation. Section 3.2.4 explains the analyses used to assess and bound the uncertainty relating to the spatial resolution, the statistical distribution, and the correction methods. The final section (3.2.5) presents an illustrative design example used to assess implications of the uncertainty related to modeling choices on stormwater infrastructure design.

3.2.1 Data

Observed data

The DDF curves are created using hourly observed precipitation data from the NOAA National Center for Environmental Information (NOAA 2016) for six cities: Birmingham, AL; Boston, MA; Boulder, CO; Pittsburgh, PA; Phoenix, AZ; and Seattle, WA. These cities were selected as a representative sample of various climate regions throughout the U.S. Each of these cities also had a weather station near it with a long data record that was selected for use. If small gaps existed in the station records, they were supplemented with data from nearby stations. Historical curves were recreated for the period from 1950 to 2013 for all cities except Pittsburgh, where only data from 1953 to 2013 were available. Table 3.1 presents the latitude and longitude coordinates, the elevation, and the historical 24-hour, 50-year event, for each weather station. The range reported for the 50-year storm represents the 10% and 90% confidence interval. The historical 24-hour, 50-year event was estimated by fitting a GEV distribution with maximum likelihood estimation (see Section 3.2.2).

Regional climate model simulations

This analysis uses simulated rainfall time series from regional climate models (RCM) from the North American Coordinated Regional Downscaling Experiment (NA-CORDEX) project (Mearns et al. 2017). NA-CORDEX is a compilation of standardized regional climate model simulations that provides simulations at an hourly time step, for two different spatial resolutions (50-km and 25-km), over the continuous period from 1950–2100. The NA-CORDEX project uses Earth System Models (ESMs)[2] (Heavens et al. 2013) to provide input conditions for the Regional Climate Models (RCMs). A single RCM "forced" by a single ESM is called an

---

[2]ESMs use more complex relationships than those in older Atmosphere-Ocean Global Circulation Models (AOGCMs). In additional to simulating atmospheric and ocean components, they include interactions between the global carbon cycle, vegetation, atmospheric and ocean chemistry, and ice sheets (Heavens et al. 2013)

Table 3.1: Characteristics of each city in this study. The range for the 50-year storm represents the 10% and 90% confidence interval.

City	Latitude	Longitude	Elevation (m)	24-hour 50-year (in)	rainfall depth (cm)
Birmingham AL	33.56	86.74	187.5	9.7 (6.15 - 17.8)	24.6 (15.6 - 45.2)
Boston MA	42.36	-71.00	3.7	7.0 (4.8 - 11.3)	17.8 (12.2 - 28.7))
Boulder CO	40.03	-105.3	1654	5.9 (3.7 - 10.5)	15.0 (9.4 - 26.7)
Phoenix AZ	33.42	-112.0	337	2.9 (2.0 - 4.7)	7.4 (5.1 - 11.9)
Pittsburgh PA	40.44	-80.02	367	4.3 (3.3 - 5.9)	10.9 (8.4 - 15.0)
Seattle WA	47.44	-122.3	113	4.5 (3.4 - 6.6)	11.4 (8.6 - 16.8)

RCM-ESM combination.

Since one of the goals of this analysis is to understand how the choice of spatial resolution of the regional climate model affects sub-daily climate-corrected DDF curves, it is necessary to use RCM-ESM combinations available at a sub-daily time step for both spatial resolutions. At the time of this analysis, three RCM-ESM combinations were available with these characteristics:

1. The RegCM4 RCM[3] driven by the Max Planck Institute Earth System Model at base resolution (MPI-ESM-LR)

2. The WRF RCM[4] driven by MPI-ESM-LR, and

3. The WRF RCM driven by the Geophysical Fluid Dynamics Laboratory Earth System Model, Modular Ocean Version (GFDL-ESM2M)

All three models considered here were run for the same emissions scenario (called RCP 8.5 (Riahi et al. 2011)); however, two are based on climate models that predict a 3.6 °C rise for a doubling of atmospheric CO_2, while the other one predicts a 2.4°C increase for the same condition. Each of these three RCM-ESM combinations was simulated at the 25-km and 50-km spatial resolution; thus a total of 6 climate model simulations were used for this analysis (3 RCM-ESM combinations and 2 spatial resolutions). Each climate model simulation (for the entire North American domain) was downloaded from the NCAR server for a historical simulation period (1950–2005) and a future simulation period (2006–2100). For each city, a time series of rainfall was extracted from the single grid cell in each climate model simulation whose centroid was closest to the latitude and longitude coordinates the city.

---

[3]The RegCM4 model was originally developed at the International Centre for Theoretical Physics (ICTP) and simulated for NA-CORDEX at the National Center for Atmospheric Research (NCAR)

[4]The WRF model (used for 2 and 3) was developed and simulated at NCAR

3.2.2 Development of updated, sub-daily DDF curves

In this study, DDF curves were created using the annual block maxima, or annual maximum series (AMS), of rainfall for each city and duration. For the 1-hour duration, the AMS were the series of maximum 1-hour rainfall depths for each year. For durations longer than 1-hour, the 1-hour time series was first aggregated to the desired duration using a moving window approach through convolution. Each AMS was then fit to a generalized extreme value (GEV) distribution using maximum likelihood estimation (MLE) fitting techniques (Scholz 1980). The GEV distribution was selected over the two parameter Gumbel distribution due to the ability to better describe the behavior of the upper tails (Overeem et al. 2008; Shahabul Alam and Elshorbagy 2015). More details about the GEV distribution can be found in Section 2.3 and in (Coles 2001).

The MLE method returns the GEV parameters that are most likely, depending on the data sample. The ML estimation error defines the confidence intervals of the fit. Return levels for a specific probability of occurrence (return period) and confidence interval are extracted using the parameters from the GEV distribution, as follows:

$$z_p(d, CI) = \mu_{CI} - \frac{\theta_{CI}}{\xi_{CI}} \left[1 - y_p^{-\xi_{CI}}\right], \text{ for } \xi \neq 0 \tag{3.1}$$

$$\mu_{CI} - \sigma_{CI} log(y_p), \text{ for } \xi = 0 \tag{3.2}$$

where $y_p = -\log(1 - p)$, p is the probability of exceedence in any year, and $z_p(d, CI)$ is the depth of rainfall for a specific duration, confidence interval, and probability. μ, θ, and ξ are the location, scale, and shape parameters of the GEV distribution, respectively, for a specific confidence interval (Coles 2001). The best-fit return level is calculated using the best-fit MLE-estimate of all three parameters; the 90% confidence interval return level is calculated using the 90% CI estimate for all three parameters. The 10% CI value is estimated in the same

manner.

Apart from this general approach to creating DDF curves, several additional modeling choices were made in order to develop the future, updated DDF curves for this research. These choices include: (i) the length of the future time period that the updated curves will present, (ii) the method for bounding the future uncertainty range, (iii) the choice of climate model spatial resolution, and (iv) the method used for spatial adjustment of the climate models to the station scale. All of these choices were tested for sensitivity; however, preliminary analysis of the future time period length and the uncertainty bounding technique indicated that findings are not significant enough to warrant additional discussion in the results section. The following subsections describe these choices in more detail.

Choice of spatial adjustment technique

Spatial downscaling, or spatial correction techniques, used to update DDF curves using climate model output, can be grouped into three main categories. The first is to create synthetic rainfall timeseries that reflect future trends, e.g., using weather generators or bias-correction (Kuo et al. 2015; Kueh and Kuok 2016), and then estimate the extreme value distribution. The second way is to alter the observed extreme events (annual maximum or partial duration series) to reflect future change, e.g., through quantile-mapping (Solaiman and Simonovic 2011), genetic programming (Hassanzadeh et al. 2013), or historical analogs (Castellano and DeGaetano 2016) and then estimating the extreme value distribution. The final and most straight-forward method is to directly adjust the historical intensity or depth using a change factor. This change factor is the change between depths estimated from historical and future climate models (Forsee and Ahmad 2011; Zhu et al. 2012a, b; Cook et al. 2017). In this analysis, one type of method from each of these categories is tested: (1) a non-parametric bias correction of the time-series, Kernel Density Distribution Mapping (KDDM); (2) a parametric transfer function fit to the annual maximum series, AMS Transfer Function (ATF); and (3) applying

a Simple Change Factor (CF) to observed DDF values, as discussed in Chapter 2 (Equation 2.3). The first two methods are both considered as forms of bias-correction, meaning the climate model data series is adjusted depending on differences with observed data; while the last method is considered to be a delta change method, where the observed value is adjusted up or down depending on the change between the historical and future climate model simulations.

Figure 3.1 presents an overview of the sequence of steps required for each method. Simulated regional climate model output is combined with observed data at the station scale to create updated DDF curves. This combination can occur when both are in the form of a continuous time series (Method 1), either before the time series is aggregated to the desired duration (Method 1A) or after the time series is aggregated (Method 1B). Only Method 1B is used in subsequent analyses due to large mean absolute error with respect to observed AMS (results shown in Appendix 5.2.6). The combination can also occur when both data inputs are represented as annual maximum series (Method 2), or when both are in the form of a depth of rainfall, after the GEV distribution has been fit to the annual maximum series. A more detailed description of the equations and processes used for each method are provided in Appendix 5.2.6.

Choice of climate model spatial resolution

This analysis evaluates the uncertainty related to the spatial resolution of the regional climate model simulations, exploring which spatial resolution leads to a larger (or smaller) prediction of extreme rainfall, or a higher (or lower) uncertainty range surrounding the prediction. The 50-year storm is used as the basis for evaluation of the two different climate model ensembles (25-km and 50-km). The climate model ensembles are compared based on the percent change from the observed (1950–2013), 50-year, best-fit storm.
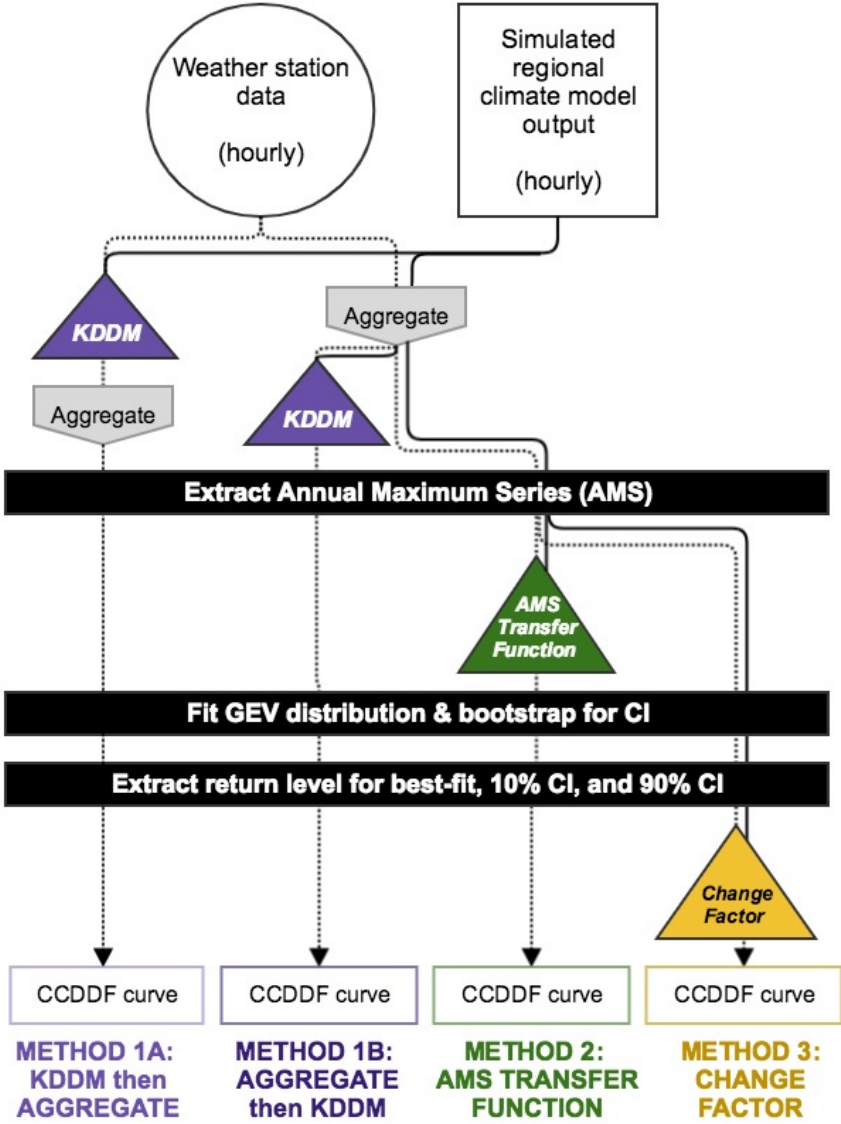
Figure 3.1: Sequence of steps carried out in each of the DDF adjustment techniques used in this study. Method 1 is Kernel Density Distribution Mapping (KDDM). Method 2 is the Annual Maximum Series (AMS) Transfer Function and Method 3 is the Simple Change Factor Method.

Selection of future time period

The models require selection of a future time period. In the present work, we selected the 80-year period from 2020 – 2099 because it represents the full range of expected precipitation in the future. Furthermore, the 80-year period had the lowest uncertainty range when compared to shorter time period lengths, but not a considerably different mid-point (results not shown).

Bounding the future range

Several types of uncertainty are introduced when multiple climate models fit to different GEV distributions are used to calculate a range of future DDF predictions. Fitting a GEV distribution to a data sample introduces parameter uncertainty, or uncertainty in the value of the distribution parameters, while climate models introduce model and scenario uncertainty. NOAA Atlas 14 represents parameter uncertainty in historical DDF curves as the 10% and 90% confidence intervals (CIs) of the GEV fit (Bonnin et al. 2006). In the future, each climate model simulation has a different GEV fit, and thus different 10 and 90% CIs. It would be highly desirable to distill this into a single range of uncertainty, which would be easier to interpret, especially for engineering stakeholders. Unfortunately, the complexity of representing the joint uncertainty of the model ensemble and the GEV distribution makes this a difficult prospect.

This analysis uses a basic approach to combine the GEV parameter and climate model uncertainties into a single range. The lower bound of the future range is calculated as the average of the 10% CIs from each model simulation, while the upper bound is the average of the 90% CI from each model simulation. The median value is the average of the "best fit" values from all climate model simulations. Other bounding techniques were examined, including taking the minimum of the 10% CIs of all climate simulations and the maximum of the 90% CIs. However, this technique was not selected because when all points were considered in a single distribution, this method led to many extreme outliers. The chosen technique was selected because it incorporates parameter uncertainty from the GEV distribution, but avoids

presentation of extreme outlier values that would likely not be considered for stormwater design (results not shown).

3.2.3   Illustrative stormwater infrastructure design example

A small design example for pipe sizing is used to evaluate how these different sources of uncertainty may affect stormwater design. The goal in the example is to design a storm sewer pipe that contains the 50-year storm for a small and large watershed. The peak discharge is calculated using the rational method: $Q_p = c\,i\,A$, where Q_p is the peak flow for the watershed, c is the runoff coefficient, i is the rainfall intensity, and A is the watershed area. The runoff coefficient is assumed to be 0.5 for a moderately urbanized watershed.

With the rational method, the duration of storm to select from the DDF curve is assumed to be equal to the time of concentration of the watershed, or the time it takes water to travel from the most distant part of the watershed to the outlet point (McCuen 2005). Two watershed areas are examined: 10 acres and 100 acres. The small, 10-acre watershed has an assumed time of concentration of 1-hour, and thus uses the 1-hour design storm. The larger 100-acre watershed has an assumed time of concentration of 24-hours, and thus uses the 24-hour design storm. Using the peak discharge, the pipe diameter is calculated with Manning's equation.

$$D_r = \left(\frac{n Q_p}{0.31 k_n \sqrt{S_0}} \right)^{\frac{3}{8}} \tag{3.3}$$

Where D_r is the pipe diameter, n is the roughness coefficient, S_0 is the channel slope, and k_n is a coefficient of the velocity versus slope relationship. The roughness coefficient is assumed to be 0.013 for ordinary concrete lining, kn equal to 1, and S_0 equal to 0.005 (or 0.5%).

The resulting diameters from Manning's equation are rounded up to the nearest standard U.S. pipe size to determine the diameter of the pipe that would be installed. The minimum diameter of storm sewer pipe recommended by the Pennsylvania Department of Transportation is 18 inches (450 mm) (PennDOT 2015). Pipes are available in increments of 3 inches (75 mm) until the 36-inch (900 mm) pipe, after which they are available in increments of 6 inches (150 mm). The pipe sizes used in this analysis are: 18, 21, 24, 27, 30, 33, 36, and 42 inches (equivalent to 450, 525, 600, 675, 750, 825, 900, and 1050 mm).

3.3   Results and Discussion

3.3.1   Change in precipitation in all cities

This section presents how the depth of rain is expected to change in the future across different durations and cities using the change factor method. Figure 3.2 shows the percent change of future (2040–2099) rainfall depth relative to the observed (1950–2013) depth for the 25-year (left) and 100-year returns periods (right). For each city, four durations (from left to right: 1, 3, 12, and 24 hours) are presented as different colored bars (light to dark). The top of the bar represents the mean of the percent change across all six climate model simulations. The range shown represents the maximum and minimum percent change across all six climate model simulations.

These results suggest that over the coming century the intensity of shorter storms (< 3 hours) is expected to increase more than the intensity of longer storms (> 12 hours). This implies that stormwater infrastructure, which is sized for a specific intensity of rainfall, should be equipped to store or convey a higher intensity of water in a short amount of time, rather than a lower intensity over a longer period. Designing for the higher intensity rainfall over a short duration leads to a design for a larger total volume of rainfall, in some cases, when designing for the same return period storm in the longer durations. The increase in the amount

of water expected during these short time intervals will, however, vary by city and return period. It is, however, expected that increases in storm depth get larger as the return period gets larger (comparing results between the two panels in Figure 3.2 for each city).

For the six cities examined, the ensemble average predicts that 1-hour precipitation is expected to increase by at least 30% for the 25-year storm, and 40% for the 100-year storm, except in Boulder, where the values are 20% and 25%, respectively. Increases are largest in Birmingham (75%, 110%), followed by Pittsburgh (50%, 75%). For longer storms, the ensemble average suggests that precipitation is expected to increase in four out of six cities, but not by as much as the shorter storms. According to the ensemble average, the 25-year 24-hour storm is expected to increase by about 25% in Birmingham, Pittsburgh, and Phoenix. In Boston, the ensemble mean suggests an increase in the 25-year storm of about 10%. Boulder and Seattle show virtually no change in the 24-hour storm according to the ensemble mean, and the predicted ranges include the potential for a negative change (i.e., smaller rainfall in a 24 hour period).

3.3.2   Influence of correction method on future rainfall predictions

The results in Figure 3.2 did not incorporate any form of bias-correction. This section evaluates how these results would change if bias-correction was considered (using KDDM or using an AMS transfer function). Figure 3.3 presents the results for the future 50-year rainfall depth using the change factor, KDDM, and transfer function methods. The left and right columns show the 1-hour and 24-hour depth, respectively, for all cities as different side-by-side pairs of panels. The observed period is shown as a black dashed line and the future period is shown as different colored lines for each method. The sloped line passes through the average of the best-fit values of the climate model ensemble (all resolutions), while the range shows the up-
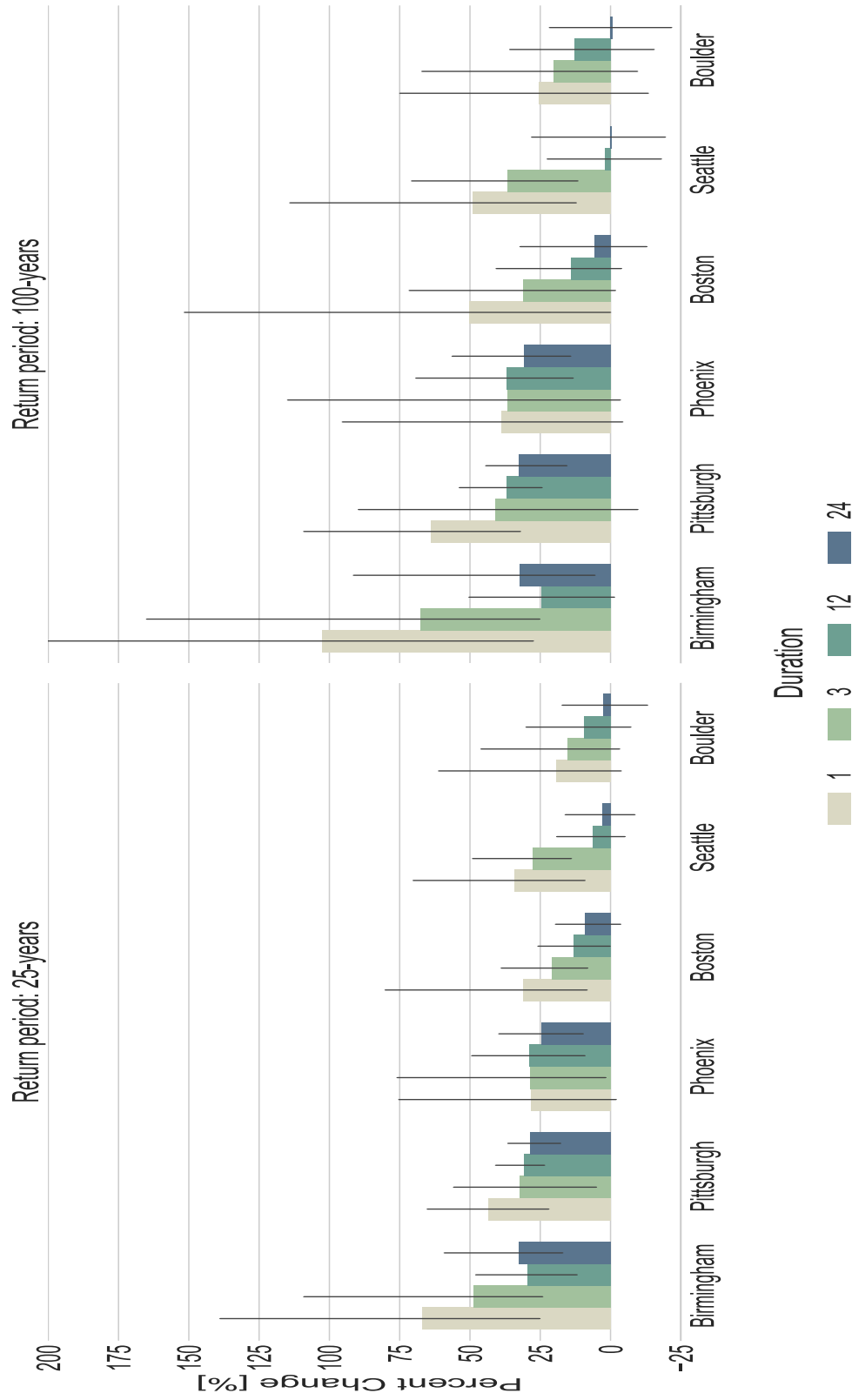
Figure 3.2: Percent change in future depth (for 2040–2099) relative to the observed depth (for 1950–2013) for each city (x-axis), using the change factor method. Four durations (1-, 3-, 12-, 24-hour) are presented as different colored bars from left to right and two return periods (25- and 100-year) in subplots. The top of the bar represents the mean of the percent change of all 6 climate model simulations, and the range shown represent the maximum and minimum of the percent change for all climate model simulations.

73

per and lower bounds defined in Section 3.2.2. The observed range shows the 10% and 90% CI for the historic station data. The upper bound on the y-axis (highlighted in red) varies in each subplot.

For the smaller return periods, the median of the ensemble for each method is relatively consistent for most cities; however, as the return period increases, the median of each method begins to diverge. In addition to differences in the median values, the uncertainty range between the three methods diverges as the return period increases. These findings are consistent with results from Sarr et al. 2015, who found that the choice of downscaling technique matters more for return periods of 10 years or longer. In the current analysis, the mean and bounds of the three methods can vary by several orders of magnitude for the larger return periods.

In these cases, it is the transfer function and KDDM methods that lead to unrealistic values. For instance, in Birmingham, the transfer function predicts the mean ensemble value for the 24-hour 100-year storm to be about 50 inches (120 cm) of rain. The historical value is around 11 inches (30 cm) - a value that is five times lower than the future prediction. This unrealistic future value is the result of the prediction of several large rainstorms in the 25-km WRF-GFDL model run that predict unrealistically large values of rainfall when aggregated to the 24-hour duration. After the transfer function is applied to the raw model output to adjust it to the station scale, these large rainstorms are amplified, and the future rainfall is predicted to be even higher. These rainfall values are considered unrealistic because the evolution of the storm within the climate model is not realistic. The storm exists for several days and moves very little during that time. In reality, storms would end and then reform, usually diurnally like the other precipitation in the region, and be a bit more widespread. The reason for this unrealistic storm could be a result of an unstable atmosphere, and the inability of the model to process this instability, given the parameterization of convection and other processes. Thus the odd behavior of the model is possibly the result of a more unstable future climate that the model cannot account for, and thus allows the storm to persist and barely move for almost a
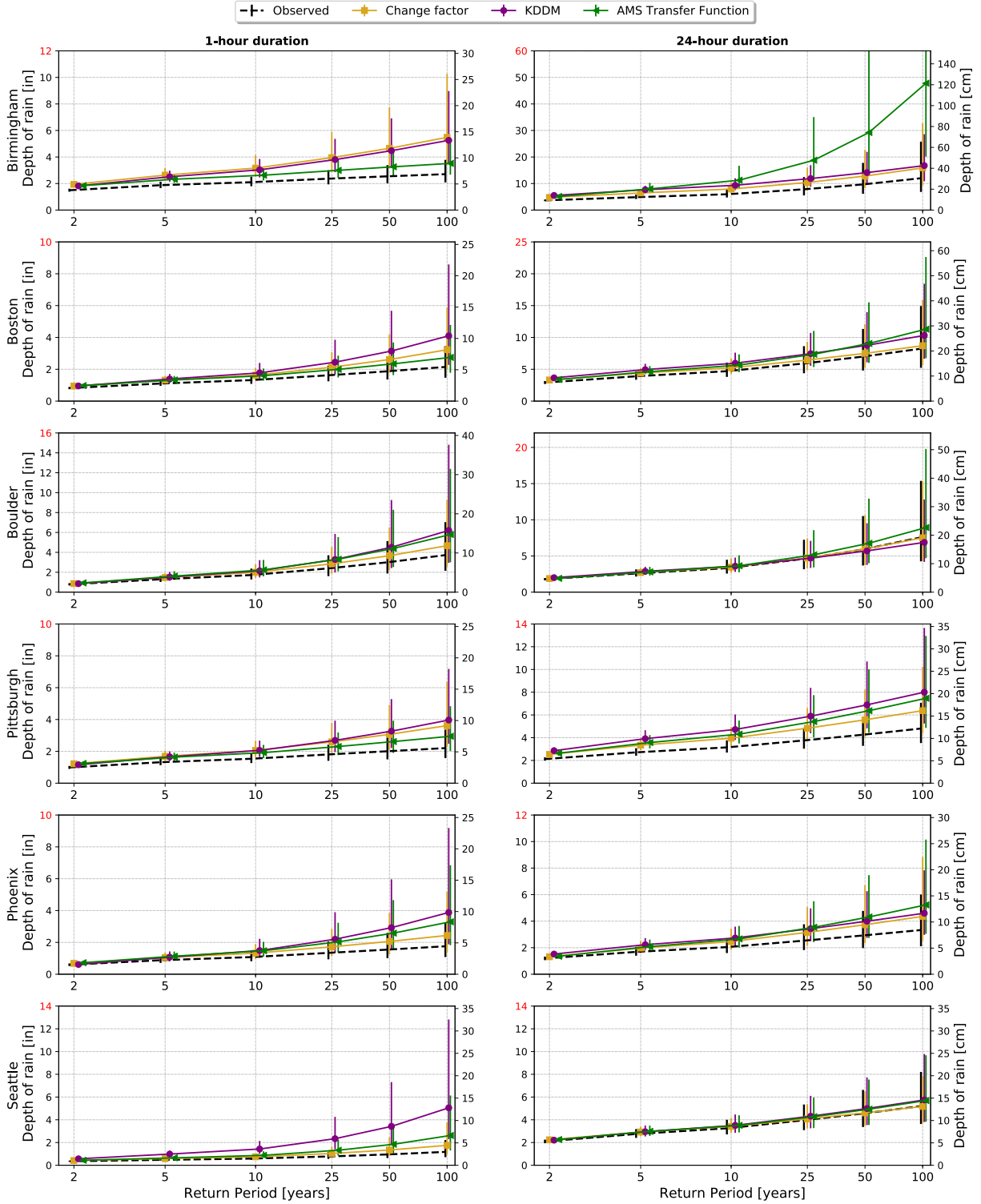
Figure 3.3: Effect of different correction methods on future rainfall depth for two durations: 1-hour (left) and 24-hour (right). Observed depth is in black; Change factor in yellow (square markers); KDDM method in purple (circle markers); and transfer function method in green (triangle markers). The sloped line passes through the average of the climate model ensemble (all resolutions); error bars represent the upper and lower bounds of this ensemble.

week.

The KDDM method also has the potential to produce erroneous results in the same fashion. Applying the KDDM method to the entire time series when extreme values exists in the raw data could distort the annual maximum series in ways that are not consistent with reality. Thus, in order to produce realistic results, these large storms outliers must be removed from the raw climate model data before using the KDDM method or the transfer function method for bias-correction, or otherwise build into the method a step that reduces the effect of extreme values on the estimation of the CDFs used in bias-correction.

The change factor method, on the other hand, does not produce as unrealistic results as the other two methods. It is relatively insensitive to the effects of unrealistic storms in the raw climate model output because it uses a multiplicative factor that is calculated after the rainfall depth has been extracted from the GEV fit of each climate model. If both the future and historical climate model simulations have unrealistic values, only the change between the two models is portrayed by the change factor method. Since the ratio of raw climate model output is used for the change, one concern of using this method may be that it underestimates future extreme rainfall at the station scale, since raw climate model output (an areal average over the grid cell) typically underestimates historical rainfall at the station scale. Results from the change factor are not, however, consistently lower than the other two methods. They are sometimes higher, sometimes lower, and sometimes in between results from the other methods.

Since there is not a single method that produces consistently higher or lower values, and median values are similar among methods, any one of these methods could be used with relative confidence (assuming extreme outliers are accounted for in the KDDM and TF methods). However, for the remaining analyses in this study, only the change factor method is used because it produces more consistent results in these six cities. While it provides more consistent results, this does not necessarily mean the results are more accurate of future conditions, just less variable with respect to the other methods. More research is needed to determine metrics

for reliability and consistency of methods.

3.3.3   Influence of climate model resolution on expected change in future

This section evaluates how the two different climate model resolutions alter the predicted 50-year design storm depth, and whether any differences subsequently alter the size of the pipe designed using this storm. This design example is oversimplified and likely would not be used in practice. However, the simplicity of this design process is desired because the example is only intended to test sensitivity of the design storm, keeping all other variables constant. The simplicity aids in evaluating the direct impact of changes to the design storm.

Figure 3.4(a) presents changes to the 50-year design storm for two durations (1-hour and 24-hour) in the future (2020–2099). The percent change is presented relative to the observed 50-year, best-fit depth (1950–2013) for each city. Colors represent the two climate model resolutions and markers show different climate model (ESM-RCM) simulations. The mean of the ensemble is shown as a black star and the grey band represents a change of +/- 25% of the mean. Figure 3.4(b) presents the change in pipe diameter if the depths of rain from the 25-km or the 50-km ensemble were used to size the pipe, instead of the observed 50-year best-fit storm. Each marker represents the design value selected from the future range for the 50-year storm: the ensemble median (star), the ensemble upper bound (upward triangle), or the ensemble lower bound (downward triangle). The two climate model resolutions are again represented by different colors.

The first trend that emerges is that the future, 50-year storm calculated with the two spatial resolutions are different from each other. In 5 out of 6 cities, the change is larger for the 50-km resolution. Only in Birmingham does the 25-km ensemble lead to a larger average percent change and a larger uncertainty range than the 50-km ensemble. Differences between the two

resolutions are large enough to lead to an increase of one pipe size in 7 out of 12 cases (using the ensemble mean for design). The majority of these cases occurred in the pipes sized for the smaller watershed, because the magnitude and range of change is larger for shorter duration storms (the 1-hour design storm was used for the smaller watershed). For the larger watershed, only Phoenix and Boston would go up by one size using the 50-km ensemble (mean) instead of the 25-km ensemble (mean).

In most design scenarios, going up or down one pipe size is not a large change and could happen easily due to expected changes in land use conditions. However, a change of one pipe size could matter more in cities where the different resolutions would lead to a decision between keeping the pipe size the same, or increasing it. Using the 24-hour design storm, this would occur in the case of Phoenix, Boston, and Seattle because in these cities, not only is the magnitude of change uncertain, but also the direction of change.

With only 6 locations examined, it is not possible to provide a general recommendation on which spatial resolution should be used to create DDF curves. However, for these 6 cities only, the 50-km models should be used if the engineer wants to be conservative when designing for extreme events. On the other hand, the 25-km models should be used if the engineer prefers the smaller uncertainty range produced by the 25-km models.

Overall, however, the differences between the resolutions remain small when the ensemble mean is used as the design storm. When the upper bound is used instead of the mean, the differences in storm depth and pipe size are larger. Designing for the upper bound increases the pipe size by at least one interval when compared to designing for the ensemble mean. In some cities, the pipe size increases by two or three step changes. The following section explores how these changes in pipe size compare to historical design values, like designing for the historical 50-year 90% CI, or the historical 100-year storm.
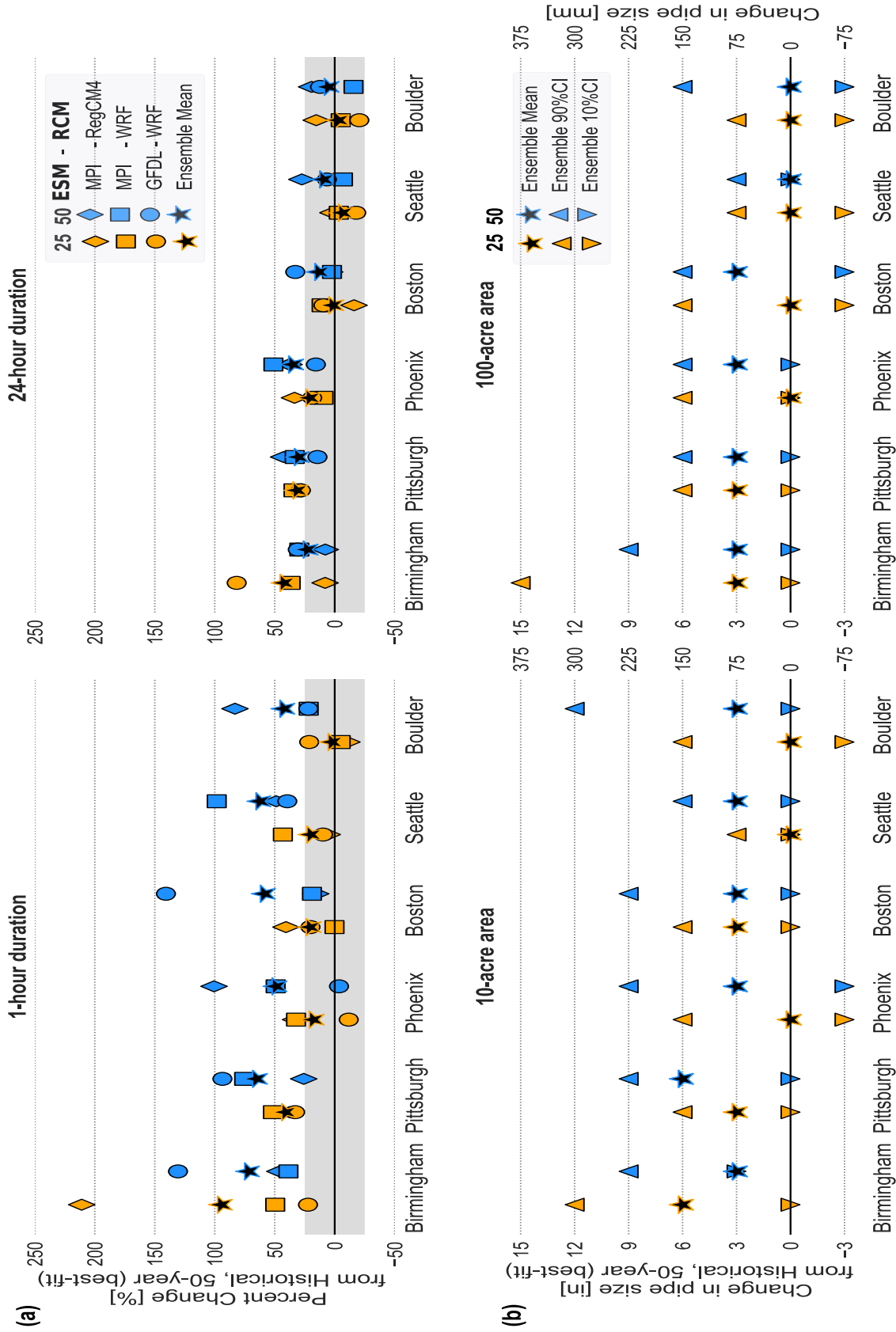
Figure 3.4: (a) Percent change in 50-year depth in future (2050–2099) relative to the historical, 50-year, best-fit depth (1950–2013) and (b) Change in pipe size above or below the historical, 50-year (best-fit) pipe size. Climate model resolution is represented by color: 25-km (orange) and 50-km (blue). The grey band represents a change of + 25%

79

3.3.4   Comparison of all design storm choices on stormwater infrastructure sizing

The previous sections identified how differences in the climate model resolution influenced the range of future rainfall depths and pipe sizes. This section identifies how these changes in pipe size compare to the pipe dimensions if the design engineer were to use values from the historical DDF curve. When selecting a historical design storm, the designers have several choices. They could assume no change will take place in the future and use the historical 50-year best-fit depth. If they wanted to assume some change in the future, but still rely on historical values, they could choose the historical 50-year upper bound, or the historical 100-year best-fit value. The resulting pipe sizes using these historical design storms are presented in Appendix A.

If the designers were to use climate projections to update the precipitation-frequency curves, they would again have several choices for a design storm: selecting the mean of the model ensemble or the upper bound. The lower bound is not recommended as a design storm choice since it could lead to an underestimation of future rainfall even if the historical rainfall conditions are maintained into the future. Figure 3.5 presents the resulting pipe diameters using various historical or future design storms for the 10-acre watershed. Each circle plot shows the results for a different city. The left side of each circle shows pipe sizes using historical (1950–2013) design values, while the right side shows future (2020–2099) design values that were calculated using the change factor technique. For the historical values, the different colors represent different return periods. For the future values, the different colors represent a different spatial resolution selected for the climate model ensemble. Solid lines represent the best-fit value, while the dotted lines show the upper bound.

These results show that increasing the design storm from the historical, 50-year (mid) storm to the historical 100-year (mid) storm does not always increase the pipe size. As a
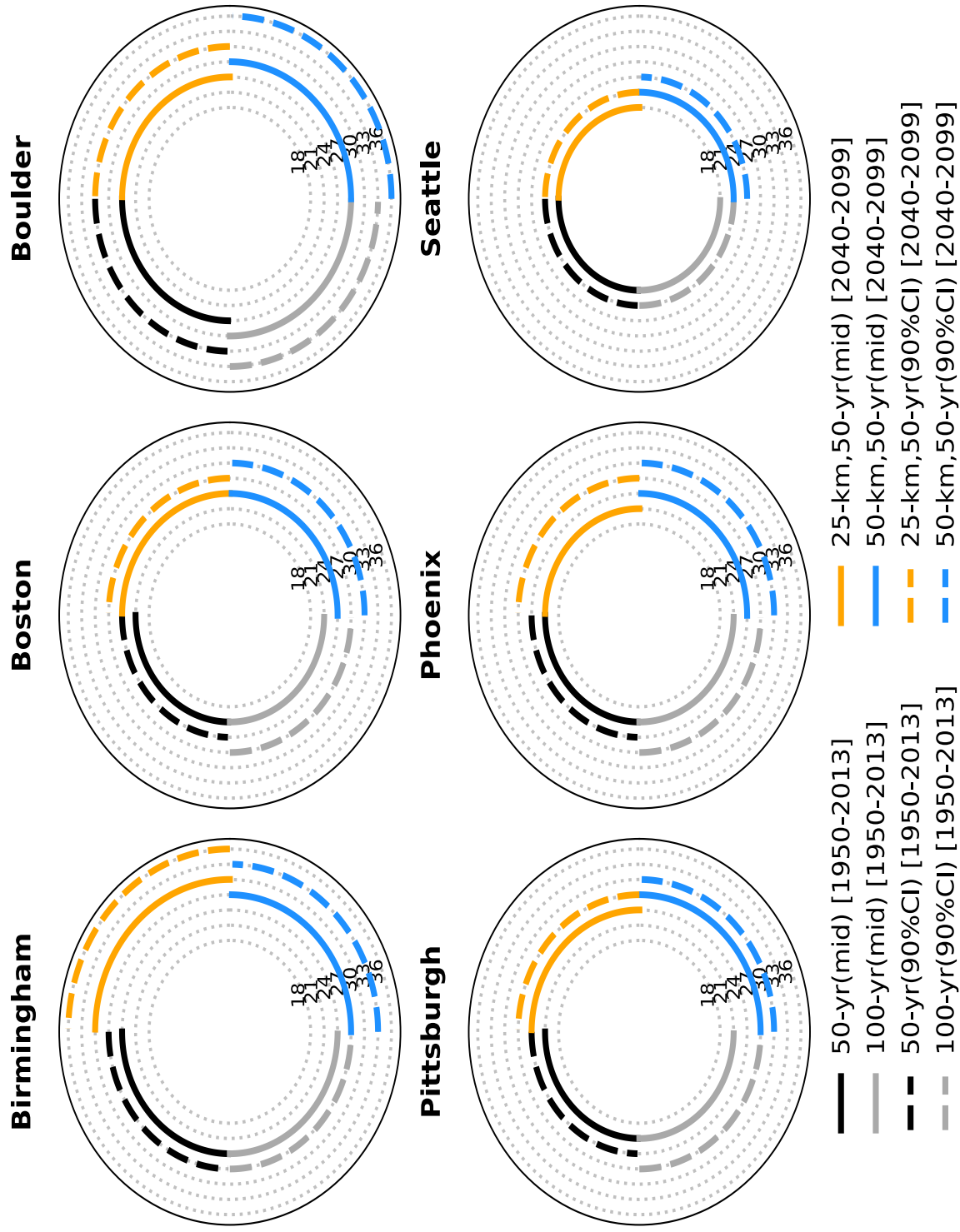
Figure 3.5: Pipe diameters in inches for the 10-acre watershed for each city and several design choices. The left side represents historical design values, while the right side shows future design values for different climate model ensemble resolutions.

result, using the 100-year storm instead of the 50-year does not provide additional protection against future extreme events. Designing for the historical, 50-year upper bound does however increase the pipe dimensions by at least one size for all cities, and thus will provide additional protection in the future.

The level of protection that using updated DDF curves will provide, relative to historical DDF curves, varies by city, climate model resolution, and by the point selected from the updated curve. For instance, in Boulder, Seattle, and Phoenix, the three cities that had small changes in extremes, the pipe dimensions do not change when the 50-year historical (mid) and 25-km based future ensemble (mid) are used for design. This means that using a design storm from the updated DDF curve does not always add additional protection in the future relative to designing for the historical, 50-year (mid) storm.

In addition, some of the historical design values are similar to those created from the climate models. In Seattle, designing with the 50-year historical, 100-year historical, and 25-km based future ensemble all lead to the same pipe dimensions; however, in Boulder, the 50-year historical storm leads to the same dimensions as the 25-km based future ensemble and the 100-year design value is consistent with the 50-km resolution ensemble. This means that without looking at results from the climate models specific to each city, it is impossible to know if the upper bound of the historical design values will provide sufficient protection in the future. In some cases, the upper bound of the historical 50-year storm will provide an equivalent level of protection as designing for the mean of the 25-km climate model ensemble, and in other cases, it will provide additional protection. For this design example and these six cities, the upper bound of the historical 100-year storm always leads to a lower level of protection than the upper bound of the 50-km climate model ensemble, and sometimes an equivalent level as the upper bound of the 25-km climate model ensemble.

3.4 Conclusions

This study investigates how different modeling choices used in the creation of climate-corrected depth-duration-frequency curves could alter design storms values and the dimensions of stormwater infrastructure that use these values. Changes were evaluated depending on (1) the choice of climate model spatial resolution, and (2) the choice of spatial adjustment technique used to apply trends from these models, in six U.S. cities: Birmingham, AL, Boston, MA, Boulder, CO, Pittsburgh, PA, Phoenix, AZ, and Seattle, WA.

Findings in the six cities demonstrate that shorter-duration storms (less than 6 hours) show a larger relative increase in intensity (and larger uncertainty range) in the future than longer duration storms (greater than 12 hours). These findings, which are consistent with previous results (Hassanzadeh et al. 2013; Kuo et al. 2015; Chandra et al. 2015; Shahabul Alam and Elshorbagy 2015), imply that stormwater infrastructure should be equipped to store or convey a higher intensity in a short amount of time rather than a lower intensity over a long period. The three correction methods evaluated (change factor, transfer function of the annual maximum series, and bias-correction of entire time series) led to nearly consistent median values in most cities and for most durations; however, differences that do occur get larger as the return period increases. A single method does not consistently over or under estimate extreme values, which suggests that any one of these methods could be used with relative confidence.

For these 6 cities, the lower spatial resolution (50-km) climate model ensemble generally provides a higher estimate of precipitation in the future than the 25-km ensemble. The differences between the 50-km and 25-km models are large enough to lead to different stormwater pipe dimensions.Thus the 50-km models should be used if the engineer wants to be conservative when designing for extreme events. On the other hand, the 25-km models should be used if the engineer prefers a smaller uncertainty range, which is what was produced by this

83

ensemble.

Overall, results from this study suggest that using the historical DDF curves for stormwater design will not be protective against future extreme events. However, if historical curves were to be used, adopting the upper bound of the historical storm depth provides more protection than doubling the return period. Nonetheless, this approach is not recommended because the additional protection provided from using the historical upper bound is independent of how much change is expected in the future. For this reason, updating DDF curves using climate model output is the recommended strategy for informing future design storms because these models are representative of expected future conditions. The resulting DDF curves and stormwater infrastructure designed from these curves are, however, sensitive to the choices made when creating these updated curves. This means that the design storm selected from an updated DDF curve is not guaranteed to provide the same level of protection in the future that a design storm from a historical curve provided under historical climate conditions. Engineers using updated DDF curves should consider these sensitivities during design and inform clients that performance of the infrastructure over its lifetime may vary as a result.

3.5   Acknowledgements

CHAPTER 4 USING RAINFALL MEASURES AND PERFORMANCE METRICS TO EVALUATE PERFORMANCE AND ADAPTATION OF GREEN INFRASTRUCTURE SYSTEMS UNDER CLIMATE CHANGE[1]

[1]Cook, L., Samaras, C., and J.M. VanBriesen (2018) Using rainfall measures and performance metrics to evaluate performance and adaptation of green infrastructure systems under climate change. In preparation.

Abstract

Climate change is expected to change the timing and magnitude of precipitation patterns in many areas, and little is known about how the performance of bio-retention basins will respond to these changing conditions. We simulate the hydrologic performance of a bio-retention system and evaluate which rainfall characteristics are most closely correlated with several performance metrics over a historical period. Metrics include percent of runoff captured, frequency and volume of discharge to the sewer, and drainage time of the surface layer. Using hourly, bias-corrected, regional climate model output from the NA-CORDEX project, we then project how performance would degrade in the future (2020–2059) and how performance correlations could change. In Pittsburgh, the system would have infiltrated 90 to 100% of runoff entering the system historically. In the future, the basin is expected to capture 5% less runoff that enters the basin (median value) because storms are expected to increase in intensity and duration. Historical performance is highly correlated with rainfall indices linked to mean rainfall intensity, and the magnitude, frequency, and total volume of extreme rainfall received by the basin. Future correlations between rainfall indices and performance metrics strengthen in most cases due to increases in extreme rainfall and variability; however, correlations weaken in some model simulations because most storms lead to consistently poor bio-retention basin performance.

Keywords: bio-retention basins, hydrologic performance, regional climate models, correlations

4.1   Introduction

Green stormwater infrastructure (GSI) is urban drainage infrastructure designed to increase infiltration and evapotranspiration of runoff, reducing inflow to sewer systems (U.S. EPA 2016). GSI is an alternative to conventional concrete or cast iron stormwater infrastructure, or grey infrastructure, which was designed to convey water quickly away from urban environments to storage, treatment, or discharge to a water body. GSI has been suggested as part of the solution to combined sewer overflows in a number of cities (Casal-Campos et al. 2015; Fischbach et al. 2017; Kloss, Calarusse, and Stoner 2006) and to help mediate the effects of increasing urban runoff expected as climatic conditions shift (Thakali et al. 2018; Demuzere et al. 2014; Foster, Lowe, and Winkelman 2011).

Bioretention basins, or "rain gardens," are one of the most widely implemented GSI, and have been identified by Leadership in Energy and Environmental Design (LEED) green building rating systems as a preferred practice for sustainable design (Davis et al. 2009). These basins can be used to abate runoff from impervious areas to comply with stormwater management targets (PWSA 2012) and/or to avoid stormwater fees put in place by local authorities, e.g., in Pennsylvania (CH2MHILL 2014; Gateway Engineers 2011), North Carolina (Environmental Finance Center at the UNC School of Government 2017), Maryland (Baltimore City Department Public Works 2017), and Ohio (Chagrin Watershed Partners 2017).

Research into the performance of bio-retention systems has shown promising results. Using on-site monitoring data, Dietz and Clausen (2005) report rain gardens in Connecticut captured 99.2% of input runoff over a period of 1-year. Davis (2008) assessed two bioretention cells near Washington DC by monitoring volume, peak flow and peak delay of runoff over 2.5 years. The timing of peak flow was delayed by a factor of 2 or more, and volume reduced by 44 to 63%. Further research by Davis et al (2012) showed that three systems in Pennsylvania, Maryland, and North Carolina completely contained small rainfall events and that discharge

from larger events was linear with respect to input volume (Davis et al. 2012). Chapman and Horner (2010) report slightly lower performance of a monitored bio-retention basin — 48 to 74% of incoming runoff was captured over a 2.5-year period (Chapman and Horner 2010).

In addition to performance observation, bio-retention basin simulation results suggest expected hydraulic behavior (Jennings, Berger, and Hale 2015). Using meteorological conditions at 35 different U.S. locations from 2012 to 2014, Jennings (2016) simulated the performance of a hypothetical bio-retention basin, considering different surface depths, storage volumes, and infiltration rates. Expected total runoff reduction ranged from 51.3 to 99.8%, with the least effectiveness along the East and Gulf Coast and most effectiveness in the Mid-Western regions. Poor performance was attributed to high rainfall totals and high intensity events; adjusting rain garden characteristics also altered performance (Jennings 2016).

While bio-retention systems show promising results in mediating effects of stormwater on urban areas, the prior research has focused on short term performance and assumed stable climatic conditions. Climate change is expected to increase the intensity of rainfall events (Brommer, Cerveny, and Balling Jr 2007; Groisman and Knight 2008; Groisman, Knight, and Karl 2012) with the largest increases expected in short duration events (less than a day) (Westra et al. 2014; Kuo, Gan, and Gizaw 2015). Predictions of future change are highly uncertain and bio-retention systems may require the use of robust or adaptable designs that allow for many possible future states to occur (Lempert 2010; Gregersen and Arnbjerg-Nielsen 2012; Lempert and Schlesinger 2001; Lempert and Groves 2010; De Neufville and Scholtes 2011). Monitoring data could be used to track performance over time; however, in the absence of monitoring data, it is possible that publicly available rainfall observations could be used as a proxy. A better understanding of the critical features of rainfall patterns (i.e., duration, intensity, inter-storm timing) that affect performance of GSI is needed in order to assess how future rainfall conditions may alter expected performance of long-lived infrastructure. In the present work, we develop a procedure to track performance of bio-retention basins over time using measures of annual rainfall. We then evaluate how these measures are expected to change in the

future and if the basin would perform as expected based on these changes.

4.2   Approach and Data

The goal of this analysis is to assess if annual measures of rainfall can be a useful indicator for tracking hydrologic performance of green infrastructure systems over time. Using rainfall measures as a way to track performance is of interest because rainfall data is usually more readily available than performance data from on-site monitoring stations. Bio-retention systems, or rain gardens, are used as an example system to test this process because they are one of the most widely implemented types of green infrastructure (Davis et al. 2009). The design characteristics of an existing rain garden system in Pittsburgh, PA are used in order to illustrate the procedure.

The first step of the analysis defines the performance metrics of interest that could be tracked over time. The second step uses observed rainfall data and continuous hydrologic simulation to estimate the historical hydrologic performance of the site over a 29-year period (1990 to 2018). The third step defines and calculates annual rainfall measures (indices) and establishes which of these indices is most indicative of annual performance. In step 4, we determine how these indices are expected to change in the future using output from climate models. The indices that are expected to change the most will be most important to track over time. Given these expected changes in rainfall, the final step evaluates if the basin would have performed as predicted based on the rainfall indices. Output from climate models is used for simulation of the bio-retention basin under future conditions (2020 to 2050). The following sections present more details about each of these steps.

4.2.1 Hydrologic performance metrics

Hydrologic performance of the rain garden was assessed using several metrics. The first metric, *Runoff Capture Efficiency*, depicts the fraction of water that does not reach the sewer system. This is calculated as the amount of water that infiltrates through the basin into the ground below, as a percentage of total stormwater that enters the site. This metric normalizes results compared to the amount of rainfall that was received in that year. It is calculated as the ratio of runoff infiltrating to the amount entering the basin.

The second metric is the *Surface Detention Time*, which calculates the length of time that water is stored on the surface of the bio-retention basin. Stormwater regulations in Pennsylvania and elsewhere state that basins must fully drain in less than 48 hours (PA DEP 2006), so this metric is used to assess how well these basins are meeting this requirement. Surface Detention Time is calculated as the continuous length of time that the depth of water on the surface is greater than 0. The basin is considered "drained" if water is not present on the surface for 30-minutes or longer. One may also be interested in the number of times per year that water ponds on the surface for longer than 48-hours, or some other threshold of interest. When the surface always drains in less time than the threshold, the maximum and average detention time for each year can be evaluated to understand how close the threshold is to being surpassed.

The third metric, *Volume of Discharge*, represents the total amount of stormwater that was discharged to the sewer system each year, either due to underdrain discharge or from excess surface runoff that was not captured by the garden. The final metric, *Frequency of Discharge*, represents the number of times per year that stormwater was discharged to the sewer system. This metric is calculated by counting the number of times that there was flow present in the outfall pipe from the system to the sewer. Both of these metrics may be of interest to stakeholders within combined sewer service areas since reducing on-site discharge could decrease the amount of total overflows to the river.

4.2.2   Simulation of historical performance

Example Site

The example site used for the simulation is located in Pittsburgh, Pennsylvania (40.46, -79.92) within the Allegheny River Basin. The site was selected because it has characteristics typical of bioretention basins, including a vegetated surface layer that allows for ponding, a subsurface soil layer that promotes infiltration, a gravel layer that provides additional subsurface storage and encourages infiltration, and an under drain to control releases to the sewer system (PA DEP 2006; Davis et al. 2009; Jennings 2016).

The site contains two rain gardens that collect water from adjacent roof and pavement areas. The first rain garden (RG1) collects runoff from an impervious area of 400 m^2. The second (RG2), which has the potential to overflow to the combined sewer system, collects water routed from RG1 as well as runoff from an additional impervious area of 377 m^2. Runoff that enters either rain garden is infiltrated and stored in a 61 cm engineered soil layer where it is available for uptake by the plants. Water that infiltrates through the soil layer enters a 30.5 cm gravel layer, which contains a 150 mm perforated underdrain. When inflow exceeds infiltration, water ponds on the RGs to a depth of 76 mm (3 in), enabled by the elevation difference between the top of the RG and the street. Ponded water can flow into a vertical surface drain that connects to the underdrain. The underdrain flow is controlled with a weir and an orifice before it can exit the basin. Figure 4.1 presents a profile view of the storage layers of the bio-retention basin.

The site is equipped with several gauges and sensors that measure rainfall, soil moisture, water level, and soil temperature data. Rain Garden 2 contains a Conductivity, Temperature, Depth (CTD) sensor, placed within a cylindrical screen at a 40-inch (1016-mm) depth, as well as two soil moisture sensors at 6-inch (152.4-mm) and 18-inch (457.2 mm) depths. Data have been collected on a 5-minute interval since July 2015.
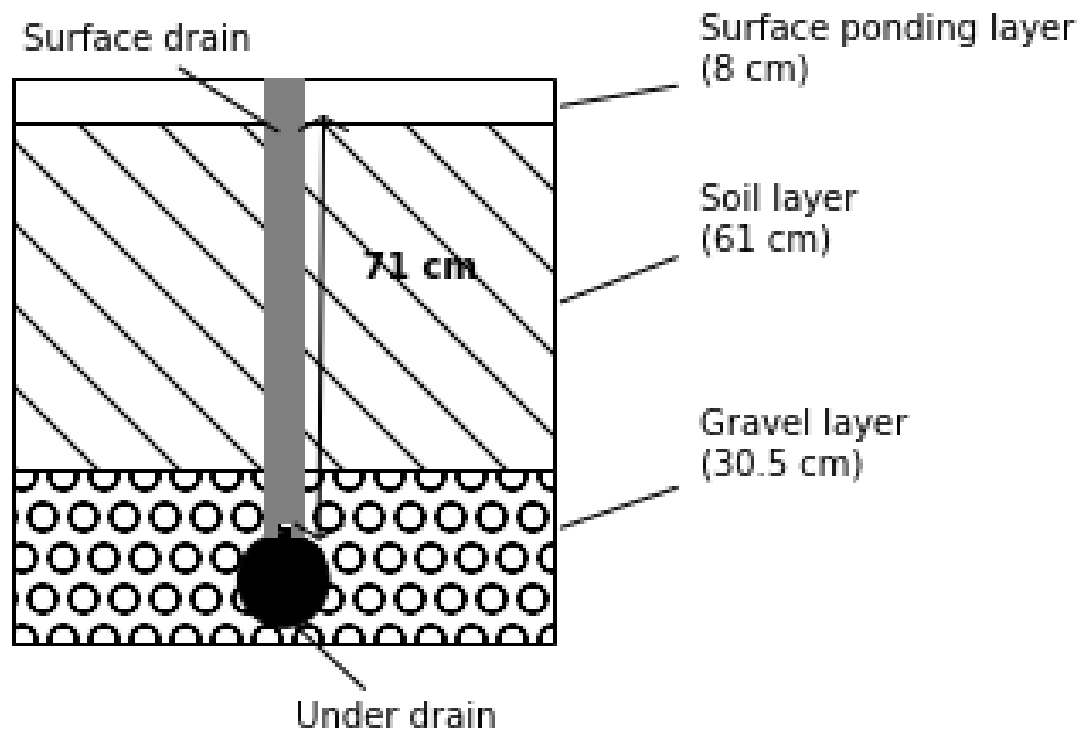
Figure 4.1: Profile view of the bio-retention layers, including surface and under drain

Observed Precipitation Data

Observed rainfall data from the period 1990 to 2018 were used to evaluate the hypothetical historical performance of the rain garden system. Hourly observed precipitation data for the period from 1990 to 2014 were obtained from the NOAA National Center for Environmental Information (NOAA 2016) for the station located at Pittsburgh International Airport. More recent data from 2014 to 2018 was obtained from a local rain gauge network maintained by 3 Rivers Wet Weather, Inc.; a station in close proximity to the airport was selected (3 Rivers Wet Weather 2018).

Continuous hydrologic simulation

The PC Stormwater Management Model (PCSWMM) Version 7.1 (Computational Hydraulics Int. 2018) is used to model the site using continuous hydrologic simulation, which is able to characterize system response over time and account for antecedent soil moisture conditions of consecutive storm events. PCSWMM was also selected for its ability to model multi-layer bioretention systems. Evapotranspiration is not considered in the model because prior studies have found that runoff reductions from evapotranspiration (ET) are negligible (Jennings, Berger, and Hale 2015; Jennings 2016). Jennings (2016) found that across 35 locations in the U.S. evapotranspiration from bio-retention basins only contributed to 0.16 to 1.06% of runoff reduction. The contribution is small because the process is only possible during daytime dry weather. The very small contribution from ET is within the bounds of uncertainty of the performance of the RG, and by excluding ET this analysis is slightly more conservative.

Eight sub-areas are modeled to represent four roof areas, two pavement areas, and two rain gardens (see Figure 4.2). The two rain gardens are modeled as bio-retention cells with an underdrain. Table B.2 in Appendix B presents the sub-basin characteristics and flow patterns. Although the project site contains two bio-retention basins, and both were modeled, in order to limit the scope and ease comprehension of results, only the basin that has the potential to drain to the sewer (RG2) will be evaluated for performance. Each performance metric is calculated annually from the simulation results for each year.

At the site, the outlet of the under drain of rain garden 1 flows into a solid HDPE pipe and then into a perforated pipe below rain garden 2. In the model, the perforated pipe is modeled as an underdrain (available with the bio-retention feature). Underdrain and surface flow of RG1 is routed to RG2. Underdrain and surface flow from RG2 is routed to a solid PVC pipe that discharges to the combined sewer system. The orifices and weirs controlling the underdrain flow are accounted for in the model using storage, orifice, and weir nodes.

Parameter values were estimated using a combination of pre-existing field sampling, as-
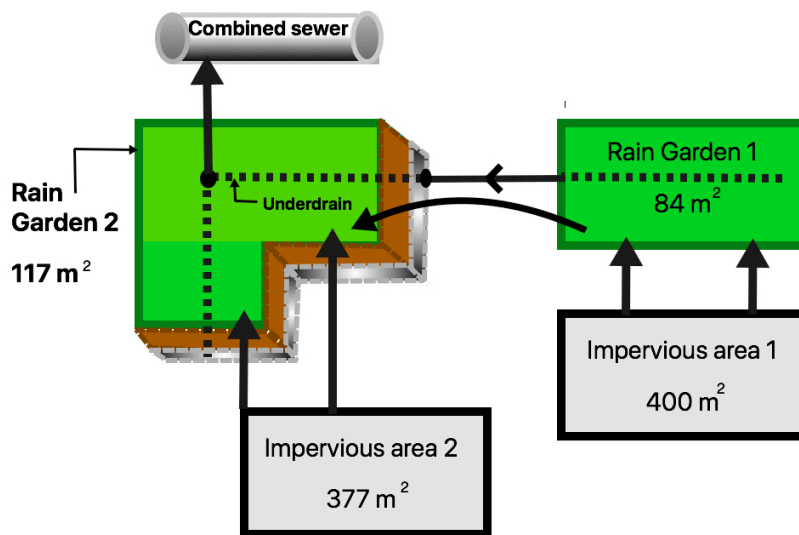
Figure 4.2: Overview of model configuration. The impervious areas drain to the surface of each garden. Rain garden 1 can drain to rain garden 2 through surface runoff or under drain flow.

built drawings, literature sources, and model default values. Post-construction soil samples of rain garden 1 report the soil as sand, with $86-88\%$ sand, 10% silt, and $2-4\%$ clay; whereas, samples from rain garden 2 classify the soil as loamy sand, with 86% sand, 8% silt, and 6% clay (A&L Great Lakes Laboratories, Inc 2015). The bio-retention basin of interest (RG2) has an assumed surface slope of 0.1% and a vegetative cover of 85%; the soil layer has an assumed conductivity of 2.5 in/hr; the gravel layer has an assumed seepage rate of 1.25 in/hr; and the under drain has an assumed discharge capacity of 0.5 in/hr. Additional parameter values related to the sub-catchments, bio-retention cell layers, and conduits are reported in Appendix B.

4.2.3 Rainfall indices most indicative of annual performance

Definition of Rainfall indices

To determine which rainfall indices could most affect rain garden performance, an exhaustive list of annual precipitation indices is defined and then calculated using the same observed rainfall data that was used in the historical simulation. Indices were calculated based on hourly and daily rainfall amounts. The list of indices was developed during exploration of the literature e.g., (Chen et al. 2015; Karl and Knight 1998) and recommendations from the Expert Team on Climate Change Detection Monitoring and Indices (Peterson et al. 2001; Karl, Nicholls, and Ghazi 1999). The initial set of indices examined for each year is listed in Table 4.1. Indices are grouped by what aspect of precipitation they would indicate, including: the central tendency of rainfall, magnitude, proportion, or frequency of extremes, or frequency of wet and dry days. The "set" column represents the different variations of each index that were tested. Each index is calculated for each year of historical rainfall data establishing a total of 28 data points for each index.

Correlation of indices and metrics

The correlation of the rainfall indices to the annual performance metrics is used to determine which indices are most indicative of annual performance. Correlation was tested using the Pearson correlation coefficient, which measures the linear relationship between two data sets that are normally distributed. The rainfall indices and performance metrics were verified to be normally distributed using the t-test (the null hypothesis could not be rejected with a p-value of 0.001). The p-value of the correlation was also calculated. This value, which is somewhat unreliable for small data sets (fewer than 500 values), represents the probability

Table 4.1: Rainfall indices considered in performance evaluation. Indices are classified as an indicator of the central tendency, magnitude of extremes, proportion of extremes, frequency of extremes, and frequency of wet and dry days

Index	Set	Description	Implication	Indicator
total	n/a	Total annual precipitation (mm)	Wet or dry year	Central tendency
nintd	n ∈ ($\mu, q50$)	Mean or median daily intensity of rain days (mm)	Intensity of rain days	Central tendency
ninthr	n ∈ ($\mu, q50$)	Mean or median hourly intensity of rain hours (mm)	Intensity of rain hours	Central tendency
maxnd	n ∈ (1, 2, 3) days	Greatest n-day total precipitation (mm)	Measure of extremes with duration less than or equal to drainage requirement	Extreme magnitude
qpd	p ∈ (90, 95, 99) quantile	pth percentile of rain day amounts (mm/day)	Intermediate to rare daily extremes	Extreme magnitude
qph	p ∈ (90) quantile	pth percentile of rain hour amounts (mm/hr)	Intermediate hourly extremes	Extreme magnitude
totqp	p ∈ (90, 95, 99) quantile	Total rain from daily qth percentile or greater (mm)	Total from intermediate to rare daily extremes	Extreme total
propqp	p ∈ (90, 95, 99) quantile	Proportion of total annual rainfall above qth percentile	Measure of intermediate to rare extremes relative to total annual rainfall	Extreme proportion
excdn	n ∈ (10, 25, 50) mm	Number of rain days with precipitation ≥ n mm (days)	Intermediate to rare daily extreme events	Extreme frequency
raind	n/a	Days per year where precipitation ≥ 0.1 mm (days)	Daily precipitation occurrence	Frequency wet & dry
ncwd	n ∈ (μ, max)	Mean or max no. of consecutive wet days where precipitation > 0.1 mm (days)	Measure of average to long-duration storms	Frequency wet & dry
ncwh	n ∈ (μ, max)	Mean or max no. of consecutive wet hours where precipitation > 0.01 mm (days)	Measure of average to long-duration storms	Frequency wet & dry
ncdd	n ∈ (μ, max)	Mean or max no. of consecutive dry days where precipitation < 0.1 mm (days)	Measure of risk of dryness and antecedent soil conditions	Frequency wet & dry

that an uncorrelated system would produce the same or larger correlation coefficient than was produced from these two datasets.

Correlations are quantified as negative (representing an inverse relationship) or positive, ranging from low (0 to 0.3), moderate (0.4 to 0.6), high (0.7 or above), and very high (0.9 or above). Indices that have high or very high correlations to the performance metrics were selected as the indices most indicative of performance. If several rainfall indices were highly correlated with the same performance metric, it is possible that these indices are correlated with each other. Correlations amongst indices were also tested at this point. If they were correlated, only one of these indices will be considered as a useful index to track.

Sensitivity analysis

Since simulation was used to calculate the performance metrics, correlations could be sensitive to the model parameters and the assumed characteristics of the installation. Most notably, the infiltration capacity of the rain garden affects how quickly rain can enter, and may affect correlations. Infiltration could decline over time due to clogging or poor maintenance and thus sensitivity was tested by decreasing the rate of infiltration into the basin. Soil conductivity was decreased to 1 in/hr, from 2.5 in/hr, and the gravel seepage rate was decreased to 0.5 in/hr from 1.25 in/hr. After simulating historical performance under these conditions, correlations of indices and performance metrics were reevaluated. Indices that had high correlations in both analyses are considered more indicative of performance than indices that were very sensitive to changes in infiltration rate. The sensitivity analysis was taken into account before the final list of indices recommended to use as indicators of performance was compiled.

4.2.4   Expected Changes in Future Rainfall

Source of Climate Model Output

This study uses output from regional climate models (RCM) from the NA-CORDEX project (Mearns et al. 2017) to evaluate anticipated future changes in rainfall. NA-CORDEX is a compilation of standardized regional climate model simulations available at an hourly time step, for two different spatial resolutions (50-km and 25-km), over the continuous time period from 1950–2100. These models were chosen over other types of downscaled climate models because of their availability at a 1-hour time step (Cook, Anderson, and Samaras 2017). A sub-daily time step is crucial for continuous hydrologic simulation models that capture highly variable, localized, and short temporal scales of rainfall-runoff interactions (Durrans et al. 1999).

The RCMs in the NA-CORDEX project use Earth System Models (ESMs) as inputs. ESMs use more complex relationships than those in older Atmosphere-Ocean Global Circulation Models (AOGCMs) (Heavens, Ward, and Natalie 2013). At the time of the present analysis, four RCM-ESM combinations were available at the 1-hour timestep (presented in Table B.5 of Appendix A.3). This study uses all four of these combinations simulated at the 50-km resolution using the RCP 8.5 emissions pathway, which is the scenario with the highest greenhouse gas emissions (Riahi et al. 2011). The 50-km resolution was chosen over the 25-km resolution as a more conservative analysis resolution since the 50-km-based simulations predicted higher rainfall intensities for the maximum annual hourly and 1-day precipitations (Cook, McGinnis, and Samaras 2018).

Bias-Correction of Climate Model Output

The raw RCM/ESM simulations are an areal average of precipitation across a grid cell (50 km x 50 km). These data are not representative of rainfall values at the station (or city) scale, and thus must be adjusted to match the scale of the observed values. The method used in this analysis to adjust the RCM-ESM simulations to the station scale is called Kernel Density

Distribution Mapping (KDDM) (McGinnis, Nychka, and Mearns 2015). Hourly data from the period 1950 to 2010 were used for this bias-correction. Data were obtained from the NOAA National Center for Environmental Information (NOAA 2016) for the station located at Pittsburgh International Airport.

This method is a type of non-parametric bias-correction that uses a relationship between the observed rainfall time series (1950–2013) and the gridded climate model time series for the historical time period (1950–2013) to adjust the entire gridded climate model time series (1950–2099) to the station scale. The relationship between the observed rainfall time series and the gridded climate model time series is defined by fitting a transfer function between their empirical cumulative probability distribution functions (CDFs). First, empirical probability density functions (PDFs) are computed using kernel density estimation; these PDFs are then integrated using the trapezoid rule to calculate CDFs. Equal points of probability from the CDFs are mapped against each other and the resultant mapping is then fitted with a spline. The equation of this spline is the transfer function between the observed data and the historical climate model simulation output. The function is then applied to the 150-year time series of the climate model simulations to obtain bias-corrected values at the station scale.

After this correction, the statistical distribution of the observations should be more consistent with the statistical distribution of the historical period of the bias-corrected data. The major benefit of using the KDDM method to convert an RCM timeseries to a station scale is that the timing of rainfall events simulated from the regional climate models at the sub-daily level are maintained in the bias-corrected data. This allows for analysis of future rain garden performance not only based on increases in volume or intensity, but also due to changes in the frequency of storm arrival that are portrayed in the climate models.

Estimating future changes

To estimate changes in the future, rainfall indices were calculated for each year of the future, bias-corrected climate data. The percent change was calculated for each index and each year with respect to the median of the historical rainfall indices over the 29-year period (1990 – 2018). Indices that are both indicative over time and expected to change substantially in the future are considered as most important to track and predict performance over time.

4.2.5   Simulating future performance

Each bias-corrected climate model simulation was used in the hydrologic simulation model to test performance in the future. Performance metrics are again calculated for each simulation. If performance changes as expected based on the changes in the rainfall indices, then it is possible to use this index to track performance over time.

4.2.6   Model calibration and time periods

Several simulations were conducted using the hydrologic simulation model, including: one validation simulation, one historical simulation, and four simulations for the future period (one for each climate model simulation). Table 4.2 presents the time periods, dates, and data sources for each analysis.

The calibration simulation uses the 5-minute rain gauge data collected on-site to simulate performance over the recent 3-year period compares. Simulated performance is compared to observed performance over the same time period, which is calculated from the water level sensors on-site. PCSWMM model parameters for the bio-retention basins were adjusted until an accuracy of +/-15% was achieved. Over the 3-year period, observed performance was estimated to be 97 – 99 percent captured and simulated performance was simulated as 84 – 94 percent captured.

Table 4.2: Time periods evaluated in analysis

Time period	Dates	Time step	Data source
Calibration	July 2015 to September 2018	5-minute	On-site rain gauge
Historical	January 1990 to September 2018	1-hour	NOAA; 3-Rivers Wet Weather
Future	January 2020 to December 2069	1-hour	4 NA-CORDEX RCM/ESM simulations

4.3   Results and Discussion

4.3.1   Simulated historical rain garden performance

The following section presents simulated performance results for a bio-retention basin with a surface layer of 3 inches, a conductivity of 2.5 in/hr in the soil layer, a seepage rate of 1.25 in/hr of gravel layer, and an under drain rate of 0.5 in/hr. Figure 4.3 presents the simulated performance of the rain garden over the historical period, 1990 to 2018. The top panel (a) shows the percent of runoff captured per year as a time-series. The color represents the proportion of total rainfall ≥ 95th quantile in each year. Panel (b) shows the maximum number of hours to drain the surface (time series) and the daily 99th quantile as the color. Panel (c) presents the frequency of overflows discharged to the sewer and the number of rain days above 10 mm as the color. For reference, the marker represents the volume of rainfall that was greater than or equal to the 95th quantile, and is the same in each subplot. The box and whisker plot at the right of the graphic summarizes the statistics of each performance metric over the 28-year period. The red line (in the box and in the plot) represents the median, the box outline shows the 25th and 75th quantiles, and the whiskers show the 5th and 95th quantiles. Black dots represent outliers beyond the 5th and 95th quantiles.

As simulated, the rain garden would have been expected to perform well, capturing a me-
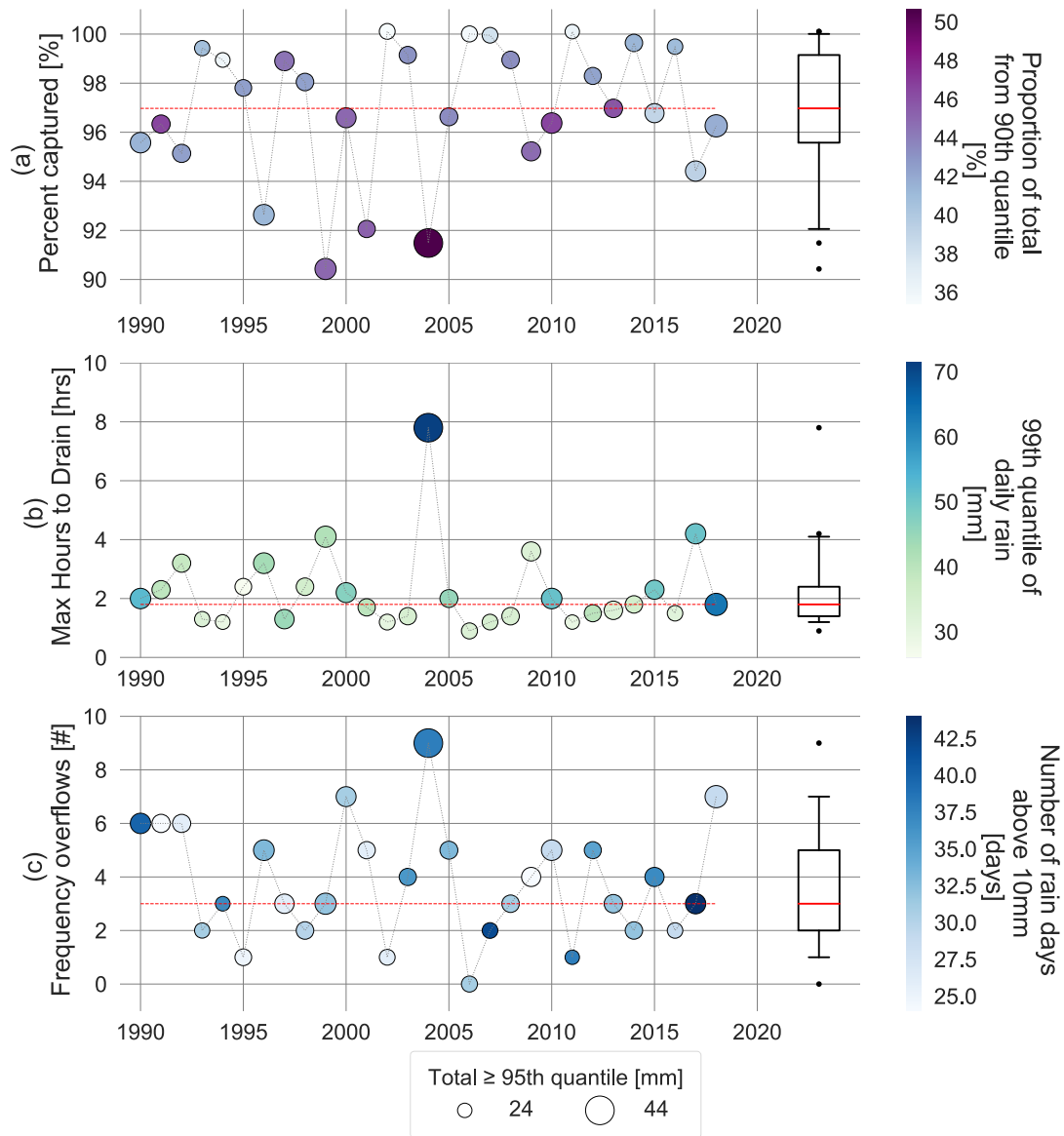
Figure 4.3: Time series of performance for historical period 1990–2018 of (a) percent of runoff captured over time, (b) maximum number of hours to drain the surface, and (c) the frequency of overflows discharged to the sewer.

dian of 97% of runoff each year, and a range of 90% to 100%. The surface of the basin was simulated to drain in less than 48-hours under all conditions, and thus would not have exceeded any regulatory thresholds for Pennsylvania. In the majority of years simulated, the surface always drained in fewer than 2-hours, and drainage time never exceeded 8-hours. Simulation results estimate that there would have been zero to nine discharges to the sewer each year (median of 3), with a median discharge volume of 13 m^3 per year. These historical simulation values can be used to set thresholds for performance degradation over time. For example, stakeholders may want to trigger adaptation when percent capture falls below 70% per year, or some other threshold that is considered not satisfactory.

Historical simulation results can also be used to determine how performance differs between metrics. For instance, the highest annual discharge volume of 67 m^3 did not occur in the year with the lowest capture rate (90%), even though capture rate in that year was one of the lowest (92%). This means that a rain garden might capture more rainfall overall if there is more rainfall to capture, but will still discharge a large amount to the sewer. Since capture efficiency is calculated relative to rainfall received, it may be the preferred metric for evaluation because it can be used for comparison to installations at other locations. However, maintaining a high percent capture does not mean that performance will remain high for the other metrics. The metrics that are most important to track will depend on stakeholder objectives and preferences.

4.3.2 Selection of rainfall indices most indicative of historical performance

In order for rainfall indices to be used as a proxy for tracking performance, they must be selected based on the individual performance metric of interest to the stakeholder. The rainfall indices that are most indicative of performance are determined by evaluating their correla-

tion to the individual performance metrics under historical conditions, taking into account sensitivity of these results to the simulation model parameters. If multiple performance metrics are of interest to the stakeholder, then multiple rainfall indices may need to be tracked. One option is to use the rainfall indices selected for each performance metric. Another option is to use a few rainfall indices that are most indicative of multiple performance metrics.

This section identifies the rainfall indices that are most indicative of individual performance metrics as well as all performance metrics as a whole. Figure 4.4 presents a heat map of the correlations between each performance metric and each rainfall index. Orange colors represent a negative correlation and greys represent a positive one.

Rain garden capture efficiency is highly, negatively correlated to many rainfall indices linked to extreme rainfall: maximum 1-day, 2-day, and 5-day rainfall, the total rainfall from ≥ 99th quantile, and the proportion of the total from ≥ the 95th and 99th quantile. This means that percent capture will decrease as the magnitude or proportion of extreme rainfall increases. All of these indices are highly correlated with each other (refer to Appendix B Figure B.4); thus they will not all be useful.

Volume of overflows to the sewer system is highly, positively correlated to the total ≥ the 95th and 99th quantile. Both of these indices are, however, correlated to each other and only one will be selected for further analysis. Overflow frequency is highly correlated to the hourly mean rainfall, as well as several indices relating to the magnitude, total, proportion and frequency of extremes. These results suggests that the more rain and more extreme rain that the basin receives, the more overflows that will occur, which is expected.

Maximum surface drainage time (indicating the longest time that the bio-retention basin took to drain) has the highest correlation out of all 5 metrics. It is most strongly correlated to very rare extreme storm metrics, like max 1-, 2-, and 5-day, 99th quantile, and the number of days with rainfall greater than 50 mm. This is expected since it is the very large (but rare) events that will lead to the longest surface detention time as rainfall enters faster than it can drain. Mean surface drainage time is only moderately correlated to many of the rainfall in-
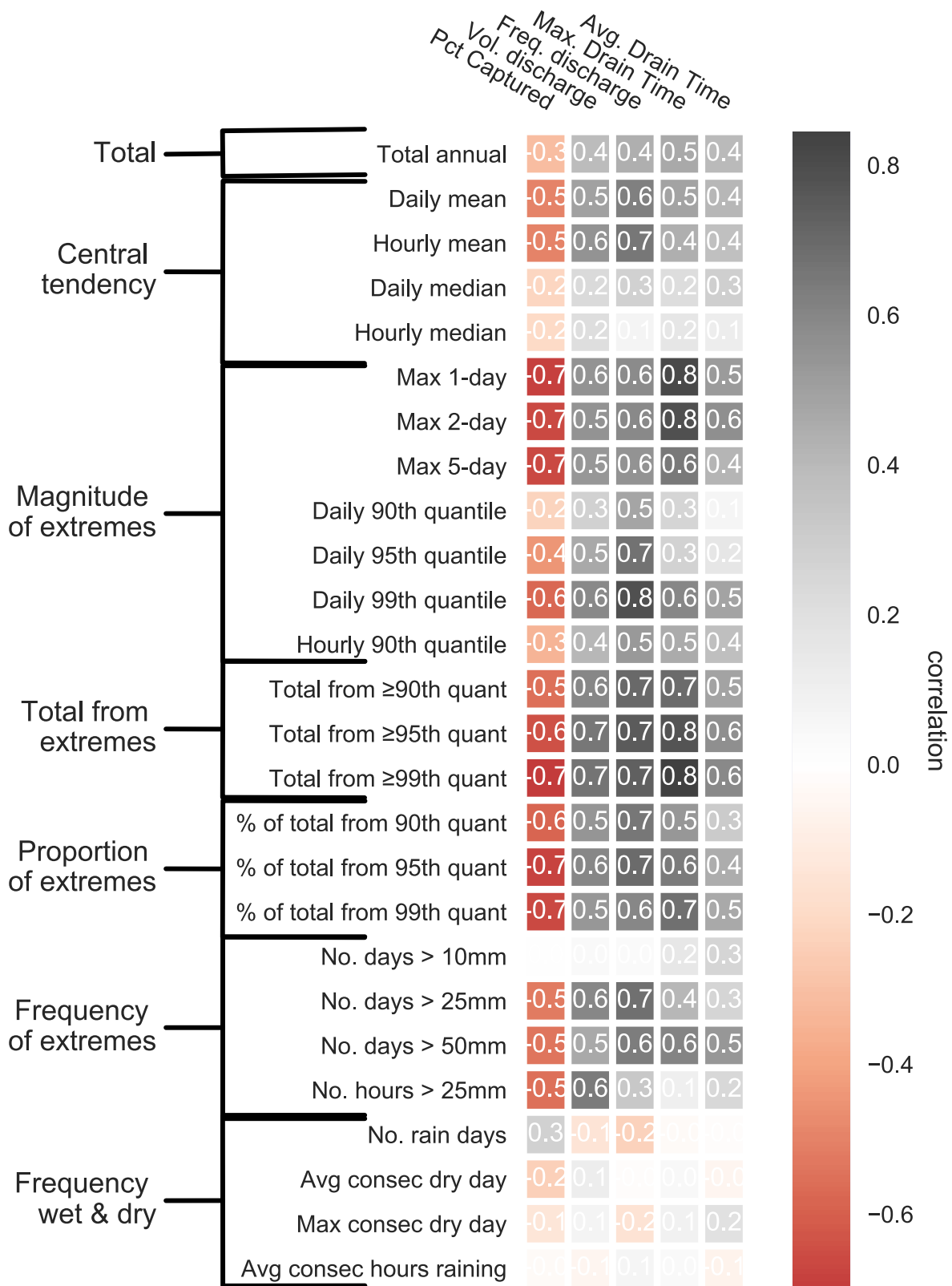
Figure 4.4: Correlations between rainfall indices and all performance metrics. Grey colors represent positive correlations, while reds represent negative correlations.

dices. The highest correlations (of 0.6) are to the maximum 2-day rainfall and the total rainfall from ≥ the 95th and 99th quantile.

The three indices with the highest, average correlation to all metrics were mean hourly, mean daily, and the daily 99th quantile of rainfall, which was very highly correlated to the total rainfall from ≥ 99th quantile and the number of rain days ≥ 40 mm. Total rainfall from ≥ 99th quantile has the largest number of very high correlations (to max 1- and 2-day, 99th quantile, and no. days ≥ 40 mm). High, positive correlations are present between the maximum 1, 2, and 5-day rainfall values, and between mean hourly and daily rainfall (as expected). Very high correlations exist between the total annual precipitation and the total rainfall from ≥ 99th quantile, the hourly mean and the hourly 90th quantile. Since the proportion of the total rainfall greater than the 90th quantile is weakly correlated to the total annual and the 90th quantile magnitude, all three of these indices will be used in the performance analysis to better understand how each one relates to performance.

Many of the indices that are highly correlated with each other are from the same group of indicators, or measures, of rainfall (presented in Table 4.1). These measures include the central tendency of rainfall, magnitude, proportion, or frequency of extremes, and frequency of wet and dry days. Only one index from each group will be recommended for use as a measure to track. Furthermore, if indices from unrelated groups are also correlated, only one or two of these may be of interest to track.

Sensitivity of historical results

Sensitivity of the results for the historical period was tested by decreasing the rate of infiltration into the basin (see Section 4.2.3). In this simulation, performance declines considerably. Capture efficiency drops from a median of 97% to a range from 65 to 91%. Volume of overflows increased from a median of 13 to a median of 120 m^3 per year. Frequency of overflows increased from a median of 3 to up to 30 events per year. The maximum drainage time dou-

bles, but still never exceeds 16-hours. As a result of these changes, many correlations between performance metrics and rainfall indices also change. Figure 4.5 presents how the correlations change. Thick outlines surround the combinations that were most highly correlated in the original simulation. Green colors represent correlations that increase while reds represent correlations that decrease. In general, correlations increase for capture efficiency and volume of overflows. Correlations for frequency of overflows mainly decrease. Correlations for mean drainage time also decrease. For maximum drainage time, correlations both increase and decrease. The following section selects the rainfall indices most related to performance during historical simulation based on these changes.

Rainfall indices most related to performance during historical simulation

Out of the five rainfall indices that were highly correlated to capture efficiency, the proportion of total from ≥ 95th quantile and the total from ≥ the 99th quantile have the highest average correlations between the two analyses. Both indices may not need to be tracked, however, since they are highly correlated to each other (0.8). The proportion of total from ≥ 95th quantile is recommended in this case since it has slightly higher average correlations.

Both of the rainfall indices that were highly correlated to volume of overflows are very highly correlated. However, the total from ≥ the 95th quantile has the highest average correlation between the two analyses. It is thus recommended for use as the index most indicative of volume of overflows.

Out of the nine rainfall indices that were highly correlated to overflow frequency, only one has a stronger correlation in the second analysis (with lower infiltration capacity). This index, which also has the highest mean correlation, is the number of days with rainfall greater than 25 mm. This index is thus the most indicative of frequency of overflows.

Out of the seven rainfall indices that were highly correlated to maximum surface drainage

107

	Percent Captured	Volume Overflows	Frequency Overflows	Max Drain Time	Mean Drain Time
Total annual	0.09	0.21	0.26	0.13	-0.30
Daily Mean	0.10	0.21	-0.02	0.21	-0.21
Hourly Mean	0.21	0.15	-0.05	0.26	-0.17
50th daily quantile	0.08	0.17	0.02	0.28	-0.29
50th hourly quantile	0.00	0.09	0.43	0.04	-0.41
Max1day	0.02	0.15	-0.48	-0.11	-0.02
Max2day	0.04	0.15	-0.39	0.01	-0.16
Max5day	0.04	0.17	-0.25	0.06	-0.12
90 daily quantile	0.18	0.32	0.12	0.15	0.14
95 daily quantile	0.16	0.24	-0.06	0.03	0.04
99 daily quantile	0.23	0.11	-0.48	0.20	-0.09
99 hourly quantile	0.26	0.19	-0.03	0.25	-0.07
Total ≥ 90 quantile	0.15	0.10	-0.20	0.01	-0.09
Total ≥ 95 quantile	0.06	0.15	-0.25	0.03	-0.16
Total ≥ 99 quantile	0.10	0.04	-0.33	-0.04	-0.09
Proportion ≥ 90 quantile	0.12	-0.05	-0.45	-0.14	0.29
Proportion ≥ 95 quantile	0.13	0.00	-0.69	-0.04	0.15
Proportion ≥ 99 quantile	0.04	0.06	-0.60	-0.07	0.04
No. days ≥ 10 mm	0.00	0.27	-0.23	0.04	-0.45
No. days ≥ 25 mm	0.18	0.10	0.03	0.09	0.04
No. days ≥ 50 mm	0.17	0.13	-0.03	0.20	-0.14
Rain hours ≥ 10 mm	-0.05	-0.34	-0.03	0.21	0.07
No. wet days	0.02	-0.05	0.43	-0.07	-0.19
Avg no. dry days	-0.04	-0.01	0.22	-0.05	0.15
Max no. dry days	-0.14	-0.17	-0.05	-0.29	-0.29
Avg consecutive wet hrs	-0.03	0.35	0.14	0.16	0.07

Figure 4.5: Changes in correlations between rainfall indices and all performance metrics when the infiltration rate is decreased during simulation. Thick box outlines represent the combinations that were most highly correlation in the original simulation. Green colors represent correlations that increase while reds represent correlations that decrease.

time, there are two indices that have comparable average correlations: the maximum 5-day rainfall and the total from $\geq$ 95th quantile. These two indices are only moderately correlated to each other (0.6) and could both be used. However, maximum 5-day rainfall is selected here because the average correlation between both analyses is slightly higher.

Out of the three rainfall indices that were highly correlated to mean surface drainage time, all three have weaker correlations when the infiltration rate is decreased. However, the total from $\geq$ 99th quantile has the smallest decrease and highest average correlation. This index is thus the most indicative of mean surface drainage time.

Overall, the four indices that have the highest average correlations between all performance metrics are: the total from $\geq$ 95th quantile, the total from $\geq$ 99th quantile, the 99th quantile, and the maximum 2-day rainfall. All of these indices are highly correlated to each other. For this analysis, the total from $\geq$ 99th quantile is selected as the most indicative of all performance metrics because its correlations are slightly higher than the others. However, selection of any one of these indices would be a decent indicator of performance. The results for all metrics are summarized in Table 4.3. The index in bold represents the index that is most indicative of the performance metric in this analysis.

4.3.3   Expected changes in future rainfall

This section summarizes expected changes in precipitation in the future. Results from all four climate model simulations suggest that total precipitation in Pittsburgh will increase, despite an expected decline in the number of rain days, and an increase in the average length of dry periods. This means that each rain storm will produce more rainfall, which is demonstrated by the increase in the maximum 1-, 2-, and 5-day storm magnitudes, as well as the magnitude of the 90th quantile. The proportion of total rainfall from extreme events (greater than the 90th quantile) is expected to increase by about 8% (median value). The annual frequency of rare and very rare events is also expected to increase.

Table 4.3: Rainfall indices most indicative of performance for historical period. The index in bold represents the index that is most indicative of the performance metric in this analysis.

Performance Metric	Rainfall Indices
Runoff Capture Efficiency	**Proportion from ≥ 95th quantile** Total from ≥ the 99th quantile
Volume Overflows	**Total from ≥ the 95th quantile** Total from ≥ the 99th quantile
Frequency overflows	**No. days where rainfall ≥ 25 mm**
Max surface drainage time	Maximum 2-day rainfall **Total from ≥ the 95th quantile** Total from ≥ the 99th quantile
Mean surface drainage time	**Maximum 5-day rainfall** Total from ≥ the 95th quantile

Subsequent analyses will more closely examine performance of only two climate model simulations: CanRCM4/ CanESM2 (CANCAN) and RegCM4/MPI-ESM-LR (MPIREG). The CAN-CAN model predicts large increases in both magnitude and variability of rainfall conditions, while the MPIREG model predicts increases in magnitude of extreme precipitation, but with less variability when compared to the other model simulation. The CANCAN model, referred to now as the *variable* simulation, has the largest annual variability in changes to total annual rainfall and proportion of total from ≥ 95 quantile, and has the largest increase in maximum 1-day rainfall. The MPIREG model, which will now be referred to as the *consistent* simulation, has the smallest range of change in total annual rainfall and proportion of total from ≥ 95 quantile, and the lowest increase in the median of the maximum 1-day rainfall.

These differences are apparent in Figure 4.6, which presents the predicted percent change of each rainfall index for the two selected climate model simulations. A comparison of the four model simulations is presented in Appendix B. The percent change was calculated for

each future year with respect to the median over the 28-year historical period. The box and whisker plots represent the range of the percent change for each year and each climate model for the future period, 2020 to 2059. The red line of the box plot represents the median over the 40-year future period; the box outline shows the 25th and 75th quantiles, and the whiskers show the 5th and 95th quantiles. Outliers are shown as orange dots (for MPIREG) and green stars (for CANCAN) The grey box represents the historical range (in terms of percent change from the median). The black dashed line at zero represents no future change, yellow bands represents a change of +/- 50%, and blue bands a change of +/- 25%. The historical median values for each index are shown at the bottom of the figure in bold for reference.

In general, the variable scenario (CANCAN) predicts higher increases in the maximum rainfall amounts, while the consistent scenario (REGMPI) predicts higher increases in 90th and 99th daily quantiles. Proportions and total volume from extremes are higher in the variable scenario. Both scenarios show similar increases in the rain days above 25 mm and 50 mm, yet the consistent scenario has smaller changes in total annual rainfall. In both scenarios, rain days above 10 mm decrease, despite an increase in days above 25 mm and 50 mm. The variable scenario loses more rain days > 10 mm than the consistent scenario. With fewer rain days > 10 mm and more rain days > 25 mm, the distribution rare event frequency in the variable scenario is shifting to the right. This shift is less pronounced in the consistent simulation. Rare and very rare rainfall is predicted to occur so much more often in the variable simulation that total annual rainfall still increases, despite a decline in the days with lower amounts of rainfall (10 – 25 mm). Overall, the consistent scenario predicts longer, intense rain bursts that are separated by longer dry periods. The variable scenario has a wider range of precipitation magnitudes, frequencies, and duration of storm events.

All of the rainfall indices most indicative of performance, including proportion from ≥ 95th quantile, total from ≥ the 95th quantile, no. days where rainfall ≥ 25 mm, and maximum 5-day rainfall, are expected to increase in both model simulations. As a result, performance would be expected to degrade. Percent capture would be expected to decline in the future

111
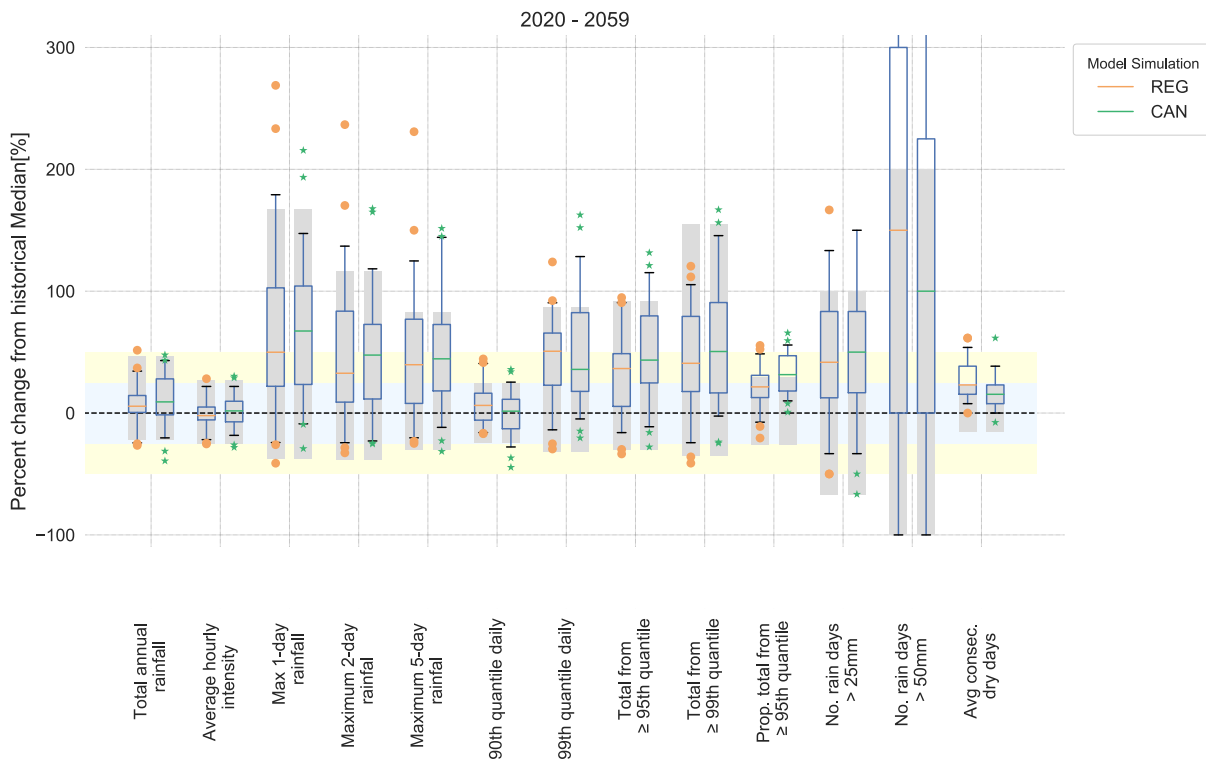
Figure 4.6: Percent change in rainfall indices for two climate model simulations, consistent (orange) and variable (green). The box and whisker plots represent the range of the percent change for each year and each climate model for the future period, 2020 to 2059. The red line of the box plot represents the median over the 40-year future period; the box outline shows the 25th and 75th quantiles, and the whiskers show the 5th and 95th quantiles. Outliers are shown as orange dots (for the consistent simulation) and green stars (for the variable simulation) The grey box represents the historical range (in terms of percent change from the median). The black dashed line at zero represents no future change, yellow bands represents a change of +/- 50%, and blue bands a change of +/- 25%. The historical median values for each index are shown at the bottom of the figure in bold for reference.

because these indices are increasing. Volume and frequency of overflows would be expected to increase, and as well drainage time.

4.3.4   Simulated future rain garden performance

This section presents the change in performance metrics for the two selected climate simulations: (1) variable increases in precipitation and (2) consistent increases in precipitation. Simulated performance of the bio-retention using these scenarios is compared to simulated historical performance. Figure 4.7 presents results for five performance metrics, including: percent of runoff captured, average and maximum hours to drain the surface, and the frequency and volume of overflows for each different simulation (historical, consistent, and variable scenarios). The y-axis shows the cumulative probability (%) and the x-axis shows the magnitude of each performance metric in percent, hours, number, and volume (m^3), respectively for the four metrics. Subplot (b) contains the average and maximum hours to drain the surface as dotted and solid lines, respectively.

Overall, the performance degrades in the future, as would have been expected based on the rainfall indices. However, despite the degradation, the basin still performs fairly well in terms of capture efficiency. Median capture efficiency decreases from 97% to 91 - 93%, depending on the simulation. In some years, performance is still high; however, in some years performance declines to less than 85% capture, but never drops below 70% in either simulation. Although the performance is fairly high, the difference between performance in the past and both future simulations is statistically significantly different (p = 0.001).

The two future simulations differ in how the percent of runoff captured is distributed annually. The consistent changes scenario, with less variability in precipitation, predicts a tighter range in percent of runoff captured than the variable scenario, which has more cases of both higher and lower performance years. These results suggest that the basin will capture less runoff overall, relative to the amount received; however, annual fluctuations in perfor-

113
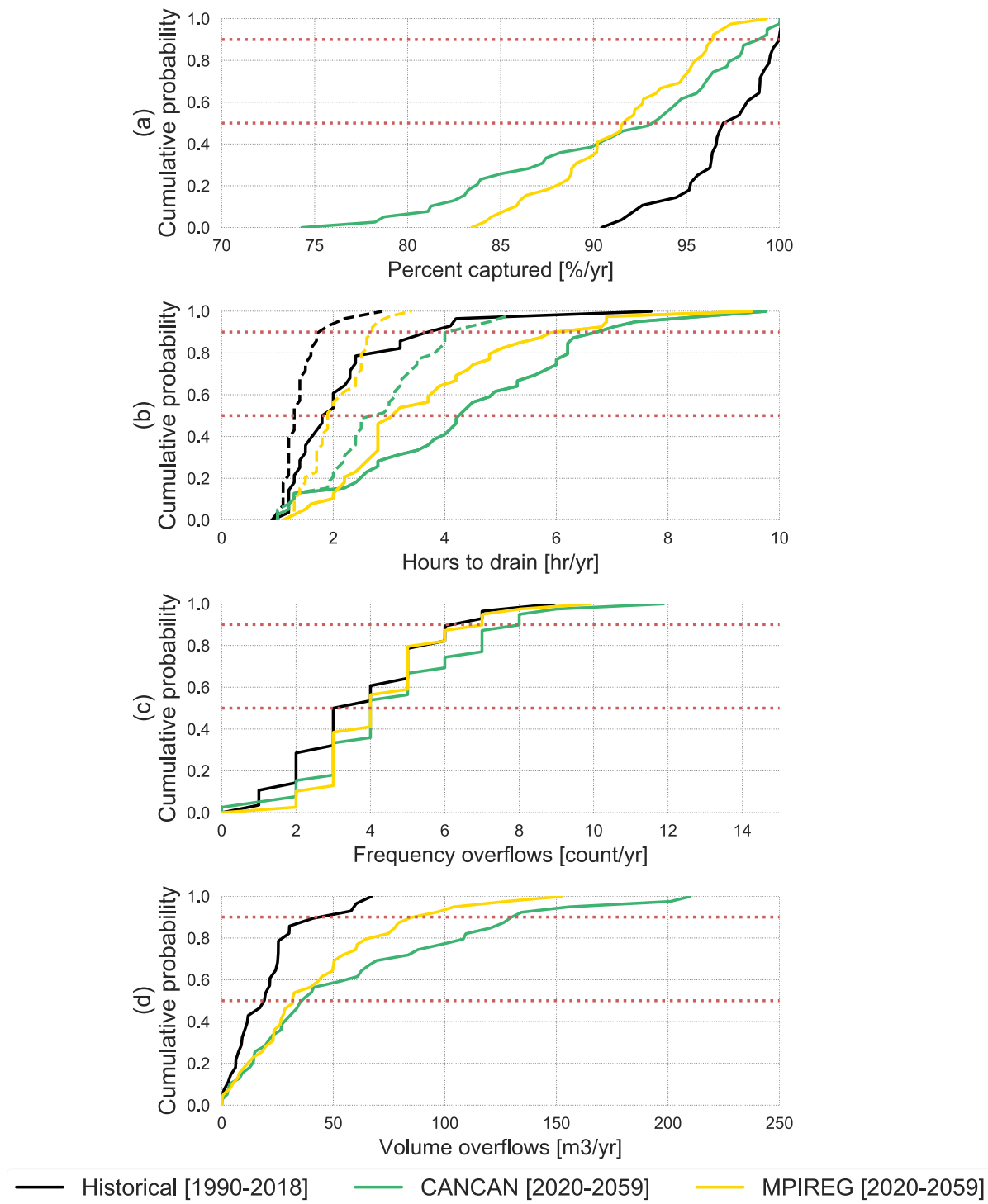
Figure 4.7: CDFs of five performance metrics: percent of runoff captured, the frequency and volume of overflows, maximum hours to drain the surface (solid lines) and mean hours to drain surface (dashed lines) for all simulations: historical (black), modest future change (yellow), and volatile future (green). The red dotted lines cross the median and the 90th quantile.

mance will depend on the variability of future rainfall.

Volume of overflows in both future simulations is also statistically significantly different from the past and each other (p = 0.001). Median overflows increase from 13 m^3 per year to about 35 m^3 per year; however, the range widens considerably more. The consistent scenario predicts as much as 150 m^3 per year and the variable scenario about 210 m^3 per year. Frequency of overflows increases slightly, from a median of 3 to 4 per year.

Under both future scenarios, the average and maximum surface drainage time is expected to increase. Median values of average annual surface drainage times double (from 1 hour to 2 hours) in the consistent future simulation, and triple (2 hours to 3 hours) in the variable simulation. The surface takes longer to drain because future storms are predicted to get longer, on average, and more intense. In both future simulations, the increases are statistically significantly different from the historical value and from each other (p = 0.001). However, these increases are not large enough to be of concern from a regulatory standpoint. The maximum surface drainage time is still far below the threshold of 48-hours; the longest time to drain across all future simulations and years is approximately 10 hours.

The increases in the rainfall indices do alter performance as expected, which suggests the rainfall indices may provide insight regarding future performance without design simulation or real time sensors for performance. This means that tracking rainfall indices over time could be an alternative to tracking on-site monitors, which may not be available. However, future research is still needed to determine exactly how much performance degrades as rainfall indices change.

4.3.5   Correlation of future rain garden performance with rainfall indices

In addition to using historical simulation results and sensitivity analysis to select indices most indicative of performance, results from hydrologic simulations using future climate model output could also be considered for index selection. Correlations between rainfall indices calculated from the future climate model output and performance metrics estimated through simulation with future climate model rainfall results as the input will show whether historical correlations in these measures are expected to remain stable as climate changes. This section presents correlations between the future simulated performance metrics and the rainfall indices calculated from the future climate model output and determines if the same indices that were selected using historical simulations would have been selected after examining future simulations. If examination of correlations between future performance metrics and future rainfall indices would not change the indices selected using only historical simulation and sensitivity analysis, then it may be acceptable to ignore looking at future simulations when evaluating indices of interest. Climate model simulations may not always be easily accessible; thus it may be preferred to rely on historical data and sensitivity analysis only.

Figure 4.8 presents the correlations between simulated performance and the rainfall indices specific to each simulation. The five performance metrics, including: percent of runoff captured, mean and maximum hours to drain the surface, and the frequency and volume of overflows, are shown as large columns. Each different simulation (historical, consistent future, and variable future) is shown within each large column.

Many of the same rainfall indices are correlated to the performance metrics in a similar manner as the historical simulation; however, some are quite different, especially for capture efficiency.

The frequency and volume of overflows in the simulated future shows the most similar
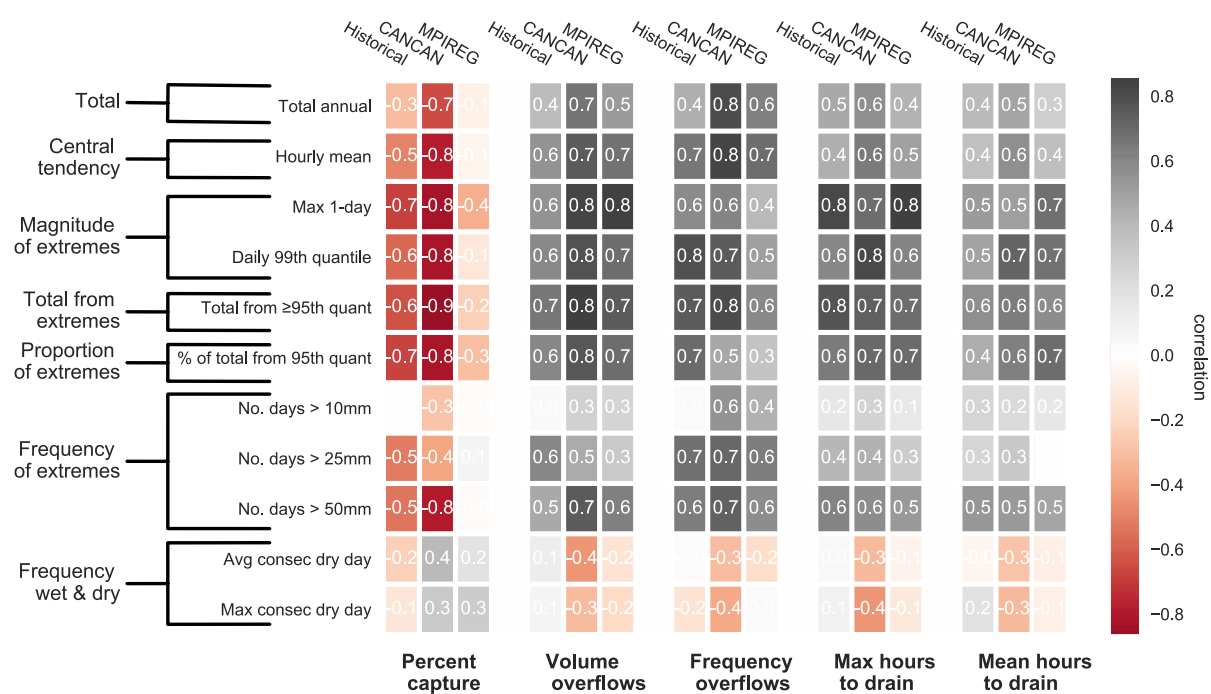
Figure 4.8: Correlation of selected rainfall indices to performance metrics for historical period and two climate model simulations in the future period. Grey colors represent positive correlations, while reds represent negative correlations. Labels on the left show the category of the rainfall indices.

correlation structure to the historical period. The volume of overflows is most strongly correlated to the daily and hourly mean rainfall, total rainfall ≥ 95th quantile, and 99th quantile, which are the same indices that were most strongly correlated in the past. Correlations to these indices are slightly higher in the simulated future than in the simulated past. This could signify that as rainfall becomes more extreme, correlations will get stronger, since more volume is being discharged.

Correlations for maximum surface drainage time in the future are weaker than they were in the historical simulations, while correlations for average surface drainage are stronger in the future. Correlations to maximum drainage time are likely weaker because the time to drain is more consistently higher in all years, meaning less variation across years. Since max time to drain is consistently high, it is more weakly correlated to the rainfall indices; changes in rainfall do not change the maximum time to drain. On the other hand, the average time to drain is more strongly correlated in the future because there is more variation, on average, in the drainage time each year.

Percent capture performance is the only metric where the models disagree on performance correlations for the future period. The consistent climate simulation is very weakly correlated with performance, whereas the variable climate simulation shows a similar correlation structure to that of the past — highly negatively correlated with extreme magnitude and frequency. Correlations are low in the consistent climate simulation because when rainfall does occur, it is in large quantities — so large, that the garden cannot infiltrate it fast enough, and water nearly always runs off. The percent capture is not correlated to the rain metrics because the garden performs consistently poorly, because the storms are all very similar. They are all short, intense events that appear after long dry spells. The variable climate simulation has the largest variety of rainfall events — from large to small. Correlations are strongest because the performance is more dependent on rainfall; it varies among rainfall events. The historical rainfall shows a similar variety of events, and thus correlations are also strong, since a variety of rainfall events leads to variability in performance. The disparities and similarities

118

in future performance over time for these two simulations is presented in Appendix B, Figure B.5.

Overall, examining the future correlations does not change the selection of the indices most indicative of performance in this analysis. Thus it may be possible to base decisions on the list of indices to track over time on the historical data without use of future simulations. However, for other locations and basins, this result could change. If bias-corrected climate simulation are readily available at a sub-daily timestep, then it is recommended to consider these results as part of the future performance modeling to inform selection of rainfall indices for regular evaluation in the future.

4.4   Summary and Conclusions

This study assessed whether annual measures, or indices, of rainfall can be a useful indicator for tracking hydrologic performance of green infrastructure systems over time. Several annual rainfall indices were established as most indicative of specific performance metrics for an example bio-retention system. The proportion of rainfall extremes from ≥ the 95th quantile was most highly correlated to the runoff capture efficiency, while the total volume of rainfall extremes from ≥ the 95th quantile was most indicative of the volume of overflows and the maximum surface drainage time. The number of days with rainfall greater than 25 mm was most highly correlated to the frequency of overflows.

Climate model simulations predicted increases in each of these rainfall indices in the future (2020 - 2059). As a result of these increases, performance degrades as expected, which suggests that rainfall indices could be used as an indicator of future performance without using real time simulation or monitoring, which may not be readily available. Although general performance degrades, capture efficiency remains high. The median remains above 90%, and the worst performance years never drop below 70% capture in either future simulation. Overflow volume increases significantly, from 13 m^3 per year to about 35 m^3 per year. The

consistent future scenario predicts as much as 150 m^3 per year of discharge to the sewer and the variable scenario about 210 m^3 per year. Frequency of overflows increases slightly, from a median of 3 to 4 per year, which is statistically significant. The surface takes longer to drain because future storms are predicted to get longer and more intense; however, these increases are not large enough to be of concern for permitting the infrastructure stand point since the surface always drains in less than 48-hours.

Although downscaled climate models were useful in this study to estimate predicted increases in specific rainfall indices, it may not always be necessary for engineers or site managers to make use of model output directly. When climate model output was used to determine which rainfall indices were most indicative of performance in future simulations, the results did not change relative to only examining historical correlations with a sensitivity analysis. Trends from climate models are still needed to understand if stakeholders should be concerned about changes in specific rainfall measures affecting future performance; however, existing analyses from the literature, the National Climate Assessment, or from NOAA's regional climate data center could be used instead of raw model output, which can be challenging to use. However, for other locations and basins, future results may be more sensitive. If bias-corrected climate simulations are readily available at a sub-daily time step, then it is always recommended to consider these results as part of the sensitivity analysis.

This study suggests that both performance metrics and rainfall indices could be used to track the need for adaptation over time. If the metric or index falls below a pre-determined threshold, then an adaptation study should be triggered. The performance metric of interest and the pre-determined threshold will depend on stakeholder preferences; however, the rainfall index should be selected based on the selected performance metric representing the objective. If multiple performance metrics align with objectives, then multiple rainfall indices could be tracked. Alternatively, the rainfall index that is most indicative of multiple performance metrics could be used. While this study shows promise for the use of rainfall indices in place of on-site sensors to track performance over time, to determine exactly how much

performance would degrade as rainfall indices change over time, more advanced statistical or machine learning models would need to be developed. These advances methods could be used to link a rate of increase in one or more rainfall indices to an quantitative decline in individual performance metrics.

4.5   Acknowledgements

CHAPTER 5   SUMMARY AND CONCLUSIONS

5.1   Conclusions

Climate change is expected to increase the intensity of rainfall across the U.S. and world-wide. These changes are a major concern for stormwater infrastructure engineers, as these systems are at risk of flooding, failure, and decreased performance. Traditional engineering design methods rely on assumptions of a stationary future based on past statistics; however, these design processes must be updated to incorporate expected future changes in rainfall patterns. Climate models can be used to understand how rainfall will change in the future; yet predictions are uncertain and not easily converted to a format useful to engineers. It is unclear how these predictions should be incorporated into the stormwater design process in order to improve infrastructure performance and resilience. The objective of this research was to determine how to use climate change projections during the stormwater design and assessment process to increase the resilience of urban drainage infrastructure under uncertain, future conditions.

To advance this objective, the second chapter developed a procedural framework for the use of climate model data in engineering analyses. The framework consists of five main steps that are necessary when using climate data for engineering applications. After defining the design decision, the first component is to understand the current methods and data requirements that have been used to create the current design process. The second step uses this information to select and extract the most appropriate downscaled climate model data source. A data source may be preferred over another due to the characteristics of the climate model output that are closest to that required in the engineering analysis. The third step involves managing the performance and uncertainty of the downscaled climate models. Uncertainty is managed by using multiple climate models (an ensemble), and one of three techniques to bound uncertainty: taking the maximum and minimum of all models, using all ensemble values, or by "culling" the ensemble through model validation with respect to historical con-

ditions. The fourth step relates to adjusting model output to the temporal or spatial resolution required for the engineering application, which can be completed using statistical downscaling techniques, bias-correction, or the change factor method. The final step discusses how to incorporate results into engineering practice by accounting for uncertainty ranges, risk levels, and changes over time.

The framework was applied to the updating of an important input parameter to stormwater design: intensity duration frequency (IDF) curves. Historical precipitation inputs for these curves are required on a sub-daily time-step, thus for this type of analysis, it is recommended to use climate model output that is available at the sub-daily level. Only regional climate models (which use dynamical downscaling) are available at this temporal resolution. Findings for updated curves in Pittsburgh suggest that depth of extreme precipitation for all durations and return periods is expected to increase in the future for Pittsburgh on the order of 10% for smaller return periods, and 25% and up for larger return periods. Results from this study suggest that it may be possible to interpret change factors as a potential climate factors of safety that could be applied to existing, stationary, depth-duration-frequency values. Based on these findings for Pittsburgh, a safety factor of 1.3 would encompass the majority of model uncertainty for depths of smaller return periods (e.g. 2 to 10 years); however, a factor of 1.3 is no longer valid when uncertainty magnifies as the return period increases to 25 years and larger.

While Chapter 2 detailed procedures and recommendations for using climate change projections to update precipitation frequency curves, many questions remained unanswered, including how different modeling choices would alter the uncertainty ranges of precipitation frequency curves, and consequently, the dimensions of stormwater infrastructure designs. Chapter 3 investigated different choices for updating depth-duration-frequency (DDF) curves and applied these to data for six U.S. cities: Birmingham, AL, Boston, MA, Boulder, CO, Pittsburgh, PA, Phoenix, AZ, and Seattle, WA. DDF curves provides the depth of rainfall expected for a given probability of occurrence and duration. Across a range of different

modeling choices, the depth of rainfall during shorter-duration storms (less than 6 hours) is expected to increase more, and predictions have a larger uncertainty range in the future than the depth of rainfall from longer duration storms (greater than 12 hours). These findings, which are consistent with previous results (Hassanzadeh et al. 2013; Kuo et al. 2015; Chandra et al. 2015; Shahabul Alam and Elshorbagy 2015), imply that stormwater infrastructure for the future should be equipped to store or convey more water in a short amount of time rather than more water over a longer period. Three adjustment techniques and two spatial resolutions of the climate models were also tested for these 6 cities. The three correction methods evaluated led to nearly consistent median values in most cities and for most durations and a single method does not consistently over or under estimate precipitation depths relative to the other methods. For the 6 cities examined, the lower spatial resolution (50-km) climate model ensemble generally provides a higher estimate of precipitation in the future than the 25-km ensemble and these differences are large enough to lead to different stormwater pipe dimensions.

Results from this study suggest that using the historical DDF curves for stormwater design in the cities analyzed will not be protective against future extreme events. However, if historical curves were to be used, adopting the upper bound of the historical storm depth provides more protection than doubling the return period. Nonetheless, this approach is not recommended because the additional protection provided from using the historical upper bound is independent of how much change is expected in the future. For this reason, updating DDF curves using climate model output is the recommended strategy for informing future design storms because these models are representative of expected future conditions. The resulting DDF curves and stormwater infrastructure designed from these curves are, however, sensitive to the choices made when creating these updated curves. This means that the design storm selected from an updated DDF curve is not guaranteed provide the same level of protection in the future that a design storm from a historical curve provided under historical climate conditions. Engineers using updated DDF curves should consider these sensitivities during

design and inform clients that performance of the infrastructure may vary over its lifetime. If the DDF updating process was completed by a single regulatory agency, ad hoc updating methods could be avoided, and the level of protection expected in the future would be more standardized.

A standardized process for creating updated DDF curves would not, however, eliminate uncertainty. Other tools and information may be needed to make a final decision on the capacity and size of the infrastructure. Chapter 4 examines how engineers can test whether stormwater infrastructure designs are robust, meaning they would perform against numerous rainfall scenarios for a range of performance metrics, using simulation. This process could be used during the planning process to iterate and improve upon designs so they account for more than a single design storm. For instance, simulation models can predict the frequency of flooding, or the volume of runoff that is discharged to the sewer system over time. Both of these metrics may be of interest to designers and local stakeholders, and specific thresholds for these and other metrics could be used as a design target.

As more rainfall observations become available over time, simulation can also be used to track progress of performance, and to determine if adaptation actions should be triggered after thresholds are exceeded. However, if the simulation model was developed during design, stakeholders in charge of assessing performance may not have access to this model or may have limited time to to rerun it each year to estimate performance. Findings from Chapter 4 suggest that in place of performance-based simulation, annual rainfall indices could also be used to predict performance degradation. Calculating changes in annual rainfall indices from observed rainfall data is a more straightforward process than calibrating and simulating hydrologic performance. To use this technique, the rainfall indices that are most indicative of each performance metric must first be established, and then changes in these indices can be tracked over time. Establishing the relationship between rainfall indices and performance metrics requires simulation and sensitivity analysis; however, this process would only need to be conducted once. It could occur during the design process, when simulation is used to

test design robustness, or as an independent study that examines index/performance metric relationships for several infrastructure installations across a region.

Downscaled climate models can be used to estimate expected increases in specific rainfall indices. If the indices expected to have large changes coincide with the same indices that are most indicative of performance, then it is particularly important to track performance of the stormwater infrastructure early on. If the largest changes do not coincide with the same indices that are most indicative of performance, then performance tracking could be postponed until there is enough data to re-evaluate changes in rainfall in 5 to 10 years. When it is important to understand exactly how performance degrades as rainfall changes, the most reliable way to do so is to install on-site sensors to collect data about performance directly. However, if monitoring equipment cannot be installed due to limited expertise or resources, then performance can be simulated using observed rainfall data, or by developing a statistical or machine-learning model that links annual rainfall indices to performance degradation.

Overall, this research contributes to the growing evidence that using existing standards of rainfall information is no longer adequate for the design of stormwater systems. Trends from climate models should be used to inform new design practices. Engineers are responsible for adopting these practices, and for communicating to clients that these changes are necessary in order to promote resilience of infrastructure systems.

5.2   Future Research Directions

5.2.1   IDF curve development and uncertainty bounding

One of the benefits of using regional IDF curves from NOAA Atlas 14 is that methods and uncertainty bounding are standardized for all regions across the U.S. Future work is still needed before IDF curves updated with climate model output can be developed in a similarly standardized fashion. Future work should explore other methods for representing and bounding

128

the range, including using bootstrapping or Monte-Carlo Markov Chains (MCMC) to produce full CDFs of each return level for a specific duration and return period. Bootstrapping uses random sampling to estimate the full distribution (Varian 2005), while the MCMC approach utilizes an algorithm to optimize over the parameter space that was generated using the joint posterior distribution from a Markov-Chain (Cheng et al. 2014). A joint distribution can then be estimated by combining the CDFs from all climate model simulations. The 10 and 90% CIs can be obtained by sampling this joint distribution.

Another area of future work could explore the fitting of the GEV distribution for extreme rainfall. It is possible that two different distributions could be fit - one to represent the more likely extreme storms (1 to 100 year return periods) - and another that represents super storms, like Sandy, that do not fit well with the first distribution. Fitting two distributions would allow for better incorporation of very large and low probability storms; however, more complexity adds to uncertainty. Trade-offs and options regarding the GEV fit still need to be explored in the future.

Further, a recommendation is still needed on which spatial resolution of climate model should be used in precipitation frequency analyses. This analysis will first need to confirm the findings (described in Chapter 3) for the entire U.S. that the 50-km resolution models generally lead to more extreme rainfall value predictions and higher uncertainty than the 25-km models. Then, a hypothesis for the cause of this difference will need to be evaluated, namely, that the lower, 50-km resolution leads to more extreme rainfall because of additional aggregation as storms move across the grid cell. For instance, in a 2-hour period, the same storm could be counted twice as it passes over a large grid cell, but only once at it passes over a station. If this is true, then a higher spatial resolution may actually be more realistic. Future work is needed to confirm this hypothesis and provide a recommendation to regulators if they are preparing to systematically update these curves.

5.2.2 Sensitivity analysis of rain garden performance

In addition to the updating of IDF curves, this dissertation also evaluated the future performance of rain gardens. However, only a single rain garden at a single location was evaluated as a demonstration of the approach. Future work is needed to test the sensitivity and robustness of the results. The first analysis should test how the results change as additional rain garden parameters change, including the volume of storage. This would inform how sensitive correlations and performance are to the initial design. A result of this analysis could be a curve of expected performance based on initial dimensions and characteristics. In addition, multiple locations should be evaluated so that curves could be developed for every state or city in the U.S. Designers could then use these curves as a guide for how much more performance they would gain under future conditions by increasing the infiltration rate or volume. To increase the efficiency of this analysis, which involves running hundreds of scenarios within the SWMM model, software engineers should develop an easy to use scenario interface for SWMM. This external interface would allow multiple scenarios to be run seamlessly without connecting through the user-interface to launch each scenario.

5.2.3 Eliciting community engagement and preferences

Several aspects of the design and adaption process will require dialogue with communities or clients. The first type of discussion is how much risk the community or client is willing to take regarding failure of infrastructure, and how much money they are willing to spend to avoid risk. This will influence how much rainfall the structure is designed to capture over its lifetime. The second type of discussion with stakeholders relates to trade-offs between different types of performance metrics, and which performance metric should be prioritized. For instance, as discussed in Chapter 4, designing for a high capture efficiency of green infrastructure may not prevent stormwater from being discharged to the sewer, or prevent standing water on the

130

surface. Once preferences are known, the final discussion to have with stakeholders is their threshold for performance degradation that will determine adaptation is triggered. Collaborative research between engineering and social sciences is needed to determine appropriate ways to elicit these preferences and how to incorporate this elicitation into the engineering planning process.

5.2.4 Thresholds for triggering adaptation

While it is important to discuss performance thresholds with stakeholders, further research is needed in this area to provide recommendations on what these thresholds could be. If rainfall indices are used to track performance over time, it is still unclear what type of changes in rainfall would trigger adaptation. It is still unclear if changes in the mean should trigger adaptation, or if these decisions should be made if a single year passes a predetermined value. Due to the variability of rainfall patterns, any one year surpassing a threshold may not be indicative of what is to come. However, how many years need to surpass the threshold in order to trigger adaption or redesign of the systems? Answers to this question are likely site specific, but they could be answered by calculating the average performance of the infrastructure over a period of time when all of these approaches are used to trigger adaptation. Furthermore, it is still unclear what should happen once adaption is triggered, and how much time is needed between adaptation triggering and infrastructure updating. More research is needed to evaluate these questions.

5.2.5 Lifecycle cost analysis of stormwater infrastructure

Another future area of research is related to the lifecycle cost of stormwater infrastructure. The performance curves discussed above could be linked with cost information, which would

present designers with estimates of how much additional gains in performance would cost to implement. For instance, increases in the storage volume is associated with additional soil or rocks needed to fill the volume, as well as additional excavation. Due to fixed costs (like renting excavation equipment or paying construction crews), adding storage volume could be only marginally more expensive. However, this will depend upon site-specific conditions. Furthermore, once capital expenditure is estimated, these values could be compared to the additional benefits that improving performance would bring. Quantifying these benefits is also an area of open research (e.g., Clark, Adriaens, and Talbot 2008; Mittman and Kloss 2015; Pruss-Ustun et al. 2008). The ultimate cost/benefit analysis would help designers to choose design parameters based on performance and cost requirements.

Cost analysis could also be used to determine benefits or consequences of over investment today in order to avoid damages in the future. This is particularly relevant for the sizing of grey infrastructure systems. Cost analysis could determine by how much the pipe diameter should be increased, and by how much the diameter could be increased before it is cost prohibitive, or before it does not make up for preventing damages in the future. This will help designers to understand the limits to increases in pipe sizes, and could also be used to convince clients that going one pipe size up would reduce damages and thus costs in the future.

5.2.6   Accounting for low flow conditions in design

While cost is one reason to not increase a pipe size, there are also additional reasons why increasing the size of the infrastructure may not alway be the solution. A pipe that is oversized can be detrimental during periods of low flow. In pipe networks, low flow conditions can cause the velocity to fall below the self-cleansing velocity (DeZellar and Maier 1980), which can lead to settling of particles, odors, corrosion and pipe deterioration (Marleni and Nyoman 2016). This is relevant for stormwater and wastewater networks. The use of continuous hydrologic simulation also has the potential to provide insights about infrastructure performance

during periods of low flow or drought conditions. These models could be used to simulate whether a selected pipe size could cause problems under periods of low rainfall in the future. This type of simulation addition to pipe networks, periods of low rainfall can dry soils, which can lead to mortality of plant species in green infrastructure systems (Jennings 2016). Future work is needed to incorporate low flow challenges into the hydrologic design process and into hydrologic simulation models.

REFERENCES

3 Rivers Wet Weather. 2018. "Calibrated Radar Rainfall Data." 2018. http://www.3riverswetweather.org/municipalities/calibrated-radar-rainfall-data.

Abatzoglou, John T., and Timothy J. Brown. 2012. "A Comparison of Statistical Downscaling Methods Suited for Wildfire Applications." International Journal of Climatology 32 (5): 772–80. https://doi.org/10.1002/joc.2312.

Adger, W Neil, Nigel W Arnell, and Emma L Tompkins. 2005. "Successful Adaptation to Climate Change across Scales." Global Environmental Change 15 (2): 77–86.

Ahmed, Sadik, and Ioannis Tsanis. 2016. "Climate Change Impact on Design Storm and Performance of Urban Storm-Water Management System-A Case Study on West Central Mountain Drainage Area in Canada." Hydrology: Current Research 7 (229): 1–11. https://doi.org/10.4172/2157-7587.1000229.

A&L Great Lakes Laboratories, Inc. 2015. "Soil Test Report for ELPC Rain Garden Soils Taken on 2/10/2015."

Allan, Richard P., and Brian J. Soden. 2008. "Atmospheric Warming and the Amplification of Precipitation Extremes." Science 321 (5895): 1481. https://doi.org/10.1126/science.1160787.

Allegheny County Sanitary Authority. 2012. "ALCOSAN Wet Weather Plan." http://www.alcosan.org/WetWeatherIssues/ALCOSANDraftWetWeatherPlan/DraftWWPFullDocument/tabid/176/Default.aspx.

Allen, M. R., and W. J. Ingram. 2002. "Constraints on Future Changes in Climate and the

Hydrologic Cycle." Nature 419 (6903): 224:32. https://doi.org/10.1038/Nature01092.

Allen, M. R., P. A. Stott, J. F. B. Mitchell, R. Schnur, and T. L. Delworth. 2000. "Quantifying the Uncertainty in Forecasts of Anthropogenic Climate Change." Nature 407 (6804): 617–20.

Allen, Robert J, and Arthur T DeGaetano. 2005. "Areal Reduction Factors for Two Eastern United States Regions with High Rain-Gauge Density." Journal of Hydrologic Engineering 10 (4): 327–35.

Anderson, Christopher J., Raymond W. Arritt, Zaitao Pan, Eugene S. Takle, William J. Gutowski, Francis O. Otieno, Renato da Silva, et al. 2003. "Hydrological Processes in Regional Climate Model Simulations of the Central United States Flood of June/July 1993." Journal of Hydrometeorology 4 (3): 584–98.

https://doi.org/10.1175/1525-7541(2003)004<0584:HPIRCM>2.0.CO;2.

Andréasson, Johan, Sten Bergström, Bengt Carlsson, L. Phil Graham, and Göran Lindström. 2004. "Hydrological Change — Climate Change Impact Simulations for Sweden." AMBIO: A Journal of the Human Environment 33 (4): 228–34. https://doi.org/10.1579/0044-7447-33.4.228.

Arnbjerg-Nielsen, K., P. Willems, J. Olsson, S. Beecham, A. Pathirana, I. Bulow Gregersen, H. Madsen, and V.-T.-V. Nguyen. 2013. "Impacts of Climate Change on Rainfall Extremes and Urban Drainage Systems: A Review." Water Science and Technology: A Journal of the International Association on Water Pollution Research 68 (1). https://doi.org/10.2166/wst.2013.251.

Arnbjerg-Nielsen, Karsten. 2011. "Past, Present, and Future Design of Urban Drainage Systems with Focus on Danish Experiences." Water Science and Technology 63 (3): 527–35.

ASCE. 2013. Minimum Design Loads for Buildings and Other Structures. Standards. American Society of Civil Engineers. http://dx.doi.org/10.1061/9780784412916.

Baltimore City Department Public Works. 2017. "Stormwater Remediation Fee Regulations." City of Baltimore. https://publicworks.baltimorecity.gov/sites/default/files/Stormwater%20Remediation%20Fee%20Regulations.pdf.

Barros, A., and J. Evans. 1997. "Designing for Climate Variability." Journal of Professional

Issues in Engineering Education and Practice 123 (2): 62–65. https://doi.org/10.1061/(ASCE)1052-3928(1997)123:2(62).

Barros, Anna P. 2006. "Rainfall Parameters for Design Storms Under a Changing Climate." In , 40. Washington D.C.: Center for Transportation and the Environment, North Carolina State University and USDoT.

Barsugli, Joseph J., Galina Guentchev, Radley M. Horton, Andrew Wood, Linda O. Mearns, Xin-Zhong Liang, Julie A. Winkler, et al. 2013. "The Practitioners Dilemma: How to Assess the Credibility of Downscaled Climate Projections." Eos, Transactions American Geophysical Union 94 (46): 424–25. https://doi.org/10.1002/2013EO460005.

Beguería, Santiago. 2005. "Uncertainties in Partial Duration Series Modeling of Extremes Related to the Choice of the Threshold Value." Journal of Hydrology 303 (1): 215–30.

Bocchini, Paolo, Dan M. Frangopol, Thomas Ummenhofer, and Tim Zinke. 2014. "Resilience and Sustainability of Civil Infrastructure: Toward a Unified Approach." Journal of Infrastructure Systems 20 (2): 04014004. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000177.

Boé, J., L. Terray, F. Habets, and E. Martin. 2007. "Statistical and Dynamical Downscaling of the Seine Basin Climate for Hydro-Meteorological Studies." International Journal of Climatology 27 (12): 1643–56.

Bonnin, Geoffry M, Deborah Martin, Bingzhang Lin, Tye Parzybok, Michael Yekta, and David Riley. 2006. "NOAA Atlas 14: Precipitation-Frequency Atlas of the United States, Volume 2, Version 3.0." U.S. Dept Commerce. http://www.nws.noaa.gov/oh/hdsc/PF_documents/Atlas14_Volume2.pdf.

Bosilovich, Michael G., Junye Chen, Franklin R. Robertson, and Robert F. Adler. 2008. "Evaluation of Global Precipitation in Reanalyses." Journal of Applied Meteorology and Climatology 47 (9): 2279–99. https://doi.org/10.1175/2008JAMC1921.1.

Brekke, Levi, Bridget Thrasher, Edwin Maurer, and Tom Pruitt. 2013. "Downscaled CMIP3 and CMIP5 Climate Projections: Release of Downscaled CMIP5 Climate Projections, Comparison with Preceding Information, and Summary of User Needs." Technical Memoradum.

Denver, CO: U.S. Department of the Interior, Bureau of Reclamation. http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/.

Brommer, David M, Randall S Cerveny, and Robert C Balling Jr. 2007. "Characteristics of Long–duration Precipitation Events across the United States." Geophysical Research Letters 34 (22).

Casal-Campos, Arturo, Guangtao Fu, David Butler, and Andrew Moore. 2015. "An Integrated Environmental Assessment of Green and Gray Infrastructure Strategies for Robust Decision Making." Environmental Science & Technology 49 (14): 8307– 14. https://doi.org/10.1021/es506144f.

Castellano, Christopher M., and Arthur T. DeGaetano. 2016. "A Multi-Step Approach for Downscaling Daily Precipitation Extremes from Historical Analogues." International Journal of Climatology 36 (4): 1797–1807. https://doi.org/10.1002/joc.4460.

CH2MHILL. 2014. "Stormwater Management Fee Policy Options and Recommendations." Green Infrastructure Advisory Committee Report. Philadelphia, PA: City of Lancaster.

Chagrin Watershed Partners. 2017. "Stormwater Utility Literature Review." Chagrin Watershed Partners. http://crwp.org/files/Literature_Review_by_Credit_Type.pdf.

Chandra, Rupa, Ujjwal Saha, and P.P. Mujumdar. 2015. "Model and Parameter Uncertainty in IDF Relationships under Climate Change." Advances in Water Resources 79 (May): 127–39. https://doi.org/10.1016/j.advwatres.2015.02.011.

Chapman, Cameron, and Richard R Horner. 2010. "Performance Assessment of a Street-Drainage Bioretention System." Water Environment Research 82 (2): 109–19.

Charles, Stephen P., Bryson C. Bates, Peter H. Whetton, and James P. Hughes. 1999. "Validation of Downscaling Models for Changed Climate Conditions: Case Study of Southwestern Australia." Climate Research 12 (1): 1–14.

Chen, Deliang, Christine Achberger, Tinghai Ou, Ulrika Postgărd, Alexander Walther, and Yaoming Liao. 2015. "Projecting Future Local Precipitation and Its Extremes for Sweden." Geografiska Annaler: Series A, Physical Geography 97 (1): 25–39. https://doi.org/10.1111/

Chen, Jie, François P. Brissette, Diane Chaumont, and Marco Braun. 2013. "Performance and Uncertainty Evaluation of Empirical Downscaling Methods in Quantifying the Climate Change Impacts on Hydrology over Two North American River Basins." *Journal of Hydrology* 479 (0): 200–214. https://doi.org/10.1016/j.jhydrol.2012.11.062.

Chen, Jie, François P. Brissette, and Robert Leconte. 2011. "Uncertainty of Downscaling Method in Quantifying the Impact of Climate Change on Hydrology." *Journal of Hydrology* 401 (3–4): 190–202. https://doi.org/10.1016/j.jhydrol.2011.02.020.

Cheng, Linyin, and Amir AghaKouchak. 2014. "Nonstationary Precipitation Intensity-Duration-Frequency Curves for Infrastructure Design in a Changing Climate." *Scientific Reports* 4.

Clark, Corrie, Peter Adriaens, and F. Brian Talbot. 2008. "Green Roof Valuation: A Probabilistic Economic Analysis of Environmental Benefits." *Environmental Science & Technology* 42 (6): 2155–61. https://doi.org/10.1021/es0706652.

Coles, Stuart. 2001. *An Introduction to Statistical Modeling of Extreme Values*. 1st ed. Springer Series in Statistics. London: Springer-Verlag London.

Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, et al. 2013. "Long-Term Climate Change: Projections, Commitments and Irreversibility." In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley, 1029–1136. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. www.climatechange2013.org.

Computational Hydraulics Int. 2018. *PCSWMM* (version 7.1). Guelph, ON: chiwater. https://www.pcswmm.com/.

Cook, Lauren M., Christopher J. Anderson, and Constantine Samaras. 2017. "Framework for Incorporating Downscaled Climate Output into Existing Engineering Methods: Application to Precipitation Frequency Curves." *Journal of Infrastructure Systems* 23 (4). https://doi.org/10.1061/(ASCE)IS.1943-555X.0000382.

Cook, Lauren, Seth McGinnis, and Constantine Samaras. 2018. "Uncertainty in Sub-Daily Climate Corrected Precipitation Frequency Curves and Effects on Stormwater Design." *In Prep*.

Cooney, Catherine M. 2012. "Downscaling Climate Models: Sharpening the Focus on Local-Level Changes." *Environmental Health Perspectives* 120 (1): a22–28. https://doi.org/10.1289/ehp.120-a22.

CSA Standards. 2012. *Development, Interpretation, and Use of Rainfall Intensity-Duration-Frequency (IDF) Information: Guideline for Canadian Water Resources Practitiioners*. Ontario, Canada: Canadian Standards Association.

Davis, Allen P, William F Hunt, Robert G Traver, and Michael Clar. 2009. "Bioretention Technology: Overview of Current Practice and Future Needs." *Journal of Environmental Engineering* 135 (3): 109–17.

Davis, Allen P., Robert G. Traver, William F. Hunt, Ryan Lee, Robert A. Brown, and Jennifer M. Olszewski. 2012. "Hydrologic Performance of Bioretention Storm-Water Control Measures." *Journal of Hydrologic Engineering* 17 (5): 604–14. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000467.

De Neufville, Richard, and Stefan Scholtes. 2011. *Flexibility in Engineering Design*. MIT Press.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, et al. 2011. "The ERA-Interim Reanalysis: Configuration and Performance of the Data Assimilation System." *Quarterly Journal of the Royal Meteorological Society* 137 (656): 553–97. https://doi.org/10.1002/qj.828.

Dee, Dick, John Fasullo, Dennis Shea, John Walsh, and National Center for Atmospheric Research Staff (Eds). 2016. "The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables." January 21, 2016. https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables.

DeGaetano, Arthur T. 2009. "Time-Dependent Changes in Extreme-Precipitation Return-Period Amounts in the Continental United States." *Journal of Applied Meteorology and Climatology* 48 (10): 2086–99.

DeGaetano, Arthur T., and Christopher M. Castellano. 2017. "Future Projections of Extreme Precipitation Intensity-Duration-Frequency Curves for Climate Adaptation Planning in New York State." *Climate Services* 5 (January): 23–35. https://doi.org/10.1016/j.cliser.2017.03.003.

Demuzere, Matthias, K Orru, O Heidrich, E Olazabal, D Geneletti, Hans Orru, AG Bhave, N Mittal, E Feliu, and M Faehnle. 2014. "Mitigating and Adapting to Climate Change: Multi-Functional and Multi-Scale Assessment of Green Urban Infrastructure." *Journal of Environmental Management* 146: 107–15.

Deser, C., A. Phillips, V. Bourdette, and H. Y. Teng. 2012. "Uncertainty in Climate Change Projections: The Role of Internal Variability." *CLIMATE DYNAMICS* 38 (3–4): 527–46. https://doi.org/10.1007/S00382-010-0977-X.

DeZellar, J. T., and W. J. Maier. 1980. "Effects of Water Conservation on Sanitary Sewer and Wastewater Treatment Plant." *Journal of Water Pollution Control* 52 (1).

Di Luca, Alejandro, Ramón de Elía, and René Laprise. 2015. "Challenges in the Quest for Added Value of Regional Climate Dynamical Downscaling." *Current Climate Change Reports* 1 (1): 10–21. https://doi.org/10.1007/s40641-015-0003-9.

Durrans, S. Rocky, Steven J. Burian, Stephan J. Nix, Ahmed Hajji, Robert E. Pitt, Chi-Yuan Fan, and Richard Field. 1999. "Polynomial-Based Disaggregation of Hourly Rainfall for Continuous Hydrologic Simulation." *JAWRA Journal of the American Water Resources Association* 35 (5): 1213–21. https://doi.org/10.1111/j.1752-1688.1999.tb04208.x.

Easterling, D. R., K. E. Kunkel, J. R. Arnold, Knutson T.R., A.N. LeGrand, L. R. Leung, R.S. Vose, D. E. Waliser, and M. F. Wehner. 2017. "Precipitation Change in the United States." In *Climate Science Special Report: Fourth National Climate Assessment, Volume I*, edited by D. J. Wuebbles, D. W. Fahey, K.A. Hibbard, D. J. Dokken, B.C. Stewart, and T.K.

Maycock, 207–30. Washington, D.C.: U.S. Global Change Research Program. doi:10.7930/J0H993CC.

Environmental Finance Center at the UNC School of Government. 2017. "North Carolina Stormwater Rates and Revenues - Preliminary Analysis." UNC Environmental Finance Center. https://efc.sog.unc.edu/sites/default/files/2017/NC%20Stormwater%20Chart%20Clean%202017_1_0.pdf.

Erdman, Jon. 2016. "18 Major Flood Events Have Hit Texas, Louisiana, Oklahoma, Arkansas Since March 2015." *Severe Weather*. *Weather Channel* (blog). August 16, 2016. https://weather.com/storms/severe/news/flood-fatigue-2015-2016-texas-louisiana-oklahoma.

Espinet, X., A. Schweikert, and P. Chinowsky. 2015. "Robust Prioritization Framework for Transport Infrastructure Adaptation Investments under Uncertainty of Climate Change." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, November, E4015001. https://doi.org/10.1061/AJRUA6.0000852.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor. 2016. "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization." *Geosci. Model Dev.* 9 (5): 1937–58. https://doi.org/10.5194/gmd-9-1937-2016.

Fadhel, Sherien, Miguel Angel Rico-Ramirez, and Dawei Han. 2017. "Uncertainty of Intensity–Duration–Frequency (IDF) Curves Due to Varied Climate Baseline Periods." *Journal of Hydrology* 547 (April): 600–612. https://doi.org/10.1016/j.jhydrol.2017.02.013.

Faturechi, R., and E. Miller-Hooks. 2015. "Measuring the Performance of Transportation Infrastructure Systems in Disasters: A Comprehensive Review." *Journal of Infrastructure Systems* 21 (1): 04014025. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000212.

Fischbach, Jordan R, Kyle Siler-Evans, Devin Tierney, Michael T Wilson, Lauren M Cook, and Linnea Warren May. 2017. "Robust Stormwater Management in the Pittsburgh Region."

Flato, Gregory, Jochem Marotzke, Babatunde Abiodun, Pascale Braconnot, Sin Chan Chou, William Collins, Peter Cox, et al. 2013. "Evaluation of Climate Models." In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group 1 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Isasc Held, Andy Pitman, Serge Planton, and Zong-Ci Zhao. Cambridge, United Kingdom and New York, NY USA: Cambridge University Press.

Forsee, W., and S. Ahmad. 2011. "Evaluating Urban Storm-Water Infrastructure Design in Response to Projected Climate Change." *Journal of Hydrologic Engineering* 16 (11): 865–73. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000383.

Foster, Josh, Ashley Lowe, and Steve Winkelman. 2011. "The Value of Green Infrastructure for Urban Climate Adaptation." *Center for Clean Air Policy* 750.

Fowler, H. J., S. Blenkinsop, and C. Tebaldi. 2007a. "Linking Climate Change Modelling to Impacts Studies: Recent Advances in Downscaling Techniques for Hydrological Modelling." *International Journal of Climatology* 27 (12): 1547–78. https://doi.org/10.1002/joc.1556.

———. 2007b. "Linking Climate Change Modelling to Impacts Studies: Recent Advances in Downscaling Techniques for Hydrological Modelling." *International Journal of Climatology* 27 (12): 1547–78. https://doi.org/10.1002/joc.1556.

Framework Convention on Climate Change. 2015. "Adoption of the Paris Agreement." United Nations. https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf.

Francis, Royce, and Behailu Bekera. 2014. "A Metric and Frameworks for Resilience Analysis of Engineered and Infrastructure Systems." *Reliability Engineering & System Safety* 121 (January): 90–103. https://doi.org/10.1016/j.ress.2013.07.004.

GAO (General Accountability Office). 2015. "Climate Information: A National System Could Help Federal, State, Local, and Private Sector Decision Makers Use Climate Information." GAO-16-37. Washington D.C.: General Accountability Office. http://www.gao.gov/products/GAO-16-37.

Gateway Engineers. 2011. "Municipality of Mount Lebanon Stormwater Fee: Credit Manual for Stormwater Fees." Mt. Lebanon Pennsylvania. https://www.mtlebanon.org/DocumentCenter/View/4075/Stormwater-Fee-Manual-APPROVED?bidId=.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux. 2008. "Performance Metrics for Climate Models." *Journal of Geophysical Research: Atmospheres* 113 (D6): n/a-n/a. https://doi.org/10.1029/2007JD008972.

Gooré Bi, Eustache, Philippe Gachon, Mathieu Vrac, and Frédéric Monette. 2017. "Which Downscaled Rainfall Data for Climate Change Impact Studies in Urban Areas? Review of Current Approaches and Trends." *Theoretical and Applied Climatology* 127 (3): 685–99. https://doi.org/10.1007/s00704-015-1656-y.

Gooré Bi, Eustache, Frédéric Monette, Philippe Gachon, Johnny Gaspéri, and Yves Perrodin. 2015. "Quantitative and Qualitative Assessment of the Impact of Climate Change on a Combined Sewer Overflow and Its Receiving Water Body." *Environmental Science and Pollution Research* 22 (15): 11905–21. https://doi.org/10.1007/s11356-015-4411-0.

Gregersen, I. B., and K. Arnbjerg-Nielsen. 2012. "Decision Strategies for Handling the Uncertainty of Future Extreme Rainfall under the Influence of Climate Change." *Water Science and Technology : A Journal of the International Association on Water Pollution Research* 66 (2): 284–91. https://doi.org/10.2166/wst.2012.173.

Groisman, Pavel Ya, and Richard W Knight. 2008. "Prolonged Dry Episodes over the Conterminous United States: New Tendencies Emerging during the Last 40 Years." *Journal of Climate* 21 (9): 1850–62.

Groisman, Pavel Ya, Richard W. Knight, David R. Easterling, Thomas R. Karl, Gabriele C. Hegerl, and Vyacheslav N. Razuvaev. 2005. "Trends in Intense Precipitation in the Climate Record." *Journal of Climate* 18 (9): 1326–50. https://doi.org/10.1175/JCLI3339.1.

Groisman, Pavel Ya, Richard W Knight, and Thomas R Karl. 2012. "Changes in Intense Precipitation over the Central United States." *Journal of Hydrometeorology* 13 (1): 47–66.

Groves, David G., and Robert J. Lempert. 2007. "A New Analytic Method for Finding Policy-Relevant Scenarios." *Uncertainty and Climate Change Adaptation and Mitigation* 17 (1): 73–85. https://doi.org/10.1016/j.gloenvcha.2006.11.006.

Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen. 2012. "Technical Note: Downscaling RCM Precipitation to the Station Scale Using Statistical Transformations–a Comparison of Methods." *Hydrology and Earth System Sciences* 16 (9): 3383–90.

Hall, Alex. 2014. "Projecting Regional Change Modeling Regional Climate Change: How Accurate Are Regional Projections of Climate Change Derived from Downscaling Global Climate Model Results?" *Science* 34 (6216): 1461–62.

Hallegatte, Stéphane. 2009. "Strategies to Adapt to an Uncertain Climate Change." *Traditional Peoples and Climate Change* 19 (2): 240–47. https://doi.org/10.1016/j.gloenvcha.2008.12.003.

———. 2014. "Decision Making for Disaster Risk Management in a Changing Climate." In *Natural Disasters and Climate Change*, 177–194. Springer. http://link.springer.com/chapter/10.1007/978-3-319-08933-1_7.

Hallegatte, Stéphane, Ankur Shah, Casey Brown, Robert Lempert, and Stuart Gill. 2012. "Investment Decision Making under Deep Uncertainty–Application to Climate Change." *World Bank Policy Research Working Paper*, no. 6193. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2143067.

Hassanzadeh, E., A. Nazemi, and A. Elshorbagy. 2013. "Quantile-Based Downscaling of Precipitation Using Genetic Programming: Application to IDF Curves in Saskatoon." *Journal of Hydrologic Engineering* 19 (5): 943–55. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000854.

Hawkins, Ed, and Rowan Sutton. 2011. "The Potential to Narrow Uncertainty in Projections of Regional Precipitation Change." *Climate Dynamics* 37 (1): 407–18. https://doi.org/10.1007/s00382-010-0810-6.

Heavens, NG, DS Ward, and MM Natalie. 2013. "Studying and Projecting Climate Change with Earth System Models." *Nature Education Knowledge* 4 (5): 4.

Hershfield, David. 1961. "Rainfall Frequency Atlas of the United States for Durations from 30 Minutes to 24 Hours and Return Periods from 1 to 100 Years." Technical Paper 40. Washington, D.C.: U.S. Department of Commerce. http://www.nws.noaa.gov/oh/hdsc/PF_documents/TechnicalPaper_No40.pdf.

Hostetler, S. W., J. R. Alder, and A. M. Allan. 2011. "Dynamically Downscaled Climate Simulations over North America: Methods, Evaluation, and Supporting Documentation for Users." US Geological Survey. http://regclim.coas.oregonstate.edu/.

Houghton, J.T., Y Ding, D.J. Griggs, M Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson. 2001. "Climate Change 2001: The Scientific Basis." *Cambridge University Press*, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, .

Hydrometeorological Design Studies Center, and NOAA's National Weather Service. 2016. "Current NWS Precipitation Frequency (PF) Documents." US Department of Commerce. http://www.nws.noaa.gov/oh/hdsc/currentpf.htm.

Infield, Elisabeth M Hamin, Yaser Abunnasr, and Robert L Ryan. 2018. *Planning for Climate Change: A Reader in Green Infrastructure and Sustainable Design for Resilient Cities*. Routledge.

IPCC. 2007. *Climate Change 2007: The Physical Science Basis*. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

———. 2012. "List of Major IPCC Reports." In *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 (IPCC)*, edited by C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, et al., 569–72. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.
146

———. 2014. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (Eds.)]*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Jennings, Aaron A. 2016. "Residential Rain Garden Performance in the Climate Zones of the Contiguous United States." *Journal of Environmental Engineering* 142 (12): 04016066.

Jennings, Aaron A, Michael A Berger, and James D Hale. 2015. "Hydraulic and Hydrologic Performance of Residential Rain Gardens." *Journal of Environmental Engineering* 141 (11): 04015033.

Karl, Thomas R, and Richard W Knight. 1998. "Secular Trends of Precipitation Amount, Frequency, and Intensity in the United States." *Bulletin of the American Meteorological Society* 79 (2): 231–41.

Karl, Thomas R, and Kevin E Trenberth. 2003. "Modern Global Climate Change." *Science* 302 (5651): 1719–23.

Katz, Richard W. 2013. "Statistical Methods for Nonstationary Extremes." In *Extremes in a Changing Climate*, 15–37. Springer.

Kennedy, Christopher, and Jan Corfee-Morlot. 2013. "Past Performance and Future Needs for Low Carbon Climate Resilient Infrastructure– An Investment Perspective." *Energy Policy* 59 (August): 773–83. https://doi.org/10.1016/j.enpol.2013.04.031.

Khan, Mohammad Sajjad, Paulin Coulibaly, and Yonas Dibike. 2006. "Uncertainty Analysis of Statistical Downscaling Methods." *Journal of Hydrology* 319 (1–4): 357–82. https://doi.org/10.1016/j.jhydrol.2005.06.035.

Kilgore, Roger T., George Herrmann, Wilbert O. Thomas Jr., and David B. Thompson. 2016. "Hydraulic Engineering Circular No. 17, 2nd Edition Highways in the River Environment — Floodplains, Extreme Events, Risk, and Resilience." Federal Highway Administration.

Kilsby, C.G., P.D. Jones, A. Burton, A.C. Ford, H.J. Fowler, C. Harpham, P. James, A. Smith, and R.L. Wilby. 2007. "A Daily Weather Generator for Use in Climate Change Studies." *Environmental Modelling & Software* 22 (12): 1705–19. https://doi.org/10.1016/j.envsoft.2007.02.005.

Kirtman, B., S.B. Power, J.A. Adedoyin, G.J. Boer, R. Bojariu, I. Camilloni, F.J. Doblas-Reyes, et al. 2013. "Near-Term Climate Change: Projections and Predictability." In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley, 953–1028. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. www.climatechange2013.org.

Kloss, Christopher, Crystal Calarusse, and Nancy Stoner. 2006. *Rooftops to Rivers: Green Strategies for Controlling Stormwater and Combined Sewer Overflows*. Natural Resources Defense Council.

Knutti, Reto, Reinhard Furrer, Claudia Tebaldi, Jan Cermak, and Gerald A. Meehl. 2009. "Challenges in Combining Projections from Multiple Climate Models." *Journal of Climate* 23 (10): 2739–58. https://doi.org/10.1175/2009JCLI3361.1.

Kueh, SM, and KK Kuok. 2016. "Precipitation Downscaling Using the Artificial Neural Network BatNN and Development of Future Rainfall Intensity-Duration-Frequency Curves." *Climate Research* 68 (1): 73–89. http://www.int-res.com/abstracts/cr/v68/n1/p73-89.

Kunkel, Kenneth E., Thomas R. Karl, Harold Brooks, James Kossin, Jay H. Lawrimore, Derek Arndt, Lance Bosart, et al. 2012. "Monitoring and Understanding Trends in Extreme Storms: State of Knowledge." *Bulletin of the American Meteorological Society* 94 (4): 499–514. https://doi.org/10.1175/BAMS-D-11-00262.1.

Kuo, Chun-Chao, ThianYew Gan, and Mesgana Gizaw. 2015. "Potential Impact of Climate Change on Intensity Duration Frequency Curves of Central Alberta." *Climatic Change*, February, 1–15. https://doi.org/10.1007/s10584-015-1347-9.

Laflamme, Eric M., Ernst Linder, and Yibin Pan. 2016. "Statistical Downscaling of Regional Climate Model Output to Achieve Projections of Precipitation Extremes." *Weather and Climate Extremes* 12 (June): 15–23. https://doi.org/10.1016/j.wace.2015.12.001.

Larkin, Sabrina, Cate Fox-Lent, Daniel A. Eisenberg, Benjamin D. Trump, Sean Wallace, Colin Chadderton, and Igor Linkov. 2015. "Benchmarking Agency and Organizational Practices in Resilience Decision Making." *Environment Systems and Decisions* 35 (2): 185–95. https://doi.org/10.1007/s10669-015-9554-5.

Larsen, Tove A., Sabine Hoffmann, Christoph Lüthi, Bernhard Truffer, and Max Maurer. 2016. "Emerging Solutions to the Water Challenges of an Urbanizing World." *Science* 352 (6288): 928. https://doi.org/10.1126/science.aad8641.

Lee, Ji Yun, and Bruce R. Ellingwood. 2017. "A Decision Model for Intergenerational Life-Cycle Risk Assessment of Civil Infrastructure Exposed to Hurricanes under Climate Change." *Reliability Engineering & System Safety* 159 (March): 100–107. https://doi.org/10.1016/j.ress.2016.10.022.

Lempert, Robert. 2010. "Robust Decision Making Approach to Managing Water Resource Risks." In *AGU Fall Meeting Abstracts*, 1:02. http://adsabs.harvard.edu/abs/2010AGUFM.H13H..02L.

———. 2013a. "Scenarios That Illuminate Vulnerabilities and Robust Responses." *Climatic Change* 117 (4): 627–646. http://link.springer.com/article/10.1007/s10584-012-0574-6.

———. 2013b. "Scenarios That Illuminate Vulnerabilities and Robust Responses." *Climatic Change* 117 (4): 627–646. http://link.springer.com/article/10.1007/s10584-012-0574-6.

Lempert, Robert J., and David G. Groves. 2010. "Identifying and Evaluating Robust Adaptive Policy Responses to Climate Change for Water Management Agencies in the American West." *Technological Forecasting and Social Change* 77 (6): 960–974. http://www.sciencedirect.com/science/article/pii/S0040162510000740.

Lempert, Robert J., Steven W. Popper, and Steven C. Bankes. 2010. "Robust Decision Making: Coping with Uncertainty." *Futurist* 50 (1): 41.

Lempert, Robert, and Michael E. Schlesinger. 2001. "Climate-Change Strategy Needs to Be Robust." *Nature* 412 (6845): 375–375. https://doi.org/10.1038/35086617.

Linkov, Igor, Todd Bridges, Felix Creutzig, Jennifer Decker, Cate Fox-Lent, Wolfgang Kröger, James H. Lambert, et al. 2014. "Changing the Resilience Paradigm." *Nature Climate Change* 4 (6): 407–9. https://doi.org/10.1038/nclimate2227.

Lopez-Cantu, T, and C Samaras. 2018. "Temporal and Spatial Evaluation of Stormwater Engineering Standards Reveals Risks and Priorities across the United States." *Environmental Research Letters*. https://doi.org/10.1088/1748-9326/aac696.

Mailhot, A., and S. Duchesne. 2009. "Design Criteria of Urban Drainage Infrastructures under Climate Change." *Journal of Water Resources Planning and Management* 136 (2): 201–8. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000023.

Marleni, Ni, and Nepi Nyoman. 2016. "Impact of Water Management Practices in Residential Areas on Odour and Corrosion in Existing Sewer Networks." PhD Thesis, Melbourne, Australia: Victoria University. http://vuir.vu.edu.au/id/eprint/32310.

Maurer, Edwin P., Levi Brekke, Tom Pruitt, and Philip B. Duffy. 2007. "Fine-Resolution Climate Projections Enhance Regional Climate Change Impact Studies." *Eos, Transactions American Geophysical Union* 88 (47): 504–504. https://doi.org/10.1029/2007EO470006.

McCuen, Richard H. 2005. *Hydrologic Analysis and Design*. 3rd ed. Pearson Prentice Hall.

McGinnis, Seth, Doug Nychka, and Linda O Mearns. 2015. "A New Distribution Mapping Technique for Climate Model Bias Correction." In *Machine Learning and Data Mining Approaches to Climate Science*, edited by Valliappa Lakshmanan, Eric Gilleland, Amy McGovern, and Martin Tingley, 91–99. Springer.

McGuffie, K., and A. Henderson-Sellers. 2001. "Forty Years of Numerical Climate Modelling." *International Journal of Climatology* 21 (9): 1067–1109.

Mearns, L. O., W. J. Gutowski, J. J. Barsugli, L. Buja, G. M. Garfin, D. P. Lettenmaier, and L. Leung. 2013. "The Development of North America CORDEX." *AGU Fall Meeting Abstracts*, December. https://na-cordex.org.

Mearns, L. O., Seth McGinnis, Daniel Korytina, Raymond Arritt, Sebastien Biner, Melissa Bukovsky, Hsin-I Chang, et al. 2017. "The NA-CORDEX Dataset, Version 1.0." Boulder, CO: NCAR Climate Data Gateway. https://doi.org/10.5065/D6SJ1JCH.

Mearns, Linda, Ray Arritt, Dave Bader, Philip Duffy, William Gutowski, William Kuo, and Ruby Leung. 2008. "Collaborative Research The North American Regional Climate Change Assessment Program (NARCCAP): Using Multiple GCMs and RCMs to Simulate Future Climates and Their Uncertainty." http://www.narccap.ucar.edu/doc/about/Project_summary.pdf.

Mendoza, Pablo A., Naoki Mizukami, Kyoko Ikeda, Martyn P. Clark, Ethan D. Gutmann, Jeffrey R. Arnold, Levi D. Brekke, and Balaji Rajagopalan. 2016. "Effects of Different Regional Climate Model Resolution and Forcing Scales on Projected Hydrologic Changes." *Journal of Hydrology* 541 (October): 1003–19. https://doi.org/10.1016/j.jhydrol.2016.08.010.

Milly, P. C. D., Julio Betancourt, Malin Falkenmark, Robert M. Hirsch, Zbigniew W. Kundzewicz, Dennis P. Lettenmaier, and Ronald J. Stouffer. 2008. "Stationarity Is Dead: Whither Water Management?" *Science* 319 (5863): 573–74. https://doi.org/10.1126/science.1151915.

Mirhosseini, Golbahar, Puneet Srivastava, and Lydia Stefanova. 2013. "The Impact of Climate Change on Rainfall Intensity–Duration–Frequency (IDF) Curves in Alabama." *Regional Environmental Change* 13 (1): 25–33. https://doi.org/10.1007/s10113-012-0375-5.

Mittman, Tamara, and Christopher Kloss. 2015. "The Economic Benefits of Green Infrastructure: A Case Study of Lancaster, PA."

Morgan, Millett Granger, Max Henrion, and Mitchell Small. 1992. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge university press.

Moss, R. H., G. A. Meehl, M. C. Lemos, J. B. Smith, J. R. Arnold, J. C. Arnott, D. Behar, et al. 2013. "Hell and High Water: Practice-Relevant Adaptation Science." *Science* 342 (6159): 696–98. https://doi.org/10.1126/science.1239569.

Mote, Philip, Levi Brekke, Philip B. Duffy, and Ed Maurer. 2011. "Guidelines for Constructing Climate Scenarios." *Eos* 92 (31): 257–58.

M'Po, Yèkambèssoun N'Tcha, Agnidé Emmanuel Lawin, Ganiyu Titilope Oyerinde, Benjamin Kouassi Yao, and Abel Akambi Afouda. 2016. "Comparison of Daily Precipitation Bias Correction Methods Based on Four Regional Climate Model Outputs in Ouémé Basin, Benin." *Hydrology* 4 (6): 58–71. https://doi.org/10.11648/j.hyd.20160406.11.

Murphy, J. 1999. "An Evaluation of Statistical and Dynamical Techniques for Downscaling Local Climate." *Journal of Climate* 12 (8 PART 1): 2256–84. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0033172619&partnerID=40&md5=fd6106d171b24b95b1b14fab06aa6e81.

Musau, J, J Sang, and J Gathenya. 2013. "General Circulation Models (GCMs) Downscaling Techniques and Uncertainty Modeling for Climate Change Impact Assessment." In *Proceedings of 2013 Mechanical Engineering Conference on Sustainable Research and Innovation*. Vol. 5.

Nakicenvoic, N, Joseph Alcamo, Gerald Davis, Bert de Vries, Joergen Fenhann, Stuart Gaffin, Kenneth Gregory, et al. 2000. *IPCC Special Report on Emissions Scenarios*. New York, NY and Cambridge, UK: Cambridge University Press.

NOAA. 2016. "National Centers for Environmental Information." October 2016. http://www.ncdc.noaa.gov/.

Olsen, J., ed. 2015. "Adapting Infrastructure and Civil Engineering Practice to a Changing Climate," May. http://dx.doi.org/10.1061/9780784479193.

Olsen, J., Bilal M. Ayyub, Ana Barros, Wayne Lei, Franklin Lombardo, Miguel Medina, Constantine Samaras, et al. 2015. "Adapting Infrastructure and Civil Engineering Practice to a Changing Climate." *American Society of Civil Engineers*, Committe on Adaptation to a Changing Climate, , May. http://dx.doi.org/10.1061/9780784479193.

Onof, C., and K. Arnbjerg-Nielsen. 2009. "Quantification of Anticipated Future Changes in High Resolution Design Rainfall for Urban Areas." *7th International Workshop on*

Precipitation in Urban Areas 7th International Workshop on Precipitation in Urban Areas 92 (3): 350–63. https://doi.org/10.1016/j.atmosres.2009.01.014.

Osborn, Liz. 2018. "United States' Rainiest Cities." Current Results. 2018. https://www.currentresults.com/Weather-Extremes/US/wettest-cities.php.

Overeem, Aart, Adri Buishand, and Iwan Holleman. 2008. "Rainfall Depth-Duration-Frequency Curves and Their Uncertainties." *Journal of Hydrology* 348 (1): 124–34. https://doi.org/10.1016/j.jhydrol.2007.09.044.

PA DEP. 2006. "Pennsylvania Stormwater Best Management Practices Manual." Harrisburg, PA: Bureau of Watershed Management. http://pecpa.org/wp-content/uploads/Stormwater-BMP-Manual.pdf.

PennDOT. 2011. "Chapter 7, Appendix A: Field Manual For Pennsylvania Design Rainfall Intensity Charts from NOAA Atlas 14 Version 3 Data." In *PennDOT Drainage Manual 2010 Edition*, 1–40. Harrisburg, PA: Pennsylvania Department of Transportation.

———. 2015. "Chapter 13, Storm Drainage Systems." In *PennDOT Drainage Manual 2015 Edition*, 13–25. Harrisburg, PA: Pennsylvania Department of Transportation. https://www.dot.state.pa.us/public/pubsforms/Publications/PUB%20584.pdf.

Pierce, David W., Tim P. Barnett, Benjamin D. Santer, and Peter J. Gleckler. 2009. "Selecting Global Climate Models for Regional Climate Change Studies." *Proceedings of the National Academy of Sciences* 106 (21): 8441–46.

Prein, Andreas F., Roy M. Rasmussen, Kyoko Ikeda, Changhai Liu, Martyn P. Clark, and Greg J. Holland. 2016. "The Future Intensification of Hourly Precipitation Extremes." *Nature Climate Change* 7 (December): 48. http://dx.doi.org/10.1038/nclimate3168.

Prudhomme, Christel, and Helen Davies. 2009. "Assessing Uncertainties in Climate Change Impact Analyses on the River Flow Regimes in the UK. Part 1: Baseline Climate." *Climatic Change* 93 (1): 177–95. https://doi.org/10.1007/s10584-008-9464-3.

Prudhomme, Christel, Nick Reynard, and Sue Crooks. 2002. "Downscaling of Global Climate Models for Flood Frequency Analysis: Where Are We Now?" *HYDROLOGICAL PROCESSES* 16 (6): 1137–50. https://doi.org/10.1002/hyp.1054.

Prüss-Üstün, Annette, Robert Bos, Fiona Gore, and Jamie Bartram. 2008. *Safer Water, Better Health: Costs, Benefits and Sustainability of Interventions to Protect and Promote Health*. World Health Organization.

Reanalysis.org. 2016. "Advancing Reanalysis." CIRES at University of Colorado Boulder. http://reanalyses.org/.

Reichler, Thomas, and Junsu Kim. 2008. "How Well Do Coupled Models Simulate Today's Climate?" *Bulletin of the American Meteorological Society* 89 (3): 303.

Riahi, Keywan, Shilpa Rao, Volker Krey, Cheolhung Cho, Vadim Chirkov, Guenther Fischer, Georg Kindermann, Nebojsa Nakicenovic, and Peter Rafaj. 2011. "RCP 8.5—A Scenario of Comparatively High Greenhouse Gas Emissions." *Climatic Change* 109 (1): 33. https://doi.org/10.1007/s10584-011-0149-y.

Rootzén, Holger, and Richard W. Katz. 2013. "Design Life Level: Quantifying Risk in a Changing Climate." *Water Resources Research* 49 (9): 5964–72. https://doi.org/10.1002/wrcr.20425.

Rosenberg, Eric A., Patrick W. Keys, Derek B. Booth, David Hartley, Jeff Burkey, Anne C. Steinemann, and Dennis P. Lettenmaier. 2010. "Precipitation Extremes and the Impacts of Climate Change on Stormwater Infrastructure in Washington State." *Climatic Change* 102 (1): 319–49. https://doi.org/10.1007/s10584-010-9847-0.

Sarr, M.A., O. Seidou, Y. Tramblay, and S. El Adlouni. 2015. "Comparison of Downscaling Methods for Mean and Extreme Precipitation in Senegal." *Journal of Hydrology: Regional Studies* 4 (September): 369–85. https://doi.org/10.1016/j.ejrh.2015.06.005.

Schmidli, J., C. M. Goodess, C. Frei, M. R. Haylock, Y. Hundecha, J. Ribalaygua, and T. Schmith. 2007. "Statistical and Dynamical Downscaling of Precipitation: An Evaluation and Comparison of Scenarios for the European Alps." *Journal of Geophysical Research: Atmospheres* 112 (D4): n/a-n/a. https://doi.org/10.1029/2005JD007026.

Scholz, F. W. 1980. "Towards a Unified Definition of Maximum Likelihood." *Canadian Journal of Statistics* 8 (2): 193–203. https://doi.org/10.2307/3315231.

Semadeni-Davies, Annette, Claes Hernebring, Gilbert Svensson, and Lars-Göran Gustafsson. 2008. "The Impacts of Climate Change and Urbanisation on Drainage in Helsingborg, Sweden: Combined Sewer System." *Journal of Hydrology* 350 (1–2): 100–113. https://doi.org/10.1016/j.jhydrol.2007.05.028.

Shahabul Alam, Md., and Amin Elshorbagy. 2015. "Quantification of the Climate Change-Induced Variations in Intensity–Duration–Frequency Curves in the Canadian Prairies." *Journal of Hydrology* 527 (August): 990–1005. https://doi.org/10.1016/j.jhydrol.2015.05.059.

Simonovic, Slobodan P., Andre Schardong, Dan Sandink, and Roshan Srivastav. 2016. "A Web-Based Tool for the Development of Intensity Duration Frequency Curves under Changing Climate." *Environmental Modelling & Software* 81 (July): 136–53. https://doi.org/10.1016/j.envsoft.2016.03.016.

Soden, Brian J., Darren L. Jackson, V. Ramaswamy, M. D. Schwarzkopf, and Xianglei Huang. 2005. "The Radiative Signature of Upper Tropospheric Moistening." *Science* 310 (5749): 841. https://doi.org/10.1126/science.1115602.

Solaiman, Tarana A, and Slobodan P Simonovic. 2011. "Development of Probability Based Intensity-Duration-Frequency Curves under Climate Change." *Water Resources Research Report* 34: 1–93. https://ir.lib.uwo.ca/wrrr/34.

Spiller, Marc, Jan H.G. Vreeburg, Ingo Leusbrock, and Grietje Zeeman. 2015. "Flexible Design in Water and Wastewater Engineering – Definitions, Literature and Decision Guide." *Journal of Environmental Management* 149 (February): 271–81. https://doi.org/10.1016/j.jenvman.2014.09.031.

Srivastav, R, A Schardong, and S Simonovic. 2014. "Computerized Tool for the Development of Intensity-Duration-Frequency Curves under a Changing Climate. Technical Manual v. 1." Water Resources Research Report 089. London, Ontario, Canada: Facility for Intelligent Decision Support, Department of Civil and Environmental Engi- neering. http://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1034&context=wrrr.

Stoner, Anne M. K., Katharine Hayhoe, Xiaohui Yang, and Donald J. Wuebbles. 2013. "An Asynchronous Regional Regression Model for Statistical Downscaling of Daily Climate Variables." *International Journal of Climatology* 33 (11): 2473–94. https://doi.org/10.1002/joc.3603.

Sun, Xiaoming, and Ana P. Barros. 2014. "High Resolution Simulation of Tropical Storm Ivan (2004) in the Southern Appalachians: Role of Planetary Boundary-Layer Schemes and Cumulus Parametrization." *Quarterly Journal of the Royal Meteorological Society* 140 (683): 1847–65. https://doi.org/10.1002/qj.2255.

Taylor, Karl E., Ronald J. Stouffer, and Gerald A. Meehl. 2011. "An Overview of CMIP5 and the Experiment Design." *Bulletin of the American Meteorological Society* 93 (4): 485–98. https://doi.org/10.1175/BAMS-D-11-00094.1.

Thakali, Ranjeet, Ajay Kalra, Sajjad Ahmad, and Kamal Qaiser. 2018. "Management of an Urban Stormwater System Using Projected Future Scenarios of Climate Models: A Watershed-Based Modeling Approach." *Open Water Journal* 5 (2): 1.

U.S. EPA. 2016. "What Is Green Infrastructure?" Green Infrastructure. September 23, 2016. https://www.epa.gov/green-infrastructure/what-green-infrastructure.

USEPA. 2015. *EPA Storm Water Management Model* (version 5.1). http://www.epa.gov/water-research/storm-water-management-model-swmm.

Varian, Hal. 2005. "Bootstrap Tutorial." *Mathematica Journal* 9 (4): 768–75.

Viner, David, and Candice Howarth. 2014. "Practitioners' Work and Evidence in IPCC Reports." *Nature Climate Change* 4 (10): 848–50. https://doi.org/10.1038/nclimate2362.

Visser, H., and A. C. Petersen. 2012. "Inferences on Weather Extremes and Weather-Related Disasters: A Review of Statistical Methods." *Clim. Past* 8 (1): 265–86. https://doi.org/10.5194/cp-8-265-2012.

Vuuren, DetlefP. van, Jae Edmonds, Mikiko Kainuma, Keywan Riahi, Allison Thomson, Kathy Hibbard, GeorgeC. Hurtt, et al. 2011. "The Representative Concentration Pathways: An Overview." *Climatic Change* 109 (1–2): 5–31. https://doi.org/10.1007/s10584-011-0148-z.

156

Walker, Warren E., Robert J. Lempert, and Jan H. Kwakkel. 2013. "Deep Uncertainty." In *Encyclopedia of Operations Research and Management Science*, 395–402. Springer. http://link.springer.com/10.1007/978-1-4419-1153-7_1140.

Walsh, J.D., Wuebbles, K, and Kossin, J. 2014. "Ch. 2: Our Changing Climate. Climate Change Impacts in the United States." In *The Third National Climate Assessment*, 19–67. U.S. Global Change Research Program.

Wastewater Committee of the Great Lakes - Upper Mississippi River. 2014. "Recommended Standards for Wastewater Facilities Policies for the Design, Review, and Approval of Plans and Specifications for Wastewater Collection and Treatment Facilities." Health Research Inc. http://10statesstandards.com/wastewaterstandards.pdf.

Wehner, MichaelF. 2013. "Very Extreme Seasonal Precipitation in the NARCCAP Ensemble: Model Performance and Projections." *Climate Dynamics* 40 (1–2): 59–80. https://doi.org/10.1007/s00382-012-1393-1.

Westra, S., H. J. Fowler, J. P. Evans, L. V. Alexander, P. Berg, F. Johnson, E. J. Kendon, G. Lenderink, and N. M. Roberts. 2014. "Future Changes to the Intensity and Frequency of Short-Duration Extreme Rainfall." *Reviews of Geophysics* 52 (3): 522–55. https://doi.org/10.1002/2014RG000464.

Wi, Sungwook, JuanB. Valdés, Scott Steinschneider, and Tae-Woong Kim. 2015. "Non-Stationary Frequency Analysis of Extreme Precipitation in South Korea Using Peaks-over-Threshold and Annual Maxima." *Stochastic Environmental Research and Risk Assessment*, November, 1–24. https://doi.org/10.1007/s00477-015-1180-8.

Wilbanks, Thomas J., Steven J. Fernandez, and Melissa R. Allen. 2015. "Extreme Weather Events and Interconnected Infrastructures: Toward More Comprehensive Climate Change Planning." *Environment: Science and Policy for Sustainable Development* 57 (4): 4–15. https://doi.org/10.1080/00139157.2015.1048134.

Wilks, D. S., and R. L. Wilby. 1999. "The Weather Generation Game: A Review of Stochastic Weather Models." *Progress in Physical Geography* 23 (3): 329–57. https://doi.org/10.1177/030913339902300302.

Willems, Patrick, Jonas Olsson, Karsten Arnbjerg-Nielsen, Simon Beecham, Assela Pathirana, Ida Bülow Gregersen, Henrik Madsen, and Van-Thanh-Van Nguyen. 2013a. "Climate Change Impacts on Rainfall Extremes and Urban Drainage: A State-of-the-Art Review." *Geophysical Research Abstracts* 15.

———. 2013b. "Climate Change Impacts on Rainfall Extremes and Urban Drainage: A State-of-the-Art Review." *Geophysical Research Abstracts* 15.

Wood, A.W., L.R. Leung, V. Sridhar, and D.P. Lettenmaier. 2004. "Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs." *Climatic Change* 62 (1–3): 189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9e.

Wood, S. J., D. A. Jones, and R. J. Moore. 2000. "Static and Dynamic Calibration of Radar Data for Hydrological Use." *Hydrology and Earth System Sciences* 4: 545–54.

Wuebbles, Donald, Gerald Meehl, Katharine Hayhoe, Thomas R. Karl, Kenneth Kunkel, Benjamin Santer, Michael Wehner, Brian Colle, Erich M. Fischer, and Rong Fu. 2014. "CMIP5 Climate Model Analyses: Climate Extremes in the United States." *Bulletin of the American Meteorological Society* 95 (4): 571–83.

Xu, Chong-yu. 1999. "From GCMs to River Flow: A Review of Downscaling Methods and Hydrologic Modelling Approaches." *Progress in Physical Geography* 23 (2): 229–49. http://ppg.sagepub.com/content/23/2/229.abstract.

Zender, Charlie, Henry Butowsky, and Wenshan Wang. 2016. "Welcome to the NetCDF Operator (NCO) Site." source forge. http://nco.sourceforge.net/.

Zhu, J. 2012. "Impact of Climate Change on Extreme Rainfall across the United States." *Journal of Hydrologic Engineering* 18 (10): 1301–9. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000725.

Zhu, Jianting, William Forsee, Rina Schumer, and Mahesh Gautam. 2012. "Future Projections and Uncertainty Assessment of Extreme Rainfall Intensity in the United States from an Ensemble of Climate Models." *Climatic Change* 118 (2): 469–85. https://doi.org/10.1007/s10584-012-0639-6.

Zhu, Jianting, Mark C. Stone, and William Forsee. 2012. "Analysis of Potential Impacts of Climate Change on Intensity–Duration–Frequency (IDF) Relationships for Six Regions in the United States." *Journal of Water and Climate Change* 3 (3): 185–96. https://doi.org/10.2166/wcc.2012.045.

APPENDIX A: SUPPLEMENTARY MATERIAL FOR THE EFFECT OF MODELING CHOICES ON UPDATED PRECIPITATION FREQUENCY CURVES AND STORMWATER INFRASTRUCTURE

A.1 Overview

This Appendix presents additional methods and results related to the development of updated depth-duration-frequency (DDF) curves (Chapter 3 of the dissertation). The first section presents a detailed overview of adjustment methods used for updating DDF curves, including a performance comparison between methods. The second section presents results of how the choice of historical design storm value could alter the pipe diameter size.

A.2 Detailed overview of adjustment methods used for updating DDF curves

The three methods used for updating are: (1) Kernel Density Distribution Mapping (KDDM), which uses non-parametric bias-correction to adjust the underlying time series from the climate model, (2) AMS Transfer Function, which adjusts the annual maximum series of the climate model using parametric transfer functions, and (3) Simple Change Factor, which uses areal reduction factors to adjust the observed depth of rain based on the ratio of change between the historical and future climate models.

A.2.1 Kernel Density Distribution Mapping

The first method uses a type of non-parametric bias-correction, called Kernel Density Distribution Mapping (KDDM) (McGinnis et al. 2015). KDDM is used to define a relationship between the observed rainfall time series and the gridded climate model time series for the historical time period (1950–2013). Using this relationship, the entire gridded climate model time series, including zero values, is bias-corrected to the station scale for the period 1950–2099 by adjusting the model values so that their statistical distribution in the historical time

period matches that of the observations. Once the time series is bias-corrected, the AMS are extracted, and the return levels are subsequently obtained using the general techniques in Section 3.2.2 for each duration, climate model, and city.

The relationship between the observed rainfall time series and the gridded climate model time series is defined by fitting a transfer function between their empirical CDFs. First, empirical PDFs are computed using kernel density estimation; these PDFs are then integrated using the trapezoid rule to calculate CDFs. Equal points of probability from the CDFs are mapped against each other and the resultant mapping is then fitted with a spline. The equation of this spline is the transfer function between the observed data and the historical climate model simulation output, which is then applied to the 150-year time series of the climate model simulations to obtain bias-corrected values at the station scale.

The time series can be bias-corrected with KDDM before or after it is aggregated to the desired duration (e.g., 24-hour). The former, referred to as method (1a), bias-corrects the 1-hour time series and then aggregates using convolution after bias-correction. The latter, referred to as method (1b), aggregates using convolution and then bias-corrects the convolved time series. Although the convolved time series contains many values that are not independent of one another, this does not present a problem for the KDDM technique, which makes no attempt to fit a parametric distribution. As long as the two datasets have similar dependency structures, the method remains applicable. Figure presents a detailed overview of the sequence of steps required for both KDDM methods (1a) and (1b). The figure also presents steps for methods 2 and 3, the AMS Transfer Function and Simple Change Factor methods, respectively.

To evaluate method (1a) and (1b), the annual maximum series of the observed data (1950–2013) was compared to the AMS of the bias-corrected climate model for the same time period. This comparison involved comparing each duration using two metrics: (i) the mean absolute error (MAE), and (ii) a Kolmogorov–Smirnov (K-S) test. The MAE (see Equation A.1) analyzes the magnitude of difference between the empirical distributions, while the KS-test checks to
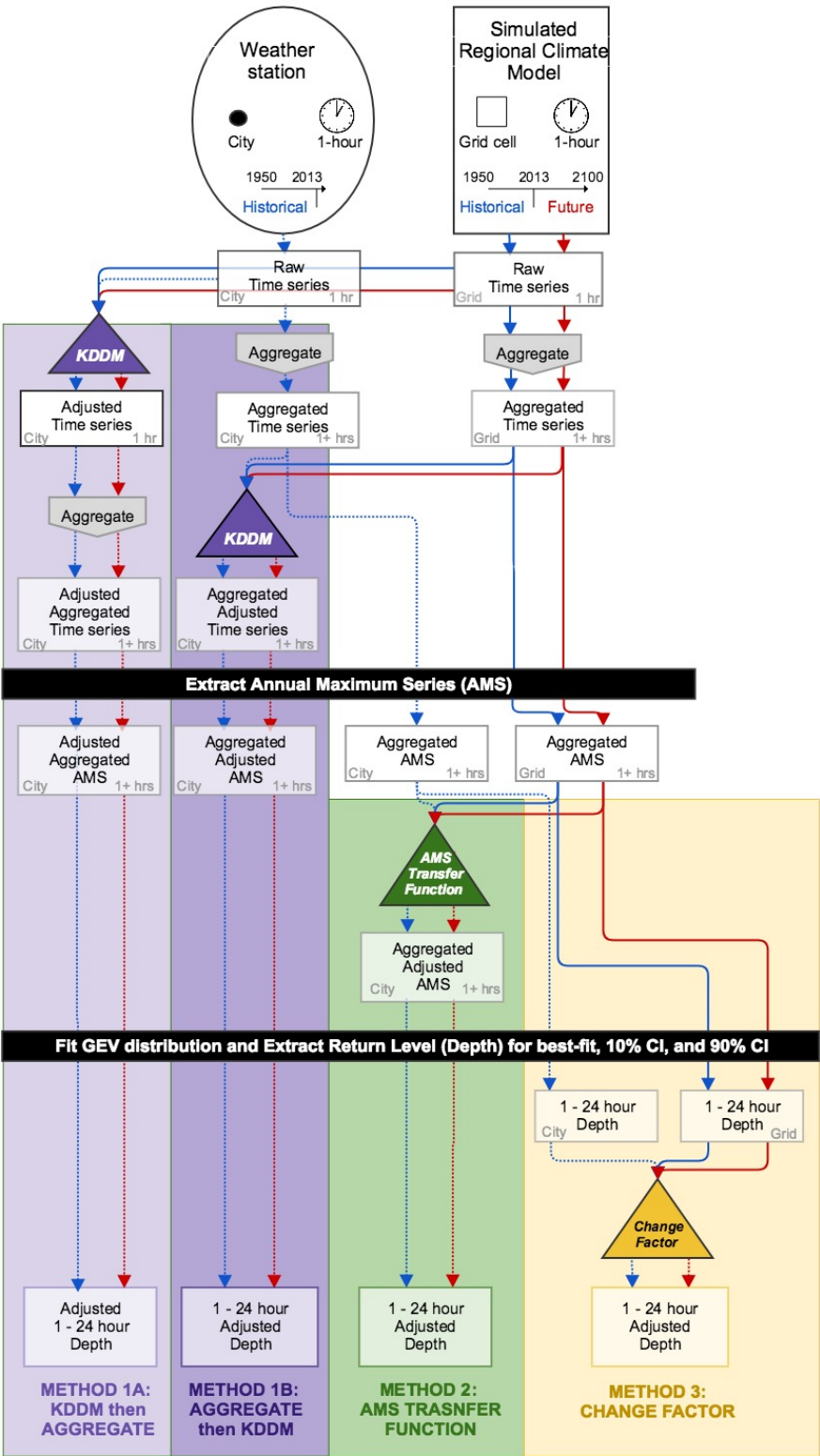
Figure A.1: Sequence of steps carried out in each of the DDF adjustment techniques used in this study.

see if the two AMS series are from different continuous probability distributions.

$$MAE_d = \frac{\sum_{q=1} |y_q - x_q|}{n} \tag{A.1}$$

The results, which are presented in Figure A.2 for Pittsburgh, show that aggregating the time series before bias-correction produces a much smaller error. This is true for all locations and durations (not shown). Method (1a) considerably overestimates the AMS compared to method (1b), and has a larger number of bias-corrected AMS that are not from the same continuous distribution as the observed AMS. Thus, only method (1b), aggregation then bias-correction, is used for subsequent analyses.

A.2.2   AMS Transfer Function

The transfer function method, or method (2), is also a bias-correction method like the KDDM method. However, instead of bias-correcting the entire time series, only the AMS is bias-corrected. This is shown in Figure 1. This method is similar to the Equidistance Quantile Matching Method developed by Solaiman and Simonovic (2011) as part of an effort to update Canadian IDF curves to reflect future extremes (Solaiman and Simonovic 2011; Simonovic et al. 2016). That method, which makes use of daily output from GCMs, uses two transfer functions: one that spatially downscales from the grid scale to the station scale, and another that temporally downscales from the daily GCM to the sub-daily level. For the present work, the temporal downscaling step was not performed. Rather, RCM output from NA-CORDEX that is already available at the sub-daily level is used, and separate transfer functions are applied for each duration. To get future DDF curves, the adjusted AMS (XFd) for each duration, climate model, and city, are fit to a stationary GEV distribution for a single, future time period (see Section 2.3.4). Return levels are extracted for this time period using Equation 1. Confidence intervals for each climate model are determined based on the MLE and bootstrapping discussed in Section 3.2.2.
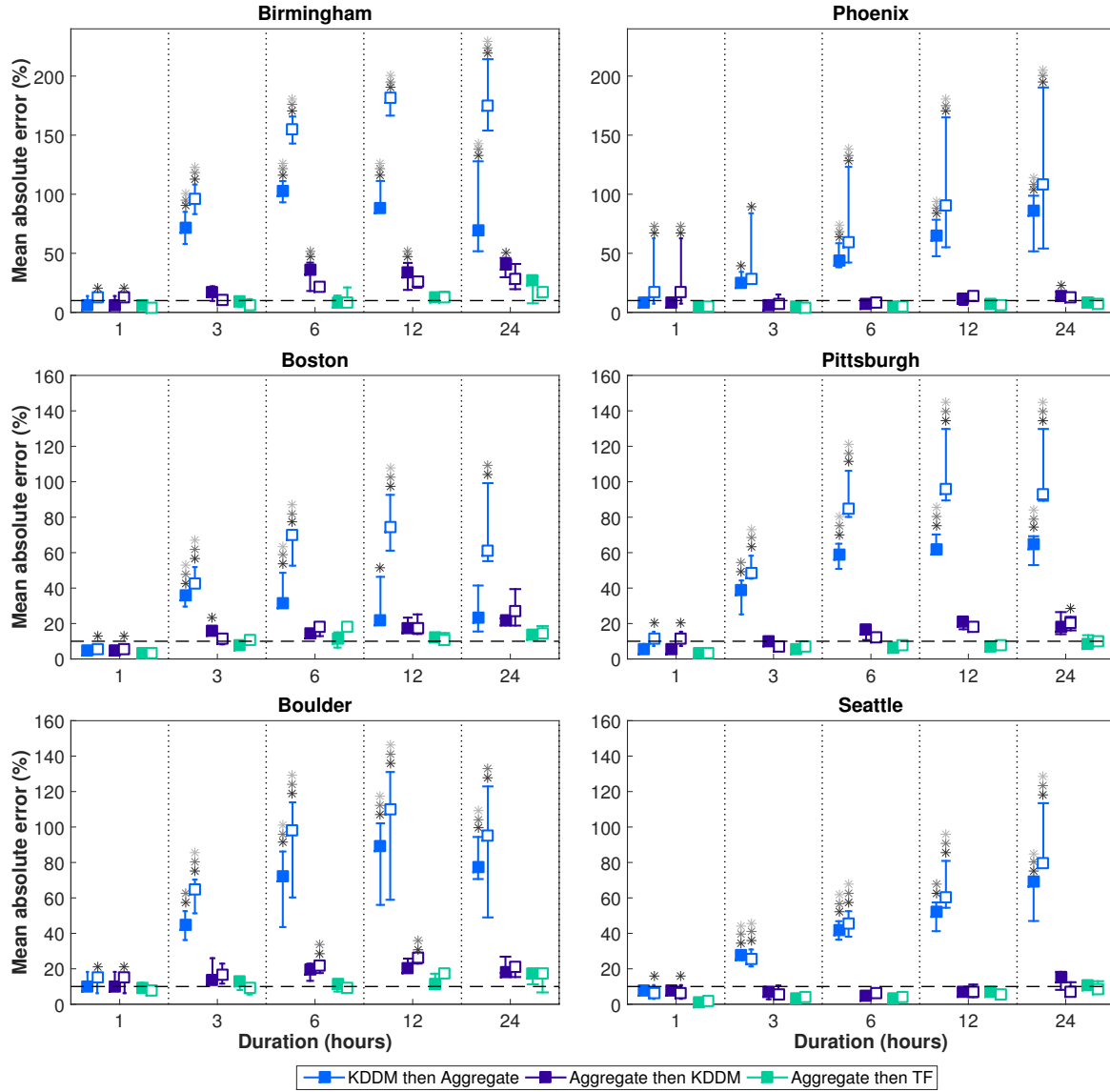
Figure A.2: Mean absolute error (MAE) of the AMS from the bias-corrected CORDEX ensemble compared to observed AMS for the historical time period (1950–2013). The different colored error bars represent the different bias correction methods: method (1a), KDDM then aggregation (light blue), and method (1b), aggregation then KDDM (dark blue), and (2) aggregation then transfer function. The median MAE from the 25-km climate model ensemble is represented by a solid, square marker, while the median of the 50-km ensemble is an open square. The maximum of the MAEs from the climate model ensemble is the upper bound of the error bar and the minimum of the MAEs is the lower bound of the error bar. The number of stars above the error bar shows the number of climate models that are not from the same continuous distribution as the observed AMS ($h_0 = 1$ from K-S test). The horizontal, dashed line represents a MAE of 10%.

Before the AMS transfer function method is applied, the AMS are first extracted from the observed data and from the climate model output for each duration. For each location, there are a total of 35 AMS (for 5 durations, 6 climate model simulations, and 1 weather station). The steps for developing the bias-corrected AMS are as follows:

1. Fit a GEV distribution, GEV_{X_d}, to the observed AMS, X_d

2. Fit a GEV distribution, GEV_{X_d} GEVYhd, to the AMS of the climate model, Y_d, for the historical period, Y_{h_d}. The number of years does not have to equal the number of years in the observed time period, but they should be similar time periods

3. Calculate $F_{GEV_{Y_{hd}}}(Y_d)$, the CDF of $GEV_{Y_{hd}}$ at the values of Y_{hd}, in order to get the probability for each value of Y_{hd}

4. Evaluate $F^{-1}GEV_{X_d}[F_{GEV_{Y_{hd}}}(Y_d)]$, the inverse CDF of GEV_{Xd} at $F_{GEV}, Y_{h_d}(Y_d)$. The result is the spatially adjusted AMS, X_{fd}

5. Repeat step 1 for each duration and city, and steps 2-4 for each climate model, duration, and city

To understand performance of this method relative to the KDDM method (1b), the AMS series of the historical climate simulation was also compared to the observed AMS time series, for all aggregated durations (3, 6, 12, 24, 48 hours), using the mean absolute error (MAE) and the K-S test. Results are shown in Figure A.2. Overall, the transfer function method has a smaller MAE than the KDDM method (1b), and never results in a rejection of the K-S test. Despite the better performance of the transfer function method, the KDDM method (1b) will still be used in subsequent analyses since the difference in error is relatively small. By continuing the analysis with both methods, the effect of this choice on performance can be evaluated.

Conceptually, methods (1) and (2) are entirely analogous: in both cases, model values are transformed into probabilities according to CDF of the model data, then transformed back into values using the CDF of the observations. The difference is in where that transformation is applied. Since the distribution of block maxima such as the AMS is known to converge to the

GEV distribution, it is not surprising that applying the transformation to the extreme values directly using the correct parametric distribution exhibits better performance than applying it implicitly using a non-parametric estimate.

A.2.3   Simple Change Factor Method

Method 3, the simple change factor method, adjusts the observed depth of rainfall for a specific duration and return period based on the change in the depth of rainfall for the same duration and return period from the historical and future climate model simulations. This method is also referred to as the areal reduction factor method (Allen and DeGaetano 2005; Zhu et al. 2012b; Cook et al. 2017). Future depths at the station scale are obtained with Equation A.2:

$$D_F^{(s)}(p,d) = D_H^{(s)}(p,d) \frac{D_F^{(g)}(p,d)}{D_H^{(g)}(p,d)} \tag{A.2}$$

where D denotes the depth of rainfall, or return level, for a given probability of occurrence (p), duration (d), and CI. This depth is either at the station scale (s) or grid scale (g), for future (F) or historical (H) time periods. Historical depths at the station scale refer to depths obtained using historical observations with methods in 3.2.2.

Depths at the grid scale for the historical or future period refer to return level depths obtained directly from the gridded regional climate model simulations. These depths are obtained by following the steps in 3.2.2 for each of the six climate model simulations for historical and future periods. For this analysis, the change factor is applied to observations of the time period 1950–2013. The historical climate model baseline is also from 1950–2013. The future time period is 2020 - 2080. Selection of the future period is discussed in Section 3.2.2).

The median future depth is estimated using the median observed return level, the median historical climate return level, and the median future return level. Confidence intervals for the future depth are obtained in the same manner, e.g., the 90% CI uses the 90% CI from the

Area (acres)	Design choice	Birmingham AL	Boston MA	Boulder CO	Phoenix AZ	Pittsburgh PA	Seattle WA
10	50-year mid	24 (600)	24 (600)	21 (525)	18 (450)	21 (525)	21 (525)
10	50-year upper	27 (675)	30 (750)	24 (600)	21 (525)	24 (600)	24 (600)
10	100-year mid	24 (600)	27 (675)	21 (525)	18 (450)	21 (525)	21 (525)
100	50-year mid	27 (675)	24 (600)	18 (450)	21 (525)	24 (600)	21 (525)
100	50-year upper	33 (825)	30 (750)	21 (525)	24 (600)	30 (750)	24 (600)
100	100-year mid	30 (750)	24 (600)	18 (450)	21 (525)	27 (675)	21 (525)

Table A.1: Diameter of stormwater pipe (in inches and mm) using historical design values for two watershed areas (10 acres and 100 acres) and all cities

observed data, historical climate model, and future climate model.

A.3   Comparison of all historical design storm choices on stormwater infrastructure sizing

This section presents the resulting pipe diameters if the design engineer were to use different values from the historical DDF curve range. The historical design value choices include selecting the 50-year, best-fit depth, the 50-year, 90% CI, and the 100-year best-fit depth. Table A.1 presents the diameter of the stormwater pipe in inches (and mm) using historical design values for both watershed areas (10 and 100 acres) and all 6 cities.

These results show that designing for the 100-year storm does not always increase the pipe

size, whereas designing for the 50-year 90% CI does. In general, for the smaller watershed, designing for the 100-year storm does would not change the pipe size, and therefore not provide additional protection against future extreme events. For the larger watershed, however, designing for the 100-year storm can provide some additional protection, but it is inconsistent across cities. Designing for the historical, 50-year 90% CI always leads to an increase of one or two pipe sizes.

APPENDIX B: SUPPLEMENTARY MATERIAL FOR RELATING RAINFALL CHARACTERISTICS TO PERFORMANCE OF RAIN GARDENS UNDER CHANGING RAINFALL PATTERNS

Table B.2: Drainage area characteristics of SWMM model

Drainage Area	Sq. Ft	Acres	Slope	% Impervious	% Zero-Impervious	Drains to
Roof 1	1440	0.03305	33%	100	100	RG 1
Roof 2	1360	0.03122	33%	100	100	RG 1
Pavement 1	1500	0.034435	2%	100	25	RG 1
Roof 3	700	0.01607	2%	100	100	RG 2
Roof 4	2500	0.05739	2%	100	100	RG 2
Pavement 2	860	0.01974	2%	100	25	RG 2
Rain Garden 1	1262	0.03	0.1%	0	25	RG 2
Rain Garden 2	907	0.02082	0.1%	0	25	Outfall

B.1   Overview

This Appendix presents additional methods and results related to Evaluation rainfall measures and performance of bio-retention systems under climate change (Chapter 4 of the dissertation). The first section presents additional methods and parameters for the SWMM model related to the sub-catchments, bio-retention cell layers, and conduits. The subsequent sections present additional results, including: pair-wise rainfall correlations, pair-wise performance metrics correlations, and

B.2   Methods and Data

B.2.1   SWMM model parameters

Eight sub-basins are modeled to represent four roof areas, two pavement areas, and two rain gardens. Table B.2 presents the sub-basin characteristics and flow patterns.

Additional parameter values related to the impervious sub-catchments are presented in Table B.3, whereas bio-retention cell layers, and conduits for Rain Garden 1 (RG1) and Rain

Table B.3: Parameters of impervious subcatchments is PCSWMM model

Parameters (units)	Value	Data Source
N-Imperv/N-Perv	0.01/0.1	SWMM default
DStore-Imperv/Perv (in)	0 /0.05	SWMM default
Percent Routed	100	Site configuration
Suction head (in) Green Ampt	2	
Conductivity (in/h)	1	SWMM default
Initial soil moisture deficit (fraction)	0	

Garden 2 (RG2) are reported in Table B.4. RG is the rain garden that has the potential to over-flow to the collection system.

B.2.2   Regional Climate Model Output

Table B.5 presents the regional climate model simulations from the NA-CORDEX (Mearns et al. 2017) dataset that were used in this study.

RegCM4 was originally developed at the International Centre for Theoretical Physics (ICTP) and simulated for NA-CORDEX at the National Center for Atmospheric Research (NCAR). WRF was developed and simulated at NCAR. CanRCM4 was developed and simulated at the Canadian Centre for Climate Modeling and Analysis.

B.3   Results

B.3.1   Historical precipitation characteristics

Rainfall characteristics in the Pittsburgh region are rainy and wet. Pittsburgh is in the top 10 rainiest cities in the U.S. On average, it rains about 146 days per year (mean value), or about every 1.6 days. With a mean total of 974 mm of rainfall per year, Pittsburgh does not make the top 10 cities with the most total rainfall (Osborn 2018). The maximum 1-day rainfall ranged

Table B.4: Parameters of the bio-retention basin layers and associated conduits for the PC-SWMM model

Parameters (units)	RG1	RG2	Data Source
Bio-retention cell – surface			
Berm height (in)	3	3	As-built drawings
Vegetation volume (fraction)	0.8	0.8	As-built drawings/ Google Earth
Roughness (Manning's n)	0.65	0.65	(McCuen 2005; Mustaffa, Ahmad, and Razi 2016)
Slope (%)	0.01	0.01	As-built drawings
Bio-retention cell – soil			
Thickness (in)	24	24	As-built drawings
Porosity (volume fraction)	0.3	0.28	Calibration
Field capacity (volume fraction)	0.15	0.15	
Wilting point (volume fraction)	0.125	0.125	
Conductivity (in/hr)	3	2.5	(Rawls, Brakensiek, and Miller 1983)
Conductivity slope	8	8	SWMM User Manual
Suction head (in)	2.04	2.4	(Rawls, Brakensiek, and Miller 1983)
Bio-retention cell – storage			
Thickness (in)	12	12	As-built drawings
Void ratio (voids/solids)	0.48	0.48	(Das 2008)
Seepage rate (in/hr)	1.25	1.25	Based on seepage rates for HDPE pipe
Clogging factor	0	0	Default
Bio-retention cell – underdrain			
Drain coefficient (in/hr)	0.5	0.5	Calibration
Drain exponent	0.5	0.5	Default
Drain offset height (in)	2	2	As-built drawings
Conduits			
Roughness (Manning's n)	0.012	0.009	(Bishop and Jeppson 1978)
Invert elevation (ft)	904.25	902.25	As-built drawings
Diameter (in)	8	6	As-built drawings

Table B.5: Regional climate model simulations from NA-CORDEX used in this study

RCM/ESM Combination	Regional Climate Model (RCM)	Earth System Model (ESM)
RegCM4/MPI–ESM–LR	RegCM4 RCM	"Max Planck Institute Earth System Model at base resolution" (MPI–ESM–LR)
WRF/MPI-ESM-LR	WRF RCM	MPI-ESM-LR
WRF/GFDL–ESM2M	WRF RCM	"Geophysical Fluid Dynamics Laboratory Earth System Model, Modular Ocean Version" (GFDL–ESM2M)
CanRCM4/CanESM2	CanRCM4 RCM	Second generation Canadian Earth System Model (CanESM2)

from 35.3 to 151.1 mm from 1990 through 2018. Out of the top 100 most populated cities in the U.S., Pittsburgh ranks around 80th in terms of daily maximum precipitation for this time period.

Figure B.3 presents two visualizations of the annual rainfall indices for the historical period (1990 – 2018) in Pittsburgh. Figure B.3(a) presents volumetric indices, including: the total annual rainfall, presented as a time series along the x-axis; the maximum 1-day precipitation, represented by the color of the marker; and the number of days with at least 25 mm of rain, shown as the size of the marker. Figure B.3(b) presents frequency indices, including: the number of rain days per year, presented as a time series along the x-axis; the average duration of wet days, represented by the color of the marker; and the maximum duration of wet days (the longest period of consecutive rainfall), shown as the size of the marker. The dashed, black line represents the arithmetic mean of the time series in each plot.

The total annual precipitation since 1990 ranged from 734 mm in 1995 to 1380 mm in 2004. Several factors contributed to 2004 having the highest total annual precipitation. Hurri-
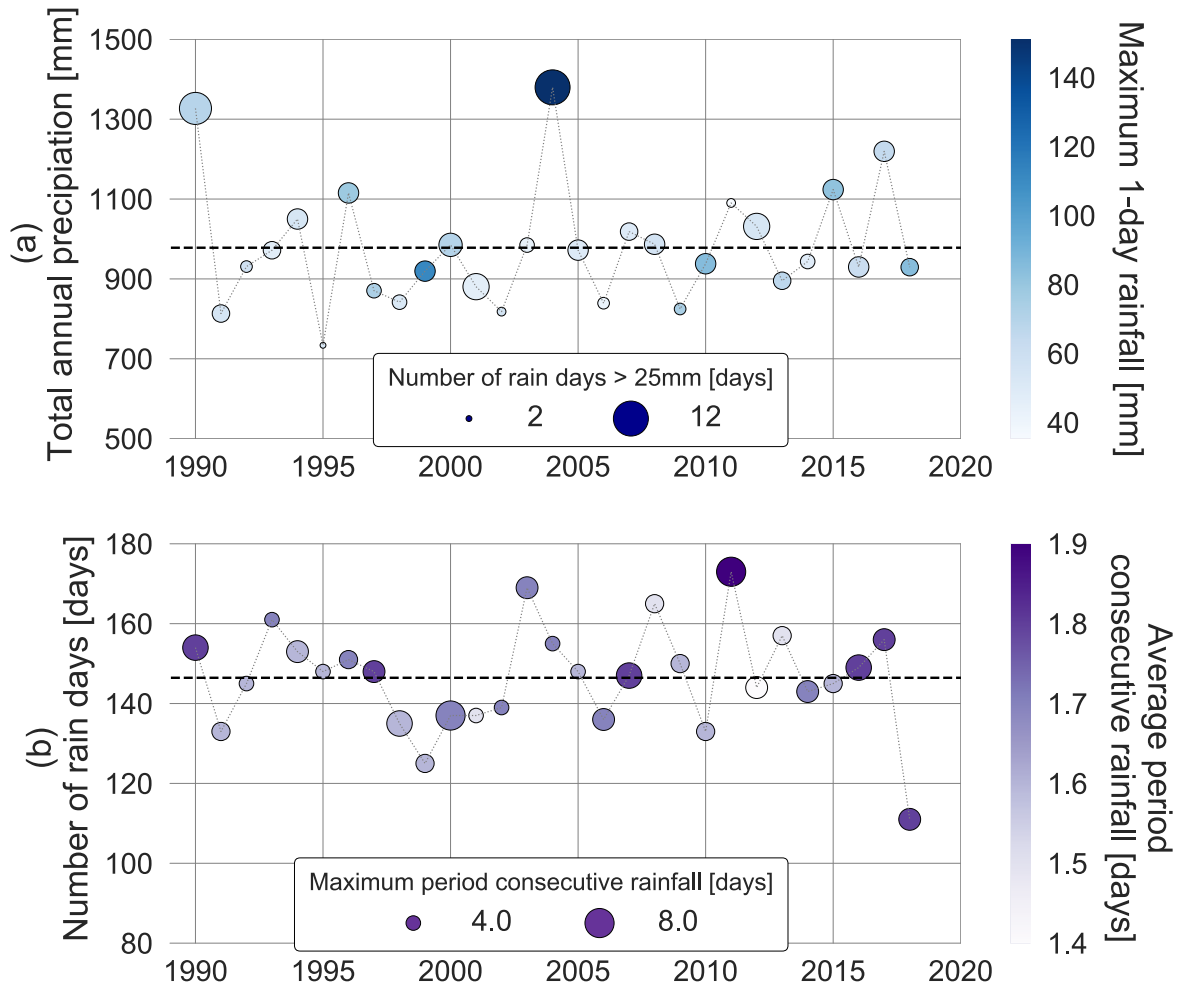
Figure B.3: Annual precipitation indices for the historical period (1990 – 2018) relating to volume (a), including: total rainfall, maximum 1-day rainfall, and no. days with ≥25 mm of rain; and relating to frequency (b), including: no. rain days per year, average and maximum duration of wet periods. Marker size represents no. days with ≥25 mm of rain (a) and maximum duration of wet days (b). Dashed line shows the time series mean.

cane Ivan hit the city in 2004, leading to the largest maximum 1-day precipitation of 151 mm, along with the largest number of days with at least 25 mm of rain (12 days). However, 1999 also had a large maximum 1-day precipitation (112 mm) during a year when the total annual precipitation was below the mean. The smallest, maximum 1-day rainfall of 35 mm occurred in 2011 and not in 1995, when total annual rainfall was lowest.

Similarly, looking at panel b, the longest maximum period of consecutive rainfall, 9 days, occurred in 2000 and 2011, not in the wettest overall year (2004). The number of rain days per year corresponds somewhat to total annual rainfall, however, there are some years, like 2003 and 2011, that had a lot of rainy days, but did not have higher total annual rainfall.

B.3.2   Historical correlations between rainfall indices

Pair-wise correlations were evaluated for all rainfall indices to determine quantitative relationships. Figure B.4 presents correlations between rainfall indices for the historical period (1990 - 2018). Results are shown as a heat map matrix. Orange colors represent a negative correlation and greys represent a positive correlation. The left and top represent the same metrics in the same order; however, the top rows use abbreviated names (see definitions in Table 4.1 in the main text).

The strongest positive correlations are between the maximum 1, 2, and 5-day rainfall values (as expected). Additional positive correlations exist between the total annual precipitation and the 90th quantile (0.7), the 90th quantile and the number of rain days above 25mm (0.7), and the number of rain days above 50 mm and the maximum 2-day rainfall (0.7). The number of days with rainfall above 25 mm has the highest number of strong correlations (apart from maximum-1, 2-, and 5-day rainfall). It is strongly correlated (0.7) to total annual rainfall, maximum 5-day rainfall, and the 90th quantile of daily rainfall.
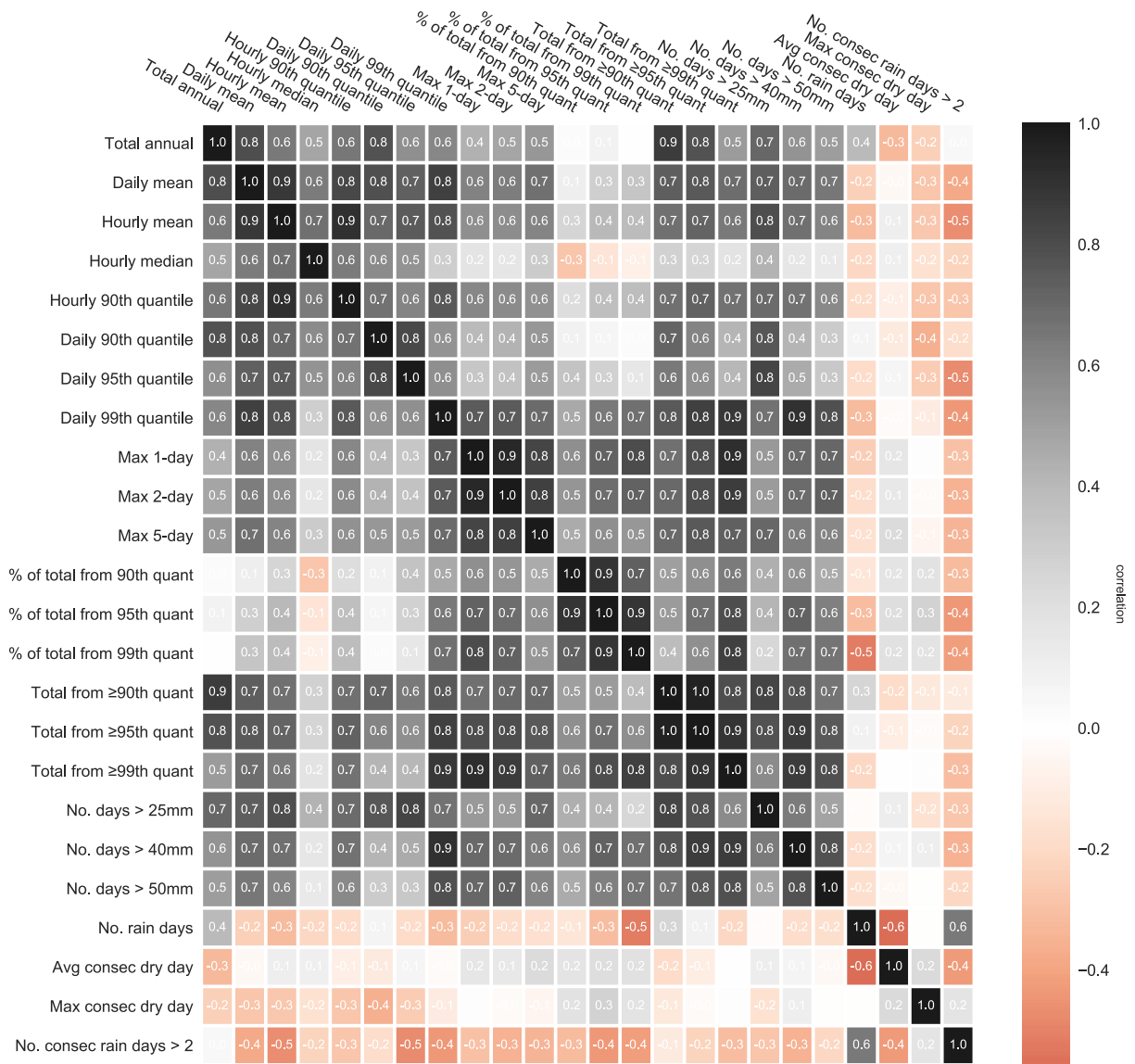
Figure B.4: Heat map of correlations among all rainfall indices for historical data. Darker colors represent a stronger correlation.

	Runoff Capture Efficiency	Volume overflows	Frequency Overflows	Max surface drainage time	Mean surface drainage time
Runoff Capture Efficiency	1	-0.83	-0.26	-0.70	-0.69
Volume Overflows		1	0.34	0.60	0.62
Frequency overflows			1	0.18	-0.10
Max surface drainage time				1	0.55
Mean surface drainage time					1

Table B.6: Correlations of performance metrics to each other

B.3.3 Historical correlations amongst performance metrics

Table presents pair-wise correlations of the performance metrics.

B.3.4 Disparities and similarities in future performance over time

Figure B.5 presents future performance over time for the two selected model simulations: consistent (left) and variable (right), and three performance metrics: percent captured (top), volume of overflows (middle), and frequency of overflows (bottom). The box plot to the left of each figure summarizes the historical range, whereas the pox plot to the right shows the range for the future simulation. The color of the markers represents the magnitude of the rainfall index with the strongest correlation to the performance metric. The size of the markers represents the total rainfall from greater than or equal to the 95th quantile of daily rainfall.
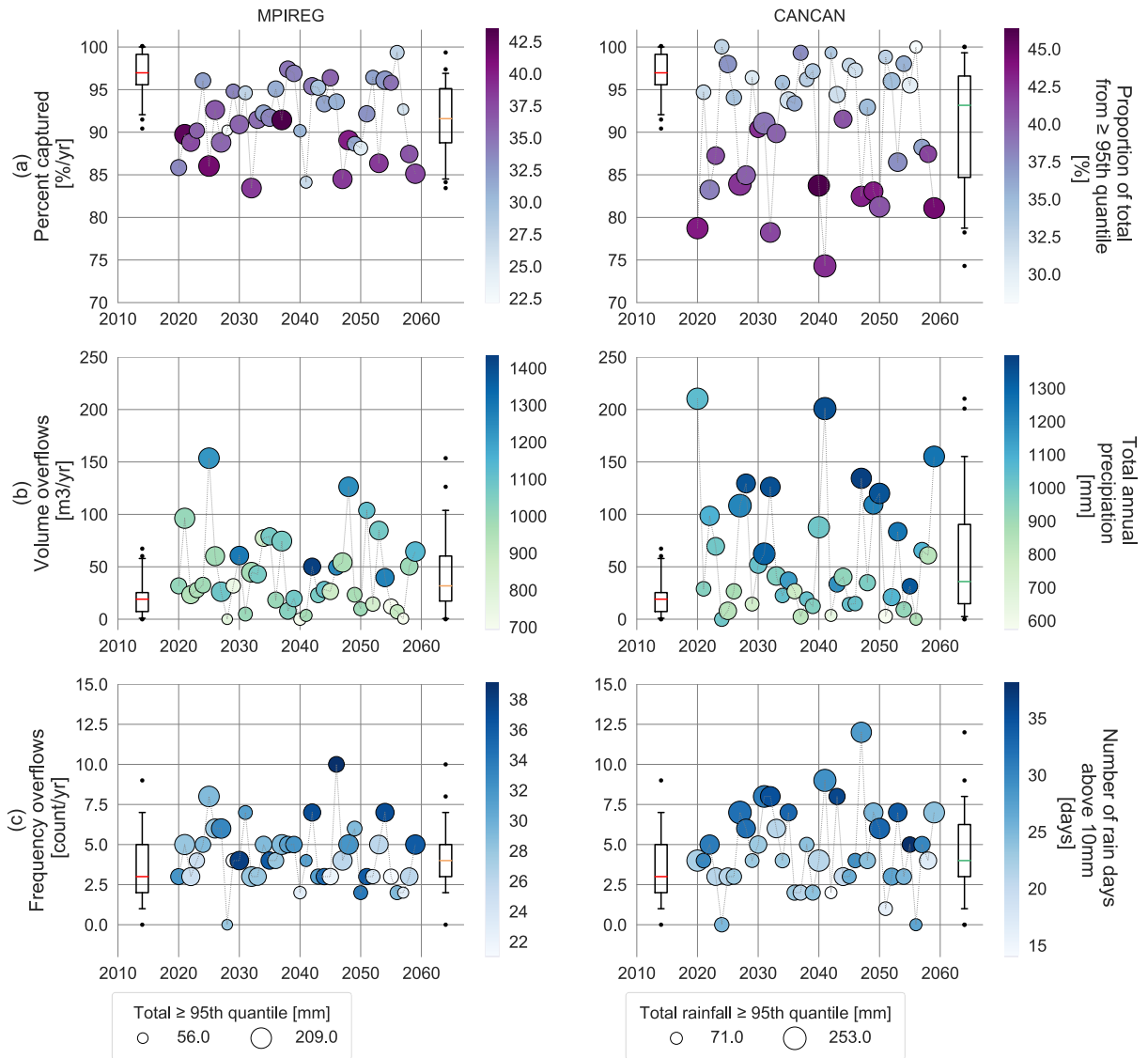
Figure B.5: Future performance over time for two selected model simulations: consistent (left) and variable (right), and three performance metrics: percent captured (top), volume of overflows (middle), and frequency of overflows (bottom). The box plot to the left of each figure summarizes the historical range, whereas the pox plot to the right shows the range for the future simulation. The color of the markers represents the magnitude of the rainfall index with the strongest correlation to the performance metric. The size of the markers represents the total rainfall from ≥ the 95th quantile.

The variation in the percent capture is clearly larger for the variable climate simulation than for consistent climate simulation. In the consistent simulation, the percent capture hovers around 92%, and is not affected by the changes in the proportion of total rainfall from the 95th quantile or above (lighter colors, meaning a lower proportion, are not related to higher capture efficiency, like they are in the variable simulation). For the other two metrics, the variability across years is large in both model simulations, which explains why correlations to these metrics are stronger and similar than for capture efficiency.