## ABSTRACT

SKARBEK, MICHAEL. Fisher Information Methods for Detecting Shifting Regimes in Water Supply Systems. (Under the direction of Dr. Emily Berglund.)

Population growth and drying climates may push dynamic water supply systems to states of continuous water shortages. Fisher information is a statistical method that quantifies the resilience of a system and can be used to identify conditions that cause system state changes. This research explores methods for calculating Fisher information to identify critical transitions in time series data. The first method that is explored in this research calculates Fisher information using the amplitude of the probability density function of states and a discrete form of the Fisher information. The second method method uses a kernel density function to approximate the probability density function of states. The two methods are applied to calculate Fisher information and compared to the theoretical value for a test data set. The two methods are applied to detect critical transitions among states of water shortages for two water supply case studies. In the first case study, Fisher information is applied for historical data describing storage in the Cantareira Reservoir System, which serves São Paulo, Brazil, and experienced a regime shift during a significant drought that began in 2011. Fisher information is interpreted to detect regime shifts, which are accurately identified to match recorded regime shifts in the Cantareira System. In the second case study, Fisher information is used to analyze the resilience of Falls Lake, which provides water supply for Raleigh, North Carolina. An agent-based modeling framework is developed to simulate a population of residential consumers as agents and their interactions with the reservoir system. The interplay of water demands and supply is simulated based on population growth, changes in housing density, diffusion of water-efficient appliances, reservoir operations, and water use restrictions. The agent-based modeling framework is applied for regional population growth, which is projected using historic growth rates. A stochastic reconstruction framework is used to simulate changes in precipitation, reservoir inflows, and evapotranspiration over a 50-year projected period. Fisher information is applied to model output to detect regime shifts in water storage for a set of selected climate change scenarios.

Fisher Information Methods for Detecting Shifting Regimes in Water Supply Systems

by
Michael Skarbek

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Civil Engineering

Raleigh, North Carolina

2019

APPROVED BY:

| | |
|---|---|
| Dr. Gnanamanikam Mahinthakumar | Dr. John Baugh |

Dr. Emily Berglund
Chair of Advisory Committee

## DEDICATION

To my parentals.

## BIOGRAPHY

I was born and raised in Danville, VA, and graduated with Bachelors of Science in Chemistry from Virginia Tech in 2010. I moved to North Carolina to work in the pharmaceutical industry for several years before returning to school to pursue Civil Engineering. After graduating, I will be applying my computing skills at Red Hat.

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Dr. Emily Berglund for all her mentorship and encouragement while pursuing a Masters. Without her support, I never could have imagined finishing this work.

I would like to thank Dr. John Baugh and Dr. Gnanamanikam Mahinthakumar for their willingness to serve on my committee and for their courses that pushed my interests in computing and systems.

Lastly, I would like to thank my parents, John and Kathy Skarbek, and Michelle for their unconditional love and support while I pursue my crazy career path.

## TABLE OF CONTENTS

**LIST OF TABLES** . . . . . **vii**

**LIST OF FIGURES** . . . . . **viii**

**1 Introduction** . . . . . **1**

**2 Background** . . . . . **4**
2.1 Water Supply Resilience . . . . . 4
2.2 Detecting Regime Shifts using Fisher Information . . . . . 5
2.2.1 Amplitude-based Discrete Method . . . . . 6
2.2.2 Kernel Density Estimation Method . . . . . 6

**3 Methodology** . . . . . **8**
3.1 Agent Based Modeling Framework . . . . . 8
3.1.1 Household Agents . . . . . 8
3.1.2 Reservoir Model . . . . . 10
3.2 Stochastic Reconstruction Framework . . . . . 10
3.2.1 Shift Factor . . . . . 10
3.2.2 Quantile Mapping . . . . . 11
3.2.3 Copula Functions . . . . . 12
3.3 Methods for Calculating Fisher Information . . . . . 13
3.3.1 Amplitude-based Discrete Method . . . . . 13
3.3.2 Kernel Density Estimator Method . . . . . 15
3.3.3 Temporal Fisher Information . . . . . 16

**4 Calculating Fisher Information for Normal Distributions** . . . . . **17**
4.1 Description of Experiments . . . . . 17
4.2 Amplitude-based Discrete Method Results . . . . . 18
4.3 Kernel Density Estimator Method Results . . . . . 21

**5 Calculating Fisher Information for the Cantareira Reservoir System, São Paulo, Brazil** . . **22**
5.1 Case Study Description . . . . . 22
5.2 Results . . . . . 23

**6 Applying the Sociotechnical Water Supply Framework for Falls Lake, Raleigh, North Carolina** . . . . . **29**
6.1 Case Study Description . . . . . 29
6.2 Climate Input Data . . . . . 29
6.3 Stochastic Reconstruction Framework Results . . . . . 31
6.4 Agent-based Modeling Framework Results . . . . . 33
6.5 Fisher Information Results . . . . . 35
6.5.1 Fisher Information for Shift Factor = 0.2 . . . . . 35
6.5.2 Fisher Information for Shift Factor = 0.8 . . . . . 38

6.5.3 Stochasticity . . . . . 41

**7 Discussion . . . . . 43**
7.1 Calculating Fisher Information . . . . . 43
7.2 Water Supply Resilience . . . . . 44

**8 Conclusions . . . . . 46**

**Bibliography . . . . . 48**

**APPENDIX . . . . . 52**
Appendix A Kernel Density Estimator Method . . . . . 53
A.1 A Numerical Method to Integrate Fisher Information for the Kernel Density Estimator Method . . . . . 53

## LIST OF TABLES

Table 6.1 Copula function information for each month. Data sets are considered independent for a p-value greater than 0.05. $C_{X,Y}$ indicates the copula function for months $X$ and $Y$. . . . . 30

## LIST OF FIGURES


Figure 3.1 Sociotechnical Water Supply Framework couples agent-based modeling with a stochastic reconstruction framework and Fisher information calculations to predict regime shifts in water supply systems. . . . . 9

Figure 4.1 Theoretical Fisher information calculated for zero mean normal distributions with standard deviation in the range [1, 50]. . . . . 17

Figure 4.2 Amplitude-based Discrete Method with disjoint bins is applied to calculate Fisher information with $k = 2$. $N$ is 10, 40, and 100 for zero mean normal distributions with standard deviation in the range [1, 50]. Error bars show the standard deviation over 30 randomly sampled data sets for each sample size and standard deviation. Theoretical Fisher information is shown as the black line. . . . . 19

Figure 4.3 Amplitude-based Discrete Method with disjoint bins is applied to calculate Fisher information for values of $k$. $N = 100$ for zero mean normal distributions with standard deviation in the range [1, 50]. Error bars show the standard deviation over 30 randomly sampled data sets for each value of $k$ and standard deviation. . . . . 19

Figure 4.4 Amplitude-based Discrete Method with overlapping bins is applied to calculate Fisher information with $k = 2$. $N$ is 10, 40, and 100 for zero mean normal distributions with standard deviation in the range [1, 50]. Error bars show the standard deviation over 30 randomly sampled data sets for each sample size and standard deviation. Theoretical Fisher information is shown as the black line. . . . . 20

Figure 4.5 Amplitude-based Discrete Method with overlapping bins is applied to calculate Fisher information for values of $k$. $N = 100$ for zero mean normal distributions with standard deviation in the range [1, 50]. Error bars show the standard deviation over 30 randomly sampled data sets for each value of $k$ and standard deviation. . . . . 20

Figure 4.6 The Kernel Density Estimator Method is applied to calculate Fisher information. $N$ is 10, 40, and 100 for zero mean normal distributions with standard deviation in the range [1, 50]. Error bars show the standard deviation over 30 randomly sampled data sets for each sample size and standard deviation. Error bars for $N = 10$ are not shown because their length exceeds the range of the y-axis. Theoretical Fisher information is shown as the black line. . . . . 21

Figure 5.1 Fisher information for Cantareira Reservoir storage volume. Lines represent calculations using the Kernel Density Estimator Method (—), Amplitude-based Discrete Method with disjoint bins (—), and Amplitude-based Discrete method with overlapping bins (—). . . . . 24

Figure 5.2 Fisher information for Cantareira Reservoir storage volume. Lines represent calculations using the Kernel Density Estimator Method (—) and the Amplitude-based Discrete method with disjoint bins (—). . . . . 25

Figure 5.3 Sliding windows for the Amplitude-based Discrete method with disjoint bins. The green shaded region indicates the temporal location of the sliding window. The inset shows the histogram of data points among system states. . . . . 27
Figure 5.4 Sliding windows for the Kernel Density Estimator Method. The blue shaded region indicates the temporal location of the sliding window. The inset shows the shape of the Kernel Density Estimate. . . . . 28

Figure 6.1 Five model simulations of inflow (acre-ft/mo/1000) for $s_f = 0.8$. . . . . 32
Figure 6.2 Mean of 50 model simulations of inflow (acre-ft/mo/1000) for $s_f = 0.8$. (—) and $s_f = 0.2$ (—). . . . . 32
Figure 6.3 Five model simulations of demand (acre-ft/mo/1000) for $s_f = 0.8$. . . . . 33
Figure 6.4 Five model simulations of storage (acre-ft/1000) for $s_f = 0.8$. . . . . 34
Figure 6.5 Mean of 50 model simulations of demand (acre-ft/mo/1000) for $s_f = 0.8$ (—) and $s_f = 0.2$ (—). . . . . 35
Figure 6.6 Fisher information for Falls Lake for 2015–2065 with $s_f = 0.2$. Lines represent calculations using the Amplitude-based Discrete Method with overlapping bins (—) and reservoir storage (—). . . . . 36
Figure 6.7 Fisher information for Falls Lake for 2015–2065 with $s_f = 0.2$. Lines represent calculations using the Amplitude-based Discrete Method with discrete bins (—) and reservoir storage (—). . . . . 37
Figure 6.8 Fisher information for Falls Lake for 2015–2065 with $s_f = 0.2$. Lines represent calculations using the Kernel Density Estimator Method (—) and reservoir storage (—). . . . . 38
Figure 6.9 Fisher information for Falls Lake for 2015–2065 with $s_f = 0.8$. Lines represent calculations using the Amplitude-based Discrete Method with overlapping bins (—) and reservoir storage (—). . . . . 39
Figure 6.10 Fisher information for Falls Lake for 2015–2065 with $s_f = 0.8$. Lines represent calculations using the Amplitude-based Discrete Method with discrete bins (—) and reservoir storage (—). . . . . 40
Figure 6.11 Fisher information for Falls Lake for 2015–2065 with $s_f = 0.8$. Lines represent calculations using the Kernel Density Estimator Method (—) and reservoir storage (—). . . . . 40
Figure 6.12 Fisher information for Falls Lake for 2015–2065 with $s_f = 0.8$. Fisher information is calculated using the Amplitude-based Discrete Method with discrete bins for five simulations of the Sociotechnical Water Supply Framework. . . . . 42
Figure 6.13 Fisher information for Falls Lake for 2015–2065 with $s_f = 0.8$. Fisher information is calculated using the Kernel Density Estimator Method for five simulations of the Sociotechnical Water Supply Framework. . . . . 42

# 1 Introduction

Urban water supply systems should reliably provide water to meet the demands of end users without disruptions. Water supply systems are sociotechnical systems, in which the technical system, social factors, and human decision-making interact to balance the relationship between water supply and demand and to sustain an expected level of service [Tel18; KZ14; Wat12]. Adaptive actors respond to the water resources system to avert immediate failure, and the system self-organizes through management actions that increase storage during wet seasons to offset impending depletion during dry seasons [Meh15]. Resilience is the ability of the system to absorb external pressures or disturbances and re-organize while retaining its functionality [Hol73; Wal04]. Resilience is a characteristic of water supply systems that emerges from adaptations in demand that respond to the timing of and variations in water supply. For example, mandatory restrictions are imposed at increasing levels by many cities in response to drought, leading to reductions in permitted demands and conservation [Wat12]. Consumers shift water consumption and reuse practices based on the severity of water shortages [Kad18], and they may increase demands during subsequent hydrologically wet periods [GA17]. At a higher level of complexity, cities create and shape adaptation pathways in response to natural disasters, political opportunities, and new information. Stakeholders select incremental decisions to improve short-term performance and major decisions to invest in significant interventions with long lifetimes [Tel18].

Water supply systems are resilient due to their ability to reorganize; however, external stresses, including population growth, climate change, and changes in land use and land cover, threaten the resilience of water supply systems [Zha18]. Water supply systems that lose resilience may transition to a new regime of self-organization, where small changes that could previously be absorbed create performance loss or system failure [Fol06]. In a new state, when the system becomes unbalanced, self-organization may require more effort and longer timelines. In extreme events, re-organization may not be possible. When the system loses resilience, the relationship between supply and demand shifts to a new regime in which demands are met with a lower frequency and the system cannot readily recover from droughts [Cou15]. For example, drought restrictions that are effective under an earlier regime fail to refill a reservoir under a new regime that is characterized by low precipitation and hardened demands. Identifying regime shifts is critical for long-term management of water supply systems [Bah17]. Water managers should be made aware that a continuous state of deficit is

imminent; they can respond by developing new sources of water, constructing new infrastructure for water reuse, or implementing management strategies, such as moratoriums on demand growth.

Assessing the stability of a dynamic system requires determining the presence of thresholds that separate different domains of stability [Bah17; Sta14; Thr09]. Identifying thresholds and delineating alternative states in a complex system is difficult, however [Thr09; Bah17; KHF08]. For systems in which regime shifts have been observed, retrospective analysis can be used to identify drivers and thresholds. Retrospective analysis is limited, however, because thresholds can be identified only when a long time series of the variables that describe the system is available, and retrospective analysis may not provide reliable insight for future management decisions. An alternative approach to locate thresholds is through analysis of experimental data describing system states. Using a simulated system, resilience can be tested by applying discrete disturbances of increasing intensity, or stress gradients, to identify the intensity required to cause a state switch, or crossing of a threshold [Thr09]. Inferential statistics approaches have been applied to detect thresholds in time series data, but they have not been applied to data sets that are characterized by noise or sparse data [And09]. Fisher information [FR22] is a concept from information theory that has been applied to detect regime shifts for time series data. The value of Fisher information is proportional to changes in the probability density function that describes the data and can be used to quantify the order in the system. When the system is disordered, the probability of the system being in any state is uniformly distributed. The system lacks order, as it may be observed in any one state from one time step to the next. The gradient of the probability distribution, and therefore, Fisher information is low, approaching zero. For an ordered system, on the other hand, the likelihood that the system is in one state is high, and the probability density function has a steep slope, with a corresponding high value of Fisher information. A shift between states can be identified when the Fisher information drops from a high value to a low value. A decline in Fisher information shows that the system is leaving a state of relative stability and is in transition among states [Eas14].

Calculating the Fisher information requires an approach to approximate the probability density function of the data. An appropriate form of the probability density function is typically unavailable for real data sets, and a few methods have been developed to calculate Fisher information by using a discrete approximation based on the amplitude of the probability density function [EC12] and a kernel density estimator to approximate the probability density function [TL17]. These methods are referred to as the Amplitude-based Discrete Method and the Kernel Density Estimator Methods, respectively. This research implements these two methods to calculate Fisher information and explores changes to the Amplitude-based Discrete Method to improve the approach used to assign data points to distinct system states. The Amplitude-based Discrete Method and the Kernel Density Estimator Methods are applied to calculate Fisher information for a set of theoretical data sets for which the true value of Fisher information is known. Fisher information calculations are then tested to detect state changes for time series of water supply reservoir data. Fisher information calculation

methods are applied for a time series of storage data for the Cantareira water supply reservoir system, which serves São Paulo, Brazil. A regime shift in water storage was documented in response to a drought and hardened demands, and the application of Fisher information accurately identifies the known shift. Fisher information calculations are integrated in a Sociotechnical Water Supply Framework to forecast regime shifts in reservoir storage for a water supply system. A dynamic model of water use, water supply, and water management is created using an agent-based modeling approach and is used to simulate feedbacks and adaptations among social and technical systems. Projections of climate change and population growth are simulated to apply increasing levels of stress on the water supply system. Climate change scenarios are simulated using a stochastic reconstruction framework, which represents shifts in average streamflow, precipitation, and evapotranspiration. The agent-based modeling framework is developed and applied to simulate water supply for Raleigh, North Carolina. The stochastic reconstruction framework is applied to simulate increasing levels of climate change to the water supply system, and projected population growth is simulated to increase community-level water demands. The Sociotechnical Water Supply Framework is applied to project reservoir storage for the period 2015–2065. Regime shifts are detected using the Amplitude-based Discrete Method and the Kernel Density Estimator Method for calculating Fisher information.

# 2 Background

## 2.1 Water Supply Resilience

The ability of water supply systems to operate satisfactorily for potential future demands and future hydrologic conditions is a critical objective for water management. Rather than base evaluations of water supply systems on cost and benefits alone, Hashimoto et al. developed a framework to assess system performance using risk-related performance criteria to understand how infrastructure may perform in an uncertain future [Has82]. Reliability, resilience, and vulnerability are based on moderate common failures, such as light droughts that do not allow reservoir systems to meet contracted water services. Resilience (or engineering resilience for clarity [Thr09]) is defined as the time required for a system to recover from failure, once failure has occurred, and mathematical formulations were developed to calculate engineering resilience for a system [Has82; SS11]. Another definition for engineering resilience was provided as the time taken for a system to return to a state of pre-disturbance [Pim84]. Engineering resilience has been applied broadly and extended to evaluate many water supply systems [SS11; Rom17; Meh15; Joã18; OVSS18; MA13; Tom18]. Dynamics in the engineering resilience of water supply systems were tested based on changes due to a non-stationary climate and population growth [Joã18; Zha18].

This research develops modeling and computational approaches to test ecological resilience, rather than engineering resilience, of water supply systems. Within the field of ecology and socio-ecological systems, ecological resilience is defined as a dynamic property of systems, rather than a recovery time [Fol06]. Management questions related to system resilience focus on thresholds and states, such as [Sta14]: How much disturbance can an ecosystem absorb before switching to another state? Where is the threshold associated with the switch between ecosystem states? A few studies of water supply systems explore thresholds that a water supply system can absorb before there is loss in system performance. These studies apply incremental reductions in climate parameters, such as rainfall, inflow, and evapotranspiration, but system dynamics are evaluated using engineering resilience and other engineering metrics, without exploring resilience based on the shift to a new system state [MA17; Ama16]. A few studies explore how water supply systems transition into new system states using states based on pre-defined states. States are defined based on observed switches in system dynamics [Cou15] or expected catastrophic levels of demand

deficit [Men15]. In this research, a new approach to testing the resilience of water supply systems is developed by quantifying the order in the water supply system to detect transitions to new system states.

## 2.2 Detecting Regime Shifts using Fisher Information

The majority of documented regime shifts for ecological systems have been inferred from time series of monitoring data, and few examples of regime shifts were found through the use of controlled experiments designed to detect the existence of alternative states [Sch05]. Methods for detecting and characterizing regime shifts have been developed and applied to identify step changes and thresholds in environmental time series using parametric and non-parametric statistical methods and time series analysis [And09; Zal08]. Fisher information [FR22] is a concept from information theory that measures the amount of information about an unknown parameter that can be obtained through observations:

$$I(\theta) = \int \left[\frac{d\,p(y|\theta)}{d\,\theta}\right]^2 \frac{d\,y}{p(y|\theta)} \tag{2.1}$$

where $I(\theta)$ is Fisher information, and $p(y|\theta)$ is the probability density for obtaining a value of $y$ in the presence of $\theta$. Eq. 2.1 can be transformed into a form that can be applied [FG07]. Here, the mean is specified as the unknown estimator, $\theta$, over a period of time, defined by:

$$\theta \equiv \frac{1}{T}\int_0^T y(t)\,d\,t \tag{2.2}$$

where $T$ is the number of observed data points. The conditional probability density $p(y|\theta)$ is assumed to be shift invariant and can therefore be written as $p(x)$ in terms of the deviation from the mean of the system, $x = y - \theta$. Substituting $p(x)$ into Eq. 2.1 gives the following form:

$$I = \int \left(\frac{p(x)}{d\,x}\right)^2 \frac{d\,x}{p(x)} \equiv \int \frac{p'(x)^2}{p(x)} d\,x \tag{2.3}$$

where $p(x)$ is the probability density function of the system being in a state $x$ or the probability of observing $x$. The value of Fisher information is proportional to the derivative of the probability density function ($p'(x)$). When the system is disordered, the probability of the system being in any state is uniformly distributed. The system lacks order, as it may be observed in any one state from one time step to the next. The probability density function is flat, and the value of Fisher information is low, approaching zero. On the other hand, for an ordered system, repeated measurements of the system show that the system has a high bias toward one state. The probability density function has a steep slope, and the Fisher information is high. The Fisher information calculation can be used to identify a shift between states when the Fisher information drops from a high value to a low value. A

decline in Fisher information shows that the system is leaving a state of relative stability and is in transition among states [Eas14].

Calculating Fisher information using the forms shown above is impractical for realistic data sets because the form of the probability density function is typically unknown. A few methods have adapted the form of Fisher information given in Eq. 2.3 to assess stability in system dynamics for modeled and real ecosystems. Two methods are described as follows.

### 2.2.1 Amplitude-based Discrete Method

The expression shown in Eq. 2.3 can be adapted to compute Fisher information analytically or numerically. A numeric approach to computing Fisher information [CF02] is referred to as the Amplitude-based Discrete Method. In this formulation, the probability density function $p(x)$ is replaced using the amplitude $q(x)$ of the probability density function ($p(x) = q(x)^2$), which minimizes calculation errors for very small values of $p(x)$. The equation is converted for use with discrete data and has been applied in this form to characterize and detect regime shifts using field data for ecological and human systems that are characterized by sparse data sets and noise. The Amplitude-based Discrete Method has been applied to identify regime shifts in output from simple dynamic models, with application to water quality in lakes [KHF08] and predator-prey populations [Eas14]. The method has also been applied for data sets of real observations. Regime shifts were detected in data sets that describe marine ecosystems using 65 climate and biological variables [KHF08], long-term diatom community dynamics [Eas16], and phytoplankton time series in a floodplain wetland in Spain [Eas16].

### 2.2.2 Kernel Density Estimation Method

A second approach was developed to estimate Fisher information for discrete data sets using a kernel density estimation to estimate the shape of the unknown distribution, $p(x)$, using a kernel, $K(x)$ [Tel10; TL17]. This approach is referred to as the Kernel Density Estimation Method. Kernel density estimation is a technique that creates a smooth curve using a set of data. The kernel density estimator provides an approximate value of a probability density function ($p(x) \approx \hat{p}(x)$) using the kernel function ($K$), which is a continuous non-negative and symmetric function. The integration of the kernel function from $-\infty$ to $+\infty$ is necessarily equal to 1.0. The kernel density estimator, $\hat{p}(x)$, has the following form (2.4):

$$\hat{p}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right) \tag{2.4}$$

where $h$ is a smoothing factor called the bandwidth; $N$ is the number of points in a data set; and $x_i$ is the $i^{th}$ data point in the set. Several kernel functions exist including uniform, triangular,

Epanechnikov, Gaussian, and others. The Epanechnikov kernel is the most efficient, however the loss of efficiency for other kernels is minimal [WJ94]. The Kernel Density Estimation Method was developed and applied to calculate Fisher information for anticipating volcanic eruptions using seismic data [Tel10].

# 3 Methodology

In this research, we build a Sociotechnical Water Supply Framework to simulate water supply and detect shifting regimes (Fig. 3.1). An agent-based modeling framework is developed to tightly couple reservoir simulation with behaviors of water consumers. A stochastic reconstruction framework is applied to construct time series of climate projections at increasing increments of stress. Fisher information is applied to quantitatively detect and characterize regime shifts in reservoir storage due to shifts in climate and population growth.

## 3.1 Agent Based Modeling Framework

The agent-based modeling framework couples a water supply reservoir, which is simulated using mass balance equations, and residential water consumers as agents that exert demands for indoor and outdoor purposes. The modeling framework is described briefly as follows, and a detailed description of agent behaviors is provided elsewhere [MA17].

### 3.1.1 Household Agents

Households are simulated as agents, with properties and behavioral rules to simulate monthly water consumption. Each agent is assigned values for a set of parameters that affect water use, including number of residents, house built year, and irrigated area. Agent behaviors include indoor water use, outdoor water use, retrofitting water appliances, and response to drought restrictions.

Indoor demands are calculated based on the number of residents in the household and the build year of the house. The daily volume of water consumed for end uses for each person in a household is calculated using data about indoor water use [Vic01; DeO11]. Household agents replace toilets, showerheads, faucets, clothes washer, and washing machines with efficient appliances that are available on the market after the appliance life span, set as 30 years, has expired. By retrofitting appliances, agents update indoor demands across the simulated horizon.

Outdoor water consumption is used for landscape irrigation, which is calculated using a theoretical irrigation model, based on a soil-water budget model. The theoretical irrigation requirement is calculated for each agent, or household, using the area of land dedicated to landscaping, evap-

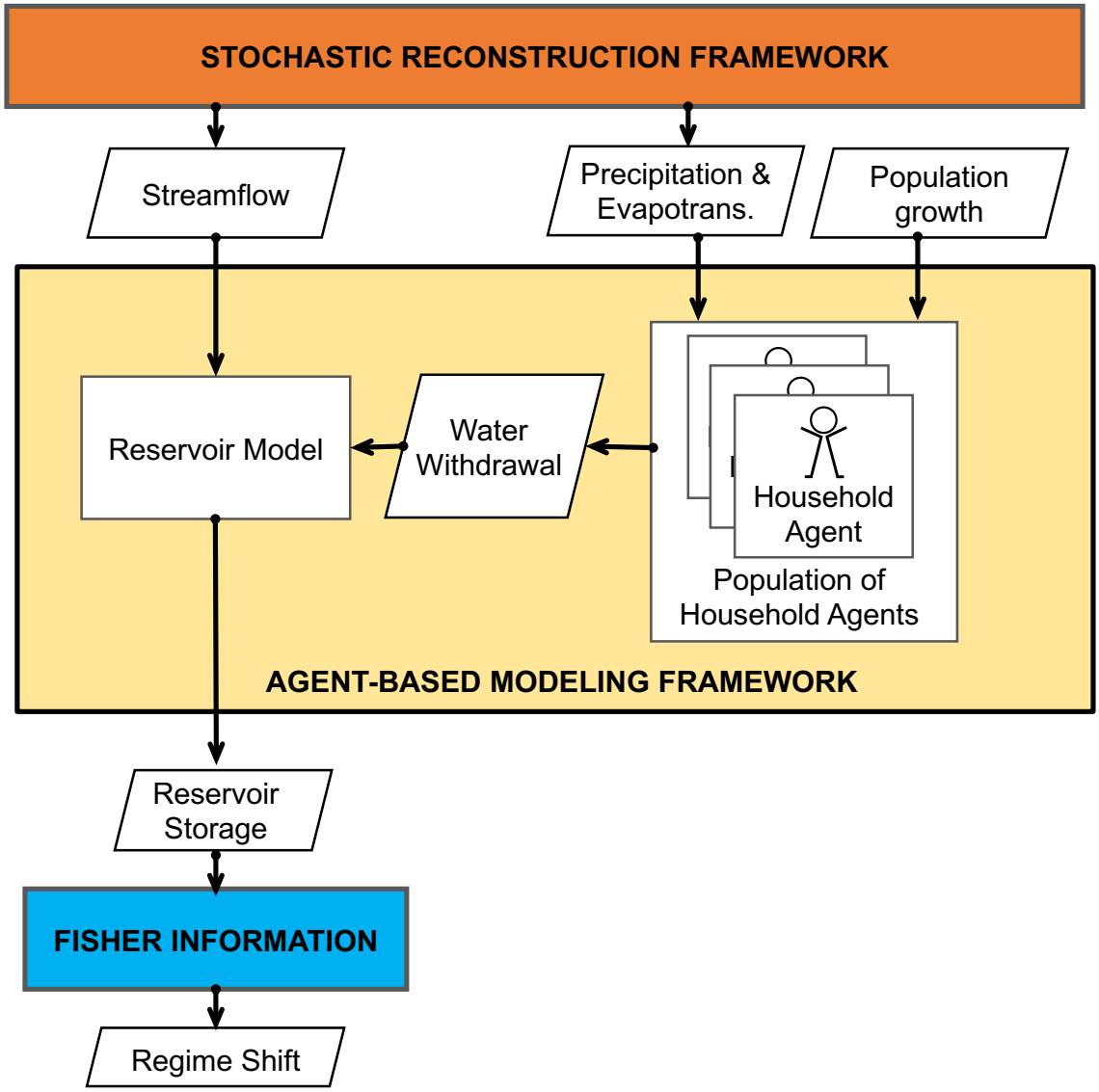


**Figure 3.1** Sociotechnical Water Supply Framework couples agent-based modeling with a stochastic reconstruction framework and Fisher information calculations to predict regime shifts in water supply systems.

otranspiration, precipitation, efficiency of irrigation technologies, and crop coefficient. Outdoor demand ($D_O, t$) is calculated as follows:

$$D_{O,t} = b \times ET_{net,t} \times \frac{A}{Eff} \times K_{crop} \tag{3.1}$$

where $D_{O,t}$ is the outdoor demand at time step $t$; $ET_{net,t}$ is the net evapotranspiration, which is the reference evapotranspiration minus the effective rainfall at time step $t$; $A$ is the irrigated area; $Eff$ is the efficiency of an irrigation technology; and $K_{crop}$ is the crop coefficient. A behavioral factor, $b$, is included to represent the frequency that agents use to water lawns; this frequency was assigned probabilistically to households using a survey of lawn watering habits in North Carolina.

### 3.1.2 Reservoir Model

Reservoir storage is calculated using the continuity equation:

$$S_t = S_{t-1} + I_t - R_t - WS_t - SP_t - Def_t \tag{3.2}$$

where $S_t$ is the volume of water stored in the reservoir at time step $t$, and time steps are simulated in months. $I_t$ is the monthly net inflow to the reservoir, which is the the inflow plus precipitation over the lake surface and minus evaporation from the lake surface. $R_t$ is the reservoir outflow at time step $t$. Reservoir releases are required to maintain downstream water quality. $WS_t$ is the total water supply, equal to the sum of the indoor and outdoor demands withdrawn from reservoir at time step $t$. $SP_t$ is reservoir spill, which is released when the water level within the reservoir exceeds a pre-defined control elevation. $Def_t$ is the reservoir deficit or shortfall. At each month, the reservoir storage is calculated using inflow, the volume of water released based on operational rules, and the volume of withdrawals from consumer agents.

## 3.2 Stochastic Reconstruction Framework

A stochastic reconstruction framework is used to generate climate scenarios that represent a drying climate [Naz13] through the use of a shift factor, quantile mapping, and copula functions. The stochastic reconstruction framework is applied to develop shifting distributions for precipitation, evapotranspiration, and inflows.

### 3.2.1 Shift Factor

The framework developed by [Naz13] is adapted here to reflect a *shifting* climate, where the distribution of hydrologic flows is updated annually. In the original framework, the shift factor is applied across a simulated time horizon and used to shift the mean of a distribution of flows to represent

statistically lower values. In this application, we model the shift factor as a target value that is reached over a simulated time horizon. The shift factor is initially set to 1.0 and reduces incrementally at each time step until it reaches the target value. The shift factor, $s_f$, can be set within the range of 0.0 to 1.0, and $s_{f,t}$ is calculated at each time step $t$ as:

$$s_{f,t} = 1.0 - (1.0 - s_f) * \frac{t}{T} \tag{3.3}$$

where $T$ is the total number of time steps in the simulated period.

### 3.2.2 Quantile Mapping

Quantile mapping is a method to create new random samples using a transformed cumulative distribution function (CDF). In quantile mapping, the original random sample $z$ with the CDF $F_Z$ is transformed to a new random sample $y$ with CDF $F_Y$ using the equation:

$$y = F_Y^{-1}(F_Z(z)) \tag{3.4}$$

The stochastic reconstruction framework applies quantile mapping to hydrologic data to represent climactic shifts by updating historical quantiles based on a shift factor, $s_f$, and the historical distribution [Naz13]. Using continuous random variable notation, $X$, to represent the hydrologic variable of interest, we use $E(X_a)$ and $\sigma(X_a)$ to represent the mean and standard deviation, respectively, of the variable at month $a$. A transformed CDF is created by updating the CDF of historical values with a new mean of $s_f E(X_a)$, where $s_f$ is the shift factor, which takes a value in the range of 0.0 to 1.0. The variance of the historical data is assumed constant and is not transformed. The updated quantile at each time step $t$ for the non-exceedance probability $x$ can be calculated as

$$X_t^*(x|s_{f,t}) = F_{X_a}^{-1}\{N_{E(X_a),\sigma(X_a)}(N^{-1}_{s_{f,t}E(X_a),\sigma(X_a)}(x))\} \tag{3.5}$$

where the $*$ indicates a "shifted" value; $t = 0, 1, 2, ..., T$; $T$, as above, is the total number of time steps; $a$ is the month corresponding to the time step ($a = (t \bmod 12) + 1$); $X_a$ corresponds to the historical values for month $a$; $X_t^*(x|s_{f,t})$ is the shifted value corresponding to the non-exceedance probability $x$ and shift factor $s_f$ at time step, $t$. $N_{\mu,\sigma}$ is the CDF of a normal distribution with mean $\mu$ and standard deviation $\sigma$; $N^{-1}$ is the inverse CDF of a normal distribution; and $F_{X_a}^{-1}$ is the inverse CDF of the historical values for variable, $X$, in month $a$. A set of shifted values, $X_t^*(x|s_{f,t})$, are created by applying Eq. 3.5 to a set of non-exceedance probabilities $x$, where $x \in [0,1]$. These shifted values generate a shifted CDF, $F_{X_t^*}$ which is sampled to generate hydrological values for use in climate simulations.

For this research, the hydrological variables of interest that would be represented by $X$ are inflow, precipitation, and evapotranspiration. For precipitation and evapotranspiration, the shifted

CDFs are randomly sampled and the resulting values are used within a simulation. The method to calculated inflows is described in the following section.

### 3.2.3 Copula Functions

Due to the large changes in inflow values throughout a given year, the temporal dependence in inflow values is important and is maintained by applying a copula technique. A copula "couples" multivariate density functions to their one-dimensional marginal distribution functions [Nel06]. If $X$ and $Y$ are a pair of continuous random variables with marginal distributions of $F_X(x) = P[X \leq x]$ and $G_Y(y) = P[Y \leq y]$ and a joint cumulative distribution, $H_{X,Y}(x,y) = P[X \leq x, Y \leq y]$, then a copula exists such that:

$$H_{X,Y}(x,y) = C(F_X(x), G_Y(y)) \tag{3.6}$$

where $x, y \in \mathscr{R}$ and $C$ is a copula function. The upper-case $X, Y$ represent continuous random variables, whereas the lower-case $x, y$ represent the value of the variables. The copula, $C$, in Eq. 3.6 is referred to as the *copula of $X$ and $Y$* and is denoted as $C_{X,Y}$.

In this research, copulas are used to generate new inflow values to capture dependence between subsequent months. For example, $X$ and $Y$ can represent the historical inflow values for January ($a = 1$) and February ($a = 2$), respectively. Given an inflow value in January, a new inflow value can be generated for February using the copula of January and February, $C_{1,2}$. New inflow values are generated using conditional probabilities. For this method, the conditional probability, denoted as $c_u(v)$, for $\text{Inflow}^*_t$ given that $\text{Inflow}^*_{t-1} \leq u$ is as follows:

$$c_u(v) = P\{F_{\text{Inflow}_a}(\text{Inflow}^*_t) \leq v | F_{\text{Inflow}_{a-1}}(\text{Inflow}^*_{t-1}) = u\} \tag{3.7}$$

where the pair $u, v$ correspond to non-exceedance probabilities. Eq. 3.7 can be read as: The conditional probability, $c_u(v)$, is the probability that the Inflow at $t$ is less than or equal to the probability $v$ given that the probability of the Inflow at $t-1$ is $u$. $\text{Inflow}_a$ corresponds to the historical inflow values for month $a$, whereas $\text{Inflow}^*_t$ corresponds to a shifted inflow value at time step $t$. To generate shifted inflow values using Eq. 3.7, the probability of $\text{Inflow}^*_{i-1}$ is determined using $u = F_{\text{Inflow}_{a-1}}(\text{Inflow}^*_{t-1})$, and a random value, $v$ is selected from a uniform distribution, $\mathscr{U}(0,1)$. Given the probabilities $(u, v)$, the conditional probability, $c_u(v)$, is determined using the inverse, $c_u^{-1}(v)$. The inverse is used to find a new $\text{Inflow}^*_t$ value:

$$\text{Inflow}^*_t = F^{-1}_{\text{Inflow}^*_t}\{c_u^{-1}(v)\} \tag{3.8}$$

## 3.3 Methods for Calculating Fisher Information

### 3.3.1 Amplitude-based Discrete Method

A time series is denoted by the set, $S$, of discrete values. The notation $x(t)$ denotes the position of a system at time, $t$, for the time interval $(1, T)$.

$$S = \{x(t), \quad t = 1, ..., T\} \tag{3.9}$$

Given a discrete set of data points $S$, Fisher information as represented using Eq. 2.3 is difficult to solve because $p(x)$ is unknown. A discretization is performed by using "bins" to represent states of a system. The bins indicate that a particular system state covers a range of values and that each $x(t)$ can be organized using these states. The bins are created based on a size of state parameter, represented as $\Delta s$. (The selection of $\Delta s$ is described in section 3.3.1.1). Bins are created and represented as $s_l$ where $l$ represents the state number. The data are organized into bins and the probability of points being located in each state, $p(s_l)$, is calculated using Eq. 3.10.

$$p(s_l) = \frac{\#\text{ points in } s_l}{T} \tag{3.10}$$

where $T$ represents the total number of data points. Probability of a state is related to amplitude of a state ($q(s)$) [FS95]:

$$p(s) = q^2(s) \tag{3.11}$$

Using the amplitude rather than the probability function reduces calculation errors that arise due to dividing by small values of $p(s)$ [EC12]. The terms in Eq. 2.3 are transformed as follows:

$$p'(s) = 2q(q'(s)) \quad \therefore \quad p'(s)^2 = 4q^2(q'(s))^2 \tag{3.12}$$

Through substitution, Eq. 2.3 is transformed to the following:

$$I = 4\int q'(s)^2 ds \quad \equiv \quad 4\int \left[\frac{dq}{ds}\right]^2 ds \tag{3.13}$$

Eq. 3.13 is discretized by replacing $dq$ and $ds$ with $dq = q(s_l) - q(s_{l+1})$ and $ds = s_l - s_{l+1}$:

$$I \approx 4\sum_{l=1}^{L}\left[\frac{q(s_l) - q(s_{l+1})}{s_l - s_{l+1}}\right]^2 (s_l - s_{l+1}) \tag{3.14}$$

Because $s_1 - s_{l+1} = 1$, Eq. 3.14 becomes:

$$I \approx 4 \sum_{i=1}^{n} [q_i - q_{i+1}]^2 \tag{3.15}$$

#### 3.3.1.1 Size of State

The size of state parameter, $\Delta s$, determines the size of bins, which is used to assign each data point to a distinct bin. $\Delta s$ is determined using Chebyshev's inequality, which states that the proportion of the values in a distribution that fall within $k$ standard deviations, $\sigma$, of the mean is $1 - 1/k^2$. This inequality is used to set the size of state:

$$\Delta s = k\sigma \tag{3.16}$$

Given a data point, $x(t)$, every other data point in $S$ that is within $\pm\Delta s$ is considered in the same state. Two methods are used to bin the data points. The bins can be disjoint or overlapping.

**Disjoint method** Disjoint bins are generated using the following steps:

1. Calculate the standard deviation $\sigma_S$ of set $S$ to calculate $\Delta s$ using Eq. 3.16.
2. Calculate the mean $\bar{x}$ of set $S$ and set this value as the center point of state $s_1$. The bin edges for state $s_1$ are $\bar{x} \pm \Delta s$. The edges of state $s_1$ include $\bar{x} - \Delta s$ and $\bar{x} + \Delta s$.
3. The edges are built away from the center bin in increments of $2\Delta s$ until the minimum and maximum values in $S$ fall into a bin.

**Overlapping method** The overlapping method uses the following steps:

1. Calculate the standard deviation $\sigma_S$ of set $S$ to calculate $\Delta s$ using Eq. 3.16.
2. The first point in $S$, $x(1)$, is set as the center of the first state. All points within $S$ that fall within $\pm\Delta s$ are included in state $s_1$.
3. The first sequential point in $S$ that falls outside of $s_1$ is set as the center of the next state, $s_2$. Every point in $S$ is again compared, and those that fall within $\pm\Delta s$ are counted in $s_2$. Some points are assigned to multiple states.
4. The above steps are repeated until all points in $S$ are counted in at least 1 state.

### 3.3.2 Kernel Density Estimator Method

The Kernel Density Estimator Method identifies the kernel-density estimate for a set, $S$, of data points. For this work, a Gaussian kernel is used which has the form (3.17):

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \tag{3.17}$$

The Gaussian kernel is used as $K$ in Eq. 2.4. Using a Gaussian kernel, the kernel density estimator and the bandwidth can be computed with a low computational complexity. Numerical methods are used to calculate Fisher information. Calculation steps are:

1. Determine optimal bandwidth, $h$. The selection of the bandwidth parameter is important to ensure that the kernel is used most efficiently [Jon96]. Several methods for identifying the value for the bandwidth have been explored, including rules-of-thumb, least-squares cross-validation, biased cross-validation, plug-in methods, and smoothed bootstrap. This framework uses a plug-in method as described by Sheather & Jones to determine the bandwidth. The performance of a kernel density estimator, $\hat{p}$, is measured against the actual probability density function, $p$, by using the asymptotic mean integrated square error (AMISE). The optimal bandwidth $h_{\text{AMISE}}$ is found using the following equation:

$$h_{\text{AMISE}} = \left[\frac{R(K)}{\mu_2(K)^2 R(p'')N}\right]^{1/5} \tag{3.18}$$

   where $R$ and $\mu_2$ are functions of a given function $g$, as follows.

$$R(g) = \int_{\text{R}} g(x)^2 dx \tag{3.19}$$

$$\mu_2(g) = \int_{\text{R}} x^2 g(x) dx \tag{3.20}$$

   Eq. 3.18 cannot be solved directly because it relies on the derivative of an unknown distribution, $p'$. A method to transform Eq. 3.18 and use a plug-in method to find the optimal bandwidth was developed and described [RD06], and the code referenced in that work are used here to find the optimal bandwidth. The value $h_{\text{AMISE}}$ is the optimal bandwidth, and $h$ is set as $h_{\text{AMISE}}$ in Eq. 2.4.

2. Generate a kernel density estimator, $\hat{p}(x)$, using the data points in $S$ and the optimal bandwidth, $h$. In this implementation, a python package is used to implement the kernel density estimator (`k = scipy.stats.gaussian_kde(S, h)`).

3. Replace $p(x)$ in Eq. 2.3 with the kernel density estimator, $\hat{p}(x)$, to generate the following:

$$I = \int_{-\infty}^{\infty} \frac{\hat{p}'(x)^2}{\hat{p}(x)} dx \tag{3.21}$$

Eq. 3.21 can be integrated using discrete numerical methods. The result of this integration is the FI value of the set, $S$. A description of the numerical method implemented in python to solve Eq. 3.21 is provided in Appendix A.

### 3.3.3 Temporal Fisher Information

#### 3.3.3.1 Sliding window

To apply Fisher information to detect regime shifts in time series data, temporal Fisher information calculations are needed [Mar99]. A “sliding window" is used to calculate temporal Fisher information along a time series for both the Amplitude-based Discrete Method and the Kernel Density Estimator Method. The sliding window, $W$, is defined by two parameters: the width, $w \in T$, and the sliding factor $\Delta \in w$. The window, $W$, is a subset of the data series $S$, defined as:

$$\begin{aligned} W(m; w, \Delta) = \{x(t), \quad & t = 1 + m\Delta, ..., w + m\Delta\}, \\ & m = 0, 1, 2, ..., M \end{aligned} \tag{3.22}$$

where $m$ is the window number, $w \leq T$, and $\Delta \leq w$. The sliding factor, $\Delta$, is smaller than $w$ so that the sliding windows overlap. For seasonal data, $w$ is set so that the sliding window covers several seasonal cycles.

#### 3.3.3.2 Size of State

The Amplitude-based Discrete Method uses a parameter, the size of state, $\Delta s$, as the size of bins, which is used to assign each data point to a distinct bin. As shown above, $\Delta s$ depends on the standard deviation of the data set. For the temporal Fisher information calculations, $\Delta s$ uses the smallest standard deviation from the set of windows. For the disjoint binning method, the bin edges are determined using the minimum, mean, and maximum values from the entire time series. The edges are determined *a priori* and used consistently within each rolling window. For the overlapping binning method, binning of points inherently requires different bins for each rolling window. The first point for each window is set as the center of the bin for the first state, $s_1$, and the approach is implemented without changes, as described above.

# 4 Calculating Fisher Information for Normal Distributions

## 4.1 Description of Experiments

The Amplitude-based Discrete Method and the Kernel Density Estimator Method were tested for data sets of known Fisher information values. Consider a normal distribution with zero mean. The theoretical Fisher information for a normal distribution is:

$$I = \frac{1}{\sigma^2} \tag{4.1}$$

This equation (Eq. 4.1) indicates that the Fisher information value for a zero mean normal distribution is only dependent on the standard deviation. A plot of theoretical Fisher information versus the standard deviation of a normal distribution is shown in Fig. 4.1.

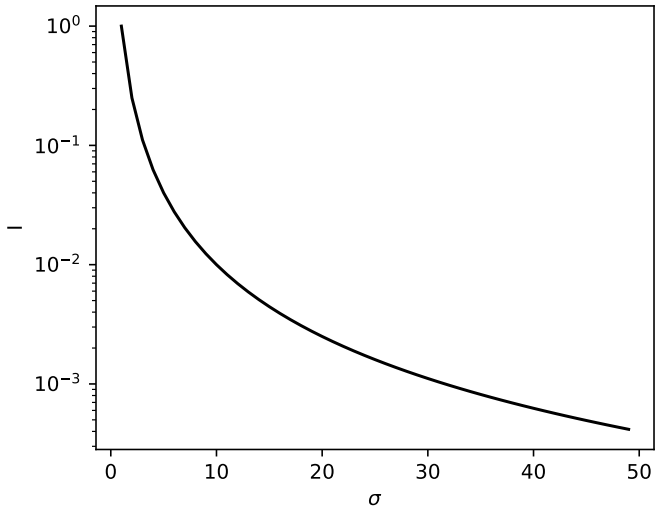


**Figure 4.1** Theoretical Fisher information calculated for zero mean normal distributions with standard deviation in the range [1, 50].

The two methods were tested to calculate the Fisher information for 50 zero mean normal distributions. Each distribution was assigned a unique standard deviation from the range [1:50] with a step of one (i.e. $\mathcal{N}(0,1)$, $\mathcal{N}(0,2)$, ..., $\mathcal{N}(0,50)$). Each of these distributions was sampled with sample size $N = 10$, 40, or 100 to create unique data sets. Each sample size was repeated 30 times to account for stochasticity in the sampling, generating a total of $50 \times 3 \times 30 = 4500$ data sets. The Fisher information was calculated for all 4500 sets using the two methods described in section 3.3.

## 4.2 Amplitude-based Discrete Method Results

The Amplitude-based Discrete Method was applied for the data sets using $k = 2$ and discrete bin states (Fig. 4.2). The calculated FI shows a flat trend across values for the standard deviation, deviating significantly from the theoretical Fisher information values. As shown in Fig. 4.2, the number of data points ($N = 10$, 40, or 100) used in calculating the Fisher information does not significantly affect the results.

The sensitivity of Fisher information values to the value of $k$ was tested. Fisher information was again calculated using $k = [1.5, 2.0, 2.5, 3.0]$ with $N = 100$ as shown in Fig. 4.3. These results show that the calculated Fisher information varies slightly with changing values of $k$, and lower $k$ values lead to lower Fisher information values. The difference between the theoretical Fisher information and the calculated value remains large. Moreover, the calculated Fisher information does not show any relationship with values of the standard deviation, unlike the theoretical Fisher information values.

Next, the Amplitude-based Discrete Method was tested using overlapping bins for $k = 2$. These results are shown in Fig. 4.4. The trend is flat across changing values of standard deviation. The number of data points, $N$, used in the Fisher information calculation does not significantly alter the results.

Several values of $k$ were used to show its influence on the calculated Fisher information values. Results for the range $k = [1.5, 2.0, 2.5, 3.0]$ with $N = 100$ are shown in Fig. 4.3. These results show that the calculated Fisher information is slightly dependent on the value of $k$, but the trend remains unchanged.

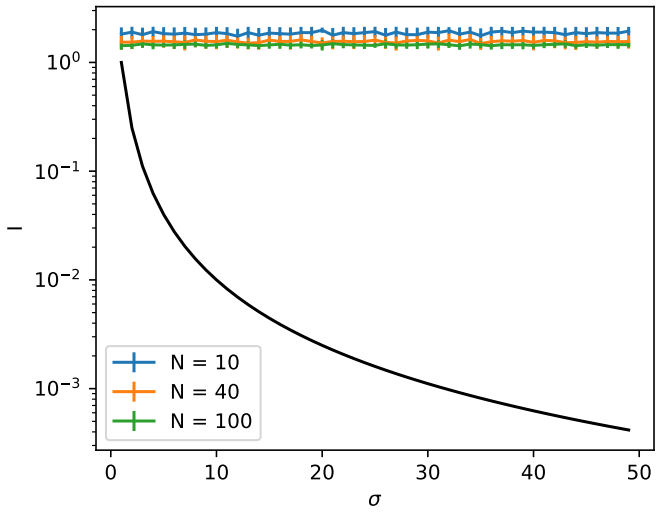


**Figure 4.2** Amplitude-based Discrete Method with disjoint bins is applied to calculate Fisher information with $k = 2$. $N$ is 10, 40, and 100 for zero mean normal distributions with standard deviation in the range [1, 50]. Error bars show the standard deviation over 30 randomly sampled data sets for each sample size and standard deviation. Theoretical Fisher information is shown as the black line.

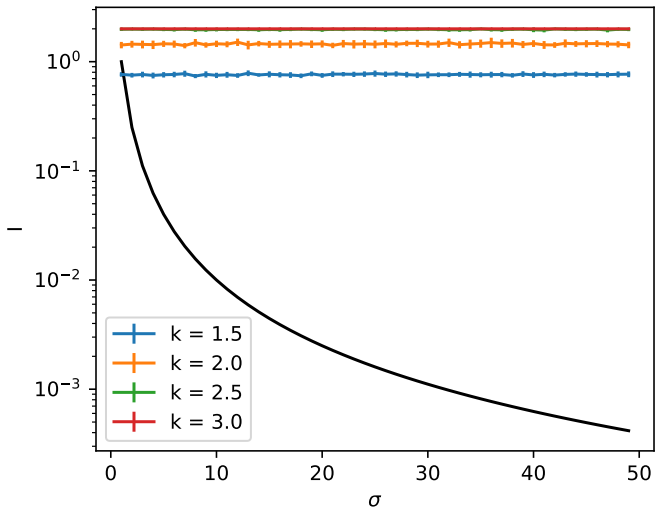


**Figure 4.3** Amplitude-based Discrete Method with disjoint bins is applied to calculate Fisher information for values of $k$. $N = 100$ for zero mean normal distributions with standard deviation in the range [1, 50]. Error bars show the standard deviation over 30 randomly sampled data sets for each value of $k$ and standard deviation.

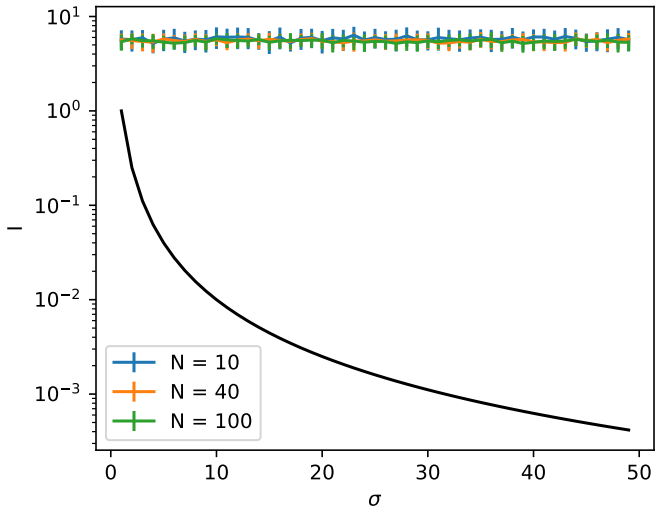


**Figure 4.4** Amplitude-based Discrete Method with overlapping bins is applied to calculate Fisher information with $k = 2$. $N$ is 10, 40, and 100 for zero mean normal distributions with standard deviation in the range [1, 50]. Error bars show the standard deviation over 30 randomly sampled data sets for each sample size and standard deviation. Theoretical Fisher information is shown as the black line.

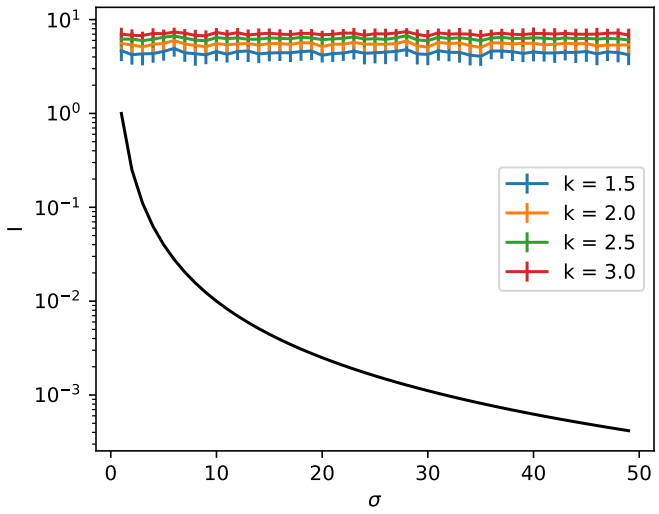


**Figure 4.5** Amplitude-based Discrete Method with overlapping bins is applied to calculate Fisher information for values of $k$. $N = 100$ for zero mean normal distributions with standard deviation in the range [1, 50]. Error bars show the standard deviation over 30 randomly sampled data sets for each value of $k$ and standard deviation.

## 4.3 Kernel Density Estimator Method Results

The Kernel Density Estimator Method was applied to calculate Fisher information (Fig. 4.6). The Kernel Density Estimator Method calculates values that in general, follow the trend of the theoretical values for Fisher information. Values from the Kernel Density Estimator Method cross the theoretical curve near $\sigma = 10$, and the deviations grow larger as $\sigma$ increases or decreases away from this intersection. The number of points in the data set significantly affects the performance of the Kernel Density Estimator Method. When $N = 10$, the variability for 30 series is much higher compared to $N = 40$ or 100, which suggests that there is a minimum number of samples required for this method to accurately calculate Fisher information.

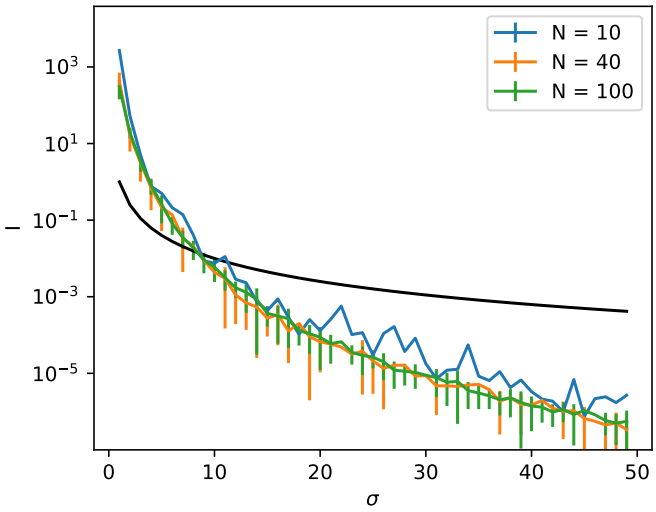


**Figure 4.6** The Kernel Density Estimator Method is applied to calculate Fisher information. $N$ is 10, 40, and 100 for zero mean normal distributions with standard deviation in the range [1, 50]. Error bars show the standard deviation over 30 randomly sampled data sets for each sample size and standard deviation. Error bars for $N = 10$ are not shown because their length exceeds the range of the y-axis. Theoretical Fisher information is shown as the black line.

# 5 Calculating Fisher Information for the Cantareira Reservoir System, São Paulo, Brazil

## 5.1 Case Study Description

The Cantareira Reservoir System in Brazil consists of five interconnected reservoirs that supply water to approximately 50% of the greater São Paulo, Brazil area, serving over 9 million people. Prior to 2013, water storage in the reservoir fluctuated with regular seasonal variation, indicative of a stable state. In mid to late 2013, precipitation in the region was significantly lower than average, resulting in a water crisis that extended into mid to late 2015. Precipitation returned to normal during the wet season in late 2014 and early 2015; the reservoir system, however, did not return to adequate supply levels. The system reached equilibrium at a new stable state, with seasonal variations in storage that fluctuated with a different set of expected values, compared to storage levels prior to 2013. The change in seasonal patterns provides empirical evidence that the water supply system had experienced a regime shift in water storage.

This system was previously examined for resilience indicators using statistical methods [Cou15], and a mathematical model was developed that shows system dynamics.The conditional variance showed an abrupt increase around the beginning of 2014 which corresponds to the same time in which the reservoir system reached its lowest stored volume. This statistical signal was used to indicate that a regime shift had occurred. The research study [Cou15] finds that reservoir systems have two stable states: high- and low-level water storage volumes as a function of precipitation and outflow. Both regimes are resilient in overlapping climatic and outflow conditions resulting in hysteretic behavior: the transitions between stable states may follow different paths. On one hand, the transition from the high- to low-level state can simply result from abnormally dry seasons or mismanagement of the water supply. On the other hand, recovery from a low level state may require an unusually wet season or a significant reduction in the reservoir outflow.

This research uses historical reservoir data from January 2004 to April 2019, as reported by Sabesp, São Paulo, Brazil [Sab]. The data includes the reservoir storage as a percentage of the useful volume and the monthly rainfall. The values for the last day of each month were used as time series data for the analysis. Monthly values were selected to match the time step of the reservoir simulation

used in Chapter 6.

## 5.2 Results

Fisher information was calculated for the Brazil storage data using three methods: 1) Amplitude-based Discrete Method with disjoint bins, 2) Amplitude-based Discrete Method with overlapping bins, and 3) the Kernel Density Estimator Method. For each method, the sliding window width is set to 47 months ($w = 47$), and the sliding factor is set to 1 month ($\Delta = 1$). The value for $k$ was selected based on the guidance provided in the development of the Amplitude-based Discrete Method [CF02]. Based on Chebychev's inequality, the proportion of the values falling within 2.0 standard deviations is 75%, and $k$ is set to 2.0. Using sliding windows, Fisher information is calculated using either the Amplitude-based Discrete Method or the Kernel Density Estimator Method. Fisher information values are plotted with the original time series of storage data by placing each Fisher information value at the end of the corresponding sliding window. This placement ensures that the Fisher information value is shown relative to observed data, rather than future data.

Fisher information values are shown for the three methods in Fig. 5.1. The values of Fisher information differ among the three methods, yet there are emergent trends that agree among the methods. The Amplitude-based Discrete Method with overlapping bins shows high variability, and the series for the Amplitude-based Discrete Method with disjoint bins and Kernel Density Estimator Method are isolated in Fig. 5.2. Results from the Amplitude-based Discrete Method with overlapping bins are difficult to interpret (Fig. 5.1). There are peaks and sharp declines in the Fisher information that correspond to the patterns calculated using the other two methods; however, the Amplitude-based Discrete Method with overlapping bins shows high variability across the time series, with other peaks that could be interpreted as regime shifts. Variability emerges because of the use of overlapping bins. A new data point that is added when the sliding window shifts by one time step can drastically change the Fisher information calculation. A new point may be in multiple states and change the calculation shown in Eq. 3.15 by a large margin, rather than a small increment, as shown for the disjoint bins below.

For the two methods shown in Fig. 5.2, lines indicate two system state changes occurred in the system, one around months 70–75, and another around months 120-130. The sharp increase in Fisher information is indicative of system organization. As the system shifts into a stable state, the Fisher information increases. After the Fisher information peaks, there is another decline, indicating a transition out of the stable state and into a period of instability. The timing of the shifts depicted by the Amplitude-based Discrete Method and the Kernel Density Estimator Method corresponds with the flood and drought that occurred in the reservoir system.

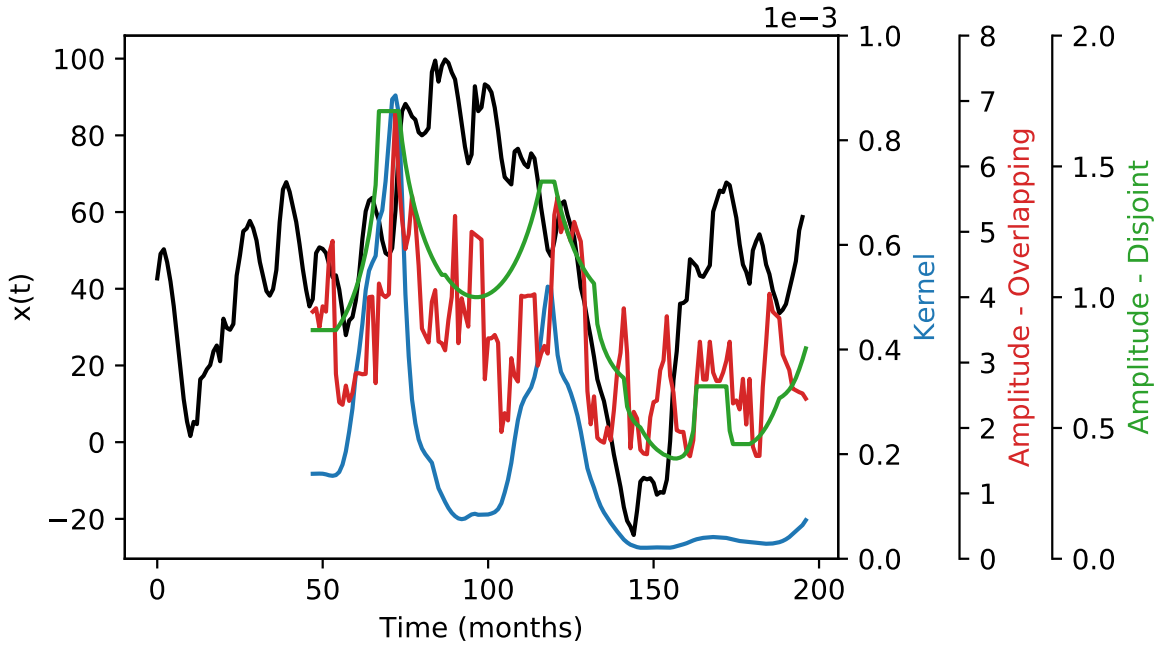


**Figure 5.1** Fisher information for Cantareira Reservoir storage volume. Lines represent calculations using the Kernel Density Estimator Method (——), Amplitude-based Discrete Method with disjoint bins (——), and Amplitude-based Discrete method with overlapping bins (——).

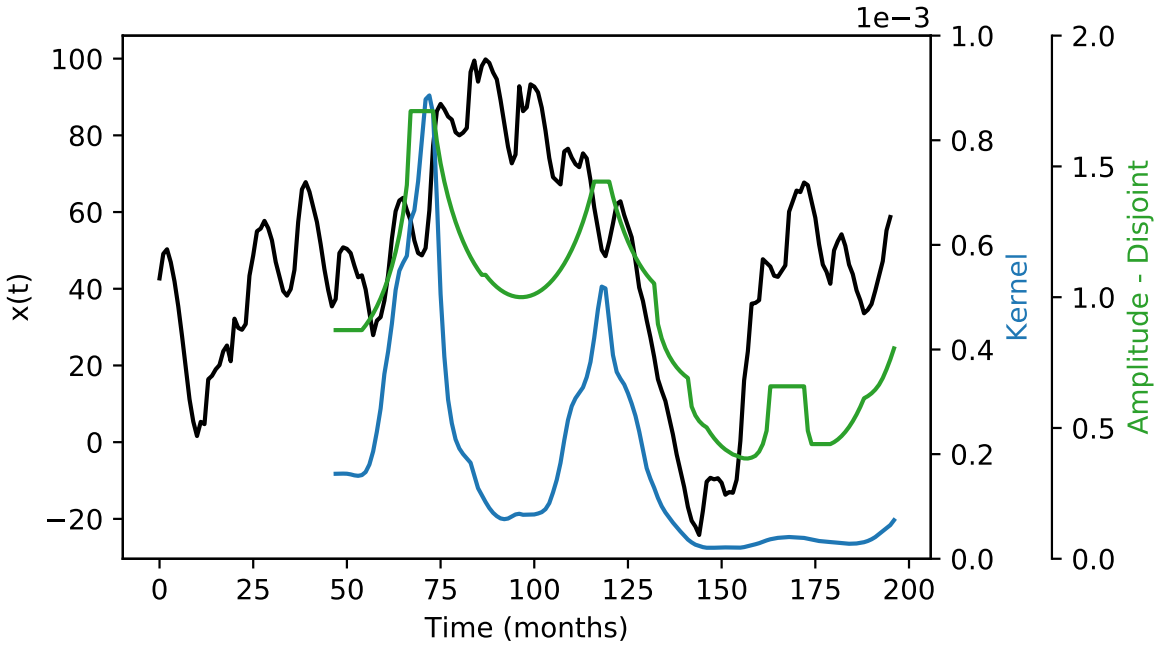


**Figure 5.2** Fisher information for Cantareira Reservoir storage volume. Lines represent calculations using the Kernel Density Estimator Method (——) and the Amplitude-based Discrete method with disjoint bins (——).

State transitions are explored through viewing changes in the distribution of data points among states. Fig. 5.3 shows three subplots of the Amplitude-based Discrete Method with disjoint bins. The sliding window moves from left to right across the data set, and the data points within each sliding window are organized within one of four states: $s_1$, $s_2$, $s_3$, and $s_4$. Each subplot (A, B, and C in Fig. 5.3) shows one instance of the sliding window, indicated by a shaded region, and an inset of the histogram of the binned data points. The Fisher information calculated for each window is shown as the green point on the right most edge of the window. The inset for Fig. 5.3A shows that the majority of data points falls into state $s_3$, and the Fisher information is correspondingly high. For the window shown in Fig. 5.3B, most of the data points have shifted to state $s_4$, and the Fisher information is also at a high value at this point in the time series. There is a dip in Fisher information between these two points, representing the transition of the system from state $s_3$ to state $s_4$. Fig. 5.3C shows a window where there is high disorder among the data, and all four system states are represented. The Fisher information value is at a low value, representing a disordered system.

State changes are also explored through the Kernel Density Estimator Method. Fig. 5.4 shows three windows in the time series, and the inset shows the shape of the kernel density estimator for the data points within the sliding window. Fig. 5.4A shows a kernel density estimator with a sharp peak, corresponding to high Fisher information. The kernel density estimator for Fig. 5.4B has a relatively sharp peak, but less so compared to Fig. 5.4A, and the Fisher information value is lower. For Fig. 5.4C, the kernel density estimator is broad, indicating low information about the rolling window.

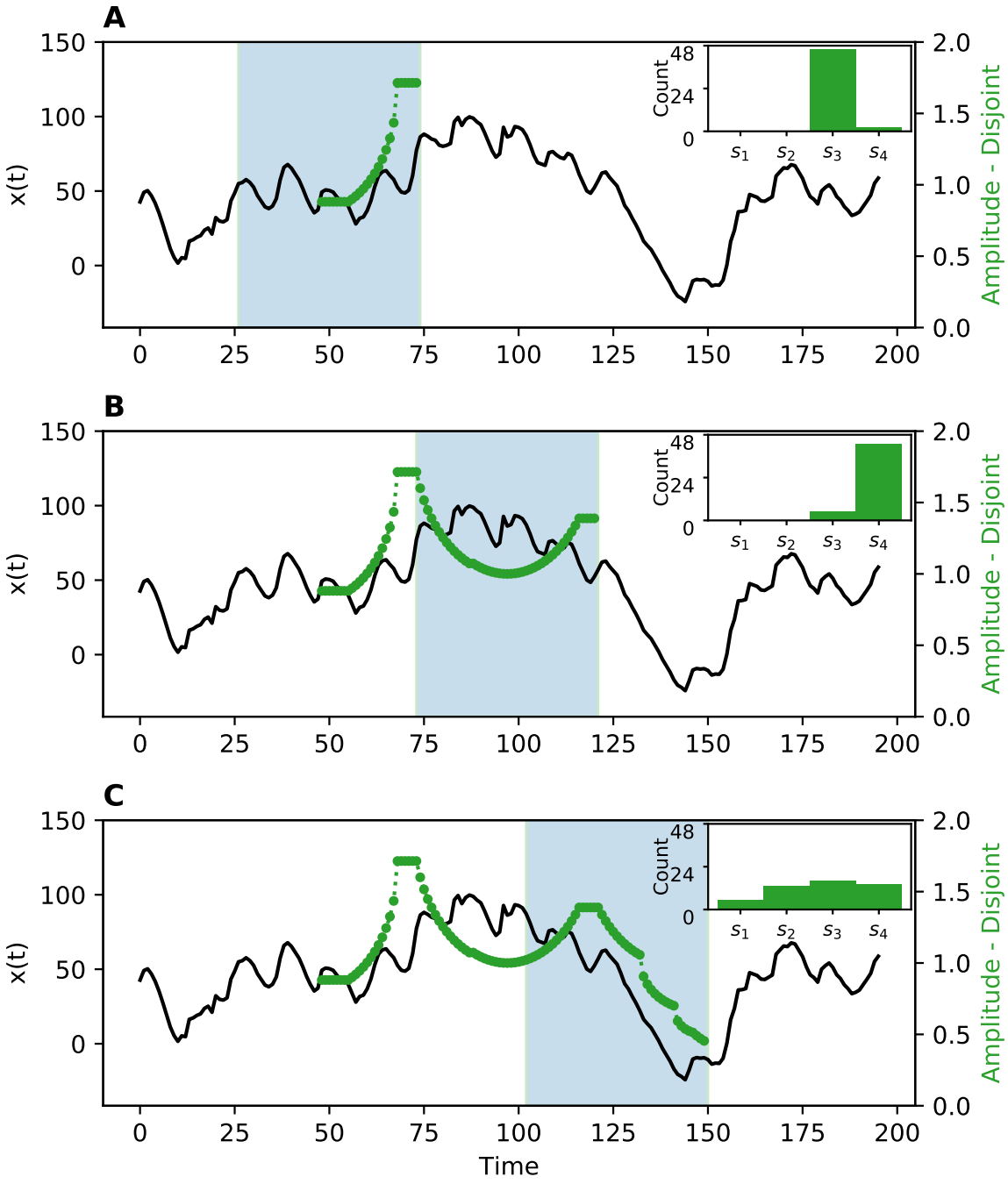


**Figure 5.3** Sliding windows for the Amplitude-based Discrete method with disjoint bins. The green shaded region indicates the temporal location of the sliding window. The inset shows the histogram of data points among system states.

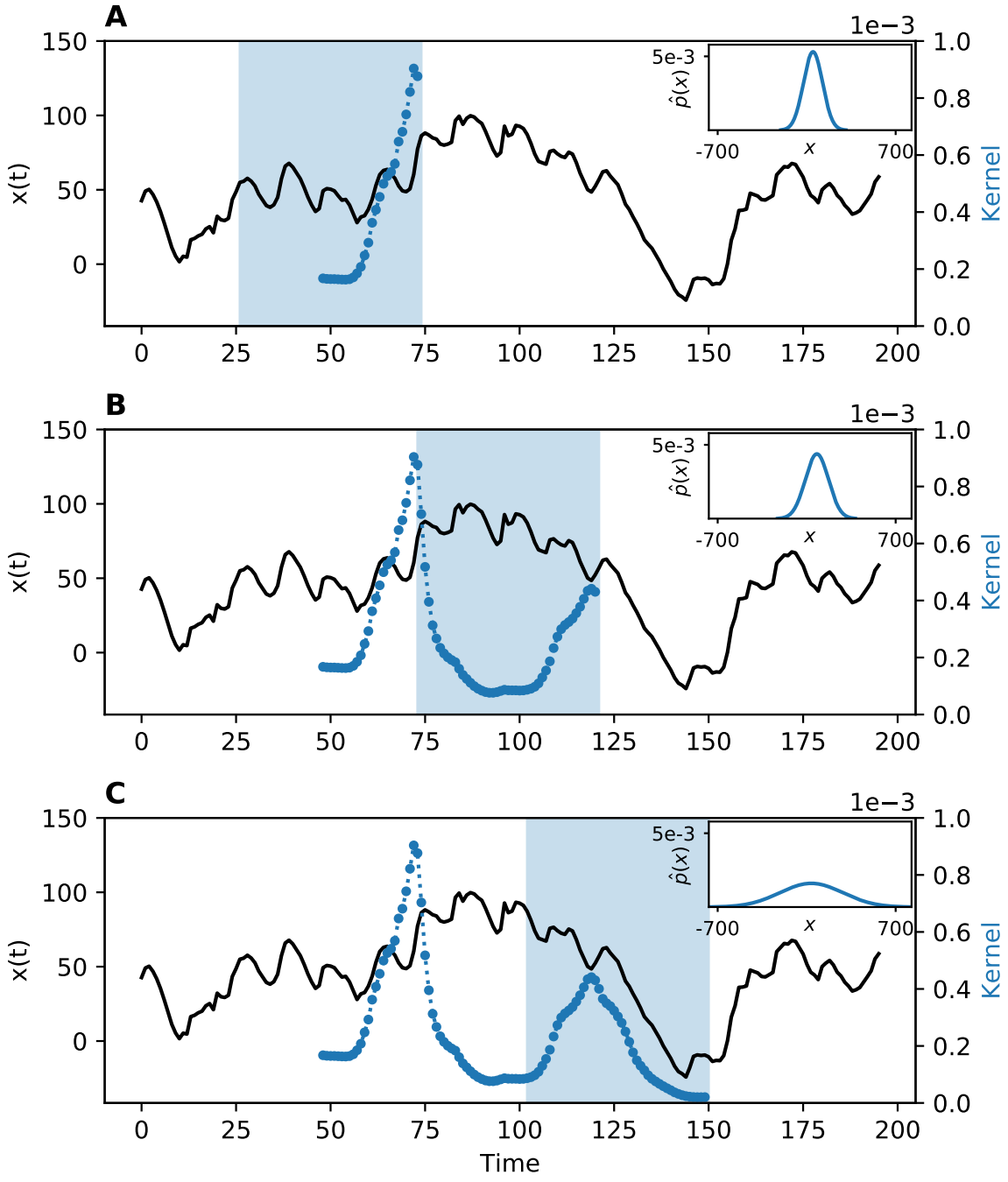


**Figure 5.4** Sliding windows for the Kernel Density Estimator Method. The blue shaded region indicates the temporal location of the sliding window. The inset shows the shape of the Kernel Density Estimate.

# 6 Applying the Sociotechnical Water Supply Framework for Falls Lake, Raleigh, North Carolina

## 6.1 Case Study Description

The City of Raleigh, North Carolina, has a population of approximately 486,000 inhabitants and covers an area of 142.8 square miles, as of 2013. According to the US Census, the population of Wake County, where Raleigh is located, grew 43.5% between 2000 and 2010. This is the highest growth of any metropolitan area in the U.S. for that period. Raleigh's primary source of drinking water is Falls Lake, which is a man-made reservoir located on the upper Neuse River and managed by the U.S. Army Corps of Engineers [TP11]. A major drought in 2007 led to a drop of 10 ft. in the water level of Falls Lake, and significant conservation efforts were enforced. The City of Raleigh began exploring options to expand its water supply in 2006 to continue to meet demands of projected urbanization and economic growth [Str19].

The Sociotechnical Water Supply Framework is applied to simulate water supply and demand dynamics for Falls Lake and the City of Raleigh for projected climate change scenarios. The agent-based model is encoded with data representing households in Raleigh, NC, and the stochastic reconstruction framework is fit using historical flow, evapotranspiration, and precipitation data. Three methods for calculating Fisher information are applied for model outputs to detect and forecast regime shifts in water supply.

## 6.2 Climate Input Data

Monthly historical data from 1983–2014 is used with the stochastic reconstruction framework to generate shifted climatic data as input for the agent based model. The historical data set is made publicly available by the U.S. Army Corps of Engineers [U.S] and includes precipitation, evapotranspiration, and inflow.

The temporal dependence of inflow values is maintained using the copula method. Copula functions to characterize temporal dependence between pairs of sequential months are found using the `VineCopula` package in `R`. The function `BiCopSelect`, which is in the `VineCopula` package, is

used to generate the appropriate copula functions for each pair of monthly data sets. Independence is checked using Kendall's $\tau$ [GF07] with a p-value of 0.05, and the copula function is determined based on Akaike Information Criterion [Nag19]. The results of the independence test and the copulas generated for each month are listed in Table 6.1. Equations for each copula functions are provided by Brechmann [Bre10] and Eschenburg [Esc13].

**Table 6.1** Copula function information for each month. Data sets are considered independent for a p-value greater than 0.05. $C_{X,Y}$ indicates the copula function for months $X$ and $Y$.

| Month ($a$) | Independence Test (p-value) | Copula function | $C_{X,Y}$ |
|---|---|---|---|
| Jan (1) | <0.05 | BB7 | $C_{12,1}$ |
| Feb (2) | <0.05 | Joe | $C_{1,2}$ |
| Mar (3) | <0.05 | Gaussian | $C_{2,3}$ |
| Apr (4) | <0.05 | Joe | $C_{3,4}$ |
| May (5) | 0.092 | *Independent* | $C_{4,5}$ |
| Jun (6) | <0.05 | Gaussian | $C_{5,6}$ |
| Jul (7) | <0.05 | Rotated Tawn Type 2 180 degrees | $C_{6,7}$ |
| Aug (8) | 0.168 | *Independent* | $C_{7,8}$ |
| Sep (9) | 0.376 | *Independent* | $C_{8,9}$ |
| Oct (10) | <0.05 | Tawn type 2 | $C_{9,10}$ |
| Nov (11) | <0.05 | Frank | $C_{10,11}$ |
| Dec (12) | 0.172 | *Independent* | $C_{11,12}$ |

The reconstruction framework generates monthly data. To generate a new inflow value for a particular month, the copula function that corresponds to that month (Table 6.1) is applied. The steps for producing projected inflow values are:

1. Create shifted CDFs $F_{\text{Inflow}_t^*}$ for $t = 1, 2, ..., T$ using quantile mapping, described in section 3.2. $T$ is the total number of time steps in the projected period.
2. Sample a random number, $x_1$, from a uniform distribution, $\mathscr{U}(0,1)$.
3. Use $x_1$ to generate the first inflow value, $\text{Inflow}_1^*$, from the inverse of the shifted CDF, $F^{-1}_{\text{Inflow}_t^*}$.
4. For monthly time steps $t = 2, 3, ..., T$:
   (a) Sample a new random number, $w$, from a uniform distribution, $\mathscr{U}(0,1)$.
   (b) Use the monthly copula to generate a new random number, $x_t$ from the inverse of the copula: $x_t = C^{-1}_{a-1,a}(w)$, where $a$ is the month corresponding to the time step $t$ ($a = t$

mod 12) + 1).

(c) Generate the next inflow value, $\text{Inflow}_t^*$, using the inverse of the shifted CDF, $F^{-1}_{\text{Inflow}_t^*}(x_t)$.

The precipitation and evapotranspiration data are not shifted through the copula method. For these data, the shifted CDFs are created using the quantile mapping method. For the precipitation and evapotranspiration data, shift factors $s_{p,t}$ and $s_{e,t}$, respectively, are correlated to inflow values. The following equations are used to modify the shift factor for precipitation and evapotranspiration data.

$$s_{p,t} = 1-(1-s_{f,t})\times r_{a,p}) \tag{6.1}$$

$$s_{e,t} = 1-(1-s_{f,t})\times r_{a,e}) \tag{6.2}$$

where $s_{f,t}$ is the shift factor at the current time step, and $r_{a,p}$ is the Pearson correlation coefficient for inflow and precipitation at month $a$, corresponding to time step $t$. Similarly, $r_{a,e}$ is the Pearson correlation coefficient for inflow and evaporation at month $a$, corresponding to time step $t$. Pearson correlation coefficients were determined from historic data, as reported by Mashhadi Ali et al. [MA17]. Quantile mapping is applied to generate projected precipitation and evaporation data using the shift factors $s_{p,t}$ and $s_{e,t}$. New values for precipitation and evapotranspiration are generated by randomly sampling a uniform distribution, $\mathcal{U}(0,1)$ at each time step, and using the random value to calculate values from the inverse CDF.

## 6.3 Stochastic Reconstruction Framework Results

The stochastic reconstruction framework is applied using a shift factor, $s_f$, of 0.8 and 0.2 to generate projections of monthly inflow data for Falls Lake, from 2015-2065, or 600 time steps. The stochasticity in inflow can be seen in Fig. 6.1, which shows model results for five random simulations using $s_f = 0.8$. These results are highly variable and do not show apparent similarities in trends or patterns. The stochastic reconstruction framework was executed for 50 random simulations using $s_f$ equal to 0.8 and 0.2. The mean of the two sets of simulations shows a seasonal pattern that mimics historical pattern (Fig. 6.2). The mean for the 0.2 shifted runs is much lower as time progresses when compared to the 0.8 shifted runs, as a result of applying the stochastic reconstruction framework.

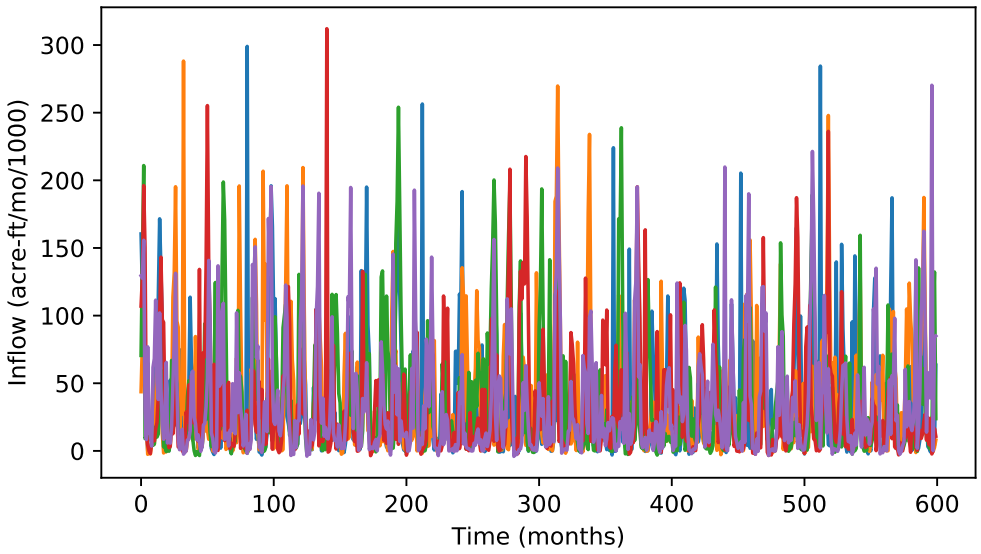


**Figure 6.1** Five model simulations of inflow (acre-ft/mo/1000) for $s_f = 0.8$.

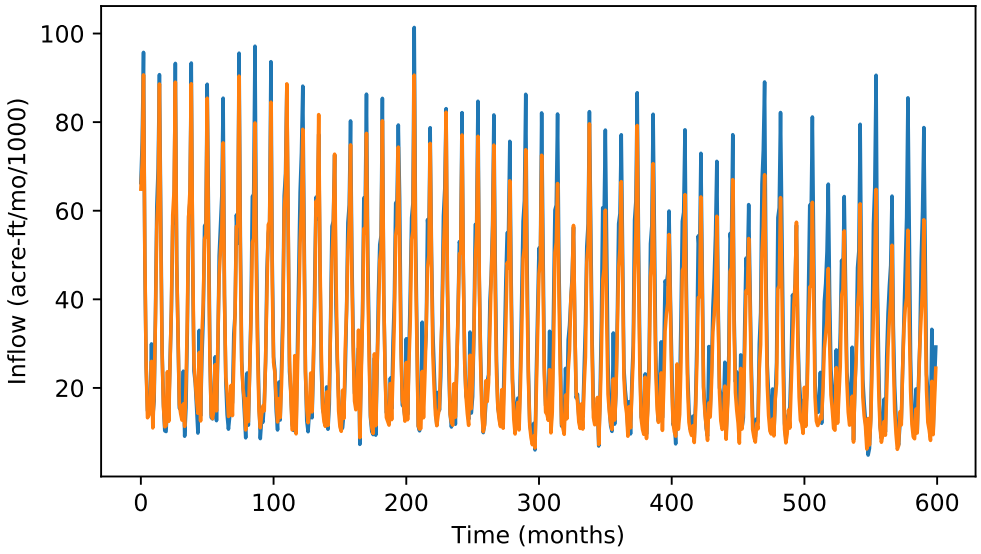


**Figure 6.2** Mean of 50 model simulations of inflow (acre-ft/mo/1000) for $s_f = 0.8$. (——) and $s_f = 0.2$ (——).

## 6.4 Agent-based Modeling Framework Results

The agent-based modeling framework was applied to simulate demands and reservoir storage using output from the stochastic reconstruction framework. Stochasticity in precipitation and evapotranspiration contribute to stochasticity in agent outdoor demand, which is a function of these climate parameters. Additional randomness in agent calculations is associated with the creation of agents and agent-level decision-making. As a result, projected demands are highly variable between model runs. The stochasticity in demands can be seen in model results for five random simulations using $s_f = 0.8$ (Fig. 6.3). Stochasticity in inflows and demands generate stochasticity in reservoir storage, and the time series of storage values for five random simulations using $s_f = 0.8$ are shown in Fig. 6.4. There is high stochasticity between runs, and each model run generates a unique set of data.

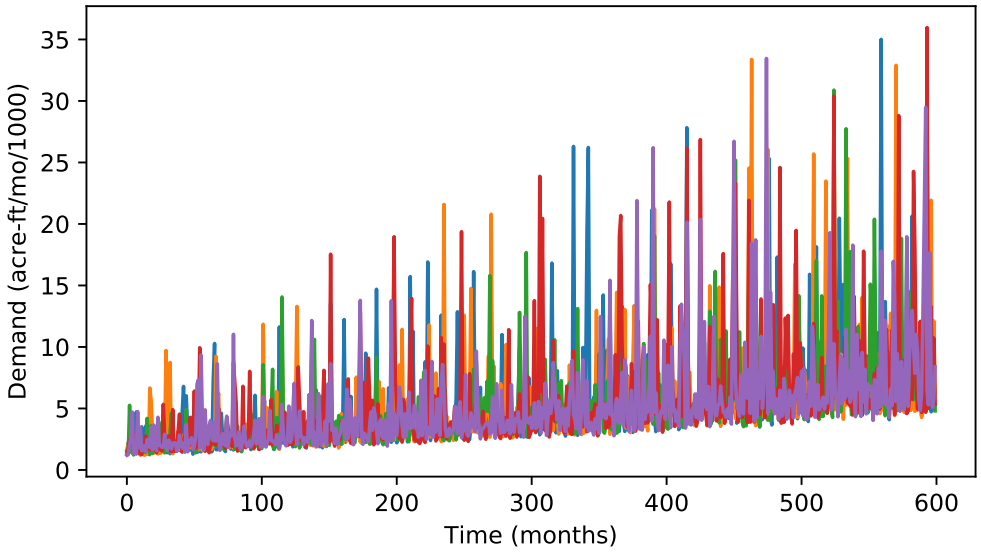


**Figure 6.3** Five model simulations of demand (acre-ft/mo/1000) for $s_f = 0.8$.

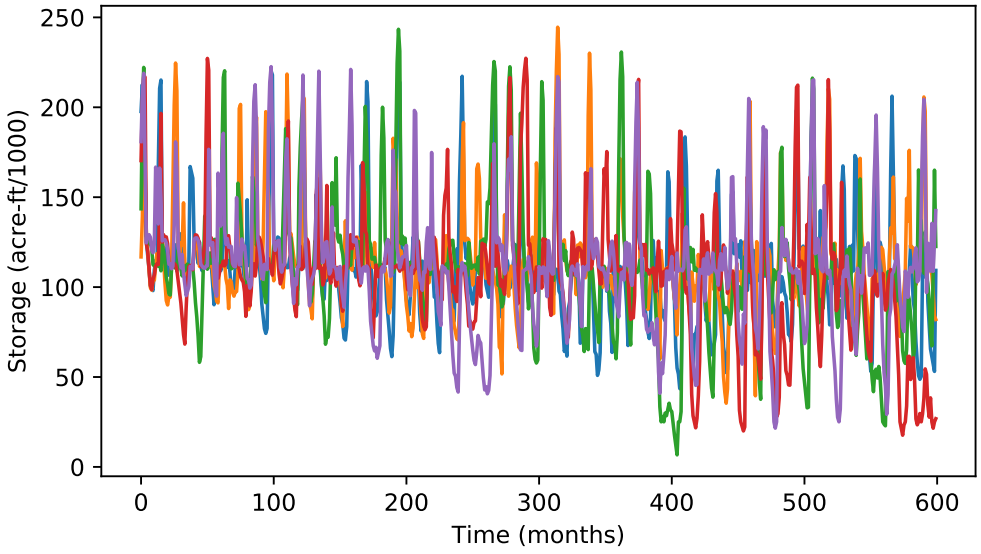


**Figure 6.4** Five model simulations of storage (acre-ft/1000) for $s_f = 0.8$.

To make temporal trends more clear, the mean demand data of 50 model outputs for $s_f$ equal to 0.2 and 0.8 are shown (Fig. 6.5). Mean demands grow throughout the simulation period due to population growth for both climate scenarios. The mean demands show only slight differences for the two climate settings, and the difference emerges more clearly at later simulation steps.

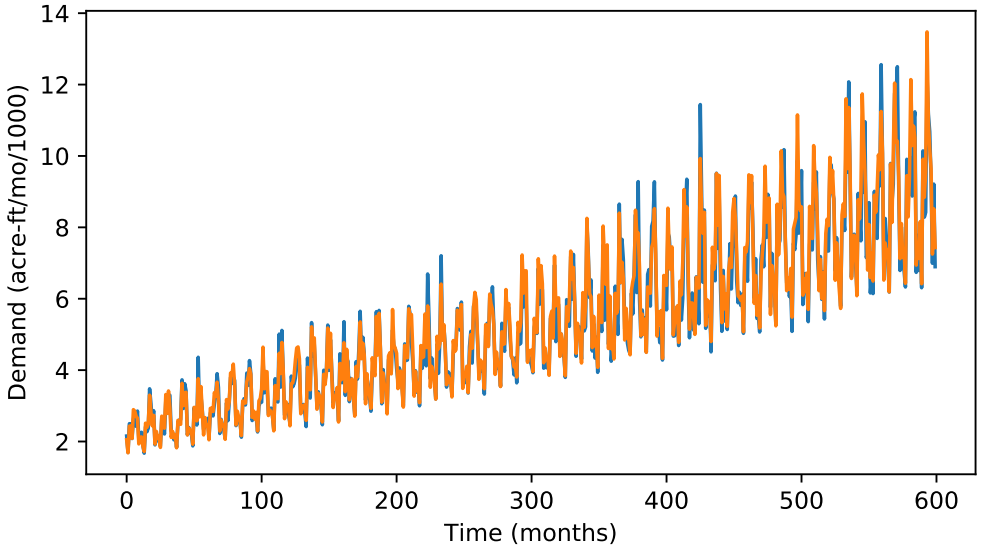


**Figure 6.5** Mean of 50 model simulations of demand (acre-ft/mo/1000) for $s_f = 0.8$ (——) and $s_f = 0.2$ (——).

## 6.5 Fisher Information Results

The Sociotechnical Water Supply Framework was applied to the Raleigh data to simulate 50 years of population growth and climate change from 2015–2065. The model was run 50 times using two shift factors ($s_f$): 0.2 and 0.8. Fisher information was calculated for the Raleigh storage data using three methods: 1) Amplitude-based Discrete Method with disjoint bins, 2) Amplitude-based Discrete Method with overlapping bins, and 3) the Kernel Density Estimator Method. For each setting, the sliding window width is set to 47 months ($w = 47$); the sliding factor is set to 1 month ($\Delta = 1$); and $k$ is set to 2 ($k = 2$). The same settings were used for the Cantareira System data. The interpretation of the Fisher information output is below. An assessment of the effects of model stochasticity on the Fisher information results are discussed in a subsequent section.

### 6.5.1 Fisher Information for Shift Factor = 0.2

A shift factor of 0.2 represents a relatively significant shift in climate and should lead to shifting regimes in the water supply system. The three methods for calculating Fisher information are evaluated as they coincide with the time series of reservoir storage, which is the output from the agent-based modeling framework for one simulation. The time series of storage data is shown in Fig. 6.6. The storage volume shows a significant decrease around 300 months, followed by a short recovery period, and another sharp decline after 450 months. The system does not appear to recover

after this second decline.

The Amplitude-based Discrete Method with overlapping bins is highly variable and difficult to interpret (Fig. 6.6). Fisher information values are highly stochastic, with peaks that could be interpreted at 100 months, 250 months, and 450 months. Similar to the results of applying this method for the Cantareira data set, results are difficult to generalize and indicate a higher number of regime shifts than would be expected from the data set.

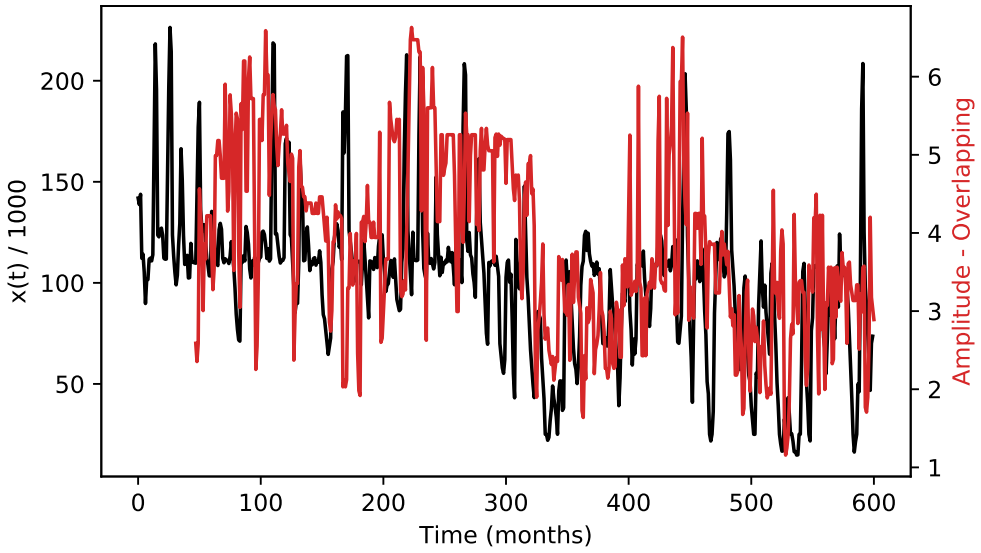


**Figure 6.6** Fisher information for Falls Lake for 2015–2065 with $s_f = 0.2$. Lines represent calculations using the Amplitude-based Discrete Method with overlapping bins (——) and reservoir storage (——).

The Amplitude-based Discrete Method with disjoint bins (Fig. 6.7) generates Fisher information values showing that the system organizes around 100 months, followed by a disorganization at 150 months. The system begins to reorganize around 190 months and peaks shortly thereafter around 205 months, suggesting that the system is within one system state during this time period. The Fisher information declines and stabilizes through 250 months, which may suggest that the system is shifting among two or more states. After 250 months, Fisher information suddenly declines over 50 months, reaching a low point at 300 months. This period suggests that the system is completely disordered. Around month 350, the Fisher information slowly rises over the following 100 months reaching another local maxima around 450 months. This peak reaches the same approximate value of Fisher information as the period 250–300, which may suggest that the system has again reorganized into two states. However, based on the storage values within this time period, the system could be in different states compared to months 200–250. The Fisher information declines sharply from the

period 450-500 and does not recover again.

The Fisher information calculated using Amplitude-based Discrete Method with disjoint bins can be interpreted and compared with system dynamics. The decline in Fisher information during time steps 250-300 correspond to a marked change in storage dynamics, during which the system approaches a new regime in the regular pattern of storage values. A second significant decline occurs between time steps 450-500, when the system again switches to a new regime of storage values. For this data set, the Amplitude-based Discrete Method with disjoint bins suggests regime shifts at appropriate times, based on visual inspection of storage data.

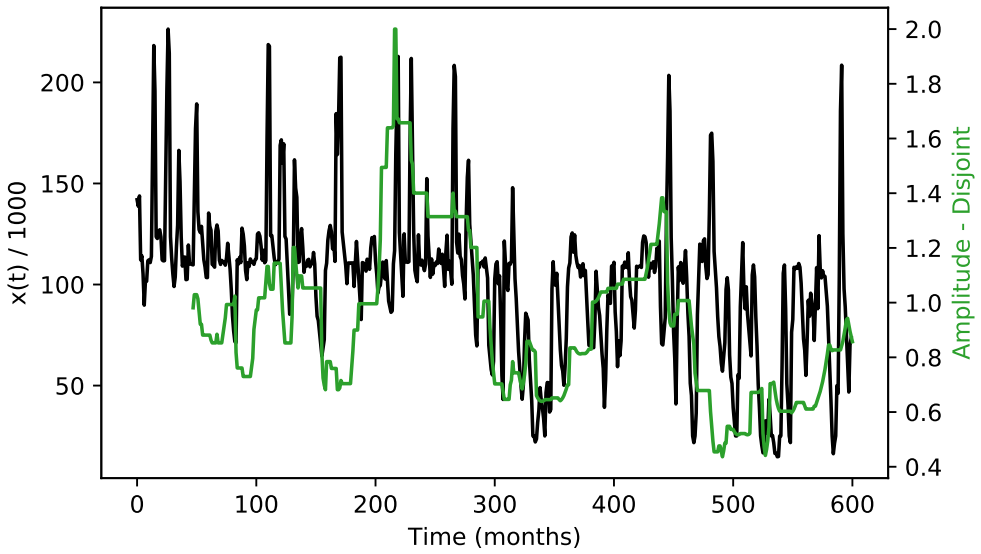


**Figure 6.7** Fisher information for Falls Lake for 2015–2065 with $s_f = 0.2$. Lines represent calculations using the Amplitude-based Discrete Method with discrete bins (—) and reservoir storage (—).

The Kernel Density Estimator Method was applied for the time series of data (Fig. 6.8). A high peak occurs just after 100 months. The peak and subsequent drop-off in Fisher information that is calculated using the Kernel Density Estimator Method do not correspond with a regime shift in storage data, but accompany a period of highly stable flows, during which the probability of the system within one state is high. Later in the time series, a period of slowly increasing Fisher information begins at approximately month 380 and continues through month 450. This increase demonstrates that the system is organizing within a small number of states, with storage oscillating consistently among a small set of values. A subsequent drop-off during time steps 450-500 accompanies a regime shift in the storage, where values oscillate between a larger set of states. The calculation of Fisher information captures a second regime shift that may be interpreted in the data (around time step

450-500), though it does not detect the first and more apparent regime shift, located approximately at time step 250-300.

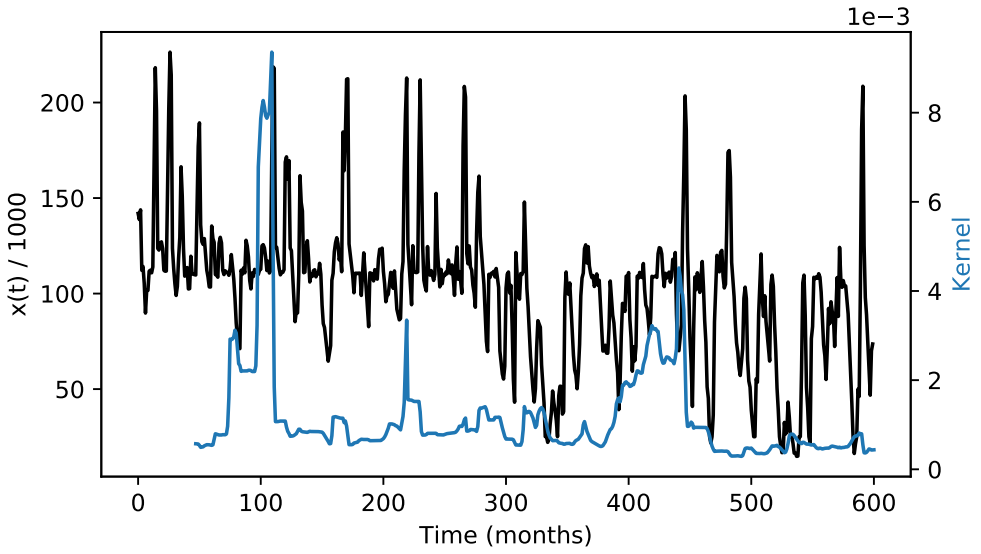


**Figure 6.8** Fisher information for Falls Lake for 2015–2065 with $s_f = 0.2$. Lines represent calculations using the Kernel Density Estimator Method (——) and reservoir storage (——).

### 6.5.2 Fisher Information for Shift Factor = 0.8

The three methods for calculating Fisher information are evaluated for a time series of storage data that was simulated using $s_f = 0.8$. The storage data for this simulation (shown in Fig. 6.9) shows a regime shift occurring during time steps 300-400, after which storage values are more frequently in a lower range. Another shift may occur beginning around time step 450, when the system begins to oscillate among a wider range of storage values.

The Amplitude-based Discrete Method with overlapping bins again shows high variability and is difficult to interpret (Fig. 6.9). There is a trend with a decreasing average during the time steps 300-400 and 450-500 corresponding to changes in the storage data. Other periods of declining trends are also apparent in the Fisher information. For example, a sharp decline is located at approximately time steps 180-200, and this does not correspond to a significant shift in the storage data.

Results from the Amplitude-based Discrete Method with disjoint bins (Fig. 6.10) can be more easily interpreted as follows. The system remains relatively stable through 150 months before losing stability. The system goes through a period of instability before transitioning into a stable state

around 250 months. Around 300 months, a significant drop in Fisher information indicates the beginning of a transition out of the stable state and the system remains unstable for 100 months. Fisher information starts to rise again around 450 months, climbing steadily until it drops off again at 500 months. The loss of Fisher information around time steps 300 and 450 corresponds with the observed shift in storage during those periods. The timing of these regime shifts correspond with the dynamics seen in the storage data through visual inspection.

The Kernel Density Estimator Method was applied to the storage data that was generated using $s_f = 0.8$ (Fig. 6.11). The kernel-based method calculates higher peaks in the Fisher information that can be more readily interpreted as regime shifts. Spikes in the Fisher information indicate that the system is highly ordered and in a stable state. When the Fisher information falls, the system has transitioned out of stability. Results shown in Fig. 6.11 suggests that the system experienced four periods of stability, followed by changes in the oscillation among system stages: 50–100 months, 150–170 months, 320–350 months, and a brief period near 500 months. These spikes in Fisher information correspond with regime shifts that are observed at time steps 300-400 and around time step 450. The regime shifts that are detected using the Kernel Density Estimator Method also correspond with regime shifts detected using the Amplitude-based Discrete Method with disjoint bins.

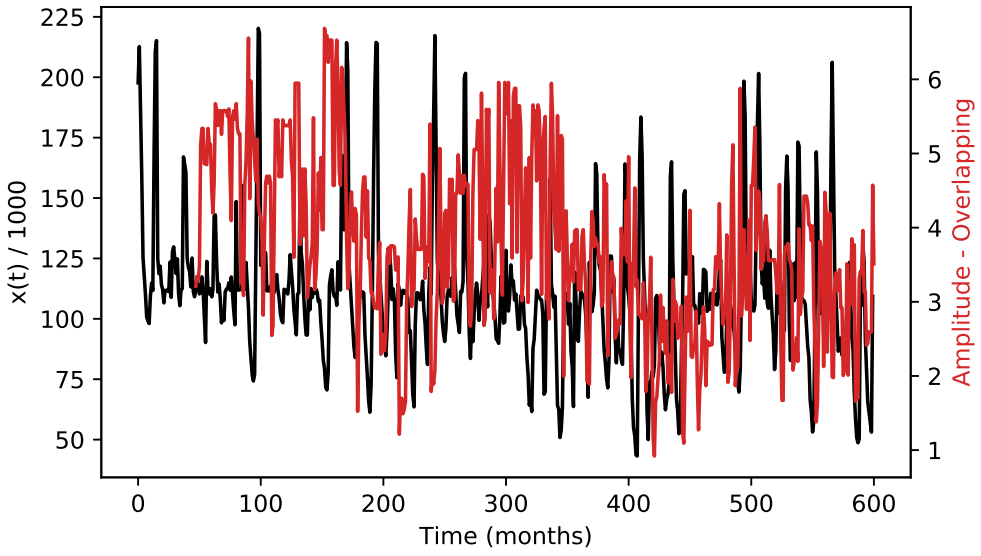


**Figure 6.9** Fisher information for Falls Lake for 2015–2065 with $s_f = 0.8$. Lines represent calculations using the Amplitude-based Discrete Method with overlapping bins (—) and reservoir storage (—).

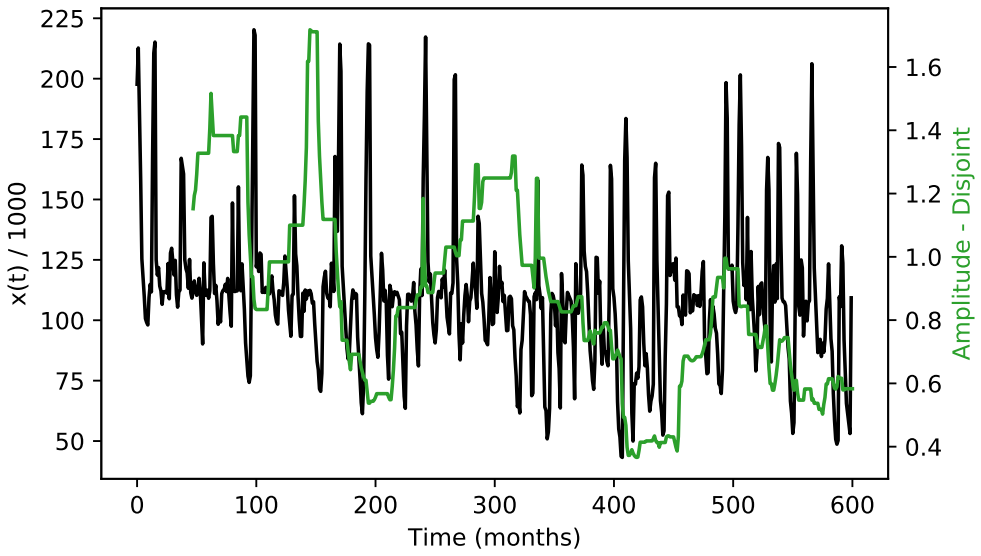


**Figure 6.10** Fisher information for Falls Lake for 2015–2065 with $s_f = 0.8$. Lines represent calculations using the Amplitude-based Discrete Method with discrete bins (—) and reservoir storage (—).

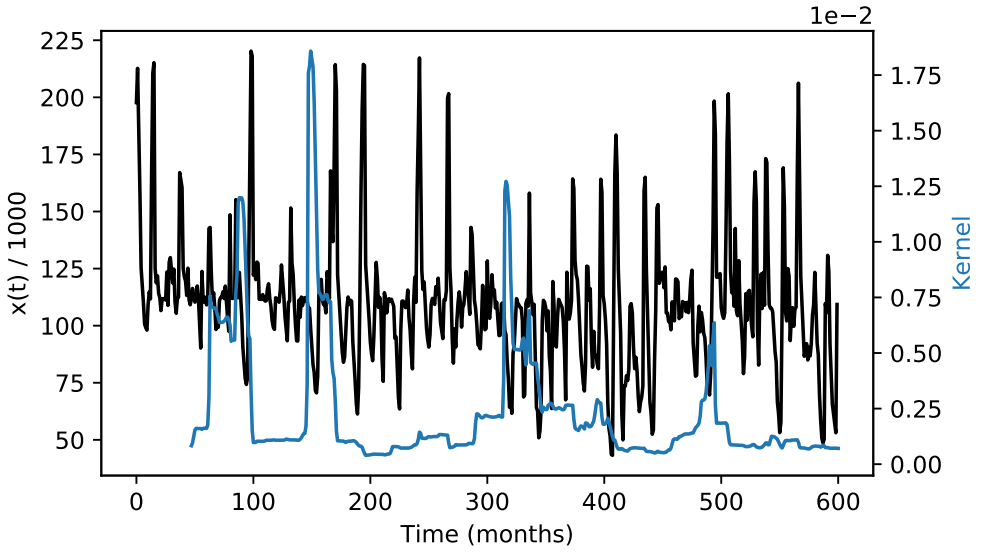


**Figure 6.11** Fisher information for Falls Lake for 2015–2065 with $s_f = 0.8$. Lines represent calculations using the Kernel Density Estimator Method (—) and reservoir storage (—).

### 6.5.3 Stochasticity

Model results that are simulated using the agent-based modeling framework and stochastic reconstruction framework are highly variable. As a result, there is a high degree of variability in the detection of regime shifts found through applying Fisher information methods. Fisher information was calculated for the same five model runs shown in section 6.2 using the Amplitude-based Discrete Method with discrete bins and the Kernel Density Estimator Approach, shown in Fig. 6.12 and Fig. 6.13, respectively. For the Amplitude-based Discrete Method, no single time period of importance emerges from the set of results. The Fisher information shows a declining value across the simulations, but discerning regime shifts may be subject to the bias of the observer. Due to gradual fluctuations in the Fisher information, visual inspection of each output is required to gain an understanding of the system.

Results generated by the Kernel Density Estimator Method Fig. 6.13 also indicate that no important periods of time emerge, but regime shifts are more apparent based on spikes in Fisher information. A more automated method may be used to detect regime shifts based on a threshold that can be applied to peak values. A threshold may be specified as a percentage of the maximum value across Fisher information series, and peaks above this threshold can be considered significant. Using a threshold of 0.01, for example, the number of regime shifts for each data set can be identified through visual inspection: three shifts for the blue line; two shifts for the red line; one shift for the green line; zero shifts for the purple line; and two shifts for the orange line. The timing of these shifts is not consistent across the data sets, though the majority of the shifts occur in the first 300 time steps of the simulated period. These numbers can be used toward gaining an understanding of system resilience through an automated approach.

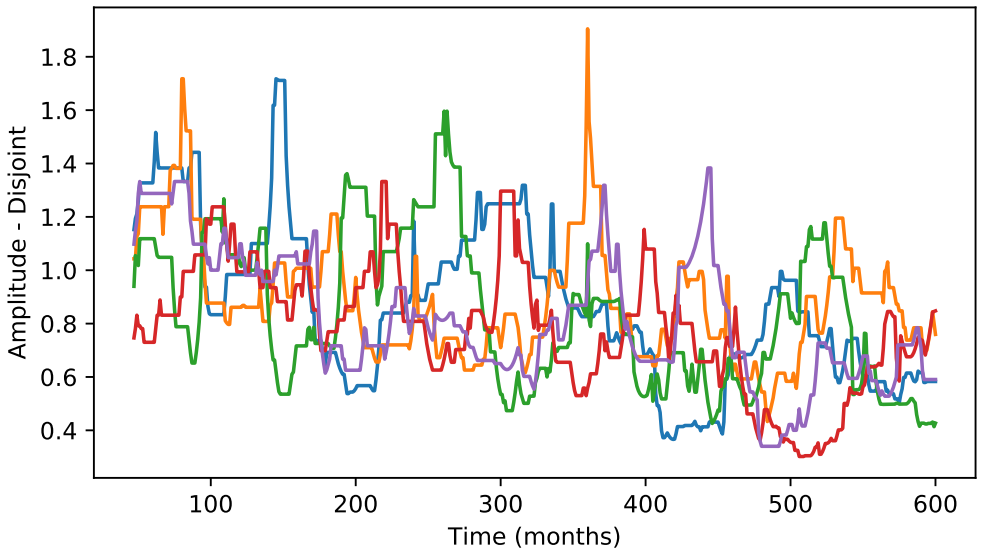


**Figure 6.12** Fisher information for Falls Lake for 2015–2065 with $s_f = 0.8$. Fisher information is calculated using the Amplitude-based Discrete Method with discrete bins for five simulations of the Sociotechnical Water Supply Framework.

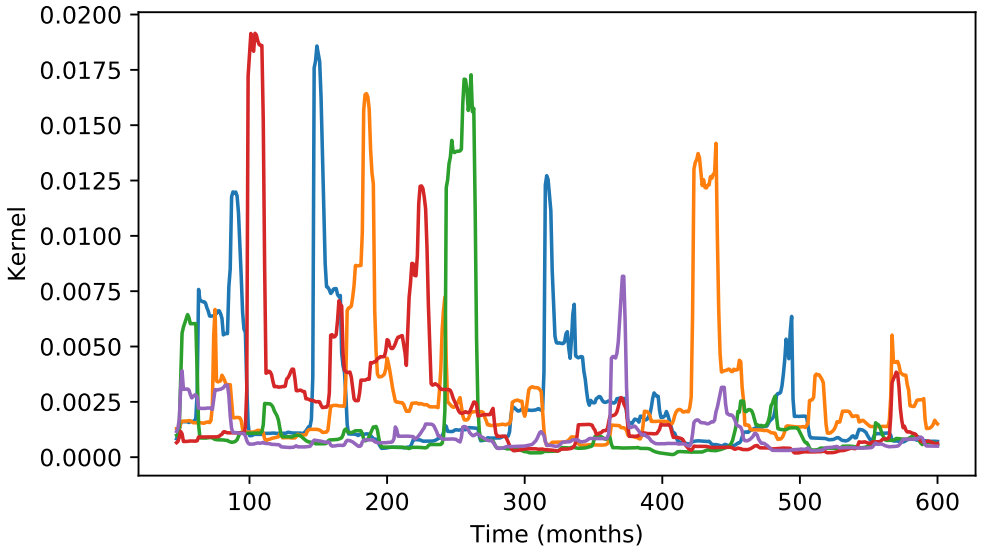


**Figure 6.13** Fisher information for Falls Lake for 2015–2065 with $s_f = 0.8$. Fisher information is calculated using the Kernel Density Estimator Method for five simulations of the Sociotechnical Water Supply Framework.

# 7 Discussion

## 7.1 Calculating Fisher Information

This research implements three methods for calculating Fisher information for a time series of data. This research did not find that one method is recommended above others for most cases, and the advantages and disadvantages of alternative methods are described as follows.

The Amplitude-based Discrete Method with overlapping bins generates a series of Fisher information values that is highly variable and difficult to interpret for the case studies explored in this research. The use of overlapping bins means that some data points may be in multiple bins, and, as a result, the calculation of the Fisher information value can change dramatically when the sliding window moves by one time step. The method of overlapping bins was developed as part of a method to discretize Fisher information calculations [CF02] and has been demonstrated with application to other data sets. This research improves the application of the discrete methodology by developing an alternative approach for binning data points. The Amplitude-based Discrete Method with distinct bins assigns each data point to only one state and improves the Fisher information output. Dynamics in Fisher information calculated using the Amplitude-based Discrete Method with distinct bins can be more easily interpreted to detect shifting regimes.

The Amplitude-based Discrete Method with distinct bins and the Kernel Density Estimator Method can be compared for application across data sets. When compared with theoretical values of Fisher information for data sets of normally distributed data, the Kernel Density Estimator Method generates Fisher information values that are significantly more accurate than the Amplitude-based Discrete Methods. The two methods generate similarly accurate results for the time series of reservoir storage observed for the Cantareira Reservoir System in Brazil. For the reservoir data simulated for Falls Lake in Raleigh, the two methods differ when they are applied for different sets of simulated data. For a high shift in climate, the Kernel Density Estimator Method misses significant shifts. This calculation relies on a period of stability to detect a shift, and some shifts may appear when the system transitions from a highly variable state to an alternative state that is also highly variable. Specifically, reservoir systems oscillate naturally during wet and dry seasons, and there may be a wide variation of storage in one state. The Amplitude-based Discrete Method with distinct bins can capture transitions in the system when states are highly variable because states are defined

based on bins that may be wide enough to capture variations within one state. The performance of the Amplitude-based Discrete Method with distinct bins, however, is expected to depend on setting the size of the bins. The Kernel Density Estimator Method and the Amplitude-based Discrete Method with distinct bins generate similar results for a moderate shift in climate. The two methods generate similar peaks to indicate similar regime shifts. The Kernel Density Estimator Method, however, generates more pronounced peaks that may be easier to interpret. As shown for a set of five randomly seeded simulation results, the Kernel Density Estimator Method may be used more readily in a threshold-based approach to automatically detect regime shifts. The ability to automatically detect regime shifts and place them in time is important to develop a high-level understanding of system resiliency under stochastic simulations of shifting climate scenarios. Further research is needed to explore and develop Fisher information-based methods that can be applied automatically without the use of visual inspection.

The research conducted here makes important contributions to the science of detecting regime shifts. As described above, the discrete method developed by Cabezas & Fath is difficult to interpret when overlapping bins are used. This research developed a binning approach that improves the application of the discrete method, which is more readily implemented than the kernel density estimator approach. The Kernel Density Estimator Method was developed and described by Telesca et al. The authors report values of Fisher information for data sets of normally distributed data that match the theoretical values more closely than the implementation reported here. Because the code was not available for their work, we could not validate our calculations, but we were able to find values that approximate the theoretical values closely. In this research, we provide the code and step-by-step algorithms to implement the Kernel Density Estimator Method. This contribution may be important for engineers who wish to apply Fisher information methods to detect regime shifts for infrastructure or natural resources systems. The work documented here is meant to provide a step-by-step guide for implementing the methods for other engineered systems.

## 7.2 Water Supply Resilience

This research demonstrates an important contribution for water supply system management. To date, resilience has been explored for water supply systems based on an engineering understanding of recovery from failure. This research creates methods that can be used to calculate resilience based on the ability to absorb shocks and maintain stability. The methods demonstrated in this research calculate transitions in system variables to detect when a system has changed states. Additional parameters, such as water deficits and reservoir release, may be included to better isolate transitions in water supply. Further research can explore methods to automatically detect transitions and calculate resilience, based on the Fisher information calculations developed here. New methods can include thresholds to indicate regions where shifting regimes occur, based on preliminary analysis

explored in this research. As shown above, the Kernel Density Estimator Method generates results that may be more easily interpreted through the use of thresholds. Resilience calculations can be applied to evaluate alternative demand-side and supply-side management strategies for improving water supply systems.

Agent-based modeling has been applied to study a range of water resources systems [Ber15]. Deciphering the output of agent-based models as a time series of dynamic data can be difficult, and changes in system parameters may be challenging to characterize. Fisher information provides an approach to characterize changes in a system and to capture transitions at a high level. Further application of Fisher information methods for the output of agent-based models can improve the knowledge and insight gained from complex simulations.

# 8 Conclusions

This research presented three methods for calculating Fisher information for time series data: an Amplitude-based Discrete Method with disjoint bins, an Amplitude-based Discrete Method with overlapping bins, and a Kernel Density Estimator method. Each method was compared against theoretical Fisher information for sets of normally distributed data. The amplitude-based methods show considerable deviations from the theoretical curve, whereas the kernel-based method deviates from the theoretical values, but follows the same general trend. The three methods were used to detect shifting regimes for two case studies for water supply systems: historic data was obtained to describe the Cantareira Reservoir System in São Paulo, Brazil, and an agent-based model was used to simulate data for Falls Lake in Raleigh, North Carolina. For the Brazil case study, each of the three Fisher information calculation methods detected two transition periods, during which the system reorganized and transitioned among states. Results from the three methods were not conclusive for the Raleigh case study. The Amplitude-based Discrete Method with overlapping bins generated highly variable Fisher information values that are difficult to interpret. The Amplitude-based Discrete Method with distinct bins and the Kernel Density Estimator Method performed similarly well to identify regime shifts. The two methods were tested for two climate change scenarios. While both methods identified regime shifts in periods where the storage data did not indicate any shift, the Kernel Density Estimator failed to detect a significant regime shift for one climate scenario. The development of the Amplitude-based Discrete Method with distinct bins provides an approach that can be easily implemented, generates output that can be more easily interpreted than overlapping bins, and detects regime shifts similar to the Kernel Density Estimator Method. For application to a large number of time series data sets, the Kernel Density Estimator Method generates output with distinct peaks to demarcate regime shifts. In this research, visual inspection is applied to identify shifts, and further research is needed to develop methods that can automatically detect shifts.

The Sociotechnical Water Supply Framework that was developed in this research incorporates a stochastic reconstruction framework to generate climate shifted scenarios. Within the framework, quantile mapping was used to maintain the historical distribution of inflows, and copulas were incorporated to preserve the temporal dependence of inflow values. This framework was adapted to reflect a shifting climate and incrementally update the distribution of hydrologic flows at each time step. The climate output from the framework was used as the input to the agent-based mod-

eling framework, which simulates supply and demand dynamics for a city with a rapidly growing population. The model output is highly variable, and the output of each run is a unique set of data that should be analyzed individually.

This research demonstrates the implementation of Fisher information calculations and a new application of these methods to detect shifting regimes in water supply systems. A new framework is developed for assessing the resilience of water supply based on the ability to absorb shocks and maintain stability. The methods demonstrated in this research calculate transitions in system variables to detect when a system has changed states and create new knowledge about dynamics in water resources that behave as complex systems.

## BIBLIOGRAPHY


[Ama16] Amarasinghe, P. et al. "Quantitative assessment of resilience of a water supply system under rainfall reduction due to climate change". *Journal of Hydrology* **540** (2016), pp. 1043–1052.

[And09] Andersen, T. et al. "Ecological thresholds and regime shifts: approaches to identification". *Trends in Ecology and Evolution* **24**.1 (2009), pp. 49–57.

[Bah17] Baho, D. L. et al. "A quantitative framework for assessing ecological resilience". *Ecology and Society* **22**.3 (2017).

[Ber15] Berglund, E. Z. "Using Agent-Based Modeling for Water Resources Planning and Management". *Journal of Water and Resources Planning Management* **141**.11 (2015), pp. 1–17.

[Bre10] Brechmann, E. C. "Truncated and simplified regular vines and their applications". PhD thesis. Technische Universität München, 2010.

[CF02] Cabezas, H. & Fath, B. D. "Towards a theory of sustainable systems". *Fluid Phase Equilibria* **194-197** (2002), pp. 3–14.

[Cou15] Coutinho, R. M. et al. "Catastrophic Regime Shift in Water Reservoirs and São Paulo Water Supply Crisis". *PLoS ONE* **10**.9 (2015), e0138278.

[DeO11] DeOreo, W. B. et al. *CALIFORNIA SINGLE FAMILY WATER USE*. Tech. rep. 2011, p. 371.

[EC12] Eason, T. & Cabezas, H. "Evaluating the sustainability of a regional system using Fisher information in the San Luis Basin, Colorado". *Journal of Environmental Management* **94**.1 (2012), pp. 41–49.

[Eas14] Eason, T. et al. "Managing for resilience: Early detection of regime shifts in complex systems". *Clean Technologies and Environmental Policy* **16**.4 (2014), pp. 773–783.

[Eas16] Eason, T. et al. "Managing for resilience: an information theory-based approach to assessing ecosystems". *Journal of Applied Ecology* **53**.3 (2016), pp. 656–665.

[Esc13] Eschenburg, P. "Properties of extreme-value copulas". PhD thesis. Technische Universität München, 2013.

[FR22] Fisher, R. A. & Russell, E. J. "On the mathematical foundations of theoretical statistics". *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **222**.594-604 (1922), pp. 309–368.

[Fol06] Folke, C. "Resilience: The emergence of a perspective for social-ecological systems analyses". *Global Environmental Change* **16**.3 (2006), pp. 253–267.

[FG07] Frieden, B. R. & Gatenby, R. A. *Exploratory data analysis using Fisher information*. English. Springer-Verlag, 2007, pp. 363, 363–xiii.

[FS95] Frieden, B. R. & Soffer, B. H. "Lagrangians of physics and the game of Fisher-information transfer". *Phys. Rev. E* **52** (3 1995), pp. 2274–2286.

[GF07] Genest, C. & Favre, A.-C. "Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask". *Journal of Hydrologic Engineering* **12**.4 (2007), pp. 347–368. eprint: https://ascelibrary.org/doi/pdf/10.1061/%28ASCE%291084-0699%282007%2912%3A4%28347%29.

[GA17] Gonzales, P. & Ajami, N. "Social and Structural Patterns of Drought-Related Water Conservation and Rebound". *Water Resources Research* **53**.12 (2017), pp. 10619–10634.

[Has82] Hashimoto, T. et al. "Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation". *Water Resources Research* **18**.1 (1982), pp. 14–20.

[Hol73] Holling, C. S. "Resilience and Stability of Ecological Systems". *Annual Review of Ecology and Systematics* **4**.1 (1973), pp. 1–23.

[Joã18] João, V. et al. "Integrated Assessment of Water Reservoir Systems Performance with the Implementation of Ecological Flows under Varying Climatic Conditions". *Water Resources Management* **32**.15 (2018), pp. 5183–5205.

[Jon96] Jones, M. C. et al. "A Brief Survey of Bandwidth Selection for Density Estimation". *Journal of the American Statistical Association* **91**.433 (1996), pp. 401–407.

[Kad18] Kadibadiba, T. et al. "Living in a city without water: A social practice theory analysis of resource disruption in Gaborone, Botswana". *Global Environmental Change* **53** (2018), pp. 273–285.

[KZ14] Kanta, L. & Zechman, E. "Complex Adaptive Systems Framework to Assess Supply-Side and Demand-Side Management for Urban Water Resources". *Journal of Water Resources Planning and Management* **140**.1 (2014), pp. 75–85.

[KHF08] Karunanithi H.; Frieden B. R.; Pawlowski, C. W.A.T. C. "Detection and Assessment of Ecosystem Regime Shifts from Fisher Information". *Ecology and Society* **13**.1 (2008).

[Mar99] Martin, M. T. et al. "Fisher's information and the analysis of complex signals". *Physics Letters, Section A: General, Atomic and Solid State Physics* (1999).

[MA17] Mashhadi Ali, A. et al. "Agent-based modeling to simulate the dynamics of urban water supply: Climate, population growth, and water shortages". *Sustainable Cities and Society* **28** (2017), pp. 420–434.

[MA13] Matonse Adão, H. et al. "Investigating the impact of climate change on New York City's primary water supply". *Climatic Change* **116**.3 (2013), pp. 437–456.

[Meh15] Mehran, A. et al. “A hybrid framework for assessing socioeconomic drought: Linking climate variability, local resilience, and demand”. *Journal of Geophysical Research: Atmospheres* **120**.15 (2015), pp. 7520–7533.

[Men15] Mens, M. J. P. et al. “Developing system robustness analysis for drought risk management: an application on a water supply reservoir”. *Natural Hazards and Earth System Sciences* **15**.8 (2015), pp. 1933–1940.

[Nag19] Nagler, T. et al. *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.2.0. 2019.

[Naz13] Nazemi, A. et al. “A stochastic reconstruction framework for analysis of water resource system vulnerability to climate-induced changes in river flow regime”. *Water Resources Research* **49**.1 (2013), pp. 291–305.

[Nel06] Nelsen, R. B. *An Introduction to Copulas*. Springer New York, 2006.

[OVSS18] O. Vieira, E. de & Sandoval-Solis, S. “Water resources sustainability index for a water-stressed basin in Brazil”. *Journal of Hydrology: Regional Studies* **19** (2018), pp. 97 –109.

[Pim84] Pimm, S. L. “The complexity and stability of ecosystems”. *Nature* **307**.5949 (1984), pp. 321–326.

[RD06] Raykar, V. C. & Duraiswami, R. *Very fast optimal bandwidth selection for univariate kernel density estimation*. Tech. rep. University of Maryland, 2006.

[Rom17] Romano, E et al. “Robust Method to Quantify the Risk of Shortage for Water Supply Systems”. *Journal of Hydrologic Engineering* **22**.8 (2017), p. 4017021.

[Sab] Sabesp, São Paulo, Brazil. *Situação dos Mananciais Sabesp - RMSP*.

[SS11] Sandoval-Solis, S et al. “Sustainability Index for Water Resources Planning and Management”. *Journal of Water Resources Planning and Management* **137**.5 (2011), pp. 381–390.

[Sch05] Schröder, A. et al. “MINI- Direct experimental evidence for alternative stable states : a review”. *Oikos* **1**.110 (2005), pp. 3–19.

[SJ91] Sheather, S. J. & Jones, M. C. “A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation”. *Journal of the Royal Statistical Society: Series B (Methodological)* **53**.3 (1991), pp. 683–690.

[Sta14] Standish, R. J. et al. “Resilience in ecology: Abstraction, distraction, or where the action is?” *Biological Conservation* **177** (2014), pp. 43–51.

[Str19] Stradling, R. “RaleighâĂŹs long search for a new water supply ends back home at Falls Lake”. *The News & Observer* (2019).

[TL17] Telesca, L. & Lovallo, M. “On the performance of Fisher Information Measure and Shannon entropy estimators”. *Physica A: Statistical Mechanics and its Applications* **484** (2017), pp. 569–576.

[Tel10] Telesca, L. et al. “Time-dependent Fisher Information Measure of volcanic tremor before the 5 April 2003 paroxysm at Stromboli volcano, Italy”. *Journal of Volcanology and Geothermal Research* (2010).

[Tel18] Tellman, B. et al. “Adaptive pathways and coupled infrastructure: seven centuries of adaptation to water risk and the production of vulnerability in Mexico City”. *Ecology and Society* **23**.1 (2018).

[TP11] The City of Raleigh & Public Utilities Department. *City of Raleigh’s Water Shortage Response Plan*. Tech. rep. 2011, pp. 1–53.

[Thr09] Thrush, S. F. et al. “Forecasting the limits of resilience: Integrating empirical research with theory”. *Proceedings of the Royal Society B: Biological Sciences* **276**.1671 (2009), pp. 3209–3217.

[Tom18] Tom, R. et al. “A Resilience-Based Methodology for Improved Water Resources Adaptation Planning under Deep Uncertainty with Real World Application”. *Water Resources Management* **32**.6 (2018), pp. 2013–2031.

[U.S] U.S. Army Corps of Engineers - Wilmington District Water Management. *Falls Lake — Neuse River Basin*. URL: `http://epec.saw.usace.army.mil/fall.htm` (visited on 07/27/2019).

[Vic01] Vickers, A. L. *Handbook of Water use and conservation*. English. Lewis Publishers, 2000 N.W. Corporate Blvd. Boca Raton, FL 33431 (USA), 2001, pp. 1–480.

[Wal04] Walker, B. et al. “Resilience, adaptability and transformability in socialâĂŞecological systems”. *Ecology and Society* **9**.2 (2004).

[WJ94] Wand, M. & Jones, M. *Kernel Smoothing*. 1st. Chapman and Hall/CRC, 1994, p. 224.

[Wat12] Watts, G. et al. “Testing the resilience of water supply systems to long droughts”. *Journal of Hydrology* **414-415** (2012), pp. 255–267.

[Zal08] Zaldívar, J.-M. et al. “Characterization of regime shifts in environmental time series with recurrence quantification analysis”. *Ecological Modelling* **210**.1-2 (2008), pp. 58–70.

[Zha18] Zhao, G. et al. “A modeling framework for evaluating the drought resilience of a surface water supply system under non-stationarity”. *Journal of Hydrology* **563** (2018), pp. 22–32.

# APPENDIX

# A Kernel Density Estimator Method

## A.1 A Numerical Method to Integrate Fisher Information for the Kernel Density Estimator Method

A discrete numerical method was implemented in python to solve Eq. 3.21. First, a sufficient range of values is needed to emulate the limits of integration, $[-\infty, +\infty]$. In this work, values which cover 99.98% of the density estimate are generated using the relationship: $\hat{F}^{-1}(0.0001) = \texttt{low}$ and $\hat{F}^{-1}(0.9999) = \texttt{high}$. An inverse cumulative function does not exist in the `scipy` kernel density estimator implementation. The `gaussian_kde` does include an `integrate_box_1d(low, high)` method which is equivalent to:

$$P(low < x \leq high) = \int_{low}^{high} \hat{p}(x)dx \tag{A.1}$$

The `gaussian_kde.integrate_box_1d(low, high)` method is used in a least-squares optimization problem to find the `low` and `high` values. The following two functions calculate the needed values:

```
from scipy.optimize import least_squares as ls

def _ls_get_value(x, k, p):
    return p - k.integrate_box_1d(-inf, x)

def _get_array_bounds(kernel, low_bound, high_bound):
    low = ls(_ls_get_value, 0, args=(kernel, low_bound))
    high = ls(_ls_get_value, 0, args=(kernel, high_bound))
    return float(low.x), float(high.x)
```

Once the `low` and `high` values are found, an array of values is created using `numpy.linspace()`

(e.g. `np.linspace(low, high, num=2 ** 11 + 1)`). This function creates `num` total, equally spaced points between `low` and `high`, inclusive. Here, the total number of points is set to $2^{11}+1$ so that the choice of integration methods is not limited.

After creating the array, the kernel is evaluated to give the probability of each point. Next, using centered differencing, the first derivative of the probabilities is calculated and then squared. The resulting values are divided by the initial probabilities, and the resulting array is integrated using either Romberg integration, trapezoidal, or Simpson's rule. Romberg integration requires $2^n+1$ evenly spaced data points. In this research, Romberg integration is used because it is the most efficient method of the three listed. These steps are summarized here:

```
low, high = _get_array_bounds(k, 0.0001, 0.9999)
vals = numpy.linspace(low, high, 2 ** 11 + 1)
probs = k.evaluate(vals)
p_prime_2 = numpy.gradient(probs, vals)
fi_integral = scipy.integrate.romb(p_prime_2 / vals)
```

The result of this integration is the Fisher information value of the set, $S$.