November 2018

Preparing Water Supply Systems for Climate Change: The Role of Hydrologic Forecasting in the Northeast

Leslie DeCristofaro
University of Massachusetts Amherst

Preparing Water Supply Systems for Climate Change:
The Role of Hydrologic Forecasting in the Northeast

A Dissertation Presented

by

LESLIE DECRISTOFARO

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2018

Department of Civil and Environmental Engineering

Preparing Water Supply Systems for Climate Change:
The Role of Hydrologic Forecasting in the Northeast

A Dissertation Presented

by

LESLIE DECRISTOFARO

Approved as to style and content by:

____
Richard Palmer, Chair

____
David Ahlfeld, Member

____
Anita Milman, Member

____
Richard Palmer, Department Chair
Department of Civil and Environmental Engineering

To my "smart non-experts"
including my mother, who gave me that advice.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Rick Palmer, for his expertise and guidance in my work and education, and for his patience and flexibility as my projects strayed from what we originally imagined six years ago.

Another thank you is owed to my committee members, David Ahlfeld, an incredible teacher, and Anita Milman, a master in the art of constructive criticism.

For insight, feedback, data, modeling, and discussions ongoing over the course of years, I have the following people to thank: at UMass, Alex McIntyre, Katie Booras, Nick Rossi, Sarah Whateley, and Scott Steinschnieder; at the NYCDEP and at the City University of New York, Adao Matonse, Allan Frei, Don Pierson, Elliot Schniederman, Jim Porter, Nachi Acharya, and Mark Zion; at Hazen and Sawyer, Kinsey Hoffman.

Funding for this work was provided by the Perrell Fellowship, the originators of which I had the pleasure to meet during my first year at UMass, and by the NOAA RISA Consortium for Climate Risk in the Urban Northeast, whose breadth of research was both fascinating and enriching.

To UMass EWRE, thank you for being the very best at what you do.

To everyone else, and there are a lot of you, I'd prefer to tell you how much you mean to me in person.

ABSTRACT

PREPARING WATER SUPPLY SYSTEMS FOR CLIMATE CHANGE: THE ROLE OF HYDROLOGIC FORECASTING IN THE NORTHEAST

SEPTEMBER 2018

LESLIE DECRISTOFARO, B.S., WASHINGTON UNIVERSITY IN ST LOUIS
M.S., UNIVERSITY OF MASSACHUSETTS AMHERST
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST
Directed by: Dr. Richard Palmer

Fresh water is a resource strongly impacted by climactic conditions. Water supply systems in the northeastern United States will see the effects of climate change on their water quality and quantity in various ways, including changes in seasonality of flows, changes in the frequency and magnitude of extreme precipitation events, and changes in the variability of precipitation and water availability. Five northeastern water supplies examined are expected to maintain at least 95% monthly reliability over a range of climates wider than the current projections. However, model results indicate that turbidity levels in New York City's Ashokan Reservoir will change with changes in mean annual precipitation and temperature.

Through a series of linked models of stochastic weather, hydrologic processes, and the supply system, Chapter 2 demonstrates the robustness of several adaptations available to the New York City Water Supply System to mitigate drought and manage water quality under climate change projections through the end of the century. Results illustrate how reducing demand and managing storage and releases based on hydrologic forecasting reduce the frequency of drought warnings and emergencies and improve system reliability in all climate change scenarios investigated. Through operations that limit turbidity propagation through the system and improvements to the Catskill Aqueduct to lower the minimum flow under conditions with high

turbidity, results demonstrate decreases lower turbidity loads and a reduction in emergency Alum use. These options demonstrate cumulative benefits when used in combination.

Chapter 3 seeks to quantify the amount of water supply system performance improvement that can be expected from improved forecasting in managing drought conditions. Using existing forecasts for Lancaster, PA, synthetic forecasts with varying quality, and a system model of the Baltimore, MD water supply system, this chapter demonstrated a method for quantifying improved system performance as a function of improved forecast quality, finding improvements in system performance to be approximately linear over a large range of forecast quality.

Chapter 4 tests a new method for the creation of statistical first-order autoregressive streamflow forecasts by conditioning the parameters and ensemble variance on a “hydrologic regime,” defined in several different ways. National Weather Service seasonal outlooks for precipitation are used as categorical forecasts of precipitation. The forecasts are found to have small positive skill, and for two of three sites, this skill is enough to result in small gains in the CRPSS of the ensemble hydrologic forecast. Utilizing perfect categorical forecasts indicates that adjusting the ensemble variance (rather than the autoregressive parameter) based on forecasted precipitation is responsible for the majority of improvements in skill for this method. The method is limited by the difficulty of long lead-time precipitation forecasting.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION: THE CLIMATE CHANGE OUTLOOK FOR NORTHEAST CITIES

1.1 Introduction

Fresh water is a resource tied directly and immediately to the climate. Large and small water supply systems throughout the northeast United States and throughout the world will see the effects of a changing climate in the coming century. The Water Utility Climate Alliance, a collection of eleven water utilities across the United States, proposes a four-step process for adapting to climate change: understanding the science, assessing the water supply system's vulnerabilities, incorporating climate change into planning, and implementing new methods, models, and adaptation strategies (Means et al. 2010). In the framework of the Water Utility Climate Alliance, this dissertation includes a review of the state of the science for climate change and its effect on water resources, and details several experiments in water resource management on the effectiveness of climate change adaptations ranging from infrastructure projects to sophisticated water supply management strategies. Sections 1.3 – 1.5 of this chapter summarize the results of three papers, currently published or in preparation, that I contributed to prior to the remaining chapters of this dissertation, following the first steps of understanding the science and assessing the effects of climate change on east coast water utilities.

1.2 The link between climate change and water resources

The Intergovernmental Panel on Climate Change lists water resources as a major impact of climate change (Cisneros et al. 2014). The National Climate Assessment discusses these impacts with regional specificity. The IPCC's technical paper on water (Bates et al. 2008) and a commentary in Science (Milly et al. 2008) were early suggestions that, for water resources planning, past conditions may not capture variability, extremes, or even average conditions in the future.

Climate change will create changing patterns of water quality and availability. Globally, a trend towards extremes is predicted, with greater prevalence of heavy precipitation and extreme drought (Bates et al. 2008). Changes in precipitation and temperature will result in alterations in seasonality of flows. Some water supply systems account for water storage in snowpack in their design, and reduced snow storage reduces the system's capacity to deliver water. Raw water quality will also be affected. Warmer temperatures will result in increased and more widespread eutrophication, while an increase in heavy precipitation events will increase sediment and pollutant loads into surface waters. The increased loads will overshadow the effect of dilution in many cases (Cisneros et al. 2014). Supply system reliability and water quality are also influenced by non-climate drivers, including land use, demand, management strategies, and agricultural practices. Changes in these additional factors make it harder to detect trends in streamflow (Cisneros et al. 2014).

Adaptation strategies can incorporate all of these drivers. No-regret, low-regret, or win-win strategies seek actions that are beneficial under most or all of the possible climate futures, and decision support tools can explore various uncertainties associated with the changing climate and the results of possible actions (Mimura et al. 2014). Risk is defined by the IPCC as the combination of exposure, vulnerability, and the likelihood of the weather event itself. Reducing the risk from a particular event involves reducing exposure (the built environment exposed to an event), and vulnerability (the adverse effects of the event on those populations) (Cardona et al. 2012).

In North America, adaptation and planning occur most often at the municipal level (Romero-Lankao et al. 2014). In the northeast United States, issues of greatest concern include the effects of changes in temperature and precipitation on water quality, the effects of rising sea levels on coastal storm risk and freshwater water resources, and aging infrastructure vulnerable to extreme events (Georgakakos et al. 2014). The northeast has experienced the largest increase in heavy precipitation in the United States (Kunkel et al. 2013) and is projected to be the first region in the country to see an increase in average temperature of 2° C, 10-20 years before the global mean temperature reaches that level (Karmalkar and Bradley, 2017). The northeast is a leader in planning for climate change and developing adaptation policies (Horton et al. 2014).

1.3 Climate stress testing Northeast water supply: an assessment of vulnerabilities and climate risk exposure

Combining projected changes in climate statistics and system-specific capacities and demands provides insight into the vulnerability of several large northeast water supplies: New York City, NY; Boston, MA; Springfield, MA; Hartford, CT; and Providence, RI. The water supplies have experienced significant drought due to interannual variability with detrimental effects on the region's water. The drought of the 1960s resulted in significant water use restrictions and remains the drought of record for much of the region.

The vulnerability of the region's water supply to climate change can be explored by modeling a range of possible changes in climate statistics in conjunction with the water supply system itself. The process includes three steps: 1) the creation of a set of weather time series with incremental changes in precipitation and temperature, 2) a rainfall-runoff model for translating those weather sequences into streamflow sequences, and 3) a model of the water supply system to simulate storage and supply. This approach first determines the climate conditions in which a system would suffer major shortfalls, and then asks whether those changes are likely based on the most recent climate models.

Whateley et al (in preparation) applied a monthly time-step model to the water supplies of NYC, Boston, Springfield, Hartford, and Providence to determine water supply reliability in the future. The systems showed greater than 95% reliability in most or all of the tested conditions, that is, in 95% or more of months, the full quantity of demand was delivered. In

Springfield, reliability drops below 95% at approximately a 16% decrease in average annual precipitation, which is outside of the range of the Coupled Model Intercomparison Project Phase 5 ensemble projection of climate change (Figure 1.1). Only in a subset of the scenarios representing a worst-case of internal climate variability did other systems drop below 95% reliability or did Springfield drop below 95% reliability in the range of climate change projections (Figure 1.2) (Whateley et al. In Preparation).

Modeling water quantity on a coarse time-step, northeast water supply systems look fairly robust to changes in annual temperature and precipitation. Therefore, a further search for the impacts of climate change on northeast water supply should explore the impacts of extreme events, utilize shorter time-steps, and incorporate water quality.

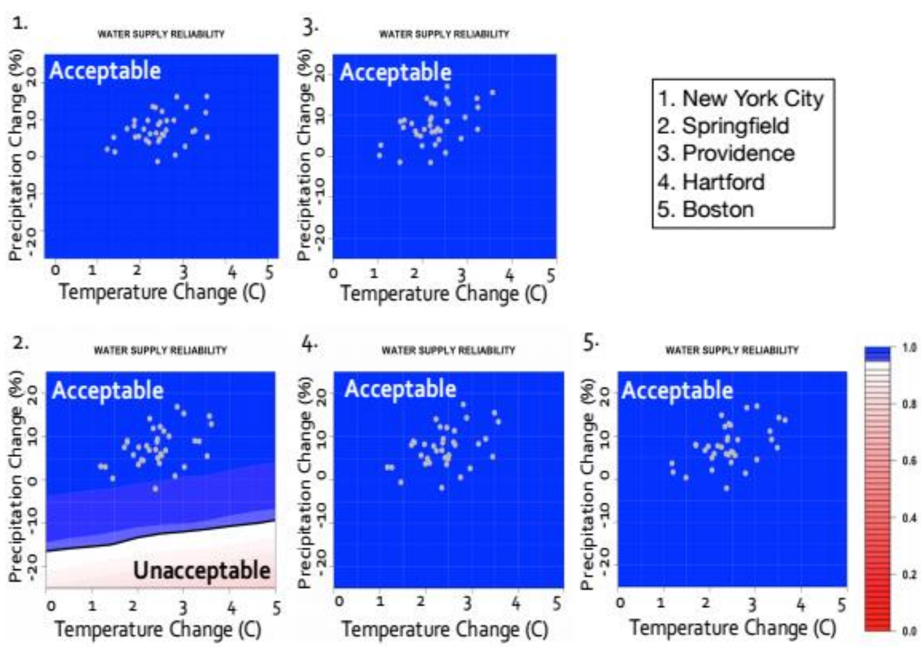

Figure 1.1 Water supply reliability for all systems with the full range of natural variability. Incremental changes in temperature and precipitation make up the x and y axes, while the gradient represents modeled reliability. Points on the plot represent individual models within the CMIP5 dataset.

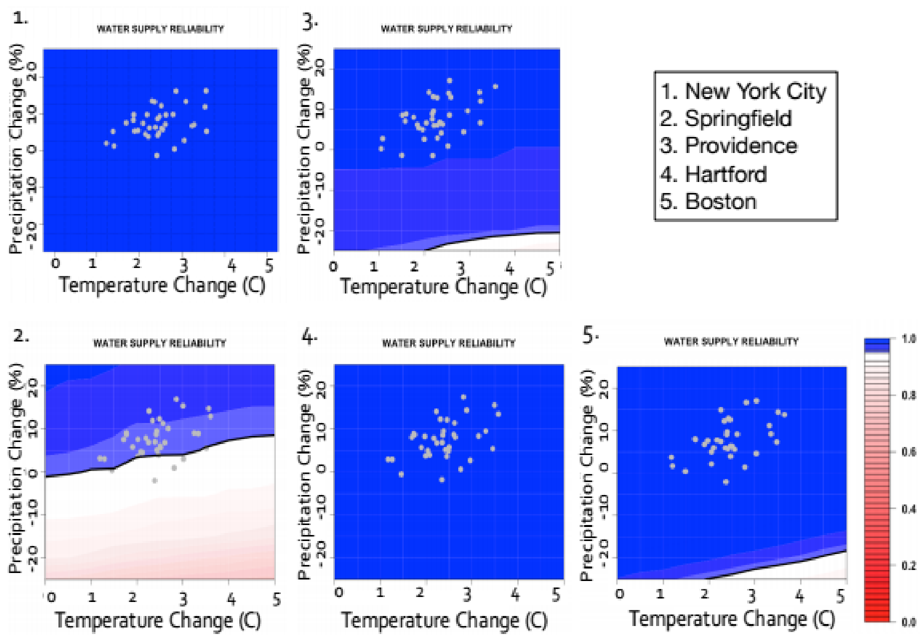

Figure 1.2 Water supply reliability in a worst-case realization of variability.

1.4 Potential Impacts of Changes in Climate on Turbidity in New York City's Ashokan Reservoir

The effect of climate change on water quality is a key issue for the New York City Water Supply System. The system is one of six water supply systems in the United States that has a filtration waiver from the EPA; they bypass an initial step of the treatment process in favor of protecting the environmental quality of their watersheds and thereby maintaining the quality of their source water. Due to the geology of the area, one watershed exhibits a high level of erosion and sediment transport to the reservoir. When high levels of turbidity are present, water is treated by alum addition in one of the city's aqueducts to allow suspended sediment to settle out before the water reaches its destination. This occurs infrequently and is considered an emergency condition, but the process is initiated by high flows, connecting it to the changing frequency of heavy precipitation.

Rossi et al. (2016) applied the same three-model approach, used in section 1.3, to examine the effects of climate change on water quality in NYC's Ashokan Reservoir. Here, a similar setup of three models in sequence translates changes in temperature and precipitation to the number of days reservoir turbidity exceeds a certain threshold. Turbidity serves as a surrogate for suspended sediment, and is modeled to have a log-log relationship (Mukundan et al. 2013) with daily flow. The use of a threshold metric represents trends in the frequency of actions needed, such as drawing more water from other source reservoirs to minimize turbidity in the mixed water downstream or, in extreme cases, the addition of alum to water withdrawn from Ashokan Reservoir.

Rossi et al. (2016) found that in-reservoir turbidity scaled with precipitation for 10 NTU and 25 NTU thresholds in both basins over the range of climates tested (Figure 1.3). Research results indicated that temperature had the greatest effect on turbidity in summer, where more water was lost to evapotranspiration in warm climates, resulting in lower flows and lower levels of turbidity. Temperature had a minimal effect on turbidity in winter (Rossi et al. 2016). The averages calculated mask the event-based nature of high turbidity. This detail, and sensitivity to changes in the range of climate change projections, warrant further study into the extreme events themselves and into appropriate management strategies.

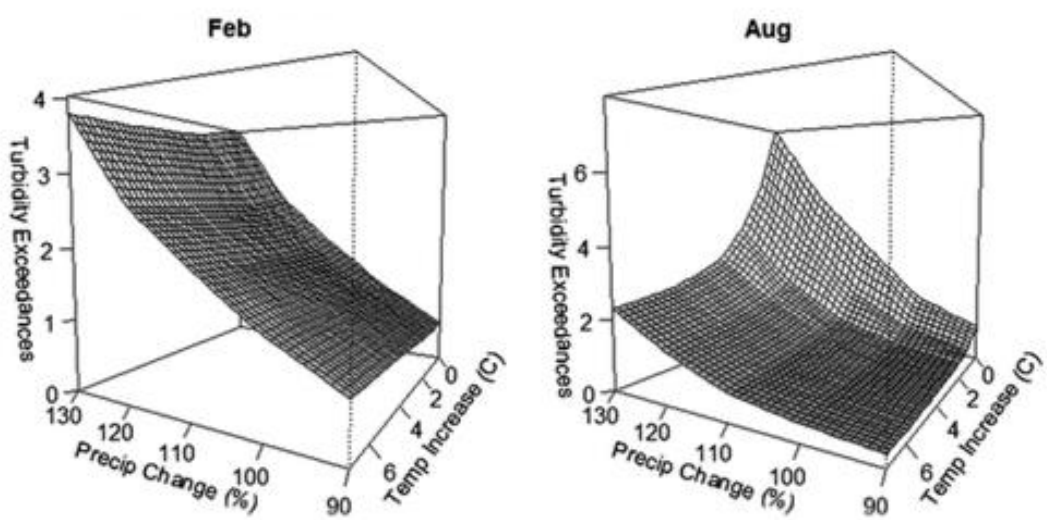

Figure 1.3 Average number of days turbidity in the west basin of Ashokan reservoir exceeds 25 NTU in February and in August.

1.5 Evaluating Stochastic Precipitation Generators for Climate Change Impact Studies of New York City's Primary Water Supply

Due to the event-based nature of high flows and high turbidity in Ashokan Reservoir, more detail should be paid to the methods of scenario generation for adaptation studies. Synthetic time series of daily precipitation totals, as used in the previous two experiments, are generated in two steps. First, a Markov chain is used to determine the occurrence or non-occurrence of precipitation. Next, a quantity of precipitation is chosen based on a distribution or sampling method. When interested in extremes, the choice of distribution makes a significant difference, as different distributions have significantly different behaviors near the tail. Acharya et al. (2017) analyzed different Markov chain orders and different statistical distributions for their behavior in matching extreme precipitation statistics. Metrics include the 95th and 99th percentile, and the annual one- and five-day maximum.

First, second, and third order Markov chains performed comparably in simulating dry and wet spell length, indicating that the most parsimonious precipitation generator should use a first order Markov chain. Using a first order Markov chain for precipitation occurrence, five parametric distributions (exponential, gamma, skewed-normal, mixed exponential distribution, and a hybrid exponential and generalized Pareto), one resampling method (k-nearest neighbor), and one 2nd order polynomial-based curve fitting method, were used to generate precipitation amount. Skewed normal, exponential, and k-nearest neighbor best captured the behavior of extreme events, with other methods under- or over-estimating the magnitude of extreme precipitation statistics when compared to the historic record (Acharya et al. 2017).

1.6 Climate change adaptations through infrastructure, management, and hydrologic forecasting

Based upon the results and insights of the research summarized in the three previously cited papers, the following chapters look in more detail at possible adaptations using case studies of east coast water supply systems and watersheds. Chapter 2 demonstrates win-win adaptations for the New York City Water Supply System, based substantially on projects already in the planning and implementation stages. These system improvements and adaptations deal both with hydrologic drought and with the effects of extreme storms on water quality. Chapter 3 proposes a method for quantifying the value of one of the adaptations included in Chapter 2: improved seasonal forecasts of streamflow for use in drought planning. The chapter uses the City of Baltimore water supply system as a case study for determining the value of hydrologic forecasts of varying quality in drought planning. Finally, Chapter 4 explores a method of improving statistical forecasts of streamflow using the National Weather Service's seasonal precipitation outlooks. The method is tested for six watersheds on the east coast, including those upstream of the New York City and Baltimore systems.

The Water Utility Climate Alliance suggests that successful climate change adaptation planning consists of four steps: understanding the science, assessing the risks, incorporating climate change into long-term planning, and implementing those plans (Means et al. 2010). The research and experiments detailed in this and the following chapters follow this path, from examining the science and vulnerabilities of water resources in the northeast in this introductory chapter, through to examining adaptations and testing a new method of ensemble hydrologic forecast creation for use in water supply management.

CHAPTER 2

WIN-WIN STRATEGIES FOR MANAGING CLIMATE CHANGE EXTREMES: A CASE STUDY OF THE NEW YORK CITY WATER SUPPLY SYSTEM

2.1 Abstract

Water supply reliability is a function of the interactions between infrastructure, system operations, water demands, and hydrologic inputs. As the short- and medium-term impacts of climate change become more apparent, operations will necessarily have to adapt to inflow and infrastructure constraints to maintain reliability. Through a series of linked models of stochastic weather, hydrologic processes, and systems modeling, this paper demonstrates the robustness of several adaptations available to the New York City Water Supply System to mitigate drought and manage water quality under climate change projections through the end of the century. Results illustrate how reducing demand and managing storage and releases based on hydrologic forecasting reduce the frequency of drought warnings and emergencies and improve system reliability in all climate change scenarios investigated. With the goal of mitigating high turbidity, operations that limit turbidity propagation through the system and improvements to the Catskill Aqueduct (to lower the minimum flow under conditions with high turbidity) demonstrate lower turbidity loads and a reduction in emergency Alum use. These options demonstrate the cumulative, combined benefits.

2.2 Introduction

Water supply management during extreme events, such as droughts and floods, is an established but evolving challenge made all the more urgent by our changing climate. Both high flows and droughts are relevant to management of the New York City Water Supply System

(NYCWSS). The region has experienced several severe droughts including the 1960s drought of record, and the NYCWSS has an additional obligation as one of the few large water supplies in the US that provides to its customers unfiltered water. The EPA regulations that establish this allowance necessitate careful management of water quality (in NYCDEP's case impacted by peak flows). The NYCDEP's extensive multi-reservoir system provides the opportunity for adaptation through operational changes, infrastructure, and conditional or permanent demand reduction, making it a prime example to demonstrate proactive long-term planning and robust adaptations to climate change.

Non-stationarity, the concept that the historic climate is no longer the only tool necessary to estimate the probability of future hydrologic events, has been broadly accepted in water supply management (Stratus Consulting and Denver Water, 2015; Milly et al., 2008). This recognition created a shift from "risk," where the probability of an event is considered known and traditional cost-benefit analyses can be performed, to "uncertainty," where probabilities of future events are unknown (Giuliani and Castelletti, 2016; Wilby and Dessai, 2010). Various methods have been explored to deal with this new challenge, including decision scaling, robust decision making, information gap analysis, and dynamic adaptive planning pathways (Brown et al., 2011; Hine and Hall, 2010; Herman et al., 2015; Haasnoot et al., 2013). These methods all require the identification of adaptation options that meet design criteria over a wide range of future climate, water management, and demand scenarios. For example, existing operational case studies test alternative reservoir operations in mitigating the effects of climate change on system metrics including flood control, hydropower, and water supply (Culley et al., 2015; Arsenault et al., 2013; Eum and Simonovic, 2010; Payne et al., 2004). Adaptations that improve system performance in present and projected future conditions are referred to as low-regrets (Field, 2012) or win-win (Lim, 2005) strategies.

Observed and anticipated increases in the frequency and intensity of climate extremes provide further challenges to water supply systems. The global historic record shows a trend in increasing severity of droughts and floods (IPCC, 2012). Further, increases in both precipitation extremes (Kunkel et al., 2013; Frei et al., 2015) and river flooding (Peterson et al., 2013) are particularly pronounced in the northeast US.

This paper explores how best to evaluate strategies that reduce the impacts of droughts and peak flows associated with projected changes in precipitation and temperature in the form of a case study of the NYCWSS. The case study uses a stochastic weather generator, a hydrologic model, and a systems operations model, similar to Borgomero et al., 2014. Additionally, the paper illustrates the effectiveness of individual and combined actions. In contrast to several of the approaches cited previously, where the adaptation is enacted at some future time, this paper assumes that actions are taken immediately, and each option is implemented permanently at the beginning of the model run. This avoids the necessity of waiting for performance degradation or a damaging threshold event to begin implementing system improvements. The adaptations considered, two for drought and two for water quality, are based predominantly on projects already under active consideration by the NYCDEP based on cost effectiveness and supported by studies of future reliability. Unique to this study is the inclusion of the full range of climate change scenarios in conjunction with different combinations of operations or infrastructure projects for this system. This chapter is organized as follows. The Setting section contains a description of the NYCWSS and the historic challenges of drought and turbidity in the system. The Methods section includes a description of the models and inputs used, the management alternatives, and the experimental design. The Results section presents the outcomes of these alternatives for different climate change projections through the end of the century. The Discussion and Conclusion sections discuss the implications of the results and future research. The primary contribution of this paper is a long-horizon examination of climate change adaptation strategies for the New York City Water Supply System. The study points to a promising outlook for the city and in the process, demonstrates a straightforward method of organizing scenarios to make such a comparison.

2.3 Setting

The NYCWSS includes nineteen reservoirs and three lakes and provides water to approximately 9.5 million people in NYC and the surrounding areas. Water demands for the system have decreased from approximately 1.4 billion gallons per day to the current water demand of 1 billion gallons per day. This decrease in water demand, during a period of increasing population, has resulted from efforts in water conservation, general changes in water use, and in changes in water priciing

The nineteen reservoirs can be considered as three subsystems: the Croton, the Delaware, and the Catskill systems. The Croton system, built between 1842 and 1906, consists of twelve reservoirs and three lakes in Westchester and Putnam counties with a total capacity of 91.6 billion gallons (346.7 MCM). A treatment plant for the Croton system was completed in 2015 and the system can provide up to 290 million gallons per day (MGD) (1.1 MCM per day) of high quality water to the city.

A majority of the city's water comes from six reservoirs west of the Hudson River (WOH), where extensive and strategic watershed protection measures have preserved the quality of the surface water source. In response to the Clean Water Act (1972), the NYCDEP, along with several other major water utilities across the country, chose to invest in watershed protection and applied for a Filtration Avoidance Determination, relieving them of the requirement to filter water and allowing them to provide only disinfection to water from their WOH reservoirs. Currently, Boston, MA; Syracuse, NY; San Francisco, CA; Seattle, WA; and Portland, OR, along with New York City, maintain their filtration waiver. The NYCDEP's permit has been renewed four times since 1993 and is currently in effect through 2027.

The WOH reservoirs are grouped into two systems. The first is the Delaware system, consisting of Cannonsville (95.7 billion gal (362.3 MCM)), Pepacton (140.2 billion gal (530.7 MCM)), and Neversink (34.9 billion gal (1321 MCM)) at the headwaters of the Delaware River and Rondout Reservoir (49.6 billion gal (187.8 MCM)) used primarily as a transfer basin. The second group is the Catskill system, consisting of Schoharie (17.6 billion gal (66.6 MCM) and Ashokan (122.9 billion gal (465.2 MCM)) reservoirs, both on tributaries of the Hudson River. Water from Schoharie reservoir is routed through the Ashokan Reservoir before traveling to New York City via the Catskill Aqueduct. Water from the Delaware Aqueduct is routed though the West Branch Reservoir, and water from both systems travels through Kensico Reservoir and then to New York City(Figure 2.1).

The Delaware system provides approximately 60% of New York City's water and has additional demands in the form of downstream releases. The NYCDEP is required to release water from Delaware reservoirs to maintain cold-water fisheries below the reservoirs, to provide water for downstream states to meet the requirements of the 1954 Supreme Court Decree (New Jersey v. New York, 1954), and to mitigate saltwater intrusion in the lower Delaware River and Delaware Bay near Philadelphia.

The Catskill reservoirs typically provide 30-40% of NYC's water, and the Catskill watersheds are the primary source of sediment and turbidity in the system. The Catskill

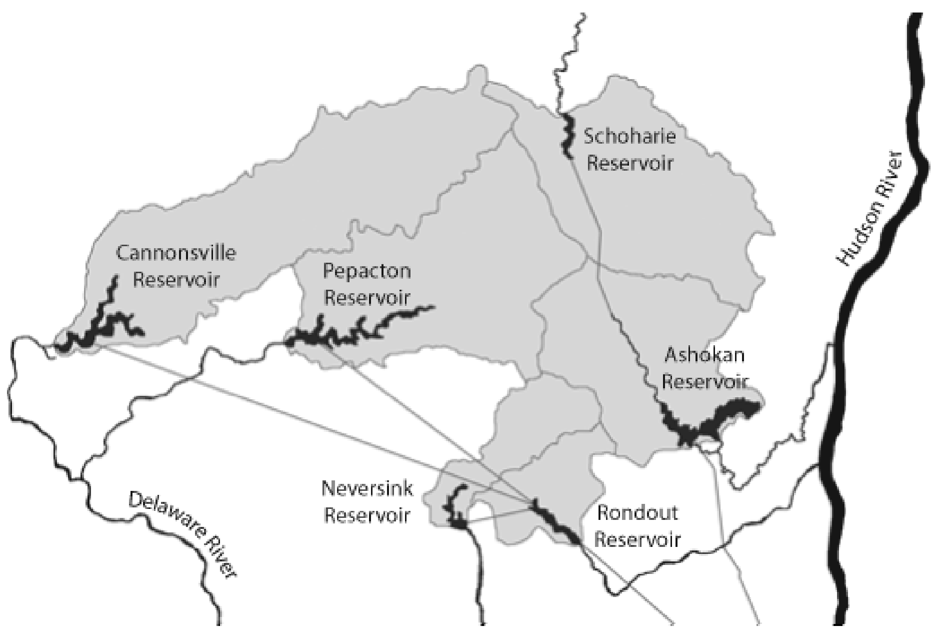

Figure 2.1 Map of the West of Hudson portion of the New York City Water Supply System.

Mountain region contains highly erodible soils in and immediately adjacent to streams, and intermittent high turbidity occurs in the streams and reservoirs of the Catskill system during high flows. Ashokan Reservoir was built with two basins and is operated to allow inorganic sediment to settle out in the west basin when necessary. Water can be withdrawn from either the west or east basin of the reservoir.

Total water demand for New York City peaked in the 1970s at approximately 1.5 billion gallons per day (5.7 MCM per day) and has been decreasing since then (NYCDEP, 2017). This decrease is due to a decrease in per capita demand and mirrors the trend and peak of water withdrawals across the United States (Maupin et al., 2014). New York City will continue to reduce demand, with a near-term goal of a 50 MGD reduction through efficiency programs and incentives, and distribution system optimization and repairs. Additionally, New York City plans to expand the scope of temporary conservation measures. Currently, conservation measures are in place to mitigate shortages caused by hydrologic drought. Future conservation measures will also address shortages related to infrastructure repair projects, with the short-term reduced demand needed for a temporary outage of the Delaware aqueduct to bypass a major leak (NYCDEP, 2015).

Both droughts and high (peak) flows are a concern for the city. The system can store about 18 months of the city's demand with no additional inflows, leaving New York City vulnerable to potential shortages during multi-year periods of low flow. In addition to the 1961-1963 drought of record, droughts of varying severity have occurred in 1980-1982, 1985, 1989, 1991, 1995, and 2002 (NYCDEP, 2017). During drought events, alternative sources are utilized, in combination with both voluntary and mandatory restrictions. These restrictions range from outdoor uses for landscaping, pools, and fountains, to the mandatory posting of "please conserve water" notifications in public and multi-unit residential buildings (NYCDEP, 2012). Voluntary conservation and restrictions on non-essential use help prevent more serious consequences of water shortages.

The New York City system is also significantly impacted by extreme precipitation and the impacts on water quality associated with high streamflows resulting from these storms. Because of the terms of the Filtration Avoidance Determination and a 5 NTU upper limit on turbidity, sediment and turbidity management in the Catskill system is essential. In the vast majority of cases, operations in anticipation of storms and subsequent settling time in the west basin allow the water withdrawn from the east basin to meet or exceed all water quality standards. However, in cases of extremely high flow, turbid water can spill from the west basin over the dividing weir into the east basin. When highly turbid water spills to the east basin, it may be necessary to treat the water withdrawn with aluminum sulfate in the Catskill Aqueduct upstream Kensico Reservoir, allowing the sediment to settle at the aqueduct's outlet before it reaches New York City. Since 1987, more than 21 million pounds (9.5 million kg) of alum has been added to Catskill Aqueduct water during ten high flow incidents (NYCDEP, 2014).

In 2010, an advanced data network and systems model, the Operations Support Tool (OST), was created for the NYCDEP to provide real-time data for more effective management of both of these challenges (NYCDEP, 2010). Position analysis (Hirsch, 1979) based on near real time data and ensemble hydrologic forecasts, has influenced reservoir release decisions and allowed the NYCDEP to meet demand, storage objectives, and downstream requirements (Porter et al., 2015). In addition to the incorporation of more sophisticated forecasts and modeling for near-term decision-making, New York City has taken an active role in preparing for climate change. The Second New York City Panel on Climate Change met in January of 2013 and subsequently published "Building the Knowledge Base for Climate Resiliency: New York City Panel on Climate Change 2015 Report." This report presents observed trends and downscaled General Circulation Model (GCM) outputs for New York City and the surrounding area. Projections were drawn from 35 GCMs and representative concentration pathways (RCPs) 4.5 and 8.5 as included in the Coupled Model Intercomparison Project Phase 5 (CMIP5) (Horton et al 2015). These projections, representing changes from historic temperature and precipitation, are used in this paper as the possible range of future climates for the 2020s through the end of the century (Table 2.1). This paper integrates water supply system modeling and the New York City climate change projections in a long-term evaluation of system performance.

Table 2.1 NYC Climate Projections

Air Temperature Baseline (1971 – 2000) 54 °F	**Low-estimate (10th percentile)**	**Middle range (25th to 75th percentile)**	**High-estimate (90th percentile)**
2020s	+1.5 °F	+2.0 °F to 2.9 °F	+3.2 °F
2050s	+3.1 °F	+4.1 °F to 5.7 °F	+6.6 °F
2080s	+3.8 °F	+5.3 to 8.8 °F	+10.3 °F
2100s	+4.2 °F	+5.8 to 10.4 °F	+12.1 °F
Precipitation Baseline (1971 – 2000) 50.1 inches	**Low-estimate (10th percentile)**	**Middle range (25th to 75th percentile)**	**High-estimate (90th percentile)**
2020s	-1 percent	-1 to +8 percent	+10 percent
2050s	+1 percent	+4 to +11 percent	+13 percent
2080s	+2 percent	+5 to 13 percent	+19 percent
2100s	-6 percent	-1 to +19 percent	+25 percent

2.3 Methods

The research framework and models used in this paper demonstrate the effects of projected climate change and adaptations on metrics that quantify the impacts of drought and turbidity on the NYCWSS. Climate altered inflows test a range of adaptations in a screening-tool level model of the system. We ask how these adaptations perform across the changes in climate projected for the next century.

2.3.1 Models

This framework incorporates appropriately selected hydrologic and systems operations models. A stochastic weather generator, a hydrologic model, and a water supply systems model operate in series. The weather generator allows variation in climate statistics, while the systems model includes current or alternative operations and infrastructure (Figure 2.2).
A stochastic weather generator model (Steinschneider and Brown, 2013) is applied to create a wide range of weather scenarios to create representative climate futures. The weather generator uses a first-order auto-regressive framework to generate annual climate, and a Markov chain and K-nearest neighbors resampling algorithm for daily weather variables to preserve correlation between locations within the study basin. The choice of a stochastic weather generator utilizing an auto-regressive lag-1 annual series is particularly appropriate for the New York region, as most multi-year wet and dry periods are the result of local atmosphere dynamics rather than tropical teleconnections (Seager et al., 2012). The model is calibrated to the gridded historic dataset created by Maurer et al. (2002) for data from January 1, 1950 to December 31, 1999. Analysis of calibration statistics used for the weather generator can be found in Rossi et al., 2016. Monthly changes in temperature (additive) and precipitation (multiplicative) are applied to the generated sequences to represent climate change scenarios (further discussed in later sections).

The synthetically generated weather sequences serve as input into a hydrologic model. The Generalized Watershed Loading Function-Variable Source Area (GWLF-VSA) model applied here is a watershed-scale model capable of simulating streamflow and water quality based on an adaptation of the GWLF to account for the dominance of saturation excess runoff (via use of VSA) over infiltration excess runoff (an assumption of the Soil Conservation Service’s Curve Number method) (Schneiderman et al., 2007). VSA, compared to the curve number method, is a more accurate representation of the shallow, permeable, well-vegetated soils in sloping watersheds typical of the northeastern US and the Catskill Mountain region (Walter et al., 2000).

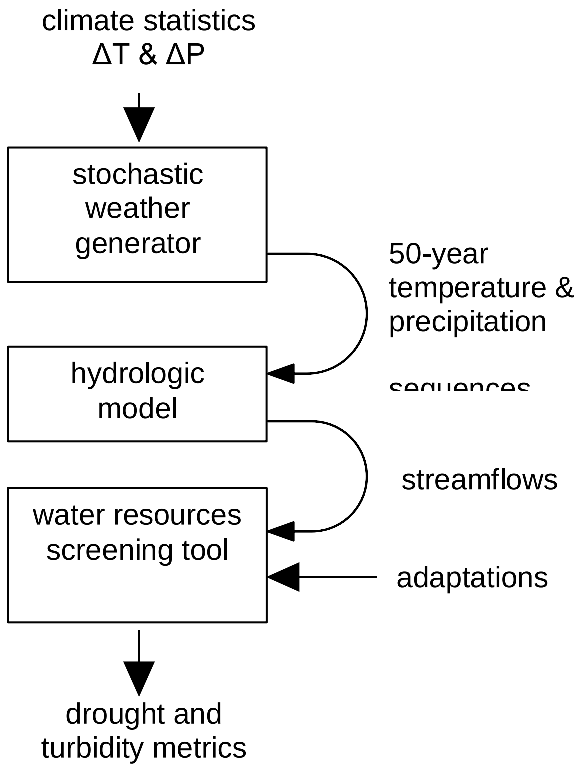

Figure 2.2 The experiment uses three models in sequence, a weather generator, a hydrologic model, and a water resources system screening tool, to determine the effects of climate change statistics and adaptations on drought and turbidity in the New York City Water Supply System.

The model was calibrated independently for the watersheds upstream of each of the six WOH reservoirs by NYCDEP researchers (Schneiderman et al., 2002). Additionally, modeled flows show expected changes in quantity and timing when driven with incremental changes in temperature or precipitation (Rossi et al., 2016). The turbidity load to Ashokan reservoir, where turbidity has historically been a concern, is modeled separately from the GWLF-VSA model as Mean Daily Turbidity (MDT) as a function of streamflow, season, and antecedent dry days and based on existing statistical analysis of 27 storm events over an 8 year period between 2003 and 2011 (Mukundan et al., 2013). Turbidity is a proxy for suspended sediment, and turbidity "load" is calculated as the product of turbidity and flow rate, MGD-NTU.

Multiple 50-year sequences are created with the weather generator representing an ensemble of climate projections. These weather sequences are then translated to reservoir inflows via the GWLF-VSA model. Due to the large number of inputs and scenarios in this method, the use of the OST was not considered possible and an alternative screening-level model of the NYCWSS was created to allow these scenarios to be tested efficiently. In the screening model the operating rules, regulations, forecasts, and sediment dynamics are simplified compared to NYCDEP's OST used in daily operations planning. The model (denoted as the Screening Tool Assessment of Turbidity and Supply, STATS) is coded as a daily water and sediment balance model in Vensim DSS (Ventana Systems Inc.) with the six WOH reservoirs modeled individually and operated conjunctively. Operations are incorporated into the model as a series of rules, rather than chosen via optimization. These rules are based on interviews with the NYCDEP staff and publicly available Federal and State regulations. Comparisons of the screening tool with the more complex OST were made to ensure reservoir storages and turbidity from the screening model were similar using correlation and Nash Sutcliffe Efficiency (Rossi, 2014). This screening model is sufficiently detailed to capture the nuances of supply operations and the basic physics of the turbidity process. Thus, changes in operations and infrastructure and the resulting changes in water quality and availability can be captured. The screening model includes multiple adaptation options.

2.3.2 Adaptations

A first category of adaptation addresses drought management. NYCWSS strives to provide ample water to the city during drought. As demands increased over time, the city's infrastructure increased in size and complexity to the system of the present day. Now, planning, monitoring, maintenance, and improvements maintain the system's reliability. Two of these drought management adaptations are tested for robustness under climate change projections through the end of the 21st century.

Adaptation 1 investigates the importance of streamflow forecasts in improving the system's ability to provide water reliably. The NYCWSS typically refills annually, after the summer drawdown. In most years, the system is near full and/or spilling by late spring or early summer, when snowmelt is completed and before high summer demands and low summer inflows begin. Late spring streamflow forecasts (based on snowpack and other antecedent conditions) offer the opportunity to advise managers on how best to balance reservoir levels across the system and to determine appropriate releases to the lower Delaware River. A primary operating objective in the past has been to forecast the likelihood of refill by June 1 of each year. In the model, forecasts are used to balance storage across the system and determine releases from the Delaware reservoirs. Reliable forecasts with lead-times of weeks to months can anticipate lower than average flows at one extreme and prevent unwanted spill on the other. In Adaptation 1, operations that utilize a perfect forecast, created by summing future model input inflows leading up to the upcoming June 1st and issued on a monthly basis, are compared to a baseline operating policy: the use of the 1950 – 2000 historic streamflow record. The decision-making structure within the system model necessitates a single trace forecast. A perfect forecast was chosen for simplicity and availability and to provide a significant contrast to the historic record.

Adaptation 2 explores the importance of demand on system reliability. Like many older water supply systems, NYCDEP's infrastructure suffers from system leaks (estimated to be 35 million gallons per day (132,000 m^3/day) from the Delaware Aqueduct alone). A leak repair program is in progress and anticipated to be completed by 2023 as part of the Water for the Future program (NYCDEP, 2015). Repairing these leaks essentially increases system capacity and enhances system performance during drought. Adaptation 2 explores the results of the Delaware Aqueduct repair and models the outcome as a 35 million gallon reduction in daily demand.

Metrics for evaluating Adaptations 1 and 2 are annual reliability and the frequency of water use restrictions (drought warnings and drought emergencies). These water use restrictions are based on reservoir storage defined by the probability of water year refill based on historic flows and demands.

The second category of adaptations is associated with high flows into Ashokan reservoir that create high concentrations of suspended sediment and turbidity. Adaptations 3 and 4 illustrate the relative importance of rules, infrastructure, and climate in determining system turbidity.

Adaptation 3 includes operations of the existing system. Recent renovations to the Ashokan Release Channel allow manipulation of storage and void. The Interim Ashokan Release Protocol, that has been in effect since 2013 (NYCDEP/ NYSDEC, 2011), identifies operation for turbidity and spill management. The proposed procedures include introducing a rule curve to create a space (or volume or void) in the reservoir for the portion of the year when high flows and high turbidity are more likely to occur, and using the Ashokan release channel to release turbid water and prevent spill to the east basin. The model simulates the release of water from the west basin of Ashokan reservoir to meet the storage objective.

Adaptation 4 addresses the improvement of existing stop shutters along the Catskill aqueduct. Installing stop shutters reduces the total flow required for adequate aqueduct pressure to serve communities along the aqueduct, but their implementation is currently difficult and time consuming. Reducing the minimum required aqueduct flow results in the ability to combine a greater amount of higher quality water from the Delaware system with a smaller amount of turbid Catskill water. Improving stop shutter operation will allow their incorporation into normal operations (NYCDEP, 2014). This is alternative is modeled as a reduction in the Catskill aqueduct minimum flow.

Metrics used to evaluate Adaptations 3 and 4 are annual days of alum use and the annual peak turbidity export (MGD-NTU) withdrawn for use from Ashokan reservoir, both averaged across each 50 year run. Here, alum use is initiated when turbidity export peaks above 5000

MGD-NTU and continues until the 5-day average export drops below 4000 MGD-NTU. All of the adaptations explored and metrics are summarized in Table 2.2.

Table 2.2 Summary of adaptations and metrics

Adaptations
Adaptation 1 - Perfect streamflow forecasting
Adaptation 2 - Leak repair (35 MGD demand reduction)
Adaptation 3 - Turbidity management operations
Adaptation 4 - Improved stop shutters
Metrics
Drought Warning – Combined Delaware system storage lower than rule curve, lower than 33% chance of water year refill based on historic flows and demands
Drought Emergency - Combined Delaware system storage lower than rule curve, likelihood that without action, system will experience shortages based on historic flows and demands
Annual Reliability - Percent of years in which all demand is met
Turbidity Export – Peak annual turbidity load (MGD-NTU) as a proxy for sediment load. 5000 MGD-NTU is an approximate upper limit before alum use
Alum Use – Annual number of days. When the turbidity and quantity of water withdrawn from Ashokan Reservoir exceeds the 5000 MGD-NTU limit, Alum is added to the water supply

2.3.3 Scenarios

The combination of adaptations and climate statistics create scenarios, and these scenarios provide insight into the impact of climate changes on system performance under different operational futures. Multiple combinations of adaptations are tested. For drought metrics, a baseline of no adaptation, Adaptation 1 alone, Adaptation 2 alone, and Adaptations 1 and 2 in combination create 4 possible drought management possibilities. For turbidity metrics, a baseline of no adaptation, Adaptation 3 alone, Adaptation 4 alone, and Adaptations 3 and 4 in combination create 4 possible turbidity management possibilities (Table 2.3).

Four sets of climate inputs reflect four different decadal ranges of multi-model ensemble climate projections. For each decade, each operational possibility is modeled, resulting in 32 scenarios, 16 for water quality and 16 for drought. For each decade, 70 sets of monthly precipitation and temperature changes from 35 climate models and 2 representative concentration pathways represent the range of climate change projections for the decade (Horton et al., 2015).

For each of these changes, ten 50-year stochastic weather sets represent and reduce the influence of internal climate variability. This framework results in 700 50-year sequences of daily weather and hydrology for each decade.

Table 2.3 Summary of scenarios

Scenarios	Metrics
Baseline	all
Adaptation 1	Drought Warning, Drought Emergency, Annual Reliability
Adaptation 2	
Adaptations 1 & 2	
Adaptation 3	Turbidity Export, Alum Use
Adaptation 4	
Adaptations 3 & 4	

2.4 Results

The results of climate change and adaptations are presented first those associated with drought, then those associated with high flow events. The modeled effects of these adaptations over multiple climate scenarios through the end of the century illustrate the adaptability of the NYCWSS and the potential strategies available for system improvements.

2.4.1 Drought

Between 1960 and 2010 in NYC, drought warnings were declared in 14 years and drought emergencies were declared in 5 years (NYCDEP, 2017). Total water demand has dropped significantly since those droughts (from a peak of 1500 MGD around 1980 to an average demand of around 1000 MGD today). In model runs utilizing stochastic weather with historic climate statistics and current water demand, the occurrence of drought warnings and drought emergencies is less frequent than the historic occurrences. This confirms that reduced demand plays a role in system reliability.

The frequency of future drought warnings and drought emergencies across all climates and adaptations illustrates several trends (Figure 2.3). The effect of climate on the occurrence of drought warnings and drought emergencies is demonstrated by a single adaptation possibility through time, thereby controlling for infrastructure and management state. Though an overall wetter climate is projected in many scenarios, a few GCMs project drier climates or seasonal shifts in precipitation that result in a greater frequency of drought, with the extreme scenarios occurring at the end of the century. This does not necessarily indicate an increase in precipitation variability of future climates, but rather an increase in the spread of climate models and greater uncertainty when projecting farther into the future.

Meanwhile, scenarios implementing one or both adaptations result in reduced frequency of drought warning and drought emergency. Adaptation 1 and the resulting operations and timely conservation are more effective than a single constant reduction in demand (Adaptation 2) when implemented alone, though a less than perfect forecast or greater reduction in water lost to leaks may change this. The joint application of both adaptations is most effective in reducing the occurrence of low storages. Together, a perfect monthly streamflow forecast with up to 11 months lead-time and a small reduction in total demand reduce the occurrence of drought emergency to less than 1 in 50 years in the majority of scenarios. Decreases in drought emergencies (the most severe classification) do not result in increases in the length or frequency of drought warning (a less severe classification). All drought levels are reduced with each adaptation. In plotting individual scenarios, storages with forecasting remained higher when experiencing extended droughts but allowed the system to draw down lower in years where refill occured, both in logical anticipation of the upcoming conditions.

For figure 2.4, a standardized metric of annual reliability allows for the comparison of water supply systems with very different infrastructure. Annual reliabilities of US water supply systems generally fall in the range of 97% - 98.5% (Vogel et al., 1999) and a 2013 study of the NYCWSS under current conditions and climate change placed the system near the top of that range (Matonse et al., 2013). The results here (Figure 2.4) match these previous probabilistic results. Here, the number of years in a 50-year trial in which the system fails to meet daily demand is modeled for each climate scenario and each alternative. In the base case, reliability remains at or above 98% in 75% of scenarios. The remaining scenarios increase in shortfall frequency with climate change; these scenarios are ones in which the climate gets drier but operations remain based on the historic, wetter, record. The two adaptations individually increase the reliability of the system. With the two adaptations combined, very few climate scenarios show reliabilities lower than 100%, even under projected 2100 climates. These special cases would be handled by managers on a case by case basis and would take advantage of water sources and management options not represented in the model. Any long term forecast provides insight on whether the system will need to implement water restrictions in order to meet demand over the longer term. A perfect forecast allows this decision with certainty, which is different in two major ways from an existing forecast. For any uncertain forecast, a more conservative decision would be made, and the ensemble would have to be analyzed on a case by case basis. The "complete drawdown" scenario (that used in a safe yield calculation) is undesirably risky.

2.4.2 Turbidity

Tudbidity in the New York City Water Supply System is a quantity that will change with climate change. Peak streamflows are linked by existing conditions, management decisions, and the modeled adaptations to the turbidity load in the Catskill Aqueduct.

The adaptation alternatives can be effective in reducing peak turbidity (Figure 2.5). Export, or turbidity as MGD-NTU withdrawn to meet demand, increases or stays approximately the same for all climate projections with all adaptations. Adaptations 3 and 4 reduce the peak turbidity by an increasingly large margin, relative to the conditions that would otherwise occur, as the projected climate gets warmer and wetter in the future. Annual averages with Adaptation 3 and Adaptation 4 remain near or below a 3000 MGD-NTU until 2050 projections. Averages for all scenarios except for the base case of no adaptation remain below 5000 MGD-NTU.
Results for historical climate, not plotted in Figure 2.4, are also noteworthy. Utilizing the historic record of flows and a different systems model, NYCDEP's OST, this combination of adaptations

was able to reduce alum use to zero (NYCDEP, 2014). Using the same methods and models of this paper, combined use of the two adaptations eliminates alum use in 88% of stochastic realizations of baseline climate, giving credence to these adaptations as "wins," or beneficial adaptations in present conditions as well.

In all climate scenarios, the combination of Adaptations 3 and 4 greatly reduced the average number of days of alum use. Adaptation 3 alone performs better than Adaptation 4. Note that these results are averages. In reality and in the model runs, alum use occurs every few years for several times the average length, rather than every year for the average length; high turbidity and the resulting alum use is episodic and correlated rather than an annual occurrence.

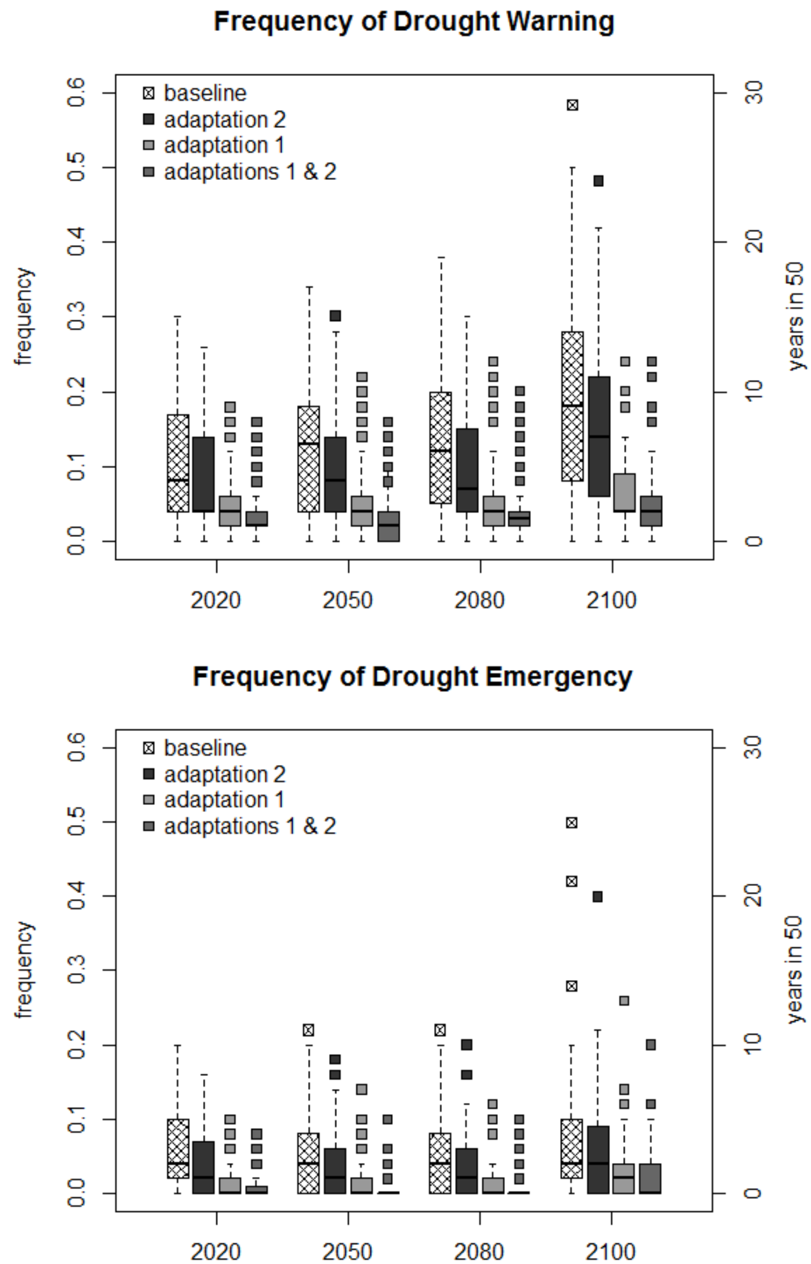

Figure 2.3 Frequency of drought warning (top) and drought emergency (bottom) statuses declared based on low reservoir storage. The range of each boxplot shows the frequency of drought warning (drought emergency) modeled as years-in-50 for each climate change scenario. For example, one outlier point represents one GCM output resulting in a higher frequency of drought warning. The same set of climate scenarios are used for each decadal grouping of alternatives. The set of scenarios change in each decade to represent climate change over time. Four adaptation scenarios are shown in each decade.

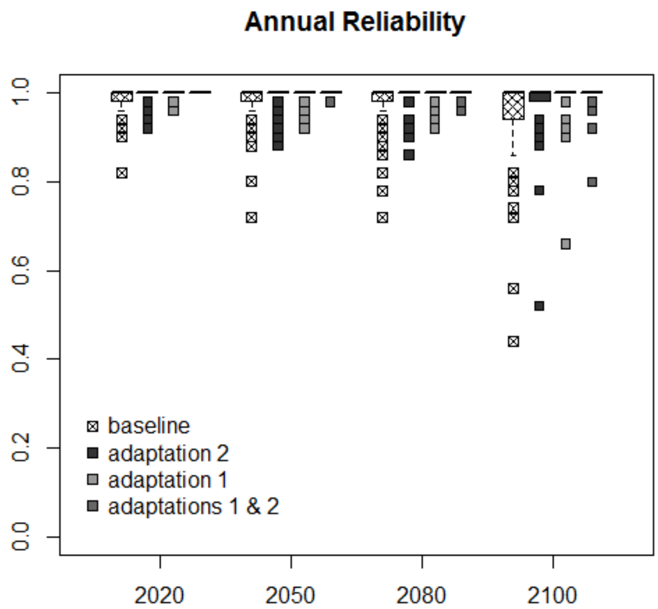

Figure 2.4 Annual reliability. The range of each boxplot shows system reliabilities in each climate change scenario. For example, one outlier point represents one GCM output resulting in lower annual reliability of the system. The set of scenarios change in each decade to represent climate change over time. Four adaptation scenarios are shown in each decade.

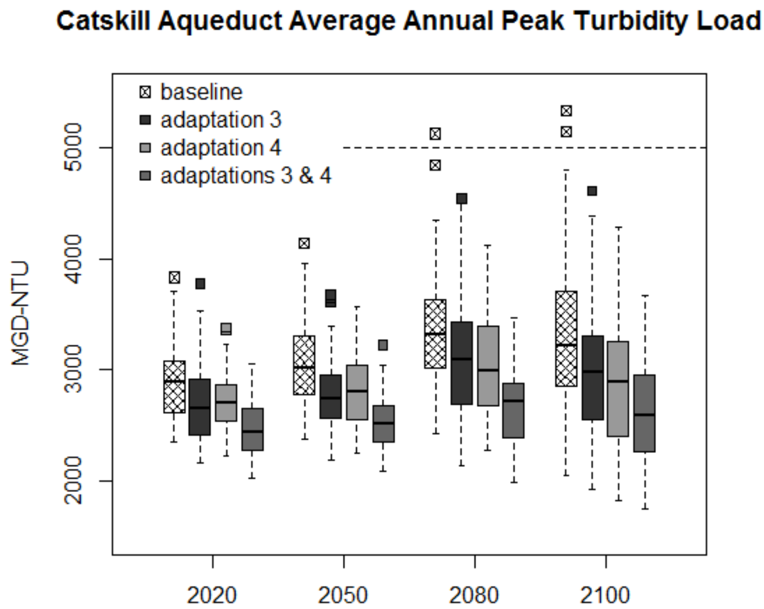

Figure 2.5 The average annual peaks of turbidity load in the Catskill Aqueduct. The range of each boxplot shows the average annual peak export for each climate change scenario. For example, one outlier point represents one GCM output resulting in a higher peak turbidity load. The same set of climate scenarios are used for each decadal grouping of alternatives. The set of scenarios change in each decade to represent climate change over time. Four adaptation scenarios are shown in each decade.

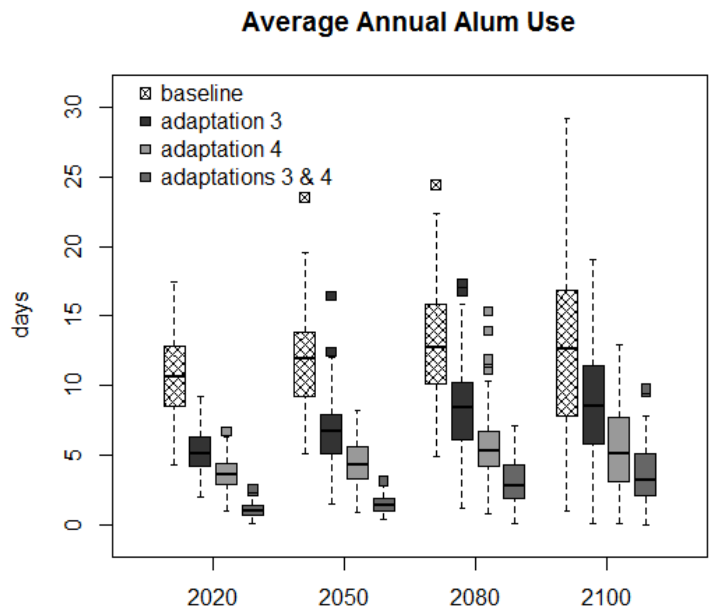

Figure 2.6 Average annual days of Alum addition in the Catskill Aqueduct for different climates and adaptations. The range of each boxplot shows the frequency of alum addition for each climate change scenario. For example, one outlier point represents one GCM output resulting in higher annual average alum use. The same set of climate scenarios are used for each decadal grouping of alternatives. The set of scenarios change in each decade to represent climate change over time. Four adaptation scenarios are shown in each decade.

2.5 Discussion

2.5.1 Implications

Climate change studies, with projections extending to 2100, show a long-term but uncertain picture of the New York City water supply system's future. It is generally accepted that climate models perform much better at estimating changes in temperature than precipitation, and changes in precipitation patterns will have significant impacts on both high flows and drought. In addition to the annual precipitation and temperature in New York State, social, political, and infrastructure changes will occur over the next 80 years. However, a long-horizon study holding these other factors constant gives information about the system that managers can utilize during a short planning horizon.

In this paper, the relative value of Adaptation 1 (forecasts) and Adaptation 2 (leak repair) for reducing drought in the Delaware River Basin reservoirs are shown to be effective in reducing the impacts of drought. The water demand reduction that has occurred in recent years is also an effective method in managing drought. Though the demand reduction modeled here represented a leak repair project in progress, initiatives to reduce the city's consumption on a municipal level are further steps towards maintaining reliability. Forecasting (here, perfect forecasting) was most effective in future climates, and the following chapters explore current forecast skill and incremental improvements in system performance for a given improvement in forecast skill in detail. This knowledge, displayed quantitatively, can encourage the development of forecasting tools at the right spatial and temporal scale for water managers.

The relative value of Adaptation 3 (operational changes) and Adaptation 4 (specific infrastructure improvements) for managing turbidity in the Catskill system are explored. The research suggests that while increased precipitation will result in increased turbidity load to the system based on historic rating curves, operations and infrastructure improvements in combination are able to reduce the length of average alum addition to levels similar to historic levels even under extreme projections. Though not included in the study, watershed and riparian land management, as discussed in multiple iterations of NYCDEP's Filtration Avoidance Determination Reports, can also serve to reduce inflow turbidity starting from its hydrogeologic source.

With the exception of perfect forecasting, this case study shows a variety of real options successful in both present and future climates, with other options unmodeled but available, showing a promising outlook for the water supply system. The monetary cost of implementing these adaptations is not considered when showing an adaptation as beneficial. Additionally, these options, particularly the operational changes utilizing the Ashokan release channel, have an effect on the wider community that would be considered in implementation. However, the dramatic improvements shown can motivate further investigation of these changes, their costs, and their benefits for NYCDEP, and encourage similar studies for other water supply managers.

2.5.2 Limitations

As in all climate studies, there are limitations and uncertainty in using climate forecasts. In evaluating adaptations, there is a tradeoff between number of scenarios tested and complexity of modeled system operations, and the choice to simplify the operations rather than reduce the scenarios was deliberate. This research contains three important simplifications that made the examination of a large number of climate scenarios possible. In reality, managers utilize forecasts and current conditions while applying their judgment in determining system operations. Here, decisions are driven by a set of deterministic rules. Additionally, perfect monthly forecasts were incorporated into the model to determine operations. The sources of error and the error structure of current River Forecast Center products is beyond the scope of this project but will be an important consideration in future work. The largest limitation in the individual models is the treatment of turbidity in Ashokan Reservoir. The model accounts for turbidity in each basin as two independent, completely mixed systems. In reality, turbidity is spatially variable throughout each reservoir and can travel in a plume from the inlet to the outlet. A second reservoir model, developed on a daily or sub-daily time-step, would improve turbidity modeling at the expense of simulation time.

Finally, utilizing three models in series allows error to propagate through the model. This error is unquantified, but effort is made to show the effects of the method of incorporating GCM inputs, show the range of outcomes, and ensure that these outcomes are in general agreement with past observations and models whenever possible. Acknowledging this, the study presents its findings as comparative results rather than absolutes.

2.6 Conclusion

Decisive action concerning large capital investments is often difficult due to uncertainties in the future. Accurate forecasts of water demands are an example of an important planning variable that has proven to be more sensitive to pricing, technological change, and consumer perceptions and values than originally estimated. Climate change, another uncertain variable, increases the magnitude of uncertainty in projections and the range of model outputs available. It is true that the effects of the most extreme projected climates in the coming century overshadow the capacity of current mitigation projects. However, these results indicate that sensible adaptation options can successfully manage drought and turbidity impacts for the majority of projected climates and that win-win strategies in anticipation of climate change are possible.

In drought management, seasonal forecasting, as monthly sums with lead-times of 1 to 11 months, proved to be most effective in reducing the number of days under drought conditions for NYCDEP when drought is defined as a reservoir storage threshold. Since this provides a large benefit under both current and projected conditions, forecast development should be an area of focus in collaborations between scientists and water managers. Meanwhile, the benefits from a small reduction in total demand demonstrate that efforts to improve efficiency and maintain infrastructure will be a crucial tool in mitigating drought should drier climates manifest.

In turbidity control in the NYCWSS, operations and infrastructure are complementary. The two options investigated, west basin operations and Catskill aqueduct stop shutter improvements, address turbidity mitigation in different parts of the process. West basin operations including the use of the Ashokan reservoir release channel manage west basin and east basin turbidity by allowing additional settling time, preventing spill, and reducing the amount of sediment transferred between basins. Meanwhile, aqueduct improvements reduce turbidity load by reducing the minimum flow rate in the aqueduct and therefore the amount of water required from Ashokan reservoir’s east basin in times of high turbidity.

Within the limitations of this study, all of the options provided the intended benefits under all climate change scenarios; none were significantly less effective or detrimental in any scenario. Additionally, for both drought and turbidity management, combining options showed cumulative benefits. Despite climate uncertainty, adaptations anticipating the general trend of projected climate change (in this geographic area, increases in both temperature and annual precipitation) will likely be robust, effective, and complementary to additional adaptations into the future.

CHAPTER 3

INTERPRETING AND EVALUATING ENSEMBLE HYDROLOGIC FORECASTS FOR WATER RESOURCE MANAGEMENT

3.1 Abstract

A challenge to the adoption of seasonal streamflow forecasts in water resource system management is the lack of a standard method to quantify improved system performance as a function of improved forecast quality. This paper proposes and tests a method of determining operational improvement with forecasts that: 1) illustrates the use of ensemble forecasts, 2) supplements the National Weather Service's Ensemble Verification System, and 3) provides flexibility in the system performance metric. The method utilizes the "Mean Capture Rate Diagram" and the "Probability Score metric" to quantify synthetic forecast quality and utilizes iterative system modeling to determine the effect of forecast quality on system performance. The City of Baltimore's Water Supply System is used as a case study. The results illustrate the relationship between 90-day forecast quality and a composite performance metric including costs and reservoir drawdown. These relationships are shown to be linear for a range of forecasts and that an upper limit exists for the benefit of forecasts, particularly in managing one- to two-year hydrologic droughts. In addition to the system-specific results in the case study, insight is gained about the method itself and its possible use for other systems bridging the gap between forecast verification and forecast adoption.

3.2 Introduction

Despite more than 30 years of experience in generating streamflow forecasts, seasonal-scale ensemble predictions are frequently underutilized by water supply managers. Many challenges in forecasting remain, such as a lack of adequate data collection (Pagano et al. 2014), inconsistent verification methods (Schaake et al. 2007), and a lack of agreement on the best methods for quantifying uncertainty. Additionally, barriers to adoption in practice are numerous and significant. Notable among these are a profession-wide disincentive to innovate (due to the high value placed on reliability and the high visibility of undesirable outcomes) and limitations in our ability to quantify improved system performance when incorporating forecasting (Rayner et al., 2005).

To address this last point, this paper proposes a methodology to explore the relationship between forecast skill and system performance, that: 1) allows the use of ensemble forecasts, rather than single-trace forecasts, to accommodate the move towards ensemble forecasts and quantified uncertainty, 2) supplements the National Weather Service's Ensemble Verification System (Brown et al., 2010), and 3) provides flexibility in its system performance metric, allowing for its application to water resources systems with a variety of goals.

With the goal of quantifying performance improvement as a function of forecast skill, this method utilizes the Mean Capture Rate Diagram and the Probability Score metric. Resampled forecasts are created with varying levels of error according to the probability score, and a system model is implemented with a predefined metric describing system performance. After iterative simulations of the system model, results characterize the relationship between forecast error and system performance, including a possible threshold of diminishing returns for improved forecasts. This approach addresses a central question associated with streamflow forecasts: what amount of water supply performance improvement (e.g. delivering water reliably at a low cost) can be expected from improved forecasting and how can this be quantified?

3.3 Background

3.3.1 Forecast Evaluation

Forecast evaluation measures, in various ways, how well a forecast of a quantity, in this case a time sequence of streamflow, reflects what will occur in the future. Forecasts may assume a variety of forms: deterministic (single trace), probabilistic (categorical with a likelihood assigned to each outcome), or ensemble forecasts (multi-trace) (a collection of multiple time sequences). Each evaluation metric is some function of the joint distribution of forecasts and the corresponding observations. For example, for a deterministic forecast, the absolute error of a forecast is the difference between the forecast and the observed value. Forecast "quality" refers to any independent metric (e.g. root mean square error - RMSE), while forecast "skill" compares a forecast to a reference forecast (e.g. Nash-Sutcliffe efficiency - NSE); a positive value of forecast skill indicates that the new forecast performs better than the reference for a given metric. The specific metrics chosen for an analysis should be appropriate to the forecast type and its intended uses (Katz and Murphy, 1997). Attempts at verification in practice require the creation of new forecasts over a past time period as though they were being issued at that time, allowing a comparison of the forecasts to the observed data. The new forecasts are verified by calculating metrics evaluating quality for both the new and previous versions and/or by calculating skill scores.

Past verification of forecast quality and skill for hydrologic forecasts is sparse and inconsistent in approach, evaluation time-frames, and metrics. Pioneering methods in data assimilation, hydrologic modeling, and statistical methods for forecasting are evaluated as they develop. For example, as part of their hydrologic forecast development process, the National Weather Service (NWS) has evaluated their hydrologic ensemble forecasts, for four basins for a 14 day lead-time (Brown et al., 2014). In general, new methods show incremental improvements over previous methods, though flow extremes and flows at long lead-times remain a challenge (Bennet et al., 2014a; Bennet et al., 2014b; Rosenberg et al., 2011; Block et al., 2009; Franz et al., 2003). Verification standards (through the Hydrologic Ensemble Prediction Experiment) (Schaake et al., 2007) and software (NWS's Ensemble Verification System, EVS) (Brown et al., 2010) have been created to standardize the process and make verification results easier to implement, share, and compare between sites and studies.

Additional research has investigated the use of forecasts in water resource systems operations by using computer models of the systems. Experiments using perfect forecasts, including the study in Chapter 2, found operational benefits, while experiments using existing or reforecast products had mixed results (Sankarasubramanian et al., 2009a; Maurer and Lettenmaier, 2004; Ritchie et al., 2004). The majority of studies considering forecast-use use synthetic imperfect forecasts created by adding error to the streamflow record, and these error-added forecasts also lead to mixed results (Anghileri et al., 2016, Sankarasubramanian et al., 2009b; Georgakakos and Graham, 2008, Mishalani and Palmer, 1988). In addition to modeling studies, a few real world examples are available. Though snowpack-based forecasts have been used successfully in the western United States, poor forecasts for the Yakima River Basin in 1977, due in part to an improperly applied hydrologic model, resulted in large unnecessary agricultural losses and subsequent litigation (Glantz, 1982). Overall, the studies demonstrate both regional and use-based differences in hydrologic forecast quality and value, partially due to infrastructure, utilization, and goals, and partially due to the predictability of climate and hydrology in that region.

3.3.2 Forecast Types

Contemporary hydrologic forecasts are delivered as collections of possible future streamflow series, termed ensembles. Forecast generation methods fall on a spectrum between forecasts based on numerical weather prediction and hydrologic models and forecasts based solely on statistical processes. In the United States, the NWS introduced ensemble forecasts utilizing a watershed model, Ensemble Streamflow Prediction (ESP) in 1985 (Day, 1985). These forecasts used observations of the weather and watershed as initial conditions and used each year of the historic weather record as equally-likely future forcings, resulting in an ensemble of future streamflow traces. This was particularly effective in basins where snow pack is considered in initial conditions during winter and spring (Shukla and Lettenmaier, 2011). More recently, these hydrologic models have utilized the outputs of numerical weather prediction models as inputs, replacing the historic record at short lead-times. Additionally, more sophisticated methods of quantifying the sources of uncertainty (sources including choice of hydrologic model parameters and imperfect observations of current conditions) are being incorporated into these forecasts. The current NWS forecasts are part of the Hydrologic Ensemble Forecasting System (HEFS) that includes ensemble bias-corrected weather forecasts, hydrologic forecasts, and verification methods (Demargne et al., 2014). This paper incorporates the hydrologic forecasts produced as HEFS forecasts.

Statistical methods for streamflow forecasts vary widely in methods and inputs. In early work in this field, Hirsch (1979) proposed creating ensemble forecasts by fitting a lag-1 auto regressive model (AR1) to the monthly streamflow record to create an ensemble forecast beginning with the observed flow and consisting of variations on the historic record. The ensemble is created by fitting an AR1 model to the historic record, running the AR1 model from the current observed streamflow, and then adding each year's error series back to the AR1 generated flows (instead of randomly generating an ensemble of error series). This process has since been expanded to a daily model to work with daily-time-step system models. In this method, each year's daily pattern of flows, in addition to the monthly error series, is incorporated into to the AR1 generated series. This paper utilizes these forecasts as AR1 forecasts.

3.4 Methods

3.4.1 The Mean Capture Rate Diagram and the Probability Score

The Mean Capture Rate Diagram (MCRD) is an illustration of the ensemble forecast's distribution (Brown et al., 2010). The diagram illustrates how much of the ensemble falls within a given range around the observation. Having more of the forecast traces within a given absolute error indicates a better forecast. The underlying metric in the MCRD is the Probability Score (PS). Wilson et al. (1999) defined the probability score as

$$PS(f_Y, x^0, w) = \int_{x^0-0.5w}^{x^0+0.5w} f_y(Y)dy \qquad \text{(Equation 3.1)}$$

for a single observation x_0, an ensemble forecast Y with PDF f_Y, and an interval w. Figure 3.1 illustrates the PS for a single forecast/observation pair. As an evaluation metric, the PS is averaged across all available forecast/observation pairs. The MCRD plots the interval "*w*" as a function of the probability score, and *w* is referred to as absolute error (Figure 3.2). The probability score is located on the x-axis and ranges from 0 to 1; the absolute error *w* is plotted on the y-axis and ranges from 0 to the largest error value found in the ensemble forecast.

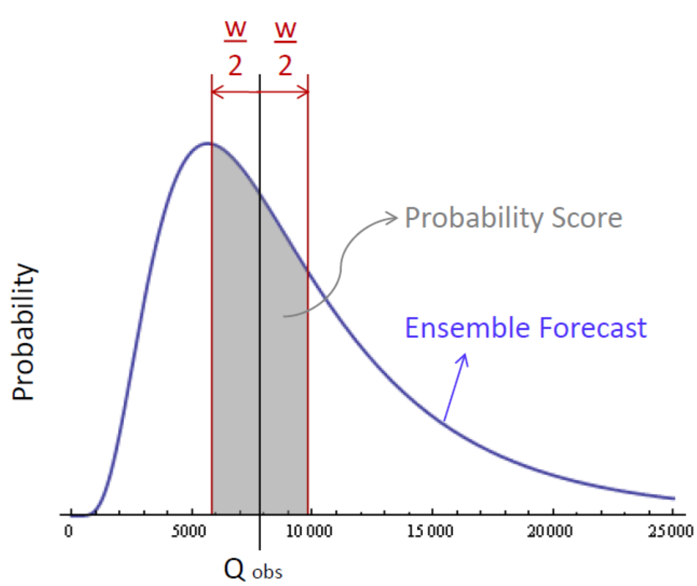

Figure 3.1 Illustration of the Probability Score. The PS calculates what fraction of the ensemble forecast falls within a given range (*w*) around the observed value.

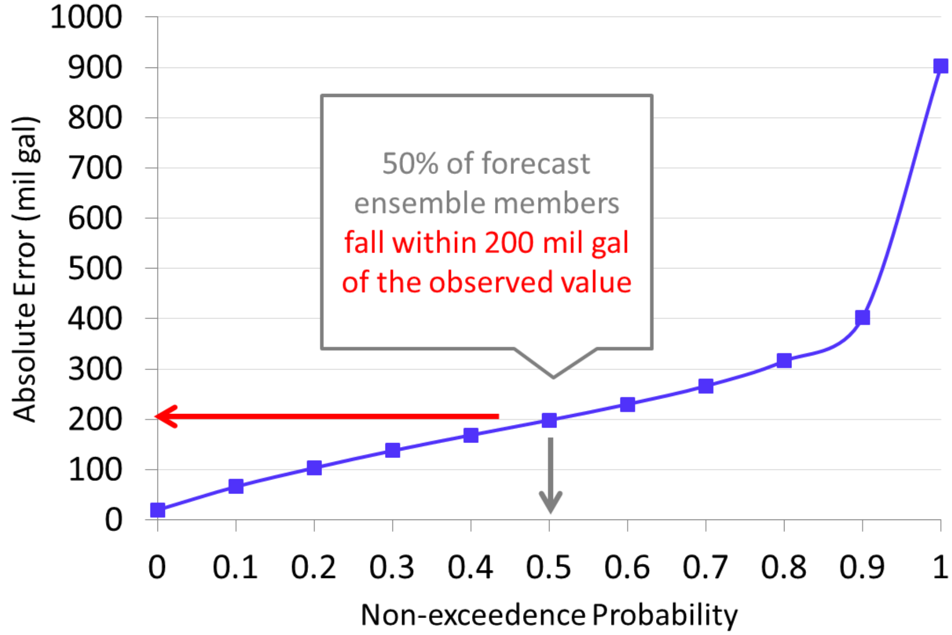

Figure 3.2 Illustration of the Mean Capture Rate Diagram. The MCRD plots the absolute error (1/2 w) as a function of the non-exceedance probability (the probability (score).

3.4.2 Determining Forecast Value

Results of the MCRD are useful for demonstrating the spread of forecasts and for comparing forecasts across sites. Two insights are gained by combining the MCRD, the probability score, and a system model: the Probability Score can become a use-specific forecast evaluation metric, and the relationship between forecasts and system performance can be quantified. To turn the Probability Score into a use-specific metric for a particular site, the parameter ½*w* (absolute error) can be defined based on target values or thresholds for a relevant system metric. Two, the same modeling approach can quantify the gains in system performance achieved through improved forecasts, which is the central intention of this paper. In this research, the experimental design and the method proposed is described in the following paragraphs.

Forecast ensembles with varying absolute error are created by sampling traces from a pool of existing HEFS and AR1 forecasts and from the historic record. Next, for each time-step, an ensemble forecast is issued. Then the system model tests multiple operating options for each forecast trace *i* and then calculates the expected value *E* of each action *j* using a single or composite metric *C* measuring system performance. A decision based on minimizing or maximizing the expected value is implemented until the next forecast is issued. The expected value of a course of action is defined as:

$$E_j = \frac{1}{n}\sum_{i=1}^{n} C_i \qquad \text{(Equation 3.2)}$$

with E_j the expected value of the *j*-th operating choice, *n* the number of forecasts, and *C* the value of the composite metric. The metric *C* is then evaluated for the entire trial period. The process is repeated using sets of forecasts with different absolute error values, resulting in a set of *C* values as a function of absolute error.

3.4.3 Case Study Setting

The City of Baltimore receives its drinking water from a surface water reservoir system and in some cases, from the Susquehanna River. Baltimore's water supply system consists of Prettyboy, Loch Raven, and Liberty Reservoirs with a total of 76 billion gallons of usable storage and a drainage area of 467 square miles in two small basins on the Chesapeake Bay between the Susquehanna and the Potomac Rivers (Figure 3.3). Annual precipitation for the region is 40 inches annually, distributed approximately equally throughout the year and with minimal snow accumulation. Surface reservoirs are operated to balance water sources and to deliver an average of 225 million gallons per day (mgd) to 2 million people. Water demand varies seasonally, with higher demands in the summer. The City of Baltimore Department of Public Works also maintains a connection to the Susquehanna River. During severe drought, the city can pump

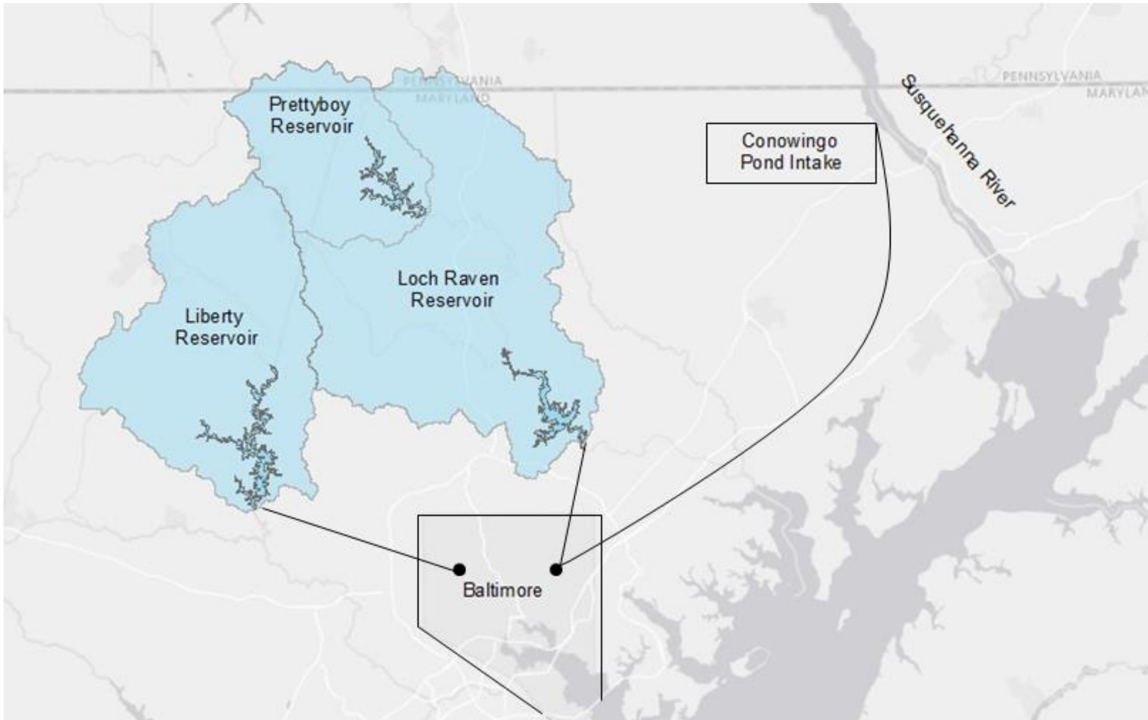

Figure 3.3 Watersheds and reservoirs of the City of Baltimore Water Supply.

water from the Susquehanna River to supplement their supply, but pumping cost, fees, and additional treatment costs make this an expensive option and it is not implemented unless it is truly necessary (Reimer, Muegge, & Associates, Inc., 2000). Medium-term drought forecasting can contribute to improving the quality of pre-treatment water, maintaining reservoir storage, and avoiding water use restrictions. This research applies HEFS and AR1 ensemble forecasts, along with the historic streamflow record as an ensemble forecast, for nearby Lancaster, PA in the context of operating the Baltimore water supply system. HEFS forecasts are not available for sites immediately upstream of the reservoirs; the section "Forecast Sources" below elaborates.

To evaluate forecast skill, this research selected Baltimore's composite metric, as defined in the previous subsection, with three terms, designed to give reservoir storage and economic losses equal weighting. This metric attempts to capture the trade-offs between the economic costs and losses from pumping and water use restrictions during drought, and the risk and poorer water quality resulting from drawing down reservoir storage levels. The equation resulting is as follows.

$$C = C\ restrictions + C\ pumping + C\ minimum\ storage\ \text{“}cost\text{”} \qquad \text{(Equation 3.3)}$$

The water demand for Baltimore consists of residential, commercial, and municipal use, 46% of which is considered elastic (City of Baltimore Department of Public Works 2002) and can be reduced at a cost by voluntary and mandatory restrictions. The cost of restrictions, *C restrictions*, is

$$C_{restrictions} = \frac{e^D}{1+\frac{1}{\eta}} \times \left[{Q_m}^{1+\frac{1}{\eta}} - {Q_d}^{1+\frac{1}{\eta}} \right] \qquad \text{(Equation 3.4)}$$

Where Q_m is the maximum demand Q_d is the demand delivered, η is the price elasticity, and D is the demand constant, including the price of water P, defined as:

$$D = \ln(P) - \frac{\ln(Q_m)}{\eta} \qquad \text{(Equation 3.5)}$$

(Jenkins et al., 2003). Monthly maximum demands, the maximum amount of water that would be used at the current price with no water-use restrictions in place (normal of non-drought conditions), are taken as average historic uses. The monthly price elasticities used were (J) -0.1; (F) -0.1; (M) -0.1; (A) -0.2; (M) -0.3; (J) -0.3; (J) -0.3; (A) -0.3; (S) -0.3; (O) -0.2, (N) -0.1; and (D) -0.1 (McIntyre et al., 2017).

The City of Baltimore Water Supply has a permit to withdraw as much as 250 MGD from the Susquehanna River and a current infrastructure capacity to pump 137 MGD. The cost of using this source is the sum of the cost of pumping the water (itself a function of energy price, flow rate, and head) and a consumptive use fee of $0.33 per 1000 gal. The cost of operating the pumps, $C_{operating}$ in equation 3.6, is $58.20 per million gallons to pump 80 MGD and $66.40 per mil gal to pump 137 MGD (McIntyre et al., 2017).

$$C_{pumping} = Q_{pumped}(C_{operating} + C_{fee}) \qquad \text{(Equation 3.6)}$$

Finally, a penalty for low storage is added to the cost equation as

$$C_{minimum\ storage\ "cost"} = W\left(1 - S_{ending}\right)^2 \qquad \text{(Equation 3.7)}$$

The weighting factor W is $\$3.9 \times 10^7$ (per fraction of storage squared) and the ending storage S_{ending} is the fraction of total storage in the reservoirs at the end of each 90-day forecast period. The weighting factor reflects modeled tradeoffs of 24 different operating policies evaluated in McIntyre and Palmer (2017), a previous study using the same system and data. In that study, outcomes ranged from a minimum storage of 9% and no economic loss in a policy that prioritized costs, to a minimum storage of 80% and 6.5% of annual revenue lost in a plan that prioritized maintaining high storage, the use of restrictions, and the use of Susquehanna water. These ranges were used to define the minimum storage “cost” term such that the composite metric would have the same value for those two extremes of experimental operating policies. A quadratic storage term resulted in more realistic actions than a linear term. Operating decisions are made based on minimizing the composite metric C, the expected total “losses.”

A model of the Baltimore system is used to determine the effect of varying absolute error in forecasts on system outcomes. The model includes reservoir water balances and several possible actions (Table 3.1), coded in R as a set of ordinary differential equations. The composite metric defined above is used as an objective. The full ensemble of forecasts is used to make an operating decision; instead of using the forecast mean or median to simplify the process, each forecast trace is run through the system model testing each possible drought mitigation action. Forecast traces are equally weighted, and the expected value of the composite metric is calculated for the 90-day period. The action with the lowest expected loss is implemented for 5 days, until the subsequent set of forecasts is issued (Figure 3.4). For this experiment, the model is run using the historic inflow record from 2/5/2001 to 12/27/2010.

The same composite metric (with minimum storage in place of ending storage) is plotted as a function of the resampled forecasts' absolute error to quantify the relationship between improved forecasts and improved system performance.

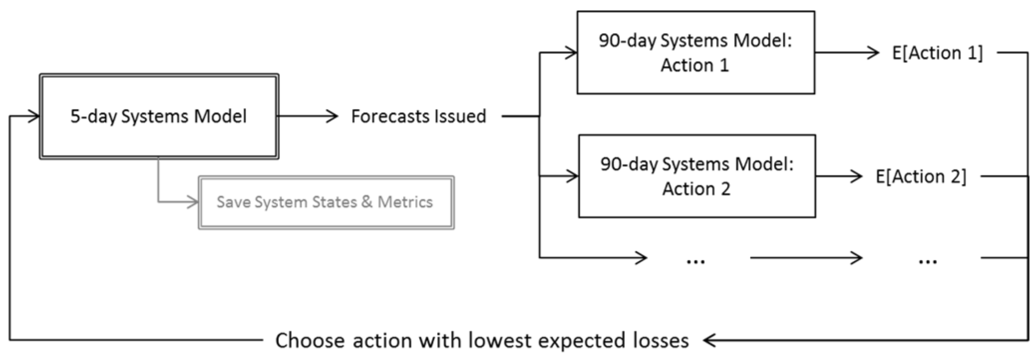

Figure 3.4 Diagram of iterative system modeling for decision-making. There are a total of seven possible actions modeled for the Baltimore water supply system.

Table 3.1 List of drought mitigation actions

Action	Susquehanna River Pumping (MGD)	Water Use Restrictions
1	0	none
2	0	voluntary
3	80	none
4	80	voluntary
5	80	mandatory
6	137	voluntary
7	137	mandatory

3.4.4 Forecast Sources

HEFS and AR1 forecasts for Lancaster, PA and their significance for operating the nearby City of Baltimore water supply are compared, as the HEFS forecasts were not available for a site upstream of that system. The HEFS forecasts at Lancaster are 37-member ensemble forecasts issued every 5 days from 2/5/2001 through 12/27/2010. The AR1 forecasts are created for the same issue dates in a 66-member ensemble using the historic streamflow record from 1929 to 1995. For forecast quality and skill analysis, forecasts are evaluated based on the record of Lancaster flow. For forecast value analysis, forecasts are used to predict the inflow of the City of Baltimore Water Supply using quantile mapping, the transfer method that performed best for these datasets when compared with Maintenance of Variance Extension, Type 1 (Mcintyre et al., 2017).

To test a range of forecast skills, 40-member ensemble forecasts are re-sampled from 199 existing forecast traces to have a specific PS. The ensemble shape is constrained at the 25th, 50th, 75th, and 100th percentile to match the average shape of the HEFS or AR1 forecasts. This requires the MCRD of the resampled forecasts to match the MCRD of the two existing forecasts types at those points. This prevents outlier forecasts in an otherwise low-absolute error forecast from disproportionately and unrealistically influencing the model's decision. Absolute error values in resampled forecasts' 90-day totals range from 1,000 MG to 30,000 MG (the equivalent of 4 to 130 days of demand for the system) with the width w corresponding to a constant PS of 0.5. The 199 existing traces include HEFS forecasts, AR1 forecasts, and the historic record.

3.5 Results and Discussion

This methodology is design to answer the question, for a specific water resource system, what amount of performance improvement can be expected from improved forecasting? To show the Mean Capture Rate Diagram in the context of water supply system use, the MCRD is plotted for the AR1, HEFS, and historic forecasts. To quantify performance improvement, resampled forecasts with a range of absolute errors are utilized in a system model of the Baltimore water supply. Operations aim to balance minimizing costs with maintaining high storage levels in the reservoirs, and the best course of action according to a composite metric was chosen based on the upcoming 90 days' ensemble forecast. To demonstrate the Probability Score as a system-specific forecast verification metric, a value of *w* is chosen based on the Baltimore case study results, and the metric is evaluated for existing forecasts for Lancaster.

3.5.1 The MCRD

The MCRD evaluates the average range (the absolute error) around the observed value necessary to capture a given amount (the non-exceedance probability) of the ensemble forecast. Points with higher y-axis values on the plot indicate that a larger error (y) is necessary to capture

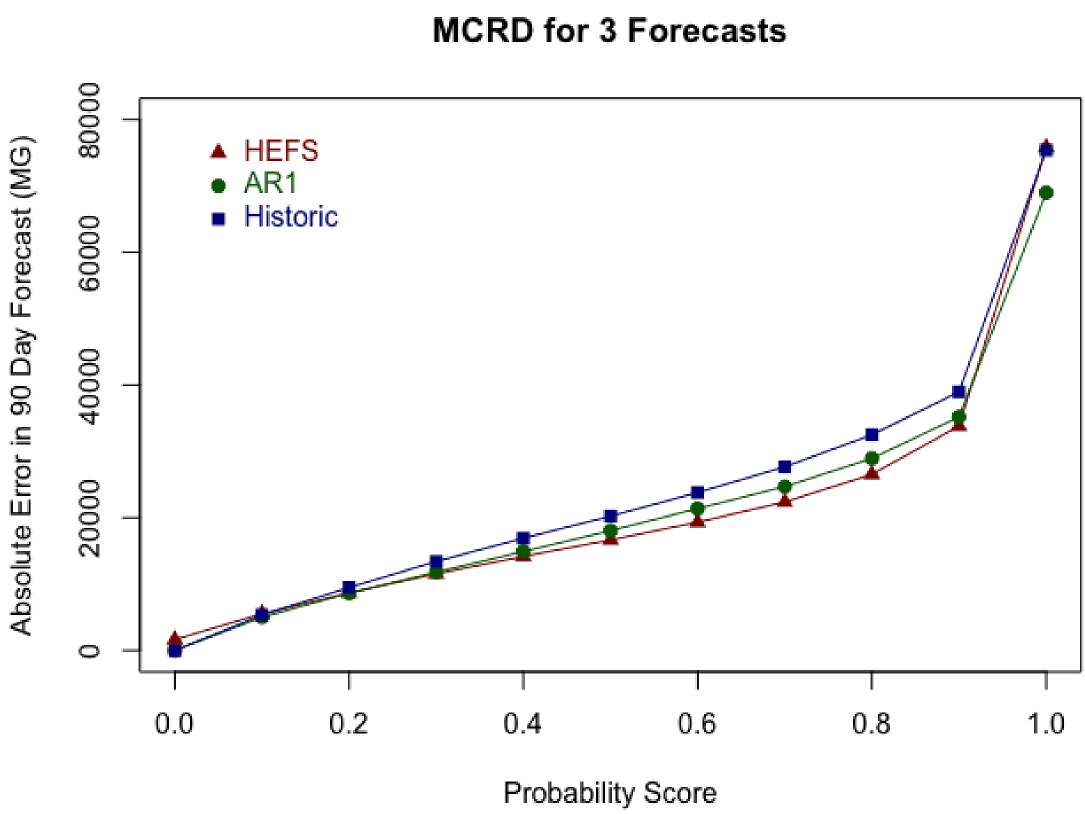

Figure 3.5 The Mean Capture Rate Diagram for three existing forecasts. The diagram plots the absolute error, w, as a function of the non-exceedance probability or the Probability Score. The plot can be read as, to capture 20% of the forecast ensemble members (x-axis), a window of 10,000 mg (y-axis) around the observed value is necessary.

a given percentage of the ensemble (x); lower y-axis values indicate a better forecast. The MCRD is plotted for HEFS, AR1, and historic forecasts as 90-day totals (Figure 3.5). For most of the range of the ensemble, HEFS forecasts perform best, having most of the ensemble members closer to the true, observed value. However, larger absolute error values at a non-exceedance probability of 1 indicate that the extreme values of the HEFS forecast ensembles are farther from the observed value than the AR1 forecasts. These differences result in the varying shapes of the distribution of the ensemble. The difference between the two forecasts is small relative to the total error of those large-error forecast members, but for both, the extreme-most 10% of forecasts are substantially poorer than the remaining 90%. This can be considered when utilizing the ensemble forecasts in practice, and a repeat of the case study giving less weight to the extreme traces may show overall better performance.

3.5.2 System Performance

To incorporate performance metrics into the MCRD, values of the composite metric, total system "losses," can be plotted as a function of absolute error for a given non-exceedance probability. Plotting "losses" as a function of forecast quality resulted in a linear trend over a large range of forecast error. This relationship is plotted for a PS of 0.5 roughly constrained to two forecasts shapes in Figure 3.6. The trend was also evident in trials defining forecasts by a PS of 0.3 and 0.7. Deviations around the trend are likely a consequence of the forecast creation method (sampling) and the decision method (a small-scale threshold-crossing effect of choosing an action based on forecasts). By testing two forecast shapes, we see that for this case study, small changes in forecast shape (here, the difference between statistical and mechanistic forecasts for the site) have less of an effect on performance than does a difference in absolute error, though this may not be true for larger differences in forecast shape. When resampled forecasts are not constrained to an ensemble shape (i.e. are only required to have 50% of traces within w of the true value and the other 50% without), the outlier traces have a larger effect on system performance.

In that case, performance was worse overall and nonlinear, with more significant improvement for a given reduction in error at the lowest absolute errors.

In this framework, a single-trace perfect forecast would be equivalent to a PS of 1 and an absolute error of 0: 100% of forecast ensemble members are within 0 MG of the observed flow. This results in 1.36% of annual revenue lost over the 2001 – 2010 period, denoted on the plot as a blue star. That is, a certain amount of drawdown, pumping, and water use restrictions are necessary for this system regardless of foresight. This sets an upper bound on the ability of forecasts to improve operations, particularly over a significant drought.

Existing forecasts are within the range of the linear trend (Figure 3.6). To have a PS of 0.5, the range of absolute errors for existing forecasts would be 16650, 18050, and 20230 MG for HEFS, AR1, and the historic ensemble, respectively.

Incremental improvements to forecasts can be made and will result in improved system performance, as long as operations are capable of utilizing ensemble forecasts. These improvements can be compared quantitatively to other options, including infrastructure changes, demand reduction, or distribution system monitoring and repairs. Here, for example, results indicated a difference of $4 million in the cost of the 2001-2002 drought, calculated using the composite metric, over the range of forecast quality tested.

The same trend is evident in the relationship between forecast absolute error and system "losses" for the 2001-2002 drought (here, the value of the composite metric over the two-year span of 2/5/2001 – 2/5/2003) (Figure 3.7). In the case of drought, any quality of forecast results in higher losses in drought years than in the record overall, indicating that even with a perfect 90-day outlook, some combination of water use restrictions, pumping, and substantial reservoir drawdown are necessary. The fact that the trend remains the same in both the drought scenario and the full record indicates that, for a reservoir system with more than a year's worth of storage, little benefit is derived from the forecasts for a multi-year drought, and that the 90-day foresight of good forecasts allows for better operations for more minor droughts elsewhere in the record.

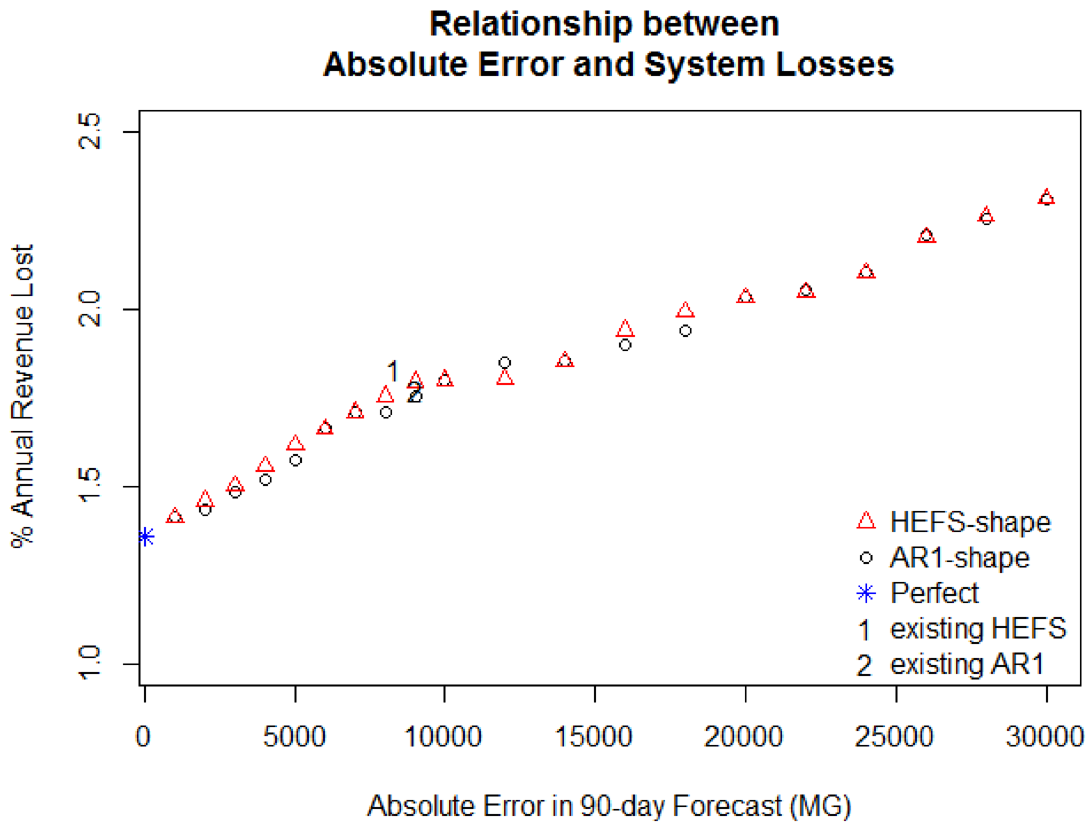

Figure 3.6 The relationship between absolute error, w, and the composite metric "System Losses" for two forecast shapes, based on a model of the Baltimore water supply using synthetic ensemble forecasts. Absolute error is varied for a Probability Score of 0.5. Perfect forecasts, and existing HEFS and AR1 forecasts are included as single points.

It is important to note that forecasts that inform drought management (as opposed to flood forecasts that inform flood management) are least valuable when reservoirs are relatively full. This is a result of the ratio of total water available between stored and future inflows and the fact that the longer-term forecasts have to be accurate for a very long period (the end of the drawdown-refill cycle).

3.5.3 Probability score as an evaluation metric

For the Baltimore system, the relationship between forecast error and system losses did not result in a threshold past which forecast improvements had diminishing returns. Instead, a target value of the composite metric can be chosen to compare the different forecast types' Probability Scores. 1500 MG for a 90-day forecast was chosen as a slightly smaller forecast spread than current existing forecasts, which would have reduced losses to 1.8% annual revenue over the trial period. Evaluating PS (w=1500 mg) resulted in PS of 0.46, 0.45, and 0.34 for HEFS, AR1, and historic forecasts. This is the equivalent of reading a non-exceedance probability from the MCRD based on an absolute error (x as a function of y). Ranking the three forecast types from best to worst by Probability Score, results are consistent with the RMSE, CRPS, and reliability of these same forecasts. This consistency suggests that the Probability Score based on a system threshold may be included in forecast verification to confirm a benefit to the forecast user.

The lack of a clear inflection point may be more widespread than just this pilot case. For many east coast water supplies, the reservoir volume is greater than a year's worth of demand, so that a multi-year drought would be necessary to substantially draw down the reservoir. In those cases, it is likely that the quality of a hypothetical two-year forecast would make a larger difference in outcomes than the quality of a three-month forecast. Further, drought mitigation actions can occur over a long period of time, such that a difference of 5 days pumping (the length of time between forecast issue days in the case study) is small relative to the total volume of water pumped over the course of a drought. For this particular case study, the multi-objective model formulation does not lend itself to a threshold-like behavior. In contrast, forecasts used in predicting high flows with the single objective of preventing downstream flooding may show an inflection point in the relationship between forecast quality and system performance. The decision of whether and how much to draw down a flood control reservoir in advance of a storm or snowmelt season is inherently a more threshold-based problem; there are no or minimal costs to a particular point, and then costs (e.g. as flood damage) begin and increase. The greatest benefit of discovering a threshold and using this variation on the Probability Score to judge and

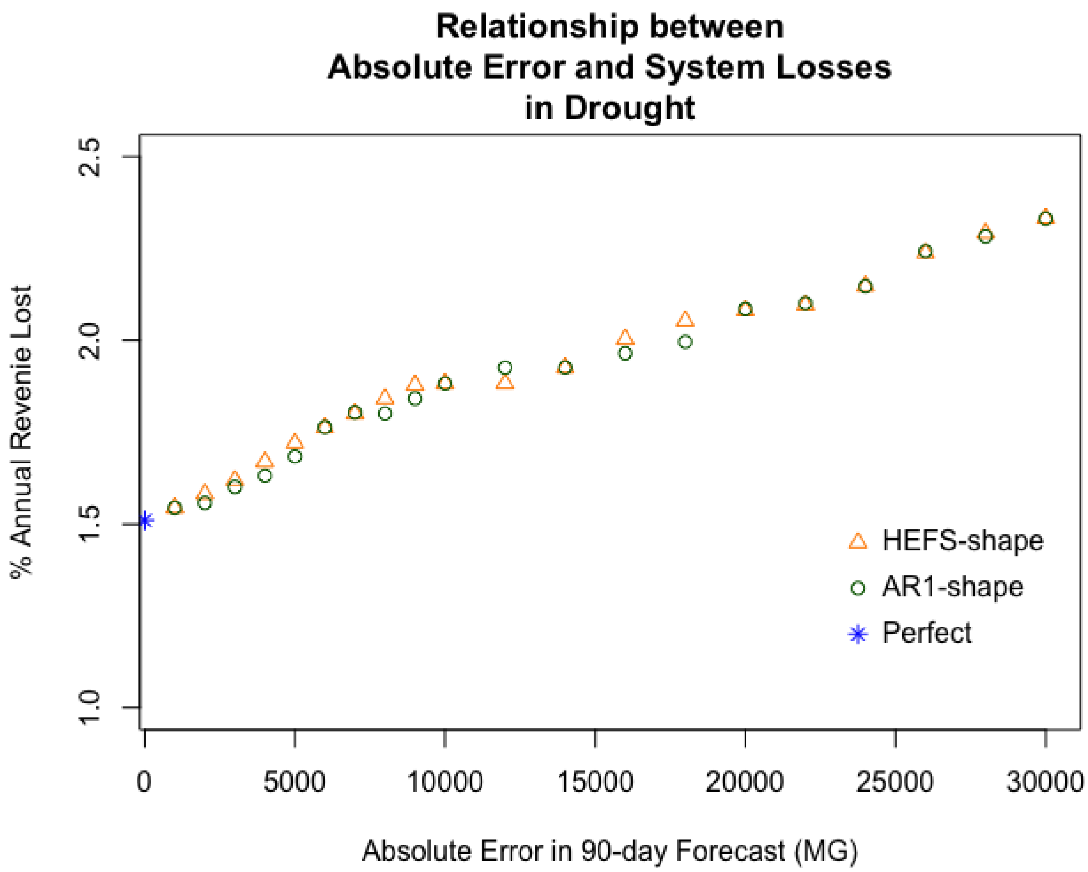

Figure 3.7 For the 2001-2002 drought, the relationship between absolute error, w, and the composite metric "System Losses" for two forecast shapes, based on a model of the Baltimore water supply using synthetic ensemble forecasts. Absolute error is varied for a Probability Score of 0.5. A perfect forecast is included as a single point.

drive forecast improvements may appear in multi-objective systems when a threshold-based function like flood control is combined with a value-providing objective such as recreation or water supply.

When there is not an obvious choice of threshold, the threshold can be defined by regulations, e.g. on a legal threshold for the total maximum daily load of a constituent, or by making a direct comparison to the benefit expected from an infrastructure project. The latter would require an additional step of estimating the cost and plausibility of that amount of forecasting improvement, itself a substantial scientific undertaking.

The method itself is generalizable to water resource systems in many cases. In cases where actions are taken based on the current and projected state of the system, the success of those actions depends on the quality of that projection. This method is intended to quantify the value of improved forecasting and to provide an additional metric that incorporates system performance. As such, the method would not be applicable to a one-time decision, such as setting a rule curve for a flood control reservoir. Other, risk-based, methods would be more appropriate in such a case.

3.6 Conclusions

This paper answers one question about the use of hydrologic forecasts for water resource system management: What amount of system performance improvement can be expected from improved forecasting? Using existing forecasts for Lancaster, PA, synthetic forecasts with varying quality, and a system model of the Baltimore, MD water supply system, this paper demonstrates a method for quantifying improved system performance as a function of improved forecast quality.

For our case study, improvements in system performance were approximately linear over a range of forecast quality, from absolute errors larger than the error in climatology through a perfect forecast. A point of diminishing returns in terms of system performance was not identified for this case study, and the inclusion of perfect forecasts in the analysis illustrated an upper limit on what is achievable through improved forecasting. In the context of a water supply's strategic planning, this method can help estimate the expected gains of an investment in forecasting, especially once pilot studies of new technologies exist. Recent advancements in remote sensing for hydrologic forecasting (Lettenmaier, 2017) may be a viable option for a water supply, placing forecasting in the same category as an infrastructure project. Additionally, with a certain source of ensemble forecasts available, there may be a best way to incorporate those forecasts into management decisions. The framework demonstrated here is also able to accommodate alternate ways of utilizing forecasts in operations, by comparing system performance with the same set of incremental quality forecasts and different indicators, triggers, and actions. In forecast verification, the current consensus is that a wide range of metrics should be used to evaluate forecasts. The results here suggest that if the ultimate use of the forecast is known, system modeling and system-specific metrics can provide additional insight to forecast quality and value.

Improved system performance is possible through improvements in hydrologic forecasting. A method similar to the one proposed here can aid in quantifying the value of forecasts and thereby aid in the adoption of state of the art hydrologic forecasts. It is “beyond the scope of this paper,” but not beyond the abilities of the profession, to see these improvements in practice.

CHAPTER 4

INCORPORATING A HYDROLOGIC REGIME IN STATISTICALLY-BASED SEASONAL STREAMFLOW FORECASTS FOR THE EAST COAST

4.1 Abstract

This research explores the use of three-month categorical precipitation forecasts for improving streamflow forecasts on the east coast of the US. Six case study sites span the east coast from Georgia to Maine. The approach taken applies the concept of "regimes," which are defined as forecasts of future wet or dry periods. For comparison, perfect categorical forecasts and recent observed precipitation totals are compared

To provide rationale for such a regime definition, correlation between streamflow and precipitation is evaluated at the sites; correlation is highest between streamflow and the precipitation that occurs concurrent to the streamflow total. To provide context and realism, the skill of National Weather Service Seasonal Precipitation Outlook forecasts is tested, as forecasts have been shown to have small positive skill on the East Coast. Finally, the forecast-based regime definitions and the two comparison cases are incorporated into a variety of AR(1) streamflow forecasts. If a precipitation forecast with actionable information is identified, the method creating that forecasts is applied to estimate the upcoming wet or dry regime at the portion of the historic streamflow record corresponding to either wet or dry years. This research suggests that, at present, there is insufficient skill in seasonal precipitation forecasts to improve streamflow forecasting in the six cases tested. This suggests that other tools or areas of operational improvement should be identified to improve water supply system performance.

4.2 Introduction

As demonstrated in the previous chapter, incremental improvements in hydrologic forecasting can result in incremental improvements in water supply system performance.

Forecasts with a seasonal lead-time have been used effectively for water supply systems on the west coast of the US (Hamlet and Lettenmaier, 2000), where snow accumulation provides actionable information related to initial-conditions and teleconnections between precipitation and Pacific oscillations. This chapter explores one possibility for improving statistical streamflow forecasting on the east coast.

Methods of forecasting streamflow continue to evolve, including statistical, model-based, and hybrid methods. These include combinations of statistical and dynamical models (Schepen et al. 2012; Rosenberg et al. 2011), models taking advantage of numerical weather prediction outputs (Bennet et al. 2014a; Bennet et al. 2014b; Wand and Fu, 2014; Coelho and Costa, 2010), and sophisticated methods for combining multiple numerical weather prediction and hydrologic models (Demargne et al., 2014; Block et al. 2009). The National Weather Service's (NWS) hydrologic forecasts include methods for quantifying the sources of uncertainty (e.g., choice of hydrologic model parameters, imperfect observations of current conditions) (Demargne et al., 2014). Currently, the quality of these forecasts varies significantly. Factors impacting forecast quality include the state of sea surface temperature oscillations, e.g. El Nino years, or the set of forecasts issued, e.g. the lowest tercile of flows. It has been suggested that improved, longer-term forecasts for the east coast might include a correlation between central pacific sea surface temperature and PDSI in the southeast and mid-Atlantic (Cole and Cook, 1998), between the Pacific/North American teleconnection index and fall, winter, and spring temperatures for the same region (Leathers et al. 1991), or between the North Atlantic Oscillation and low-flow streamflow statistics in the northeast (Steinschnieder and Brown, 2011). In general, teleconnections for the east coast are weaker than for the western United States. Nonetheless, the possibility of longer-term memory and wider climactic conditions influencing streamflow, such as in the Columbia River Basin in the Pacific Northwest (Hamlet and Lettenmaier, 1999), provides motivation for investigating the possibility of regime-based hydrologic forecasting on the east coast.

For the primary method investigated here, the current climate condition or the "regime" is predicted using the National Weather Service's Seasonal Precipitation Outlook, a three-month categorical forecast of precipitation. The categories predict "above average precipitation" (for the time of year), "below average precipitation," or "an equal chance of above or below average precipitation." Six sites located along the east coast of the United States are used here to test regime-based hydrologic forecasts. Section 4.3 details the sites, statistical forecasting method, and evaluation metrics used in this study. Section 4.4 presents an assessment of both the underlying assumptions in the forecasting method and the method itself. Several possible regime definitions and the categorical precipitation forecast skill are tested at the specific study sites. Ultimately, the NWS seasonal precipitation forecasts and two comparison regime definitions are used to create conditional first-order autoregressive forecasts for six case study sites on the east coast. The comparisons are used to determine if there is actionable information in either the observed or forecasted precipitation and how the method of incorporating a regime influences the forecast's quality. Section 4.5 discusses the results and their implications for east coast forecasting and water resources management. Section 4.6 provides the chapter conclusions.

4.3 Methods

4.3.1 Data

Six east coast stream gages serve as pilot locations, two southeastern sites, one mid-Atlantic site, and three northeastern sites (Figure 4.1). Characteristics of the sites chosen include: long streamflow records, no upstream regulation of flows, and similar watershed size and flow magnitudes. One site for which NWS HEFS forecasts exist, Lancaster, PA, is included to facilitate future comparisons and utilization studies.

For each site, three data series are required. Daily streamflows were acquired from USGS stream gages. Monthly precipitation records were obtained from the National Climatic Data Center (NCDC) from weather stations near each of the gages (Table 4.1). The categorical seasonal precipitation forecasts were obtained from the NWS. These forecasts are issued monthly and archived, and forecasts were obtained from the archived seasonal outlook maps. The entirety of the overlapping precipitation and flow record was utilized for each of the sites. Seasonal precipitation forecasts are available beginning in October 1995.

A preliminary analysis of the datasets, including the correlation between precipitation and streamflow (as monthly totals) and the skill of the categorical NWS precipitation forecasts, is included in the results to provide rationale and insight to the forecasts' performance. For correlations, the entire overlapping portion of the streamflow and precipitation record was used. For evaluating the hydrologic forecasts, the entire overlapping precipitation and streamflow record was used to create hydrologic forecasts for the 1995 – present time frame, when NWS forecasts are available to use as a regime.

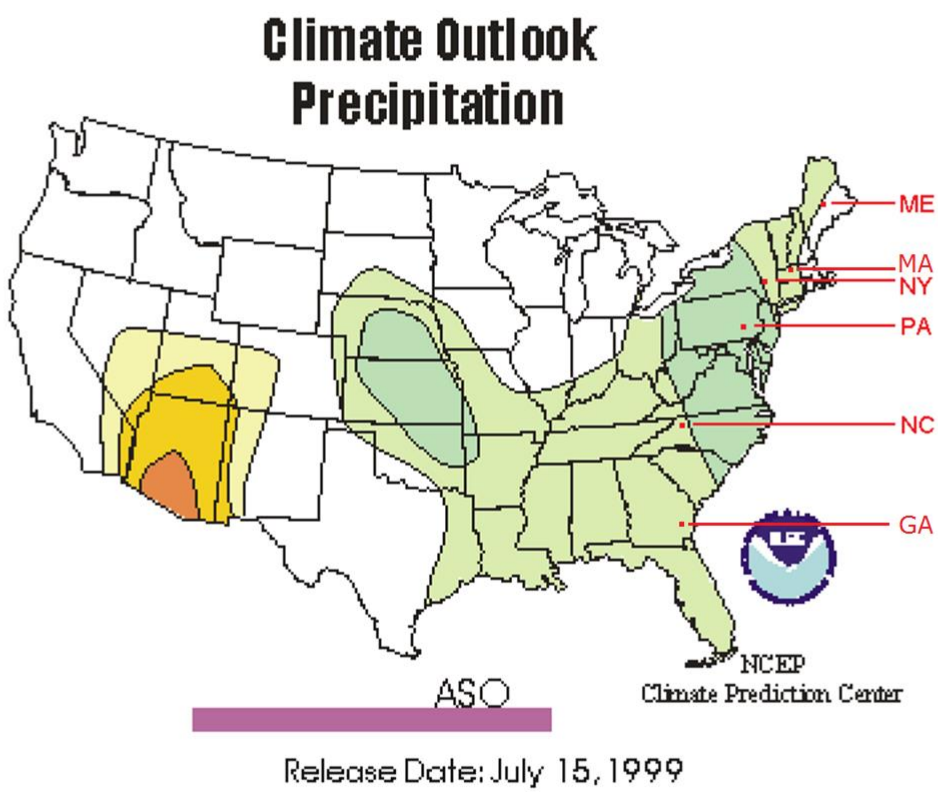

Figure 4.1 Case study sites plotted on a NWS seaosnal precipitation forecast map. The seasonal forecast for August/September/October, issued on July 15, predicts wetter-than-average conditions for most of the east coast.

Table 4.1 Case study sites for AR1/seasonal forecasting testing

Gage (USGS)	Name	Average Flow	Watershed Area	Record Length
02387000	Conasauga River at Tilton, GA	200 – 1050 cfs	687 sq. mi.	1964 - present
02112000	Yadkin River at Wilkesboro, NC	400 – 1000 cfs	504 sq. mi.	1928 - present
01576500	Conestoga River at Lancaster, PA	150 – 600 cfs	324 sq. mi.	1948 - present
01423000	West Branch Delaware River at Walton, NY	70 – 1400 cfs	332 sq. mi.	1956 - present
01166500	Millers River at Erving, MA	100 – 1800 cfs	372 sq. mi.	1950 - present
01031500	Pisquatis River near Dover-Foxcroft, ME	70 – 1000 cfs	298 mi sq	1947 - present

4.3.2 Forecast Creation

Streamflow forecasts are created using an autoregressive model, and a baseline model is modified to incorporate the current hydrologic regime. The baseline forecasts are created as follows. The historic streamflow record is aggregated by month, normalized via a log-transform, and standardized to a mean of 0 and standard deviation of 1. With $Q_1, Q_2, \ldots, Q_n$ as monthly inflow totals, the normalization and standardization proceeds as follows.

$$Y_t = \ln(Q_t) \qquad \text{(Equation 4.1)}$$

$$Z_t = \frac{Y_t - Y_{month,mean}}{Y_{month,std}} \qquad \text{(Equation 4.2)}$$

An AR(1) model is applied to the normalized standardized series Z_t with a different autoregressive parameter ϕ_{month} for each month, leaving an error series ε_t from each year of the record. A time sequence of observed ϕ is calculated from the normalized standardized data.

$$\phi_t = Z_t * Z_{t-1} \qquad \text{(Equation 4.3)}$$

A monthly ϕ parameter (12 total) is created as the average of all observed ϕ s for that month.

$$\phi_{month} = average(\phi_t, month) \qquad \text{(Equation 4.4)}$$

The error series from each year of the observed record is extracted from the data using the autocorrelation equation and parameters from the previous steps. The error series are saved as part of the process.

$$\varepsilon_t = Z_t - \phi_{month} Z_{t-1} \qquad \text{(Equation 4.5)}$$

The monthly AR(1) model is initiated using the past month's streamflow and runs into the future for up to 12 months. Each year's error series is returned to the AR(1)-generated flows, creating variance in the forecasts based on variance in the historic record. The ensemble size is the number of years of record available. The standardization and normalization of the data series is then reversed, and the daily variations reapplied to create an ensemble forecast at a daily time-step.

Regime-based forecasts differ from the baseline forecasts in the following ways. When precipitation is predicted or observed to be above- or below-average, the regime-based case calibrates the monthly autoregressive parameter ϕ based on historic years with the same regime. In those same cases, only a portion of the historic variance, as ε_t, is included in the ensemble. When the regime is above or below average, the ensemble size is 1/3 that of the original AR(1) forecasts. When the forecast predicts an "equal chance" of above or below average precipitation, the forecast utilizes the full record (i.e. the forecast is not constrained to the middle tercile when "equal chance" is forecast). To further investigate the process, two variations are also included, the first using just regime-based ϕ, and the second using just regime-based ensemble variance.

Three methods of defining the regime are explored. The central experimental method of the paper uses NWS Seasonal Precipitation Outlook forecasts, defining the regime as wetter than average, drier than average, or an equal chance of being above or below average. This maintains the same categories as the forecasts. For the east coast sites included here, forecasts of above or below average precipitation are only made in 15-30% of months with the remaining being forecasts of equal chance. A second method uses the previous three months' precipitation to define the regime retrospectively into the same three categories. When the observed precipitation is in the middle tercile, a regime of "equal chance" is forecast. Regime designation using this observed data is distributed more evenly into the three categories than with the NWS predictions. Finally, a hypothetical case of perfect categorical forecasts is included in the analysis. This forecast correctly designates the regime into above- or below-average categories based on the upcoming precipitation.

4.3.3 Evaluation Metrics

This method for conditioning streamflow forecasts on a precipitation-based regime has assumptions: streamflow is correlated with precipitation, and for the first of three regime definitions, the precipitation forecasts provide actionable information. The preliminary analysis to test these assumptions uses correlation, accuracy, and the Heidke Skill Score. Accuracy is defined as the fraction of correct categorical forecasts.

$$accuracy = \frac{correct\ forecasts}{total\ forecasts} \qquad \text{(Equation 4.6)}$$

The Heidke Skill Score (HSS) is a measure of the forecast accuracy compared to random chance. A skill score gives further insight into the information available in precipitation forecasts.

$$HSS = \frac{NC-E}{T-E} \qquad \text{(Equation 4.7)}$$

where *NC* is the number of correct forecasts, *T* is the total number of forecasts, and *E* is the number of forecasts expected to be correct by chance. The HSS can range from -∞ to 1, with 0 equaling no skill, 1 being a perfect forecast, and negative skill scores indicating that the forecast performs worse than a random guess. For categorical forecasts, the predicted tercile (with an "equal chance" forecast designated as a forecast of the middle tercile) is compared to the observed tercile. The numerical values are not considered in analyzing the quality of categorical forecasts in this case.

The hydrologic forecasts themselves are evaluated with the Continuous Ranked Probability Score and Skill Score. The CRPS integrates the squared difference between the ensemble and the observation for each ensemble forecast/observation pair.

$$CRPS = \int_{-\infty}^{\infty} \left(F_{forecast} - F_{observation}\right)^2 dx \qquad \text{(Equation 4.8)}$$

with $F_{forecast}$ the CDF of the forecast, and $F_{observation}$ the CDF of the observation (a step function). The skill score compares the regime-based forecasts with the baseline forecasts, with 1 being a perfect score, 0 indicating no skill, and negative values indicating that the regime-based method is not an improvement over the baseline. The skill score is calculated by comparing the mean CRPS for each forecast type over the available record.

$$CRPSS_{method} = 1 - \frac{CRPS_{method,mean}}{CRPS_{baseline,mean}} \qquad \text{(Equation 4.9)}$$

4.3.4 Experimental Design

The experimental design has three steps. First, several possible choices for the regime definition are chosen by analyzing the streamflow and precipitation records. The correlation between upcoming streamflow and a regime definition is used as an indicator of a promising regime definition. Second, the quality and skill of the categorical precipitation forecasts are assessed using Accuracy and the Heidke Skill Score. This creates a context for evaluating results. Finally, multiple, different forecasts are created, comparing regimes, methods, and the case study sites. The regime and method combinations result in nine cases for each case study site (Table 4.2). These forecasts are assessed using the CRPSS. The results from this experimental design establish the chapter's conclusions.

Table 4.2 Regime definition and methods combine to create nine hydrologic forecast cases

<table>
<tr><th colspan="2"></th><th colspan="3">Regime Definition</th></tr>
<tr><th colspan="2"></th><th>NWS Forecast</th><th>Perfect Forecast</th><th>Retrospective</th></tr>
<tr><th rowspan="3">Hydrologic Forecast Creation Method</th><th>conditional ϕ and ε_t</th><td>case 1</td><td>case 2</td><td>case 3</td></tr>
<tr><th>conditional ϕ only</th><td>case 4</td><td>case 5</td><td>case 6</td></tr>
<tr><th>conditional ε_t only</th><td>case 7</td><td>case 8</td><td>case 9</td></tr>
</table>

4.4 Results

4.4.1 Correlation between streamflow and precipitation

Creating a regime-based forecast requires some method of defining the regime. This trial uses precipitation totals either prior to or concurrent with the desired streamflow forecast. The choice of regime definition discussed up to this point is chosen in the following analysis. Four candidate regime definitions are tested: total precipitation in the three months preceding the forecast, total precipitation for one month immediately preceding the forecast, total observed precipitation in the first month of the forecast time period, and total observed precipitation in the three months concurrent with the forecast. When the regime is based on preceding precipitation, as in the first two "retrospective" regime definitions above, the regime is known with certainty at the time of hydrologic forecast creation. When the regime is based on upcoming precipitation, the regime would be available as predicted categories only, though at this stage, the correlation is ested using observed precipitation values from the historic record. This test of correlations is a necessary first step in rationalizing the proposed method of the paper, because if the highest correlation were found to be between streamflow and prior precipitation for the case study sites, there would be no need to rely on precipitation forecasts.

Expectedly, three-month streamflow totals had the highest correlation with precipitation in those months, followed by one month of concurrent streamflow, three months of preceding streamflow, and one month of preceding streamflow. Three-month precipitation concurrent with the hydrologic forecast is chosen as the first experimental regime definition. To examine a rule that does not rely on a forecast quantity, three-month preceding precipitation is chosen as a possible alternative. (A perfect categorical forecast of three-month precipitation concurrent with the hydrologic forecast is chosen as a comparison case.) One rule per trial is used to define the regime, to avoid overly reducing the pool of data used in generating the streamflow forecasts. (If both preceding precipitation and forecast precipitation were used in the regime definition, the nine categories created would result in a very small number of ensemble members for the resulting hydrologic forecasts.) The different rules used to define the regime are therefore: 1) NWS forecasted tercile of precipitation concurrent with the forecast, 2) tercile of preceding three months' precipitation, and 3) a perfect categorical forecast of the precipitation tercile concurrent with the streamflow forecast.

Figure 4.2 presents monthly correlations at each site for each possible regime definition, and the average correlation over all months and sites for each possible regime definition. The streamflow total for a January data point is the sum of January, February, and March streamflow. The one-preceding-month regime definition sums the precipitation in December, and the three-preceding-month regime includes the precipitation totals for the previous October, November, and December.

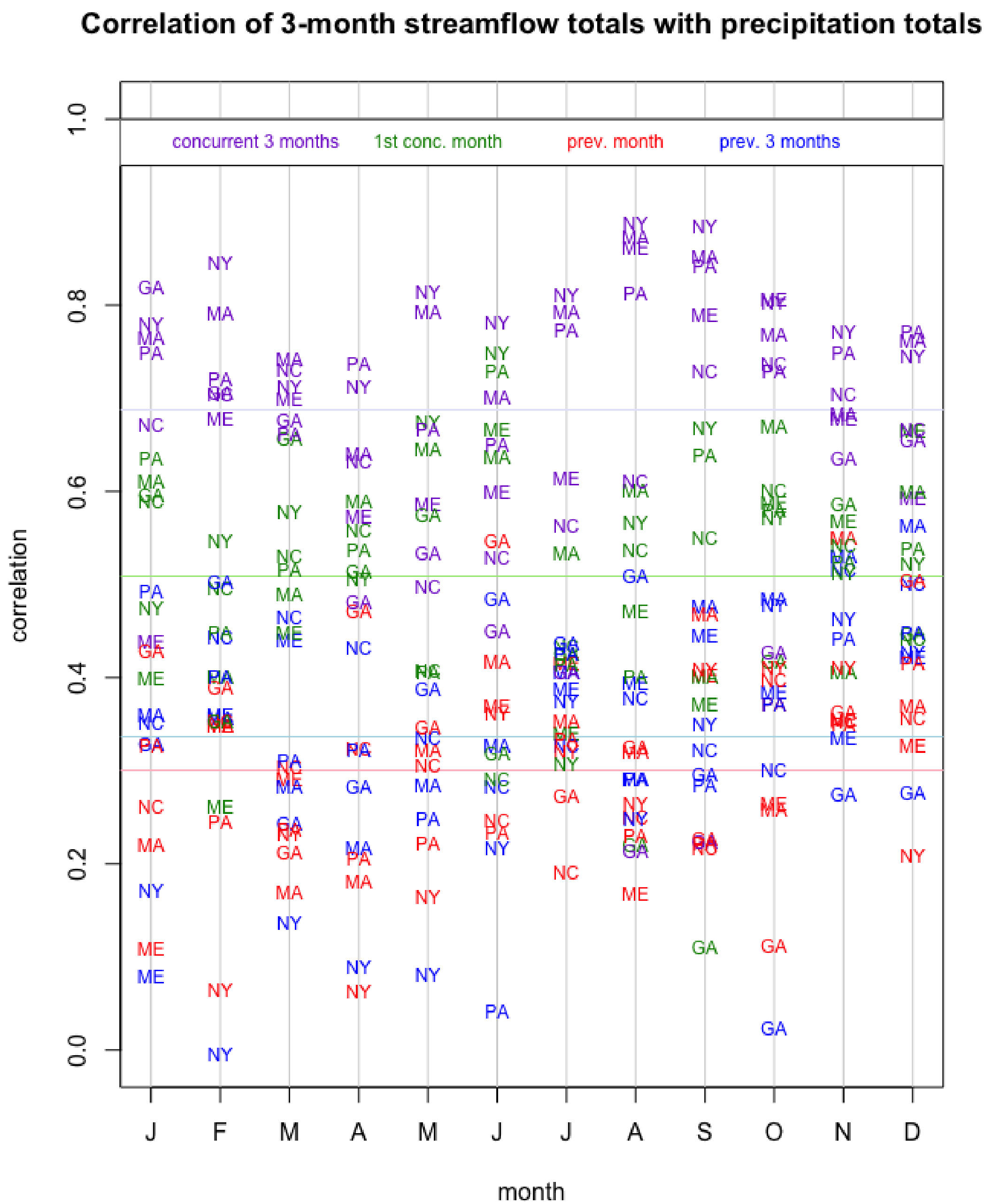

Figure 4.2 Monthly correlation of three-month streamflow totals to concurrent three-month precipitation (purple), first concurrent month precipitation (green), one previous month precipitation (red), and three previous months' precipitation (blue), for six sites designated by state abbreviation. Horizontal colored lines indicate averages over all sites and all months.

4.4.2 Heidke Skill Score for precipitation forecasts

The Heidke Skill Score compares a categorical forecast to a random guess. In general, forecasts had a small positive amount of skill, with the southeastern and mid-Atlantic performing better than the northeastern sites (Table 4.3). Three cases had a negative skill score: monthly forecasts for Maine (-0.001), monthly forecasts for Massachusetts (-0.003), and seasonal forecasts for New York (-0.009).

Table 4.3 Categorical precipitation forecasts' accuracy and skill

Site	Monthly		Seasonal	
	Accuracy	HSS	Accuracy	HSS
GA	0.40	0.094	0.39	0.079
NC	0.40	0.077	0.48	0.098
PA	0.38	0.055	0.55	0.125
NY	0.33	0.025	0.28	-0.009
MA	0.30	-0.003	0.41	0.059
ME	0.29	-0.002	0.46	0.017

4.4.3 Regime based hydrologic forecasts

Though the central method of the experiment is the utilization of NWS Seasonal Precipitation Forecasts in statistical streamflow forecasts, a comparison of several rules and methods for regime-based forecasts is necessary to determine whether there is information in the forecasts and whether regime based-forecasts are an overall improvement in predicting 90 day streamflow totals. The differences in hydrologic forecasts using three regime definitions differentiate whether it is more useful to use uncertain forecasts or known retrospective observations in defining the regime, "Regime Definition" in Tables 4.4 – 4.9. In addition, three variations incorporating the regime into statistical forecasts are compared to a baseline forecast and to each other to determine how the regime influences the process. In these variations, either the autoregressive parameter, the ensemble error terms, or both, are conditioned on the regime, "Conditional ϕ" and "Conditional ε_t" (Tables 4.4-4.9). In general, none show a substantial improvement over the baseline forecast.

Conditioning the autoregressive parameter on the regime is detrimental in most cases, even when a perfect forecast of the regime is used (forecast cases 1, 2, and 3 with conditional ensemble variance, and cases 4, 5 and 6 alone). Instead of wet and dry regimes having distinct patterns in their autocorrelation, the autocorrelation parameter behaves erratically in the conditional cases. For example, a plot of the unconditional and conditional AR(1) for the Maine site shows that the parameter does not show any distinct patterns (Figure 4.3). The smaller subset used in the conditional forecasts may hurt the model's ability to best predict the autocorrelation.

Table 4.4 Georgia hydrologic forecast skill

Forecast Case	Conditional ϕ	Conditional ε_t	Regime Definition	CRPSS
1	Y	Y	NWS	-0.09
2	Y	Y	perfect	0.04
3	Y	Y	retrospective	-0.39
4	Y	N	NWS	-0.04
5	Y	N	perfect	-0.02
6	Y	N	retrospective	-0.26
7	N	Y	NWS	-0.05
8	N	Y	perfect	0.07
9	N	Y	retrospective	-0.37

Table 4.5 North Carolina hydrologic forecast skill

Forecast Case	Conditional ϕ	Conditional ε_t	Regime Definition	CRPSS
1	Y	Y	NWS	0.04
2	Y	Y	perfect	0.07
3	Y	Y	retrospective	0.007
4	Y	N	NWS	0.04
5	Y	N	perfect	0.07
6	Y	N	retrospective	-0.05
7	N	Y	NWS	0.12
8	N	Y	perfect	0.13
9	N	Y	retrospective	0.03

Table 4.6 Pennsylvania hydrologic forecast skill

Forecast Case	Conditional ϕ	Conditional ε_t	Regime Definition	CRPSS
1	Y	Y	NWS	0.07
2	Y	Y	perfect	0.13
3	Y	Y	retrospective	0.10
4	Y	N	NWS	0.000
5	Y	N	perfect	0.12
6	Y	N	retrospective	0.07
7	N	Y	NWS	0.07
8	N	Y	perfect	0.13
9	N	Y	retrospective	0.14

Table 4.7 New York hydrologic forecast skill

Forecast Case	Conditional ϕ	Conditional ε_t	Regime Definition	CRPSS
1	Y	Y	NWS	-0.03
2	Y	Y	perfect	0.09
3	Y	Y	retrospective	-0.05
4	Y	N	NWS	0.007
5	Y	N	perfect	-0.12
6	Y	N	retrospective	-0.11
7	N	Y	NWS	-0.007
8	N	Y	perfect	0.25
9	N	Y	retrospective	0.01

Table 4.8 Massachusetts hydrologic forecast skill

Forecast Case	Conditional ϕ	Conditional ε_t	Regime Definition	CRPSS
1	Y	Y	NWS	-0.02
2	Y	Y	perfect	0.16
3	Y	Y	retrospective	-0.03
4	Y	N	NWS	-0.008
5	Y	N	perfect	-0.05
6	Y	N	retrospective	-0.07
7	N	Y	NWS	0.001
8	N	Y	perfect	0.24
9	N	Y	retrospective	-0.005

Table 4.9 Maine hydrologic forecast skill

Forecast Case	Conditional ϕ	Conditional ε_t	Regime Definition	CRPSS
1	Y	Y	NWS	-0.04
2	Y	Y	perfect	0.22
3	Y	N	NWS	-0.002
4	Y	N	perfect	0.004
5	N	Y	NWS	-0.04
6	N	Y	perfect	0.22

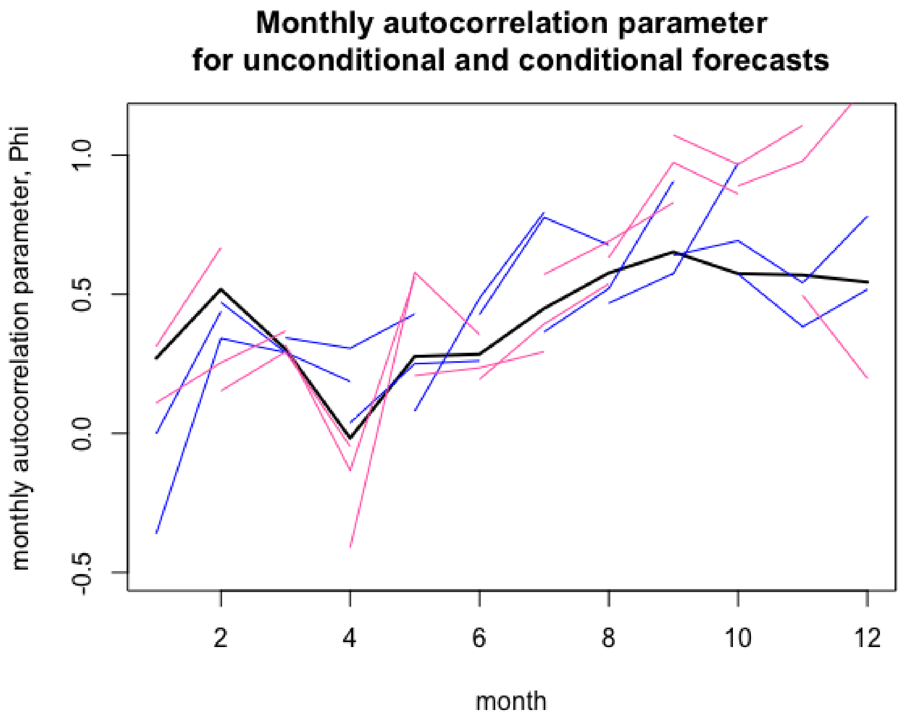

Figure 4.3 The monthly autocorrelation parameter fit to the full (unconditional) historic record, black, and to the wettest tercile of the historic record, blue, and the driest tercile of the historic record, pink, in each month. The parameters for wet and dry conditions are calibrated in a series of three-month blocks starting in each month.

Conditioning the ensemble variance, as a portion of the full record of error traces, based on the regime has the most influence over the forecast's skill (forecasts cases 7, 8, and 9, in tables above). When the regime is defined by perfect precipitation forecasts, the method is consistently an improvement over the baseline. Meanwhile, for the other two regime definitions, the method either improved or worsened the hydrologic forecasts.

Perfect forecasts performed best in almost every case; in most cases, there was no difference between the skills of forecasts defining the regime based on the NWS seasonal precipitation outlooks and forecasts defining the regime based on retrospective observations of precipitation. Similarly, there is no pattern between the relative correlation between precipitation and streamflow over these periods with the definition's effectiveness as a regime definition. Accuracy and the HSS for forecasts indicated very little information in the precipitation forecasts, and this lack of information was observed in the streamflow forecasts. Additionally, seasonal variations in both precipitation forecast skill and hydrologic forecast skill were not included in the analysis, and these may contribute to a lack of pattern in results.

4.5 Discussion

The cases and sites included in this experiment were chosen to answer several questions about streamflow forecasting on the east coast, including questions regarding the skill of precipitation forecasts, regional differences in efficacy, and methodology choices all in support of determining whether NWS precipitation forecasts could be used to improve statistical streamflow forecasts on the east coast. Differences in skill of the methods were discussed briefly in the results section with two key points. Artificially limiting the dataset in the conditional ϕ case, by calculating the autoregression parameter based on a regime-defined subset of years, resulted in poorer performance of the statistical forecasting method. Users of statistically-based hydrologic forecasts should be aware of the size of their dataset when testing any future method that deliberately limits the dataset. Meanwhile, disaggregating the ensemble variance was more successful in improving streamflow forecasts based on a regime when the predicted regime was correct, as in the perfect precipitation forecast cases. While the record as a whole is standardized and normalized to a mean of 0 in this method, the degree to which a historic year was wet or dry is preserved in the individual error traces, e.g. error traces from wet years will have average totals well above 0. It is these error traces that store the range, from wet to dry, of possibilities in the method used here. The location of the gage, categorized as southeastern, mid-atlantic, and northeastern, has a small effect on whether statistical streamflow forecasts result in improvement through utilizing a hydrologic regime. Southeastern and mid-Atlantic precipitation forecasts from NWS perform better than those from northeastern sites. Weather patterns originating from the southern Atlantic may be more predictable at this coarse seasonal scale than weather patterns coming across the continent. Lower correlations between streamflow and concurrent precipitation for several months in Georgia result in the streamflow forecasts for that site decreasing in skill regardless of the use of perfect forecasts. It is unlikely that water storage in snowpack is a driving factor in lower streamflow forecast skill in the northeast. If that were the case, a seasonal pattern showing lower correlation between streamflow and precipitation in winter or spring would be observed. Based on the results across all sites, for this regime-based method to perform better than baseline, both high correlations between precipitation and streamflow, and precipitation forecast skill above zero are necessary.

Regardless of location, a major impediment to improved streamflow forecasting is the uncertainty in future climactic conditions. Lacking good precipitation forecasts, defining the hydrologic regime based on observations or forecasts that are currently available did not result in improved hydrologic forecasts. This likely reflects the reality on the east coast, where hydrologic forecast uncertainty is dominated by the uncertainty in upcoming conditions (Shukla & Lettenmaier, 2011) and there is a lack of actionable skill in precipitation forecasts. This is, in itself, valuable information in predicting an upper limit of hydrologic forecast skill. Combined with the methods in Chapter 3, this allows a utility to decide whether or not to invest in improving the forecasts themselves or to put resources towards other methods of improved operations or infrastructure. Currently, formally incorporating a regime into long-term forecast generation for east coast sites is not recommended, though methods other than the one tested here may show better results. Possible improvements in the global observation network may improve longer term precipitation forecasts over time (Lettenmaier, 2017), and shorter term precipitation forecasts and the resulting hydrologic forecasts for flood warnings are still a valuable resource in water resource management (Hapuarachchi et al., 2011).

High quality three-month hydrologic forecast can be used in seasonal drought planning combined with other indicators, such as demand and current storage, the latter of which is a strong indicator of potential shortages (Booras et al., 2017). With perfect categorical precipitation forecasts, using a subset of ensemble variance consistently improved upon the baseline hydrologic forecasts. If precipitation forecast information is good, an even simpler method may be best for use in water supply system operation. If the hydrologic forecast is created using the baseline method from this experiment with no regime-based conditioning, the decision-maker can choose to focus on some portion of that ensemble based on predicted precipitation while still having the full record of historic variance available. Changes in operations utilizing the level of forecasts that are available, may also result in improved operations for water supply, but managers should be conscious of the forecast's skill and limitations. In terms of resource allocation for a municipal water supply, a potential approach may range from clearly recognizing the limits of current forecasts and investing in more sophisticated and locally generated forecasting methods.

4.6 Conclusion

In literature and in practice, hydrologic forecasting for water supply management on the east coast is a more difficult task than for the western United States and a barrier to more widespread implementation. In this research, a method of improving statistical ensemble streamflow forecasts based on defining a hydrologic regime is tested for three-month flow totals for six sites on the east coast, with their intended use being for water supply management. Based on these experiments, there is insufficient skill in NWS seasonal precipitation forecasts and limited correlation with preceding precipitation for a regime-based approach to be an improvement in practice. Knowing the initial basin conditions is helpful, but uncertainty associated with longer-term precipitation forecasts results in the forecasts not having a positive impact on predicting streamflows. The research illustrates that with better precipitation forecasts, utilizing a subset of the baseline statistical forecasts would be the most successful method of incorporating a categorical regime. While the utilization of forecasts in water supply management can be beneficial, improving longer-lead-time forecasts themselves is a difficult task.

CHAPTER 5
CONCLUSION

The research in this dissertation, summarized and included in full, examines the science and vulnerabilities of water resources in the northeast, compares climate change adaptations for water quality and quantity management for the New York City Water Supply System, quantifies the value of small incremental improvements in hydrologic forecasting for use in drought management for the City of Baltimore Water Supply System, and searches for those incremental improvements by modifying an existing statistical ensemble forecasting method.

Chapter 1 details three studies performed prior to the research that comprises the remaining three technical chapters. First, turbidity is modeled in the New York City Water Supply System using stochastic weather sequences, a hydrologic model, and a system model of the New York City Water Supply System. Incremental changes in precipitation and temperature (spanning those projected by CMIP5) are shown to have varying effects on the in-reservoir turbidity depending on the season. Second, different statistical methods for generating weather sequences (to be used in climate change planning or extreme events management) are tested. A lag-1 Markov chain process coupled with either skewed normal, exponential, or KNN-resampling distributions are determined to be the most parsimonious and best approach to capture the statistics of extreme precipitation events. Finally, the New York City Water Supply system is included in a study of climate change outcomes for northeastern cities. On a monthly time-step, with a large amount of reservoir drawdown, the NYCWSS will remain reliable (in terms of water quality) across the range of current climate projections. These experiments provide the basis and the inspiration for the remaining chapters.

Chapter 2 investigates adaptations for the New York City Water Supply system in more detail, modeling both water quantity and turbidity on a daily time-step for climate projections through the end of the century and testing possible adaptations. A stochastic weather generator, hydrologic model, and system model incorporating adaptations are used in series, with climate change projections for temperature and precipitation from CMIP5. Results illustrate how reducing demand and managing storage and releases based on hydrologic forecasting reduce the frequency of drought warnings and emergencies and improve system reliability. In addition operations that limit turbidity propagation through the system and improvements to the Catskill Aqueduct (to lower the minimum flow under conditions with high turbidity) result in lower turbidity loads and a reduction in emergency Alum use. The city's existing infrastructure, combined with upcoming projects to improve the system and reduce demand, provide a promising outlook for the system's future reliability.

Chapter 3 expands on one of the adaptations included in Chapter 2: the use of hydrologic forecasting at three-month lead-times. Prompted by the need to quantify improved system performance with forecasting (cited as one barrier to the implementation of hydrologic forecasting in water resource engineering), a method is applied that creates forecasts with incremental differences in quality to determine the effect on system performance of improved forecasts. The method created a set of 90-day outlook ensemble hydrologic forecasts issued every 5 days with varying skill according to the Probability Score. Each ensemble forecasts (representing a particular forecast quality) is evaluated using a water supply systems model. Each trace was considered equally likely, and the expected value (a composite metric of economic costs and reservoir storage) of each possible action is calculated. The best option was implemented until the next forecast is issued. For a case study site of the City of Baltimore Water Supply System, system performance improved linearly with forecast quality. Because of an imbalance between water demand, system capacity, and inflows, even a perfect forecast could not prevent the necessity of restrictions, pumping, or reservoir drawdown over the 2001-2002 drought. Nonetheless, the method can play a useful part in a water resource system's long term planning.

Finally, Chapter 4 examines a possible method for improving statistically-based streamflow forecasts on the eastern US coast by defining a hydrologic regime based on forecast or observed precipitation. Several methods for both defining the regime (as wetter, dryer, or average according to the historic precipitation terciles used in the National Weather Service's Seasonal Precipitation Outlook maps) and creating the forecasts (an AR(1) monthly process for streamflow with daily data and variability preserved from the historic record) were tested and compared. Overall, improvements in the hydrologic forecasts were limited by the skill of the precipitation forecasts at a three-month lead-time. A regime-based forecast was not a significant improvement over the baseline unconditional AR(1) process. On the east coast, hydrologic forecast uncertainty is dominated by uncertainty in the upcoming conditions, and the results in Chapter 4 reinforced this.

Future work following from all three chapters should include the application of the methods herein to other sites or systems on the east coast or elsewhere. Comparisons with systems from across the country would provide further information on the unique challenges and opportunities in water resources engineering on the east coast, compare large and small systems, and further refine the methods themselves.

The results in Chapter 4, in particular, would benefit from a comparison with west coast sites for this method and with further variations on the method itself for all sites. A comparison with west coast sites would explore differences in the ranges and seasonality of precipitation and streamflow at each site, the skill of the precipitation forecasts, and at the effectiveness of the method itself especially at sites with higher precipitation forecast skill. This would provide broader insights into the effectiveness of regime-based AR(1) forecasts and perhaps identify a threshold of forecast skill or a type of hydrology where this or a similar method becomes a viable improvement over baseline forecasts. Several variations on the process are also possible, including utilizing individual precipitation forecast traces of the ensemble forecast used to make the Seasonal Outlook maps if that data becomes available. If this results in an improvement, the added complexity in the process may be a worthwhile change to the method.

From Chapter 3, the method for quantifying system performance improvement can also include the change in system performance as a function of operation or infrastructure changes, adding a third dimension to the results and providing insight to where improvement stands to be gained. For example, the operating policies in Booras et al. (2017) can be adapted to utilize ensemble forecasts, and a more thorough analysis of the system and options can be completed. Similarly, the comparison of infrastructure and operating policies for the New York City Water Supply System in Chapter 2 can be expanded with incremental forecast quality from the methods utilized in Chapter 3. The methods in Chapters 2 and 3 combined can create a full picture of options and outlooks for infrastructure, forecasting, and the effects of climate change on the system. All of the results and methods can be updated as new datasets and climate change projections become available.

Through system-specific studies and wise utilization of the existing science, researchers and water managers will continue to pursue best management practices to provide reliable and high quality drinking water to the cities of the northeast. This dissertation is a small part of that large effort.

BIBILIOGRAPHY

Acharya, N., Frei, A., Chen, J., DeCristofaro, L., & Owens, E. M. (2017). Evaluating Stochastic Precipitation Generators for Climate Change Impact Studies of New York City's Primary Water Supply. Journal of Hydrometeorology, 18(3), 879-896.

Alemu, E. T., Palmer, R. N., Polebitski, A., & Meaker, B. (2010). Decision support system for optimizing reservoir operations using ensemble streamflow predictions. Journal of Water Resources Planning and Management, 137(1), 72-82.

Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., & Lettenmaier, D. P. (2016). Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. Water Resources Research, 52(6), 4209-4225.

Arsenault, R., Brissette, F., Malo, J. S., Minville, M., & Leconte, R. (2013). Structural and non-structural climate change adaptation strategies for the Péribonka water resource system. Water resources management, 27(7), 2075-2087.

Bates, B.C., Z.W. Kundzewicz, S. Wu and J.P. Palutikof, Eds. (2008). Climate Change and Water. IPCC Secretariat, Geneva, 210 pp.

Bennett, J. C., Robertson, D. E., Shrestha, D. L., Wang, Q. J., Enever, D., Hapuarachchi, P., & Tuteja, N. K. (2014a). A System for Continuous Hydrological Ensemble Forecasting (SCHEF) to lead times of 9days. Journal of hydrology, 519, 2832-2846.

Bennett, J. C., Wang, Q. J., Pokhrel, P., & Robertson, D. E. (2014b). The challenge of forecasting high streamflows 1–3 months in advance with lagged climate indices in southeast Australia. Natural Hazards and Earth System Sciences, 14(2), 219-233.

Block, P. J., Souza Filho, F. A., Sun, L., & Kwon, H. H. (2009). A streamflow forecasting framework using multiple climate and hydrological models. JAWRA Journal of the American Water Resources Association, 45(4), 828-843.

Booras, K., McIntyre, A. R., Joshua Weiss, W., Howells, C., & Palmer, R. N. (2017). Incorporating Streamflow Forecasts with Aggregate Drought Indices for the Management of Water Supply. Journal of Water Resources Planning and Management, 144(1), 04017078.

Borgomeo, E. et al., 2014. Risk-based water resources planning: Incorporating probabilistic nonstationary climate uncertainties. Water Resour. Res., 50(8): 6850-6873. DOI:10.1002/2014wr015558

Brown, C., Werick, W., Leger, W., and Fay, D. (2011). "A Decision-Analytic Approach to Managing Climate Risks: Application to the Upper Great Lakes." JAWRA Journal of the American Water Resources Association, 47(3), 524–534.

Brown, J. D., Demargne, J., Seo, D. J., & Liu, Y. (2010). The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. Environmental Modelling & Software, 25(7), 854-872.

Brown, J. D., He, M., Regonda, S., Wu, L., Lee, H., & Seo, D. J. (2014). Verification of temperature, precipitation, and streamflow forecasts from the NOAA/NWS Hydrologic Ensemble Forecast Service (HEFS): 2. Streamflow verification. Journal of hydrology, 519, 2847-2868.

Cardona, O.D., M.K. van Aalst, J. Birkmann, M. Fordham, G. McGregor, R. Perez, R.S. Pulwarty, E.L.F. Schipper, and B.T. Sinh, 2012: Determinants of risk: exposure and vulnerability. In: Managing the Risks of Extreme Events and Disasters to Advance

Cisneros, J., B.E., T. Oki, N.W. Arnell, G. Benito, J.G. Cogley, P. Döll, T. Jiang, and S.S. Mwakalila. (2014). Freshwater resources. In: 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 University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 229-269.

City of Baltimore Department of Public Works. (2002). "Baltimore City Core Water Supply System Operating Plan."

Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 65-108.

Coelho, C. A., & Costa, S. M. (2010). Challenges for integrating seasonal climate forecasts in user applications. Current Opinion in Environmental Sustainability, 2(5), 317-325.

Cole, J. E., & Cook, E. R. (1998). The changing relationship between ENSO variability and moisture balance in the continental United States. Geophysical Research Letters, 25(24), 4529-4532.

Culley, S., Noble, S., Yates, A., Timbs, M., Westra, S., Maier, H. R., ... & Castelletti, A. (2016). A bottom-up approach to identifying the maximum operational adaptive capacity of water resource systems to a changing climate. Water Resources Research, 52(9), 6751-6768.

Day, G. N. (1985). Extended streamflow forecasting using NWSRFS. Journal of Water Resources Planning and Management, 111(2), 157-170.

DeCristofaro, L., and Palmer, R.N. (in review). Win-win Strategies for Managing Climate Change Extremes: A Case Study of the NYCWSS. Journal of Water Resource Planning and Management.

Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., ... & Schaake, J. (2014). The science of NOAA's operational hydrologic ensemble forecast service. Bulletin of the American Meteorological Society, 95(1), 79-98.

Eum, H. I., & Simonovic, S. P. (2010). Integrated reservoir management system for adaptation to climate change: the Nakdong River Basin in Korea. Water Resources Management, 24(13), 3397-3417.

Field, C. B. (Ed.). (2012). Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change. Cambridge University Press.

Franz, K. J., Hartmann, H. C., Sorooshian, S., & Bales, R. (2003). Verification of National Weather Service ensemble streamflow predictions for water supply forecasting in the Colorado River basin. Journal of Hydrometeorology, 4(6), 1105-1118.

Frei, A., Kunkel, K. E., & Matonse, A. (2015). The seasonal nature of extreme hydrological events in the Northeastern United States. Journal of Hydrometeorology, 16(5), 2065-2085.

Georgakakos, A., P. Fleming, M. Dettinger, C. Peters-Lidard, Terese (T.C.) Richmond, K. Reckhow, K. White, and D. Yates, 2014: Ch. 3: Water Resources. Climate Change Impacts in the United States: The Third National Climate Assessment, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 69-112. doi:10.7930/ J0G44N6T.

Georgakakos, K., and N. Graham (2008), Potential benefits of seasonal inflow prediction uncertainty for reservoir release decisions, J. Appl. Meteorol. Climatol., 47(5), 1297–1321.

Giuliani, M., & Castelletti, A. (2016). Is robustness really robust? How different definitions of robustness impact decision-making under climate change. Climatic Change, 1-16.

Glantz, M. H. (1982). Consequences and responsibilities in drought forecasting: the case of Yakima, 1977. Water Resources Research, 18(1), 3-13.

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & ter Maat, J. (2013). Dynamic adaptive policy pathways: a method for crafting robust decisions for a deeply uncertain world. Global environmental change, 23(2), 485-498.

Hamlet, A. F., Huppert, D., & Lettenmaier, D. P. (2002). Economic value of long-lead streamflow forecasts for Columbia River hydropower. Journal of Water Resources Planning and Management, 128(2), 91-101.

Hamlet, A. F., & Lettenmaier, D. P. (1999). Columbia River streamflow forecasting based on ENSO and PDO climate signals. Journal of water resources planning and management, 125(6), 333-341.

Hamlet, A. F., & Lettenmaier, D. P. (2000). Long-range climate forecasting and its use for water management in the Pacific Northwest region of North America. Journal of Hydroinformatics, 2(3), 163-182.

Hapuarachchi, H. A. P., Wang, Q. J., & Pagano, T. C. (2011). A review of advances in flash flood forecasting. Hydrological Processes, 25(18), 2771-2784.

Herman, J. D., Reed, P. M., Zeff, H. B., & Characklis, G. W. (2015). How should robustness be defined for water systems planning under change?. Journal of Water Resources Planning and Management, 141(10), 04015012.

Hirsch, R. M. (1979). Synthetic hydrology and water supply reliability. Water Resources Research, 15(6).

Hirsch, RM (1981). Stochastic hydrologic model for drought management. J. Water Resour. Plann. and Manage. 107(2), 303-313.

Hine, D., and Hall, J. W. (2010). “Information gap analysis of flood model uncertainties and regional frequency analysis.” Water Resources Research, 46 (1).

Horton, R., Bader, D., Kushnir, Y., Little, C., Blake, R., & Rosenzweig, C. (2015). New York City Panel on Climate Change 2015 Report Chapter 1: Climate Observations and Projections. Annals of the New York Academy of Sciences, 1336(1), 18-35.

Horton, R., G. Yohe, W. Easterling, R. Kates, M. Ruth, E. Sussman, A. Whelchel, D. Wolfe, and F. Lipschultz, 2014: Ch. 16: Northeast. Climate Change Impacts in the United States: The Third National Climate Assessment, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 16-1-nn.

Intergovernmental Panel on Climate Change (2012). “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.” [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp.

Jiménez Cisneros, B.E., T. Oki, N.W. Arnell, G. Benito, J.G. Cogley, P. Döll, T. Jiang, and S.S. Mwakalila, 2014: Freshwater resources. In: 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 (ed)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 229-269.

Jolliffe, I. T., & Stephenson, D. B. (Eds.). (2012). Forecast verification: a practitioner's guide in atmospheric science. John Wiley & Sons.

Karmalkar, A. V., & Bradley, R. S. (2017). Consequences of Global Warming of 1.5 C and 2 C for Regional Temperature and Precipitation Changes in the Contiguous United States. PloS one, 12(1), e0168697.

Katz, R. W., & Murphy, A. H. (1997). Economic value of weather and climate forecasts.

Kunkel, K. E., Karl, T. R., Brooks, H., Kossin, J., Lawrimore, J.H., Arndt, D., Bosart, L., Changnon, D., Cutter, S.L., Doesken, N. & Emanuel, K. (2013). Monitoring and understanding trends in extreme storms: State of knowledge. Bulletin of the American Meteorological Society, 94(4), 499-514.

Leathers, D. J., Yarnal, B., & Palecki, M. A. (1991). The Pacific/North American teleconnection pattern and United States climate. Part I: Regional temperature and precipitation associations. Journal of Climate, 4(5), 517-528.

Lettenmaier, D. P. (2017). Observational breakthroughs lead the way to improved hydrological predictions. Water Resources Research, 53(4), 2591-2597.

Lim, B. (Ed.). (2005). Adaptation policy frameworks for climate change: developing strategies, policies and measures. Cambridge: Cambridge University Press.

Matonse, A. H., Pierson, D. C., Frei, A., Zion, M. S., Anandhi, A., Schneiderman, E., & Wright, B. (2013). Investigating the impact of climate change on New York City's primary water supply. Climatic Change, 116(3-4), 437-456.

Maupin, M.A., Kenny, J.F., Hutson, S.S., Lovelace, J.K., Barber, N.L., and Linsey, K.S.. (2014). Estimated use of water in the United States in 2010: U.S. Geological Survey Circular 1405. Available at http://dx.doi.org/10.3133/cir1405.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen (2002). "A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States." J. Climate., 15(22), 3237–3251Maurer, E. P., & Lettenmaier, D. P. (2004). Potential effects of long-lead hydrologic predictability on Missouri River main-stem reservoirs. Journal of Climate, 17(1), 174-186.

McIntyre, A., Booras, K., Weiss, W.J., Hoffman, k., Howells, C., and Palmer, R.N. (2017). Analyzing Streamflow Forecasts in the Context of System Performance: A Case Study of the City of Baltimore Water Supply. Journal of Water Resources Planning and Management.

Means, E., Laugier, M., Daw, J., Kaatz, L., & Waage, M. (2010). Decision support planning methods: Incorporating climate change uncertainties into water planning. Water Utility Climate Alliance White Paper.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Zbigniew, W., Lettenmaier, D. P., and Stouffer, R. J. (2008). "Stationarity Is Dead : Whither Water Management ?" Science, 319(February), 573–574.

Mimura, N., R.S. Pulwarty, D.M. Duc, I. Elshinnawy, M.H. Redsteer, H.Q. Huang, J.N. Nkem, and R.A. Sanchez Rodriguez, 2014: Adaptation planning and implementation. In: 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 University Press, Cambridge, United Kingdom and New York, NY, USA, pp 69-898.

Mishalani, N. R., & Palmer, R. N. (1988). Forecast uncertainty in water supply reservoir operation. JAWRA Journal of the American Water Resources Association, 24(6), 1237-1245.

Mukundan, R., Pierson, D. C., Schneiderman, E. M., O'Donnell, D. M., Pradhanang, S. M., Zion, M. S., and Matonse, a. H. (2013). "Factors affecting storm event turbidity in a New York City water supply stream." Catena, 107, 80–88.

New York City Department of Environmental Protection. (2010). "New York City's Operations Support Tool (OST) White Paper Prepared for The Delaware River Basin Supreme Court Decree Parties." Available at http://water.usgs.gov/osw/odrm/documents/OST_White_Paper.pdf

New York City Department of Environmental Protection/ New York State Department of Environmental Conservation (2011). "Interim Ashokan Release Protocol" Available at www.nyc.gov/html/dep/pdf/ashokan_interim_release_protocol_from_dec_10-18-11.pdf

New York City Department of Environmental Protection (2012). Drought Management and Contingency Plan. Available at http://www.nyc.gov/html/dep/pdf/droughtp.pdf

New York City Department of Environmental Protection. (2014) Draft Scope for the Modification of the Catalum SPDES Permit, April 9, 2014. CEQR No. 14DEC001. Available at http://www.dec.ny.gov/docs/water_pdf/catalumdraftscopeeis.pdf

New York City Department of Environmental Protection. (2015). Water for the Future: Upstate Water Supply Resiliency, Final Scope of Work. Available at http://www.nyc.gov/html/dep/pdf/reviews/upstate_water_supply_resiliency/wff-upstate-water-supply-resiliency-final-scope-of-work-091615.pdf

New York City Department of Environmental Protection. (2017). History of Drought and Water Consumption. Available at http://www.nyc.gov/html/dep/html/drinking_water/droughthist.shtml

New York State Department of Environmental Conservation and New York City Department of Environmental Protection. (2011). Interim Ashokan Release Protocol. Available at http://www.nyc.gov/html/dep/pdf/ashokan_interim_release_protocol_from_dec_10-18-11.pdf

Pagano, T. C., Wood, A. W., Ramos, M. H., Cloke, H. L., Pappenberger, F., Clark, M. P., ... & Verkade, J. S. (2014). Challenges of operational river forecasting. Journal of Hydrometeorology, 15(4), 1692-1707.

Payne, J. T., Wood, A. W., Hamlet, A. F., Palmer, R. N., & Lettenmaier, D. P. (2004). Mitigating the effects of climate change on the water resources of the Columbia River basin. Climatic change, 62(1-3), 233-256.

Peterson, T. C., Heim Jr, R. R., Hirsch, R., Kaiser, D. P., Brooks, H., Diffenbaugh, N. S., Dole, R.M., Giovannettone, J.P., Guirguis, K., Karl, T.R. and Katz, R.W. (2013). Monitoring and understanding changes in heat waves, cold waves, floods, and droughts in the United States: state of knowledge. Bulletin of the American Meteorological Society, 94(6), 821-834.

Porter, J. H., Matonse, A. H., & Frei, A. (2015). The New York City Operations Support Tool (OST): Managing Water for Millions of People in an Era of Changing Climate and Extreme Hydrological Events. Journal of Extreme Events, 2(02), 1550008.

Rayner, S., Lach, D., & Ingram, H. (2005). Weather forecasts are for wimps: why water resource managers do not use climate forecasts. Climatic Change, 69(2), 197-227.

Reimer, Muegge, & Associates, Inc. 2000. Baltimore City “Core” Water Supply System Operating Plan, prepared for the City of Baltimore, Department of Public Works, Bureau of Water & Wastewater.

Ritchie, J. W., Zammit, C., & Beal, D. (2004). Can seasonal climate forecasting assist in catchment water management decision-making?: A case study of the Border Rivers catchment in Australia. Agriculture, ecosystems & environment, 104(3), 553-565.

Romero-Lankao, P., J.B. Smith, D.J. Davidson, N.S. Diffenbaugh, P.L. Kinney, P. Kirshen, P. Kovacs, and L. Villers Ruiz, 2014: North America. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, 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 University Press, Cambridge, UK, and New York, NY, USA, pp. 439-1498.

Rosenberg, E. A., Wood, A. W., & Steinemann, A. C. (2011). Statistical applications of physically based hydrologic models to seasonal streamflow forecasts. Water Resources Research, 47(3).

Rossi, N., DeCristofaro, L., Steinschneider, S., Brown, C., & Palmer, R. (2016). Potential Impacts of Changes in Climate on Turbidity in New York City’s Ashokan Reservoir. Journal of Water Resources Planning and Management, 04015066.

Rossi, N. (2014). “Potential Impacts of Changes in Climate on Water Quality in New York City’s Ashokan.” *Environmental and Water Resources Engineering Masters Projects*. 61. Retrieved from https://scholarworks.umass.edu/cee_ewre/61

Sankarasubramanian, A., Lall, U., Devineni, N., & Espinueva, S. (2009). The role of monthly updated climate forecasts in improving intraseasonal water allocation. Journal of Applied Meteorology and Climatology, 48(7), 1464-1482.

Sankarasubramanian, A., U. Lall, F. A. Souza, and A. Sharma (2009b), Improved water allocation utilizing probabilistic climate forecasts: Short-term water contracts in a risk management framework, Water Resour. Res., 45.

Schaake, J. C., Hamill, T. M., Buizza, R., & Clark, M. (2007). HEPEX: the hydrological ensemble prediction experiment. Bulletin of the American Meteorological Society, 88(10), 1541-1547.

Schepen, A., Wang, Q. J., & Robertson, D. E. (2012). Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall. Journal of Geophysical Research: Atmospheres, 117(D20).

Schneiderman, E. M., Pierson, D. C., Lounsbury, D. G., and Zion, M. S. (2002). "Modeling the Hydrochemistry of the Cannonsville Watershed with Generalized Watershed Loading Functions (GWLF)." Journal of The American Water Resources Association, 38(5), 1323–1347.

Schneiderman, E. M., Steenhuis, T. S., Thongs, D. J., Easton, Z. M., Zion, M. S., Neal, A. L., and Walter, M. T. (2007). "Incorporating variable source area hydrology into a curve-number-based watershed model." Hydrological Processes, 21(25), 3420–3430.

Seager, R., Pederson, N., Kushnir, Y., Nakamura, J., & Jurburg, S. (2012). The 1960s drought and the subsequent shift to a wetter climate in the Catskill mountains region of the New York city Watershed. Journal of Climate, 25(19), 6721-6742.

Shukla, S., & Lettenmaier, D. P. (2011). Seasonal hydrologic prediction in the United States: understanding the role of initial hydrologic conditions and seasonal climate forecast skill. Hydrology and Earth System Sciences, 15(11), 3529-3538.

Steinschneider, S., and Brown, C. (2013). "A semiparametric multivariate, multisite weather generator with low-frequency variability for use in climate risk assessments." Water Resources Research, 49(11), 7205–7220.

Steinschneider, S., & Brown, C. (2011). Influences of North Atlantic climate variability on low-flows in the Connecticut River Basin. Journal of hydrology, 409(1-2), 212-224.

Stratus Consulting and Denver Water. 2015. Embracing Uncertainty: A Case Study Examination of How Climate Change is Shifting Water Utility Planning. Prepared for the Water Utility Climate Alliance (WUCA), the American Water Works Association (AWWA), the Water Research Foundation (WRF), and the Association of Metropolitan Water Agencies (AMWA) by Stratus Consulting Inc., Boulder, CO (Karen Raucher and Robert Raucher) and Denver Water, Denver, CO (Laurna Kaatz). May 12.

T. C., Wood, A. W., Ramos, M. H., Cloke, H. L., Pappenberger, F., Clark, M. P., ... & Verkade, J. S. (2014). Challenges of operational river forecasting. Journal of Hydrometeorology, 15(4), 1692-1707.

Vogel, R. M., Lane, M., Ravindiran, R. S., & Kirshen, P. (1999). Storage reservoir behavior in the United States. Journal of Water Resources Planning and Management, 125(5), 245-254.

Walter, M.T., M.F. Walter, E.S. Brooks, T.S. Steenhuis, J. Boll, and K. Weiler (2000). "Hydrologically sensitive areas: Variable source area hydrology implications for water quality risk assessment," Journal of Soil and Water Conservation, 55(3), 277-284.

Wang, H., & Fu, X. (2014). Utilization of climate information and soil moisture estimates to provide monthly and sub-monthly streamflow forecasts. International Journal of Climatology, 34(13), 3515-3527.

Whateley, S., DeCristofaro, L., Pica, J., Palmer, R., and Brown, C. (In Preparation). Climate stress testing Northeast water supply: an assessment of vulnerabilities and climate risk exposure.

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

Wilson, L.J., Burrows, W.R. and Lanzinger, A. (1999) A strategy for verification of weather element forecasts from an ensemble prediction system. Monthly Weather Review, 127, 956-970.

Zhao, G., & Davison, M. (2009). Valuing hydrological forecasts for a pumped storage assisted hydro facility. Journal of hydrology, 373(3), 453-462.