THE Water Research FOUNDATION

PROJECT NO.
4703

Long Term Vulnerability Assessment and Adaptation Plan for the San Francisco Public Utilities Commission Water Enterprise - Phase I

San Francisco Water Power Sewer

Long Term Vulnerability Assessment and Adaptation Plan for the San Francisco Public Utilities Commission Water Enterprise - Phase I

Prepared by:

Baptiste François, Alexa Bruce, Khanh Nguyen, Dong Kwan Park, and David Rheinheimer
University of Massachusetts, Amherst

Umit Taner
University of Massachusetts, Amherst; and Deltares

Sungwook Wi and Hassaan Khan
University of Massachusetts, Amherst

Alexis Dufour and David Behar
San Francisco Public Utilities Commission

David Yates and Caspar Ammann
National Center for Atmospheric Research

Marjolijn Haasnoot
Deltares

Casey Brown
University of Massachusetts, Amherst

Co-sponsored by:

San Francisco Public Utilities Commission

2021

THE Water Research FOUNDATION

The Water Research Foundation (WRF) is a nonprofit (501c3) organization that provides a unified source for One Water research and a strong presence in relationships with partner organizations, government and regulatory agencies, and Congress. WRF conducts research in all areas of drinking water, wastewater, stormwater, and water reuse. The Water Research Foundation's research portfolio is valued at over $700 million.

WRF plays an important role in the translation and dissemination of applied research, technology demonstration, and education, through creation of research-based educational tools and technology exchange opportunities. WRF serves as a leader and model for collaboration across the water industry and its materials are used to inform policymakers and the public on the science, economic value, and environmental benefits of using and recovering resources found in water, as well as the feasibility of implementing new technologies.

For more information, contact:

The Water Research Foundation

1199 North Fairfax Street, Suite 900
Alexandria, VA 22314-1445
P 571.384.2100

6666 West Quincy Avenue
Denver, Colorado 80235-3098
P 303.347.6100

www.waterrf.org
info@waterrf.org

WRF ISBN: 978-1-60573-565-8
WRF Project Number: 4703

Prepared by University of Massachusetts, Amherst; San Francisco Public Utilities Commission; National Center for Atmospheric Research; and Deltares

Acknowledgments

The research team wishes to thank the Water Research Foundation (WRF) and the San Francisco Public Utilities Commission (SFPUC) for funding this project. Many individuals within SFPUC contributed to this effort by providing data and insights about the SFPUC regional water system. This is especially the case of Alexis Dufour who dedicated a significant amount of time working with the team in all topics and items developed throughout the project. Others have also significantly contributed; Matt Moses, Amod Dhakal, Chris Graham, Reggie Walters, Erin Franks, Mike Brown, Kristina Cordero, Enio Sebastiani, Nicholas Waelty, Victor Zarazua and David Behar. The research team also thanks Steve Ritchie, Ellen Levin, David Behar, Caspar Amann, Raymond Mah, Katherine Jacobs, Leon Basdekas, Brent Burton, Tirusew Asefa and Robert Lempert for their valuable comments on the draft of this report. The research team acknowledges the contribution from Deltares and especially from Marjolijn Haasnoot, Dana Stuparu and Williem van Deursen. The team also acknowledges the contrition from David Yates (NCAR) and from the RAND corporation for organizing the climate information workshop at the beginning of the project, especially Robert Lempert, Sandra Berry and Stephanie Tanverakul. The team thanks the contribution from all participations to the climate information workshop, who are Joe Barsugli (University of Colorado, Boulder), Melissa Bukovsky (NCAR), Daniel Cayan (Scripps Institution of Oceanography), Benjamin Cook (National Aeronautics and Space Administration), Michael Dettinger (U.S. geological Survey), Ethan Gutmann (NCAR), Andy Jones (Lawrence Berkeley National Laboratory), Karen McKinnon (UCLA) and Michael Wehner (Lawrence Berkeley National Laboratory).

Research Team

Principal Investigators:
Casey Brown, PhD
University of Massachusetts, Amherst

Baptiste François, PhD
University of Massachusetts, Amherst

Project Team:
Alexa Bruce, MSc
University of Massachusetts, Amherst

Hassaan Khan, PhD
University of Massachusetts, Amherst

Khanh Nguyen, MSc
University of Massachusetts, Amherst

Don Park, MSc
University of Massachusetts, Amherst

David Rheinheimer, PhD
University of Massachusetts, Amherst

Umit Taner, PhD
University of Massachusetts, Amherst

Sungwook Wi, PhD
University of Massachusetts, Amherst

WRF Project Subcommittee or Other Contributors

Alexis Dufour, PE
San Francisco Public Utilities Commission

David Behar
San Francisco Public Utilities Commission

David Yates, PhD
National Center of Atmospheric Research

Caspar Ammann, PhD
National Center of Atmospheric Research

Marjolijn Haasnoot, PhD
Deltares

Katherine Jacobs
University of Arizona

Leon Basdekas, PhD
Black & Veatch

Brent Burton
Metro Vancouver

Tirusew Asefa, PhD
Tampa Bay Water

Robert Lempert, PhD
RAND Corporation

WRF Staff

John Albert, MPA
Chief Research Officer

Kenan Ozekin, PhD
Research Unit Leader

Abstract and Benefits

Abstract:

Climate change and other changes to external conditions may jeopardize the future ability of the San Francisco Public Utilities Commission's (SFPUC's) Regional Water System (RWS) to meet the desired level of service. To help better understand the potential vulnerability of the RWS to uncertain future conditions, SFPUC partnered with The Water Research Foundation and University of Massachusetts Hydrosystems Research Group (HRG) to develop a long-term vulnerability assessment (LTVA) of the RWS. The goal of this study was to quantitatively and qualitatively assess to what extent climate change will be a threat to the Regional Water System in comparison to, or in combination with, other external drivers of change over the next 50 years (2020-2070). To answer these questions, the computational analysis used decision scaling, whereby vulnerabilities are identified via multi-dimensional sensitivity analysis with the uncertain external factors and available predictors or projections are used to assess the level of concern associated with the identified vulnerabilities. Computer simulation models were developed to simulate changes in climate and weather (climate-weather generator), hydrology of each of the contributing watersheds, operations of the physical and management systems, long-term demand, water quality, and finance. In addition, a small set of specific scenarios was designed to investigate the effects of regulatory changes to environmental flow requirements and failure of key infrastructure components. The results serve as the basis for adaptation planning by revealing the system's vulnerabilities and assessing the level of concern that accompanies those weaknesses using climate projections and other future projections.

Benefits:

- This project demonstrates a multiple stressor vulnerability analysis including climate change, water demand, regulatory change, and infrastructure failure.
- The project illustrates the innovative decision scaling methodology, a method designed to make the best use of uncertain climate projections, to address the long-term concerns of the SFPUC.
- The project illustrates vulnerability assessment in categories not typically included in climate change analyses, such as regulatory change, infrastructure failure, financial impacts, and changes in precipitation, temperature, variability, and water demand.
- The analysis identified the specific climate vulnerabilities of the SFPUC water system, in terms of both water supply and finance, and conclusions are not dependent on climate projections.
- An expert elicitation was conducted with national climate scientists to assess the level of concern of climate projections and the vulnerabilities that were identified.
- The findings serve as the basis for future work in adaptation planning to explore adaptation pathways and identify low-regret investments and long-term options to increase resilience of SFPUC infrastructure.

Keywords: climate vulnerability, vulnerability analysis, climate change, water supply reliability, water demand, regulatory risk, infrastructure risk, water resources systems analysis, water finance, water quality, watershed hydrology

Contents

Tables

Figures

Acronyms and Abbreviations

ACDD	Alameda Creek diversion dam
AF	Acre-feet
AIC	Akaike information criterion
AS	Alameda siphons
CA	Calaveras Reservoir
CAPEX	Capital expenditure
CCSF	City and County of San Francisco
CFS	Cubic feet per second
CliWxGen	Climate Weather Generator (developed by UMass)
CMIP5	Coupled Model Intercomparison Project (Phase 5)
CONUS	Continental United States
CRT	Coast range tunnel
CS	Crystal Springs Reservoirs
BC	Bias correction
BCSD	Bias Correction and Spatial Disaggregation method
BIC	Bayesian information criterion
DBP	Disinfection byproduct
DP	Don Pedro
GCM	General circulation model
HH	Hetch Hetchy Reservoir
HHLSM	Hetch Hetchy/Local System Model
HRG	The Hydrosystems Research Group at the University of Massachusetts
HTWTP	Harry Tracy Water Treatment Plant
IFR	Instream flow requirement
ISG	Individual supply guarantee
KGE	Kling-Gupta efficiency
KNN	K-nearest neighbors
LG	Lagrange
LOS	Level of service
LP	Linear programming
LTVA	Long-term vulnerability assessment
MGD	Million gallons per day
NAE	Environmental narrative
NAI	Infrastructure narrative
NAD	Demand narrative
NCAR	National Center Atmospheric Research
NTU	Nephelometric turbidity units
NSE	Nash-Sutcliffe efficiency
OCPs	Other climate parameters
OPEX	Operating expenditure

PRMS	Precipitation Runoff Modeling System
QRNN	Quantile Regression Neural Network
RCP	Radiative concentration pathways
RWS	Hetch Hetchy Regional Water System
SAC-SMA	Sacramento Soil Moisture Accounting Model
SD	Standard deviation
SFPUC	San Francisco Public Utilities Commission
SFWSM	San Francisco Water System Model (developed by HRG)
SJPL	San Joaquin pipelines
SVWTP	Sunol Valley water treatment plant
TR	Technical report
TOC	Total organic carbon
UTREP	Upper Tuolumne River ecosystem program
WARM	Wavelet autoregressive model
WET	Water executive team
WSIP 2018	Water system improvement program
WG	Weather generator
WTP	Water treatment plant
WSA	Water supply agreement

Executive Summary

ES.I Key Findings

Overall, the study found that:

- Climate change exacerbates impacts from other external drivers of change and is not the single most important driver of vulnerability for the San Francisco Public Utilities Commission's (SFPUC's) Regional Water System (RWS).
- The RWS at a baseline demand of 227 mgd is resilient to changes in climate and other external drivers.
- The RWS water supply performance declines with reductions in mean precipitation, but is mostly insensitive to increases in temperature.
- The RWS is more vulnerable to changes in demand and instream flow requirements than changes in mean annual temperature and precipitation.
- The RWS is vulnerable to changes to mean climate when demand or instream flow requirements increase.

ES.2 Background and Objectives

To help better understand the potential vulnerability of the SFPUC RWS to uncertain future conditions, SFPUC partnered with The Water Research Foundation to develop a long-term vulnerability assessment (LTVA) of the RWS. The study was conducted by University of Massachusetts Hydrosystems Research Group (HRG) with input from National Center for Atmospheric Research (NCAR), other climate scientists, and Deltares.

The goal of the LTVA is to help quantitatively and qualitatively assess to what extent climate change will be a threat to the RWS in comparison to, or in combination with, other external drivers of change over the next 50 years (2020-2070). More specifically, the assessment aims to answer the following questions:

- Under what conditions and when will the RWS no longer meet system performance criteria?
- Is climate change the most important driver of vulnerability for the RWS and if not, what is?

While climate change is the driver of change that triggered this study, the intent is to understand the effects of climate change in the context of effects from other drivers of change on the RWS.

ES.3 Project Approach

The LTVA was performed using the decision scaling approach whereby vulnerabilities are first identified and used as a basis for assessing risks. The analysis includes a multi-dimensional quantitative stress test and qualitative scenarios in which sources of vulnerability are revealed through testing against changing conditions. A suite of computer models of the RWS was created, calibrated, and used to estimate the system performance under a range of future and/or unexpected conditions. Models were developed to simulate changes in climate and weather, hydrology of each of the contributing watersheds, operations of the RWS, long-term demand, raw water quality, and finance. Narratives or qualitative scenarios were designed to investigate the effects of instream flow requirements and failures of key infrastructure components. A series of performance metrics and targets was used to reveal vulnerabilities of the RWS water supply.

Boundaries of temperature and precipitation changes were set for the stress test from reviewing climate projections and findings from an elicitation workshop of climate science experts organized by SFPUC. Temperature changes of up to +7 °C above historical baseline (1986-2005) and precipitation changes from -40% to +40% of historical baseline were examined. A range of demand changes above baseline was evaluated. The narratives developed included five new instream flow requirements (IFR) and four critical infrastructure outages across different parts of the RWS.

ES.4 Results

- According to climate projections and expert elicitations, there is a central tendency of warming of +2°C and +4°C by 2040 and 2070 (Representative Concentration Pathway [RCP] 8.5), respectively, with no clear direction of change in mean annual precipitation over the planning horizon.
- In the Upcountry region, by 2040, most projections and elicitations of warming estimate between +1°C and +4°C, and precipitation changes between -5% and +5% compared to historical baseline; and by 2070 estimates of warming range between +3°C and +6°C, and precipitation changes range between -15% and +15% (RCP8.5).
- Changes in hydrology due to climate change affect the RWS's ability to meet water supply targets[1]. At 227 mgd baseline demand, the RWS can sustain up to +4°C and -5% precipitation change before failing to meet targets for delivery reliability, frequency of 20% rationing, storage reliability, and duration of rationing.
- Precipitation change is an important driver for RWS performance. A decrease by 10% or more will cause RWS targets to be missed. The climate projections and expert elicitations show that such a change in precipitation is possible by 2040, although unlikely. The likelihood of this change increases toward 2070.
- The RWS shows minor sensitivity to temperature change for the metrics evaluated in this study. Most metrics stay above target under warming conditions. However, warming conditions often magnify the loss in system performance if precipitation or demand change.
- Demand change appears to be a major driver of future RWS performance. An increase in demand by 15% (265 mgd) will lead to failure to meet rationing frequency targets under current climate conditions. At 265 mgd demand, the rationing frequency targets would be met if there is an increase in precipitation of 10%. If demand increases by 30%, the rationing target cannot be met even when precipitation increases by 40%, which is believed plausible but unlikely over the planning horizon.
- The RWS is particularly vulnerable to the state-adopted new IFR below Don Pedro Dam (State WQCP), which represents a significant reduction in water available. At a demand of 227 mgd, the effect of state-amended WQCP under current conditions is equivalent to a reduction in mean annual precipitation of about 15% in terms of the water delivery reliability (reliability around 85%, rationing in 1 out of 6.5 years on average).

ES.5 Benefits

There is great concern among the public and water professionals regarding the potential implications of a changing climate on the future of water supply. Yet, the water supply industry is currently struggling to characterize the effects of future climate change on systems and to incorporate climate change considerations into their long-term planning. This report provides a detailed case study of the decision scaling methodology, which provides a systematic approach for addressing climate change concerns while also incorporating non-climate considerations. The case study illustrates the application to the

[1] Due to the limitations of the models, these results are likely understating the impact of climate change. Next steps will include improvements on these models to address the bias in the current versions.

SFPUC RWS, a large, complex water supply system spanning the width of California. The result is a vulnerability assessment that reveals clear thresholds of climate change that are problematic while using climate projections to inform rather than drive the analysis. The report demonstrates methods for using climate stress testing and qualitative scenarios of demand change, new regulation, and infrastructure failure to reveal vulnerabilities singly and in combination. The report shows how climate change projections and expert elicitation from climate experts can be used to assess a level of concern associated with vulnerabilities. The analysis reveals the vulnerability of the system to specific climate changes and shows the relative effect of climate change versus other uncertain factors. The result for SFPUC is a clear vision of priority considerations for long-term planning.

ES.6 Related WRF Research

- An Integrated Modeling and Decision Framework to Evaluate Adaptation Strategies for Sustainable Drinking Water Utility Management Under Drought and Climate Change (4636)
- Impacts of Climate Change on Honolulu Water Supplies and Planning Strategies for Mitigation (4637)
- Improving Tradeoff Understanding in Water Resource Planning Using Multi-Objective Search (4941)
- Mapping Climate Exposure and Climate Information Needs to Water Utility Business Functions (4729)
- Water Utility Planning Strategies to Mitigate Impacts of Climate Change in Central Ohio (4585)

CHAPTER 1

Introduction

1.1 Regional Water System Overview

The Hetch Hetchy Regional Water System (RWS) is a municipally owned utility operated by the SFPUC, a department of the City and County of San Francisco and serves both retail and wholesale customers (Figure 1-1). The RWS supplies water from the Tuolumne River watershed and from local reservoirs in the Alameda and Peninsula watersheds, delivering high-quality drinking water to 2.7 million residents and businesses in the Bay Area. The Upper Tuolumne River flow is partitioned between Turlock and Modesto Irrigation Districts (the Districts) and the SFPUC at the location of La Grange, a few miles downstream of New Don Pedro Dam. The river flow at La Grange is reconstructed to obtain the unimpaired natural flow before it is partitioned between water right holders. The RWS draws an average of 85% of its supply from the Tuolumne River watershed, collected in Hetch Hetchy Reservoir in Yosemite National Park. This water feeds into a system of tunnels and pipelines delivering water by gravity 167 miles to Bay Area reservoirs and customers. The remaining 15% of the RWS supply is drawn from local surface waters in the Alameda and Peninsula watersheds. The split between these resources varies from year to year depending on the water year hydrology and operational circumstances.

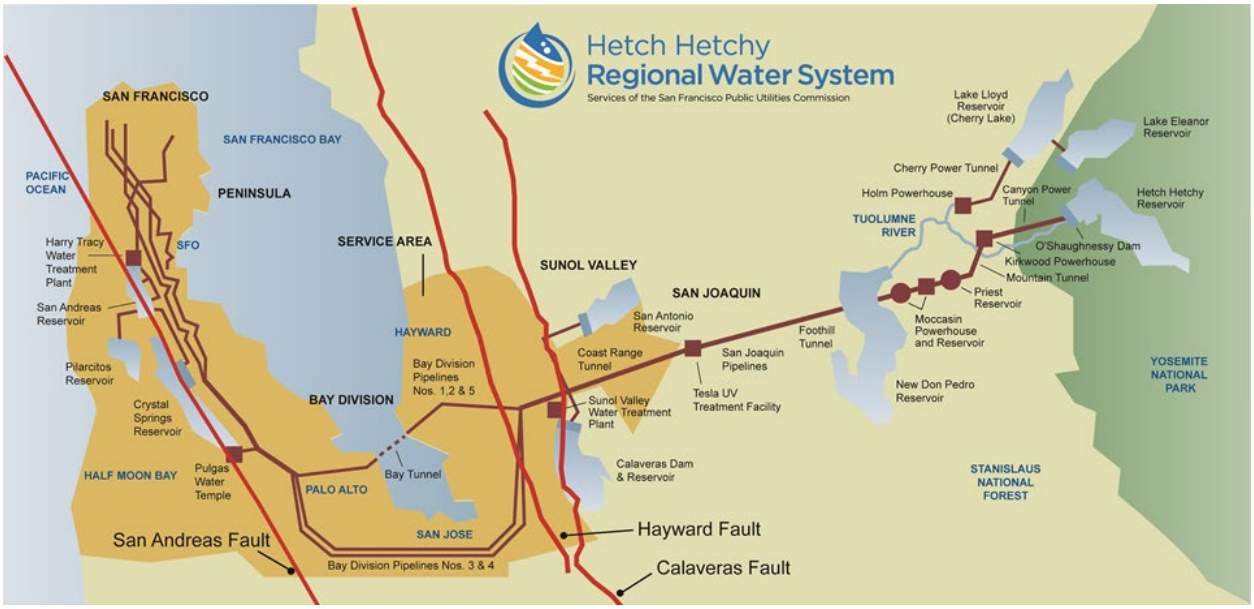


Figure 1-1. Hetch Hetchy Regional Water System Owned and Operated by SFPUC.
Source: Courtesy of SFPUC.

The RWS is a complex water supply network divided into the Hetch Hetchy (Upcountry), East Bay (Alameda or Sunol Valley), and Peninsula Systems. The in-city Retail Customer distribution system, owned and operated by the SFPUC, is not part of the RWS, though is supplied by water from the RWS.

In the Upcountry System, water is diverted from Hetch Hetchy Reservoir into a series of tunnels and large pipelines from the Sierra Nevada Mountains and through the San Joaquin Pipelines that cross the San Joaquin Valley to the Coast Range Tunnel, which connects to the Alameda system at the Alameda East Portal. The Upcountry System water is disinfected without filtering at the Tesla Treatment Facility.

The Upcountry System has a surface water filtration waiver from State and Federal agencies due to a pristine watershed producing exceptional water quality.

The East Bay System includes two reservoirs, San Antonio Reservoir and Calaveras Reservoir plus conveyance facilities connecting the Upcountry System and East Bay water sources to the Peninsula System. Calaveras Reservoir collects water from the upper Alameda, Arroyo Hondo, and Calaveras Creek watersheds. San Antonio Reservoir collects water from San Antonio Creek watershed. San Antonio Reservoir can also receive water from the Upcountry System via the San Antonio Pump Station. The Bay Division Pipelines (BDPLs) cross the South Bay to the Peninsula System delivering water to SFPUC customers along the pipeline route. The Sunol Valley Water Treatment Plant (SVWTP) filters and disinfects water supplied from San Antonio and Calaveras Reservoirs. In rare situations during which Hetch Hetchy water must be treated (e.g., when turbidity levels exceed filtration avoidance waiver), Hetch Hetchy water is transferred from the Alameda Siphons to SVWTP to be treated.

The Peninsula System includes conveyance facilities connecting the BDPLs to the in-City distribution system and to other SFPUC customers on the Peninsula. Two reservoirs, Crystal Springs and San Andreas, collect runoff from the San Mateo watersheds, and Crystal Springs also receives water from Hetch Hetchy Reservoir. Pilarcitos Reservoir collects runoff from the Pilarcitos watershed. The Harry Tracy Water Treatment Plant (HTWTP) filters and disinfects water supplied from Crystal Springs and San Andreas Reservoirs before it is delivered to the Peninsula customers and to the in-City distribution system. The City Distribution System is not included in the scope of the LTVA.

1.2 Planning for Climate and Other Changes

Climate change is having a profound impact on California's water resources, as evidenced by changes in temperature, precipitation, snowpack, and river flows (Vicuna et al. 2007). A recent study conducted by the UMass HRG for the California Department of Water Resources found that 2 ℃ of warming alone causes a 20% reduction in water deliveries for the California State Water Project (CA DWR 2019). CMIP5 projections of climate change for the region show a consistent expectation of warming between 1-3℃ by mid-century (1.5). Precipitation projections range from decreases of 20% relative to mean annual average to increases of more than 20% also by mid-century (1.5).

In 2012, the SFPUC and Turlock Irrigation District (TID) sponsored a hydrologic modeling study to determine streamflow sensitivities to possible increases in temperature and changes in precipitation due to climate change in the Tuolumne watershed (Hydrocomp et al. 2012). For the hydrologic study, the likelihood of any specific climate future was not assessed, and the report did not seek to comprehensively frame all the changes climate scientists expect from global warming. The goal of that study was simply to assess the sensitivity of reservoir inflows to a range of changes in two climate variables, temperature and precipitation. For that purpose, a physically-based conceptual hydrology simulation model was calibrated against past conditions and used to assess potential changes in the timing and volume of runoff that may occur for increases in temperature of up to 5.4℃ (9.7℉) and changes in precipitation ranging from between -15% to +6% relative to existing conditions. A review of the literature and consultation with climate science experts informed the selection of climate scenarios that encompassed a range of temperature and precipitation changes that may be experienced through 2100 so that potential changes in watershed runoff could be simulated and analyzed. With regards to temperature alone, median annual runoff at Hetch Hetchy Reservoir were projected to decrease by 0.7-2.1% from present-day conditions with increases in temperature between +0.6 and 1.7℃ (+1 and +3℉) and decrease by 2.6-10.2% from present-day with increases in temperature between +2 and 5.4℃ (+3.5 and +9.7℉). When changes in precipitation are accounted for on top of temperature, the median annual runoff at Hetch Hetchy Reservoir were forecasted to decrease by 7.6-8.6% from present-day conditions

with a 5% decrease in precipitation, and by 24.7-29.4% from present-day conditions with a 15% decrease in precipitation. Low runoff years are critical to evaluate water supply reliability. In the 2012 modeling study, climate change effects are exacerbated in low runoff years and decreases in runoff will be larger in dry years by a factor of between 1.5 and 3. Water supply analysis shows that the RWS would be vulnerable to temperature increases above +3.5℃ (6.3℉) without a change in precipitation and for temperature increase above +1.7℃ (3℉) when combined with decreases in precipitation greater than 5%.

Lessons learned highlighted the difficulty of selecting climate scenarios and that climate change cannot be studied on its own but must be included in the broad context of water system planning to advise making decisions. The 2008 economic downturn, the 2012-2016 severe drought and the 2018 adopted regulatory actions for minimum instream flow requirement at the State level[2] illustrate real stresses to water supply reliability other than climate change. These recent events emphasized the need for an integrated approach including climate change and other sources of vulnerability, singly and in combination.

1.3 Purpose of the Study

There are many uncertain factors such as climate change, changing regulations, water quality, growth and economic cycles that may create vulnerabilities for the San Francisco Public Utilities (SFPUC) Regional Water System's (RWS) ability to meet levels of service. Traditionally, water system planning at SFPUC is done using a single demand projection and the historical hydrologic sequence. However, this traditional planning approach is challenged when multiple stressors are to be combined to explore vulnerabilities of the RWS. A system vulnerability is a set of uncertain conditions, also known as events or hazards, resulting in unacceptable performance ("failure" of the system). The risk is the combination of the likelihood of an event and the magnitude of its consequence. The uncertainties associated with the degree to which these factors will occur and how much risk they present to the water system is difficult to predict, but nonetheless they need to be considered in SFPUC water system planning.

To address this planning challenge, the SFPUC partnered with The Water Research Foundation (WRF) to develop and test a new multi-stressors assessment framework for water system planning using a vulnerability-based planning approach (project 4703, *Long Term Vulnerability Assessment and Adaptation Plan for the SFPUC Water Enterprise – Phase I*). The University of Massachusetts Hydrosystems Research Group (HRG) was selected for this work with the support of the National Center for Atmospheric Research (NCAR), other national climate scientists and Deltares. The approach is to explore a range of future conditions to identify vulnerabilities and assess the risks associated with these vulnerabilities over the next 50 years (2020 – 2070). The work attempts to answer two questions:

- Under what conditions and when will the RWS no longer meet water supply performance criteria?
- Is climate change the most important driver of vulnerability for the RWS and if not, what is?

The second question is acknowledgment that water system planning must contend with uncertainty and that non-climate uncertainty can outweigh uncertainty related to future climate. This is an attempt to develop and test a multi-stressor assessment framework.

To scope the uncertainties and key system vulnerabilities to be considered in this study, SFPUC Water Executive Team (WET), SFPUC staff from Finance and the HRG held a workshop (Brown and Dufour

[2] State Water Resources Control Board Water Quality Control Plan for the San Francisco Bay/ Sacramento-San Joaquin Delta Estuary (Bay-Delta Plan) amendments adopted in December 2018 for the Lower San Joaquin River basin.

2017). SFPUC staff were asked to identify key sources of vulnerability that they considered could significantly impact the SFPUC's ability to meet level of service (LOS) goals and rank them in terms of importance and degree of uncertainty. Importance was defined based on the possible impacts each source could have on the SFPUC's ability to achieve the LOS objectives, with higher impacts corresponding to higher importance. Uncertainty was defined as how much we know about the impact of each source. The results of this exercise are shown in Figure 1-2. Among a wide range of sources of vulnerability deemed important/uncertain, climate change was identified as highly uncertain, but of medium importance, while demand was deemed of medium uncertainty and medium importance. Some of the more important uncertainties included regulatory factors, infrastructure, and natural hazards. Each of these sources can be considered and, in some cases, analytically explored using the approach described herein to discover the key concerns and possible vulnerabilities that will challenge SFPUC in the years ahead.

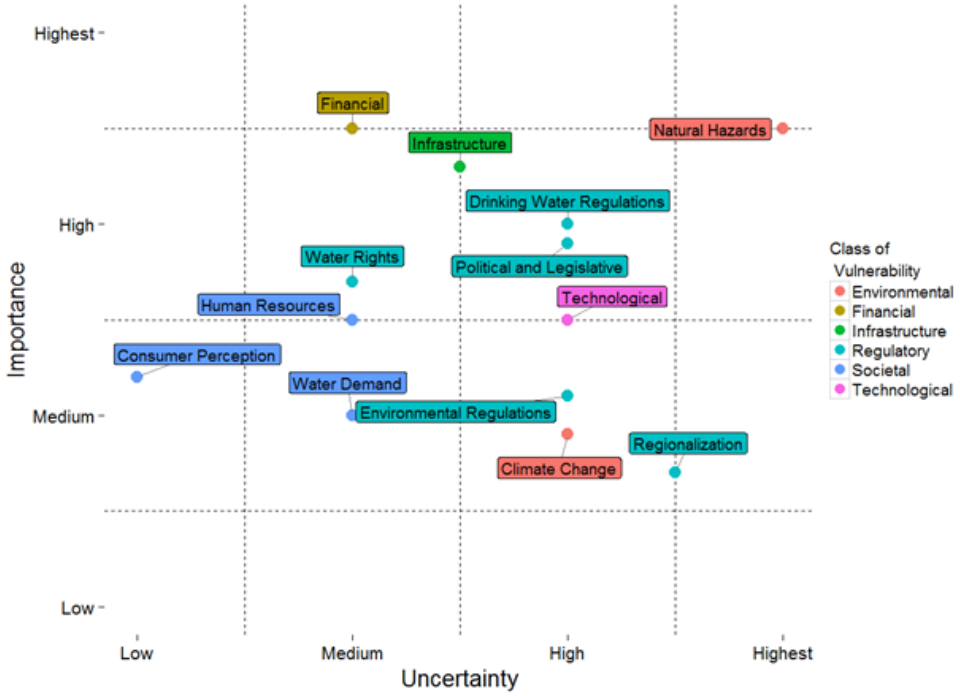


Figure 1-2. Importance and Uncertainty Associated with Sources of Vulnerability Identified by SFPUC Leadership.

1.4 Vulnerability Assessment Objectives and Scope

The LTVA generally aims to provide the tools, analytical framework and important insights needed to plan for an uncertain future by assessing vulnerability stemming from climate and other drivers of change, with the goal of informing the adaptation planning process.

Based on tailored simulation models and narratives, the vulnerability assessment 1) identifies potential vulnerabilities to the SFPUC water enterprise through an exploration of uncertain future conditions, 2) assesses the risks associated with vulnerabilities through integration of the best available information from climate models and expert opinion from SFPUC and others. The aim of this assessment is to provide guidance for water supply decisions that address the risk of system vulnerabilities over the next 50 years and provides a technical platform for further assessment. This phase of the assessment (Phase 1) will help the SFPUC identify its needs to address identified vulnerabilities. A future phase of study builds on the assessment to develop and evaluate an adaptation plan consisting of a portfolio of options that together are flexible and robust to a wide range of futures (Phase 2).

1.4.1 Decision Scaling

The assessment was performed using the concept of decision scaling (Brown et al. 2012; Brown and Wilby 2012; Poff et al. 2015) whereby vulnerabilities are first identified and used as a basis for assessing risks, then later integrated with more detailed assessments of vulnerability pathways to define adaptation plans (Haasnoot et al. 2013; Wise et al. 2014). The general analysis framework includes a multi-dimensional stress test approach, or structured sensitivity analysis that preserves the correlations and dependencies of key uncertain hydro-meteorological factors executed via high power computing resources. It includes the creation of a stochastic weather generator, new hydrologic models, a new water system simulation model, new water demand model and modules devoted to finance and raw water quality estimations. It also includes scenario narratives to address key uncertainties associated with changes in instream flow requirements and failure of important infrastructure facilities. The stress test is implemented in two ways that offer different and complimentary perspectives on system vulnerability: 1) *step*, where changes in climate and demand do not occur over time but instead occur over the entirety of the time series*, and 2) trend* (also called *transient* in the climate change literature), whereby changes in climate and demand occur gradually over time. While 1) is useful to estimate system performance and/or vulnerability metrics that require long time series of stationary forcing to converge, 2) is required for planning of adaptation options (which will be evaluated during Phase 2). More details regarding the design of the vulnerability assessment are provided in Chapter 3.

1.4.2 Spatial Scope

The spatial scope of this assessment generally encompasses 1) all critical facilities of the RWS and 2) all 27 wholesale customers part of the Bay Area Water Supply & Conservation Agency (BAWSCA) and the City and County of San Francisco (for 28 urban demand nodes in total). In addition, there are also 8 suburban retail customers and Cordilleras Mutual Water Company (MWC) a wholesale customer not part of BAWSCA. The analysis includes New Don Pedro Reservoir, including La Grange, to account for the operational connection between Don Pedro Reservoir operations in the Water Bank scheme. The analyses thus span the Upcountry, East Bay and Peninsula regions and delivery to the wholesale customers and the City of San Francisco (including its suburban retail customers) but do not explicitly include within-city water distribution systems.

1.4.3 Temporal Scope

The vulnerability assessment is carried out over a 50-year period from 2020 to 2070. The time step of the assessment varies, depending on 1) the specific analytic component and 2) the performance metric of interest, each of which is described further below.

1.5 Report Organization

This report is organized as follows. First, the regional water system and the vulnerability assessment are described in more detail in this introduction. Second, the summary of the climatology and of the climate projections across the SFPUC regions is given in Chapter 2. Third, the methods used to quantitatively model the response of the RWS to probable ranges in drivers of change are described in Chapter 3. Fourth, the long-term vulnerability assessment design is described in Chapter 4. The outcomes of the vulnerability assessment are described quantitatively in Chapter 5. The vulnerability assessment includes analysis of system performance with stress tests related water supply, environmental stewardship, water quality, and finance, and with narratives related to instream flow requirements and infrastructure failures. Chapter 6 presents a proof-of concept adaptation planning study using dynamic adaptation policy pathways. Finally, the vulnerability assessment results are discussed qualitatively for better interpretation in Chapter 7. Apart from the vulnerability assessment itself and subsequent discussion, much of this report aggregates and organizes text, figures, and tables from a range of other reports and

documents prepared for the LTVA, with some editing and additions for consistency and clarity; these reports are referenced explicitly as relevant and are available as supplementary materials. These reports are:

- HRG TR1 (HRG 2018). *Technical Report 1: Weather Generator Module*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.
- HRG TR2 (HRG 2021a). *Technical Report 2: Hydrologic Modeling Module*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.
- HRG TR3 (HRG 2020a). *Technical Report 3: Urban Water Demand*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.
- HRG TR4 (HRG 2021b). *Technical Report 4: San Francisco Water System Model*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.
- HRG TR5 (HRG 2021c). *Technical Report 5: Raw Water Quality Model*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.
- HRG TR6 (HRG 2020b). *Technical Report 6: Finance Module*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.
- Stuparu D., A. Dufour, W. van Deursen, M. Haasnoot (2019) Piloting Adaptation Pathways for San Francisco Water. Deltares
- Lempert, R., Berry, S., Tanverakul, S. (2019) Climate Information Workshop Goals, Process and Results, The RAND Corporation.

CHAPTER 2

Historical Trends and Climate Projections

The approach of decision-scaling, used in this study, is to determine when the performance of a water system becomes unacceptable due to changes from external drivers such as climate change. Due to the large and inherent uncertainty in the future climate, the approach selected incorporates the full plausible range of climate change projections without regard to likelihood to explore a comprehensive set of possible outcomes. In meeting this approach for climate change, the following factors need to be defined: 1) the climatic stressors that are relevant, 2) the appropriate method to generate synthetic future climate inputs to be used in the stress test and lastly 3) the range of change in the climatic stressors. For this study, climate change is limited to changes in mean annual precipitation and mean annual temperature in SFPUC watersheds and service area. A weather generator is also used to generate realistic synthetic time series of precipitation and temperature. To develop the weather generator signals of climate change in the SFPUC watersheds and an understanding of the natural variability of annual precipitation is necessary. This chapter covers the information needed to develop the weather generator and establish the range of climate change explored in this study. The chapter also presents a detailed analysis of climate change projections by global circulation models for Central California by the National Center for Atmospheric Research (NCAR) and findings from an elicitation workshop of climate science experts organized by SFPUC. Findings from projections and elicitations were used to set bounds for the stress test (Chapter 4) and to judge the seriousness of any potential vulnerabilities identified by the vulnerability assessment (Chapter 5).

2.1 Historical Trends and Variability

2.1.1 Temperature Trend

Across California, temperatures have shown a warming trend in the past century. As a whole, the state experienced an increase of 0.6 to 1.1°C (1.1 to 2°F) in mean temperature over the past century (i.e., from early 1900s to 2010s) (CA DWR 2015). Seasonal trends indicate a greater temperature increase in summer months than in winter months. Studies of precipitation trends in the state have yielded inconclusive results. While northern California shows increases in both mean annual precipitation and number of rainfall days, southern California shows small or no change, although none of these change are significant at 5% (Killam et al. 2014). Given these well-documented trends in the 20th century in California's climate, the HRG conducted time-series analyses on the trends for annual and seasonal temperature and precipitation across the SFPUC watersheds to determine the nature of the trends and understand their spatial and temporal variations.

The HRG has assessed annual and seasonal trends in minimum temperature (T_{min}) and maximum temperature (T_{max}) from the year 1956 to 2011. For the trend analysis, stations with long-enough records were used (i.e., over the 56-year analysis period). These stations are: Half Moon Bay (HMB) and San Francisco Airport (SFO) for the Peninsula region; Mt Hamilton (HML), Livermore (LVK) and Newark (NEW) for the East Bay region; and Moccasin (MCN), Early Intake (IN), Cherry Valley (CHV) and Hetch Hetchy (HTH) for the Upcountry region respectively. For annual trend analysis, minimum and maximum temperature values were averaged from each station over each water year (October to September).

Annual trends in T_{min} over the Upcountry, Peninsula, and East Bay regions is shown in Figure 2-1. Positive trends are observed in all three regions. They are all significant at 5% using the Mann-Kendall test[3]. The trend in East Bay region is the strongest. Annual trends in T_{max} over the same three regions are shown in Figure 2-2. Unlike T_{min}, none of these trends are significant at 5% using the Mann-Kendall test.

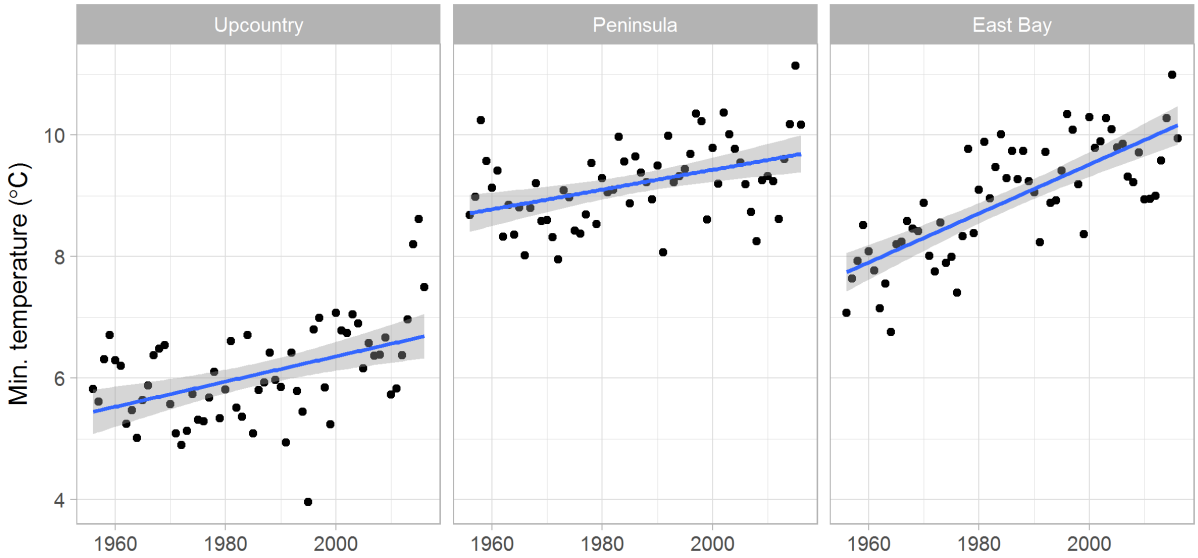


Figure 2-1. Annual Averages of Daily Minimum Temperatures (T_{min}) across the Upcountry, Peninsula, and East Bay Regions.

Results are shown for the analysis period from the year 1956 to 2011. Blue lines indicate the linear trend fitted to the underlying data. The shaded regions show the 95% confidence interval for the associated trend line.

[3] The purpose of the Mann-Kendall test is to statistically assess if there is a monotonic upward or downward trend of the variable of interest over time.

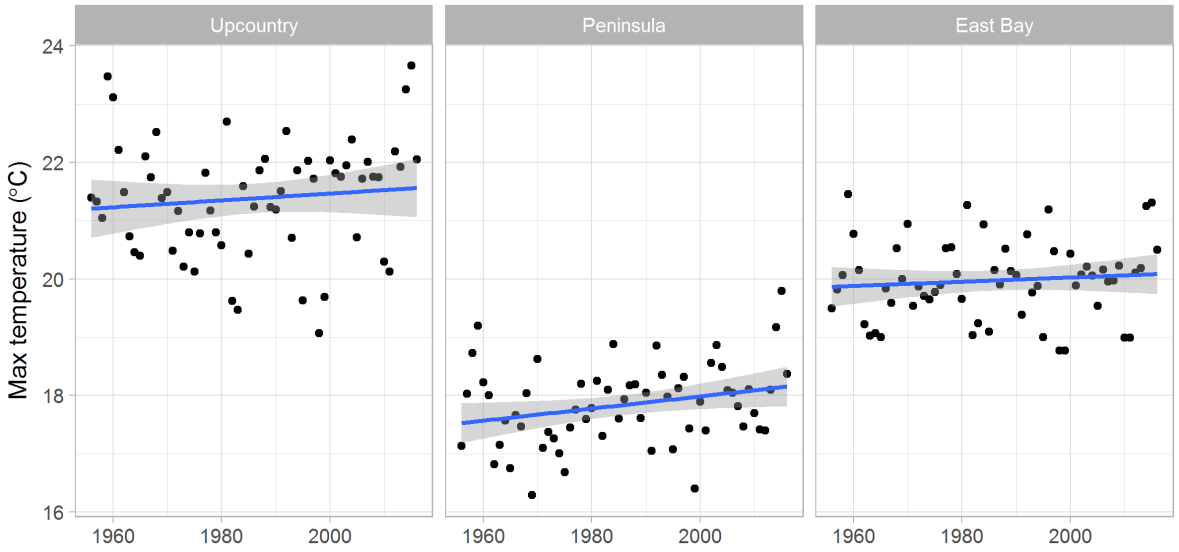


Figure 2-2. Annual Averages of Daily Maximum Temperatures (T_{max}) across the Upcountry, Peninsula, and East Bay Regions.

Results are shown for the analysis period from the year 1956 to 2011. Blue lines indicate the linear trend fitted to the underlying data. The shaded regions show the 95% confidence interval for the associated trend line.

Historical trends in T_{min} and T_{max} from the same gauges are also evaluated for the dry season (i.e., from April through September) and wet season (i.e., from October through March). T_{min} exhibits a positive trend in all three regions and both seasons (Figure 2-3). All are significant at 5% using the Mann-Kendall test. The strongest trends are observed in the East Bay region. In contrast for T_{max}, no trend is significant at 5% using the Mann-Kendal test (Figure 2-4). More details regarding the analysis of the trends in temperature variable are provided in the Technical Report 1: Weather Generator Module (HRG 2018).

Overall, the T_{min} and T_{max} trends shown in Figure 2-1 through Figure 2-4 are important to understand the possible range of changes in the spatial and temporal availability of water resources. The positive trends detected for the Upcountry region for T_{min} are especially important because this can affect the phase of precipitation, the volume of snowpack, and the timing of flows into the Upcountry reservoirs (e.g., change in precipitation from snow to rain). Earlier analyses have shown that over the last several decades, rising temperatures in the Sierra Nevada and northern California trigger decreasing snowpack and earlier snowmelt (Barnett et al. 2008).

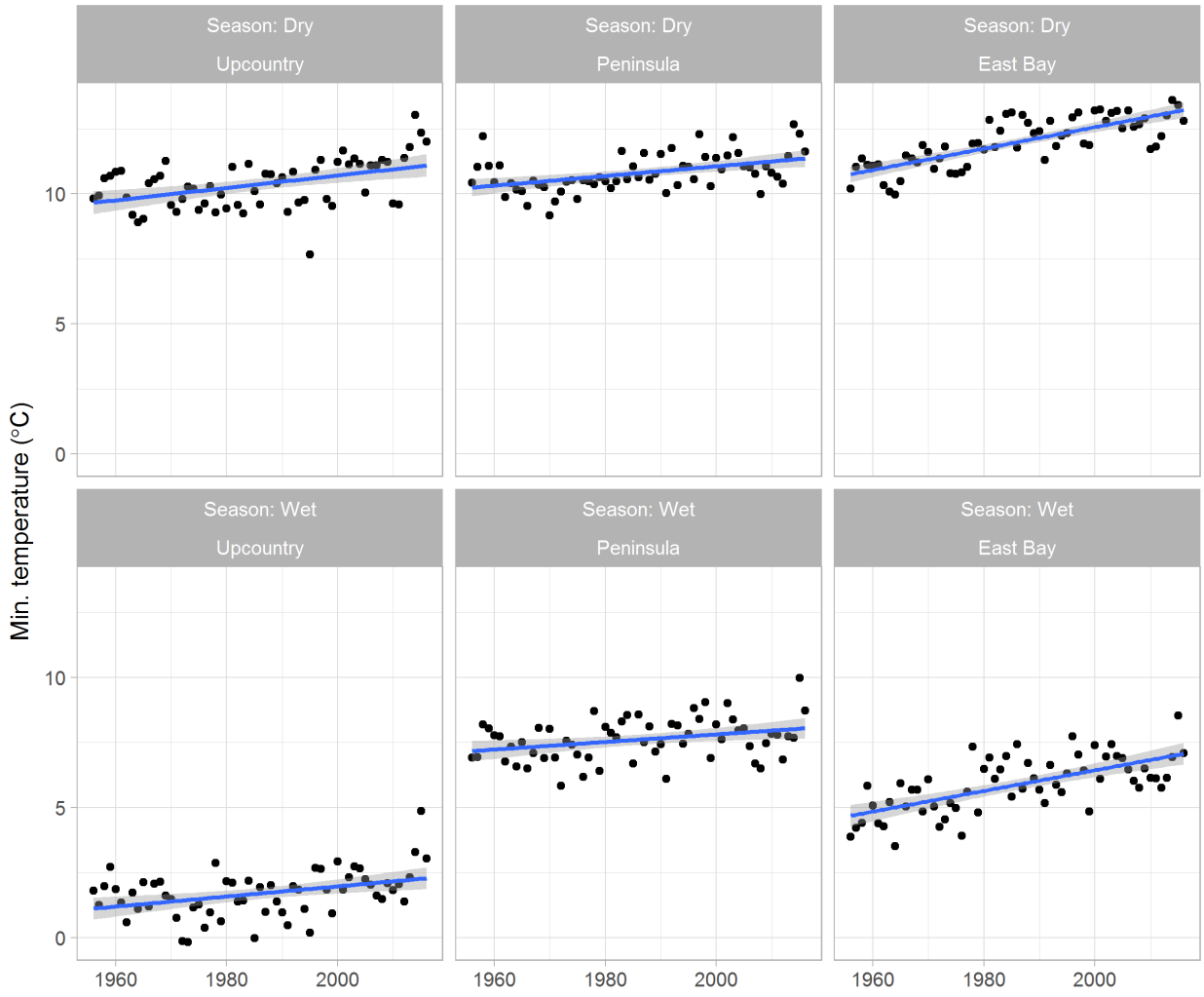


Figure 2-3. Dry Season (April through September) and Wet Season (October through March) Averages of Daily Minimum Temperatures (T_{min}) across the Upcountry, Peninsula, and East Bay Regions.
Results are shown for the analysis period from the year 1956 to 2011. Blue lines show linear trend fitted over the data. Shaded regions show 95% confidence interval for the fitted trend line.

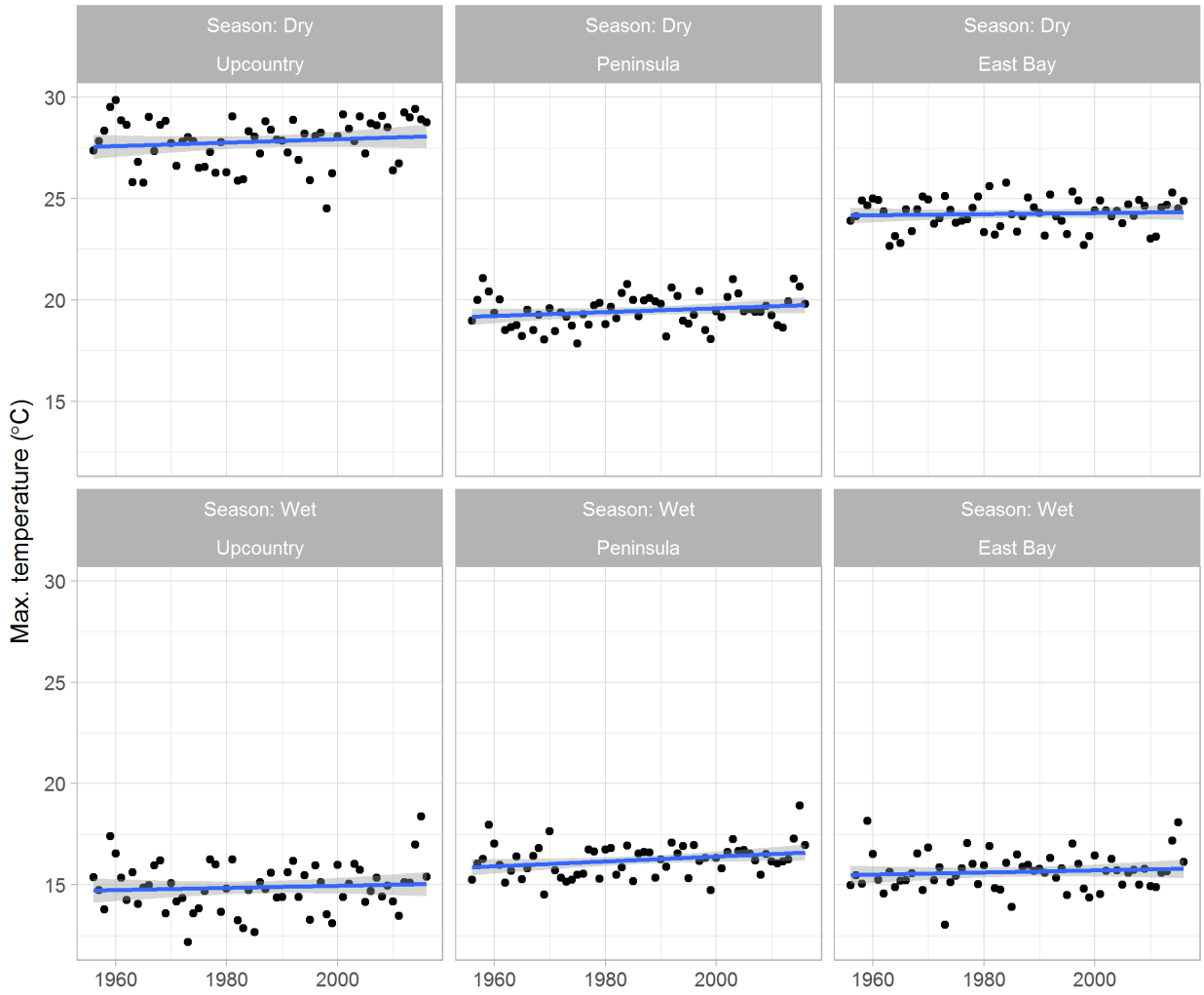


Figure 2-4. Dry Season (April through September) and Wet Season (October through March) Averages of Daily Maximum Temperatures (T_{max}) across the Upcountry, Peninsula, and East Bay Regions.
Results are shown for the analysis period from the year 1956 to 2011. Blue lines show linear trend fitted over the data. Shaded regions show 95% confidence interval for the fitted trend line.

These results generally agree with the previous studies done by SFPUC (Hydrocomp et al. 2012), which concluded that the average daily temperatures have increased over the 79-year period from 1930 to 2008, although increases were not consistent. According to SFPUC previous studies, there is no apparent trend in average daily temperatures from about 1930 to 1960. From about 1960 to 2017, average daily temperatures at Upcountry locations, Hetch Hetchy (HTH) and Cherry Valley (CHV) have increased, with the increase being mostly driven by an increase in daily minimum temperatures, which is consistent with the results shown in Figure 2-1 through Figure 2-4. Previous SFPUC studies have also shown that temperature records at Moccasin at 938 ft. elevation do not show preferential increases in daily minimum temperatures relative to daily average or daily maximum temperatures. A similar result was obtained for this station in this study (HRG 2018).

These findings are similar to the results from other climatic studies in the region. Behnke (2011) found that daily minimum temperatures in the Sierra have generally increased since 1900, with most of the increase occurring before 1930 and since 1960. Bonfils et al. (2008) found that daily minimum winter

temperatures in the Sierra increased over 1.5°C (2.7°F) between 1950 and 1999, while winter average daily maximum temperatures increased over 0.8°C (1.4°F).

2.1.2 Annual Precipitation Trend

For the precipitation trend analysis, stations with long-enough records were used (i.e., over the 56-year analysis period). These stations are: Half Moon Bay (HMB), San Francisco Airport (SFO), Upper (UCS) and Lower (LCS) Crystal Springs Reservoirs, San Andreas Reservoir (SA) and Pilarcitos Dam (PLD) for the Peninsula region; Mt Hamilton (HML), Livermore (LVK), Newark (NEW), Sunol (SUNO) and Calaveras (CAL) for the East Bay region; and Moccasin (MCN), Early Intake (IN), Cherry Valley (CHV), Hetch Hetchy (HTH), Sonora (SON) and Yosemite (YOS) for the Upcountry region respectively.

Figure 2-5 illustrates the results of the trend analysis for annual precipitation, while Figure 2-6 show results of the trend in both dry and wet seasons. None of the regions shows statistically significant trend at 5% level using the Mann-Kendal test, a finding that is consistent with the literature (Killam et al. 2014). More details about the trend analysis can be found in the Appendix of the Technical Report 1 (HRG 2018).

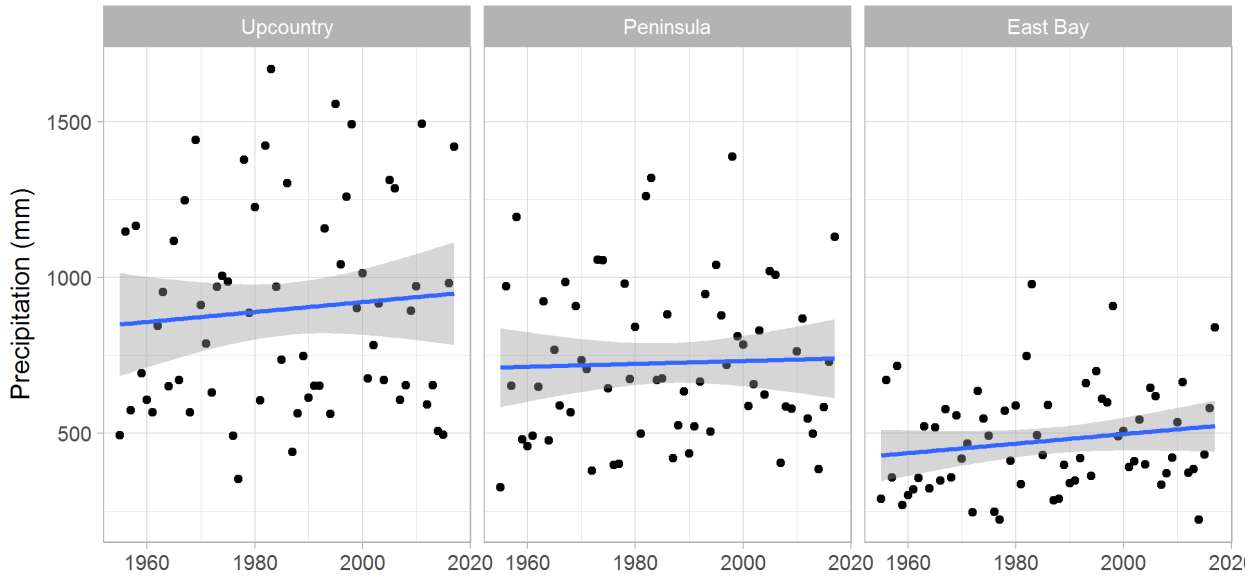


Figure 2-5. Annual Precipitation across the Upcountry, Peninsula, and East Bay Regions.
Results are shown for the analysis period from the year 1956 to 2011. The blue lines indicate the linear trend fitted to the underlying data. The shaded regions show the 95% confidence interval for the associated trend line.

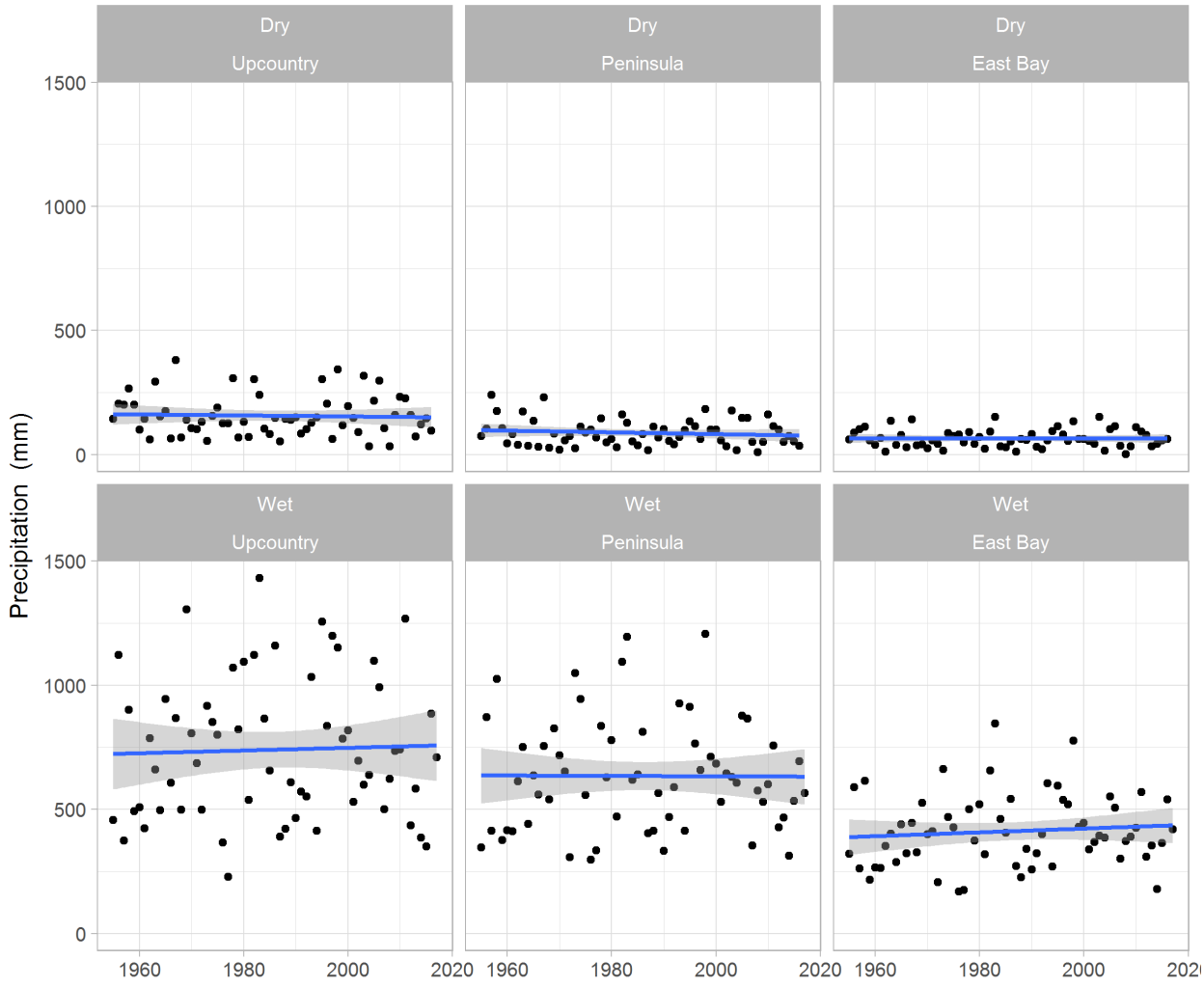


Figure 2-6. Dry Season (April through September) and Wet Season (October through March) Sum of Daily Precipitation across the Upcountry, Peninsula, and East Bay Regions.
Results are shown for the analysis period from the year 1956 to 2011. Blue lines show linear trend fitted over the data. Shaded regions show 95% confidence interval for the fitted trend line.

2.1.3 Variability in Annual Precipitation

Deser et al. (2012) defines natural climate variability (also known as internal climate variability) as the "variability of the climate system that occurs in the absence of external forcing, and includes processes intrinsic to the atmosphere, the ocean, and the coupled ocean-atmosphere system". The term "external forcing" refers for instance to the increasing greenhouse gas concentration in the atmosphere, which alters the earth-atmosphere energy balance and leads to warming temperature. Natural climate variability is acknowledged having two components whose effects are noticed across different temporal and spatial scales. The first component is deemed a "noise" (it does not offer a signal) that is noticeable from local to regional scales. The second component is associated with low-frequency oscillations at regional to sub-global scales that are often driven by the large-scale climate patterns. Examples of such a patterns include El-Nino Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO) and the Atlantic Multidecadal Oscillation (AMO). Understanding the effect of those large-scale climate patterns on regional climate is important because they can be important factors to explain, among other things,

the frequency and/or severity of hydro-meteorological extremes such as droughts and floods or the occurrence of other natural hazards such as wildfires (e.g., Cardil et al. 2020).

The impact of large-scale climate patterns on precipitation in California is well documented (Dettinger et al. 1998). For instance, ENSO has been shown to influence precipitation in the State, especially in Southern California (evidence for the effect of ENSO on precipitation in the Central California region is inconclusive). Across the state of California, Shukla et al. (2015) highlighted a complex relationship between climate patterns, such as ENSO and PDO, and the occurrence of meteorological drought. Regarding streamflow variable, Andrews et al. (2004) showed that magnitude of flooding is negatively correlated with ENSO in Northern California, and positively correlated in Southern California. Note that correlation is not significant in central California.

The HRG investigated the historical annual precipitation records across the Hetch Hetchy RWS for presence of a low frequency variability signal using a method called Wavelet Analysis (Torrence and Compo 1998). Broadly speaking, Wavelet Analysis allows identification of the different modes of variability within a time series. Here, the "modes of variability" refers to specific periods (durations in years) at which a signal oscillates (repeat itself). For instance, important modes of variability for the solar radiation at the top of the atmosphere over a specific location are daily and annual. The former is due to the diurnal cycle caused by the rotation of earth around its axis, while the latter results from the seasonal cycle resulting from the orbit of earth around the sun. The Sunspot (≈11 years) and de Vries (≈200 years) cycles are modes controlling the variability of solar radiation at larger temporal scales and result from periodic variations of the activity of the sun. Wavelet Analysis allows identification of these modes in a time series of observations, although they might be initially masked in the time series by local weather and climate conditions. Technical details regarding the Wavelet Analysis are provided in the Technical Report 1 (HRG 2018).

Figure 2-7 shows the global wavelet spectrum obtained for the Hetch Hetchy and Pilarcitos rain gauges. The global wavelet spectrums (black) show peaks of power values to indicate the major modes of variability within the analyzed signal. To verify the significance of the identified modes of variability, the global wavelet spectrum obtained from the wavelet analysis is compared with the one obtained for a random process (red). If the power value (on the horizontal axis) obtained for a given period is larger than the one obtained for a random process, the mode of variability is deemed significant.

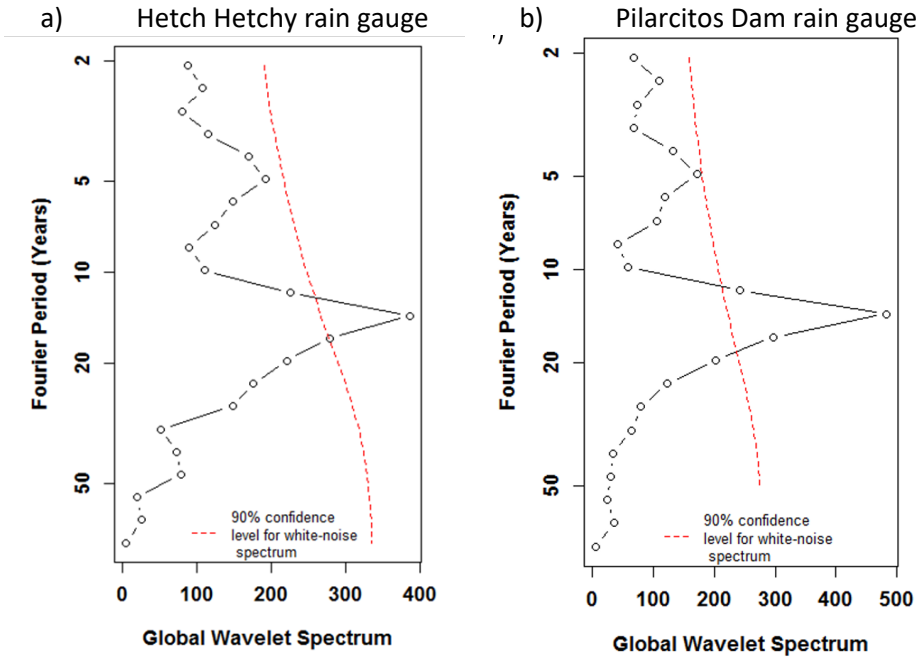


Figure 2-7. Global Wavelet Spectrum for the Hetch Hetchy (a) and Pilarcitos Dam (b) Rain Gauges (Black Curves). Rainfall records at both locations spans from 1930 through 2016. Annual precipitation time series are used as input time series to the wavelet analysis. As such, the shorter mode of variability (Fourier Period) that can be detected is 2 years. The dash-red curves show the theoretical power spectrum obtained for a red-noise at a 90% confidence level interval.

Regarding the annual precipitation time series observed at Hetch Hetchy (Figure 2-7a) and Pilarcitos (Figure 2-7b) rain gauges, the power spectrums show a peak at both locations at a 5-year period. This mode of variability could be associated with ENSO whose period of variability is often deemed ranging from 4 to 6 years. However, the power values obtained from the wavelet analysis for this period are lower than the one obtained with a random process (red curves). As such, the 5-year period cannot be deemed significant. This result is consistent with the lack of evidence of the impact of ENSO on precipitation in central California (e.g., Dettinger et al. 1998).

On the other hand, the power values observed at both locations are larger than the one obtained from a random process at a period close to 15 years. This quasi-periodic 15-year cycle in the precipitation signal for the Hetch Hetchy and Pilarcitos stations has been identified previously in the literature but surprisingly has received little further attention. This signal is also visible in the paleo-records for the past 200 years, but not before that (Meko et al. 2014). The climatic patterns responsible for this signal are not currently well understood. The cause of this phenomenon is unknown, however, including this quasi-periodic 15-year cycle allows a realistic representation of the low frequency variability of precipitation in the region. Some analysts would not include such a factor when the causation is unknown, but it is included here based on our assessment.

2.2 Climate Projections

In this section, results from recent climate projections are evaluated across the SFPUC region based on the data from the Global Coupled Model Intercomparison Project phase 5 (CMIP5, Taylor et al. 2012). In addition, the SFPUC convened nine climate science experts representing a variety of perspectives and disciplines to provide climate information relevant to the climate vulnerability of the RWS. The gathered information is then considered to evaluate the possible range of mean changes in long-term

precipitation and temperature for perturbing the natural variability realizations presented in the previous section. The work presented here is a joint effort of the HRG and the NCAR. The analysis is structured into four parts: 1) choice of Representative Concentration Pathway (RCP) of greenhouse gases (GHG) for the study, 2) description of temperature and precipitation model projections, 3) findings from the elicitation workshop of climate science experts, and 4) a summary of projected changes for use in the decision-scaling analysis and in particular with the weather generator.

2.2.1 Choice of RCP

General Circulation Models (GCMs, sometimes also called Global Climate Models) provide a scientific basis for helping climate scientists develop and test theories of the climate system and explore its complexities, and thus provide a useful tool for understanding climate change and the projections of future climate and related impacts. The Inter-Governmental Panel on Climate Change (IPCC) developed a Fifth Assessment Report (AR5) that guided the development of the community Coupled Model Intercomparison Project, Phase 5 (CMIP5) for which datasets of GCM output are available to the community for assessing regional climate change (IPCC, n.d.). The CMIP5 is a collaborative climate modelling process, coordinated by the World Climate Research Programme (WCRP), which resulted in a database of GCM simulations from the main climate modeling centers of various scientific institutions. A standard set of assumptions and procedures were followed by each modeling center, resulting in a consistent set of GCM simulations of the past, current, and future climate under various assumptions of greenhouse gas emissions.

The emission scenarios adopted by the IPCC and CMIP5 community in AR5 to characterize future greenhouse gas (GHG) concentrations are the Representative Concentration Pathways (RCP). The quantity of GHG releases in the atmosphere by human-induced activities or by natural feedbacks over the 21st century cannot be projected; the uncertainties are too vast. This is the reason why the IPCC developed RCP scenarios, with each RCP having a prescribed increase in radiative forcing by 2100 relative to preindustrial levels. The increase in GHG in the atmosphere increases the radiative forcing which is the additional energy in the Earth's climate system. The word "representative" signifies that each RCP provides only one of many possible emission scenarios that would lead to the specific radiative forcing characteristics in the year 2100. The term "pathway" emphasizes that not only the long-term concentration levels are of interest, but also the trajectory taken over time to reach that outcome (Moss et al. 2010). Four pathways were used for climate modeling and research for the AR5 and each describe different climate futures, all of which are considered possible depending on the volume of greenhouse gases (GHG) emitted in the years to come. The RCPs are labeled as RCP2.6, RCP4.5, RCP6, and RCP8.5 after a possible range of radiative forcing values in the year 2100 (2.6, 4.5, 6, and 8.5 W/m^2, respectively).

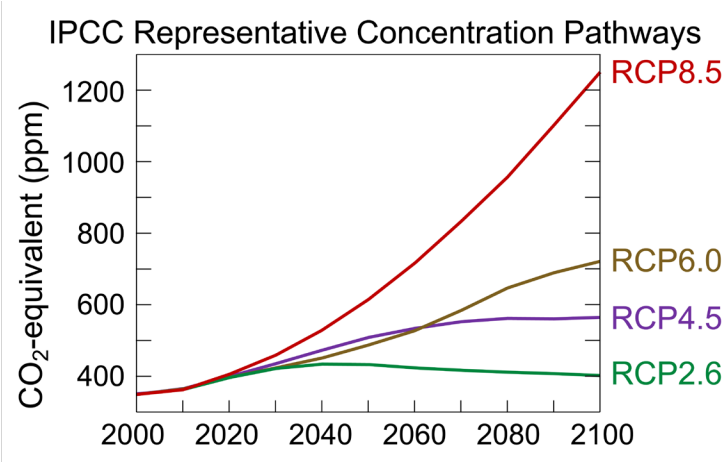


Figure 2-8. Trends in Concentrations of CO_2 Concentration for the Four RCPs of IPCC AR5.
Source: Efbrazil 2020.

Generally, the different assumptions of each RCP tend to result in similar climate projections to mid-century, which then diverge to the end-of-century (Figure 2-8). The RCP8.5 scenario results in the greatest amount of warming by the end-of-century and is the highest emissions scenario. The planning horizon for this study is 2070 or just beyond the midpoint of the 21st century. For that reason, it was decided that the RCP8.5 scenario has the greatest potential warming impacts and therefore, it allows an assessment of the outer bounds of RWS impacts for 2070.

2.2.2 Global Circulation Models Projections

2.2.2.1 Temperature Projections

Figure 2-9 and Figure 2-10 summarize the GCM multi-model spatial projection of minimum and maximum surface temperature changes for winter (DJF) and summer (JJA) seasons across the western US (including the SFPUC region) around the 2040 and 2070 periods using a 30-year averaging interval for each. The largest changes are seen in areas with snow feedback (winter in high-latitudes and continental interiors) and high elevations (all seasons), while the smallest changes are projected over the ocean and coastal regions. The magnitude of this warming by mid-century (~2040) varies slightly between models: in winter 1.5 to 2 °C in the Upcountry region and closer to 1 °C closer to the ocean; in summer, ~2 to 2.50°C in the Upcountry region and 1.5 to 2°C closer to the ocean in the East Bay and Peninsula regions. There is strong intensification of warming into the second half of the 21st century with warming about 4.5 to 5°C. The seasonal variations of this robust change are more modest along the coastal regions and greater in the interior regions.

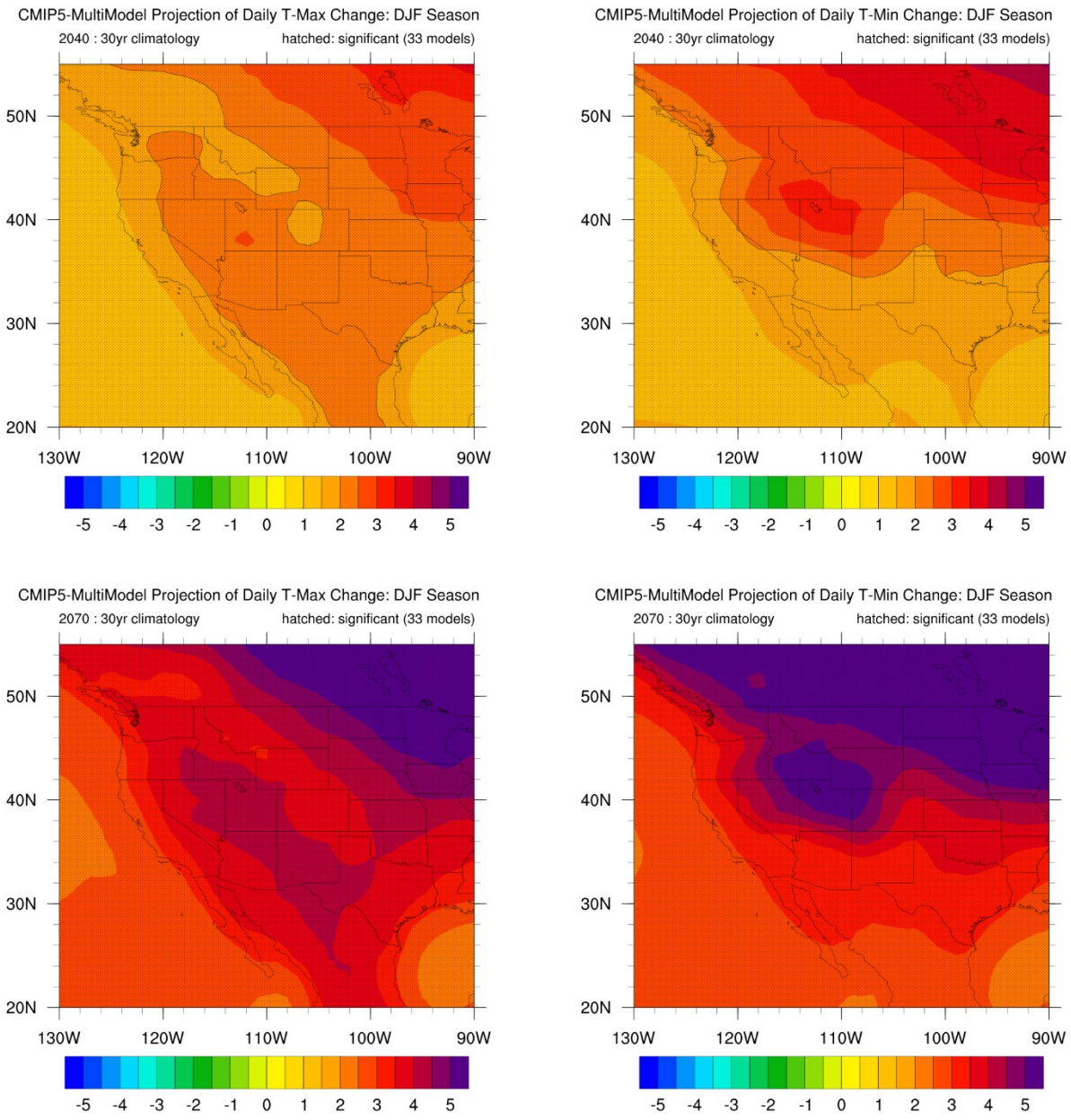


Figure 2-9. The Projected Changes in Maximum (Left) and Minimum (Right) Surface Temperature (in ^{o}C) for the Winter Season (DJF) as the CMIP-5 Multi-Model Average, for the 2040 (Top) and the 2070 (Bottom) for the 30-Year Averaging Intervals Relative to Historical Averaging Interval (1965-1994).

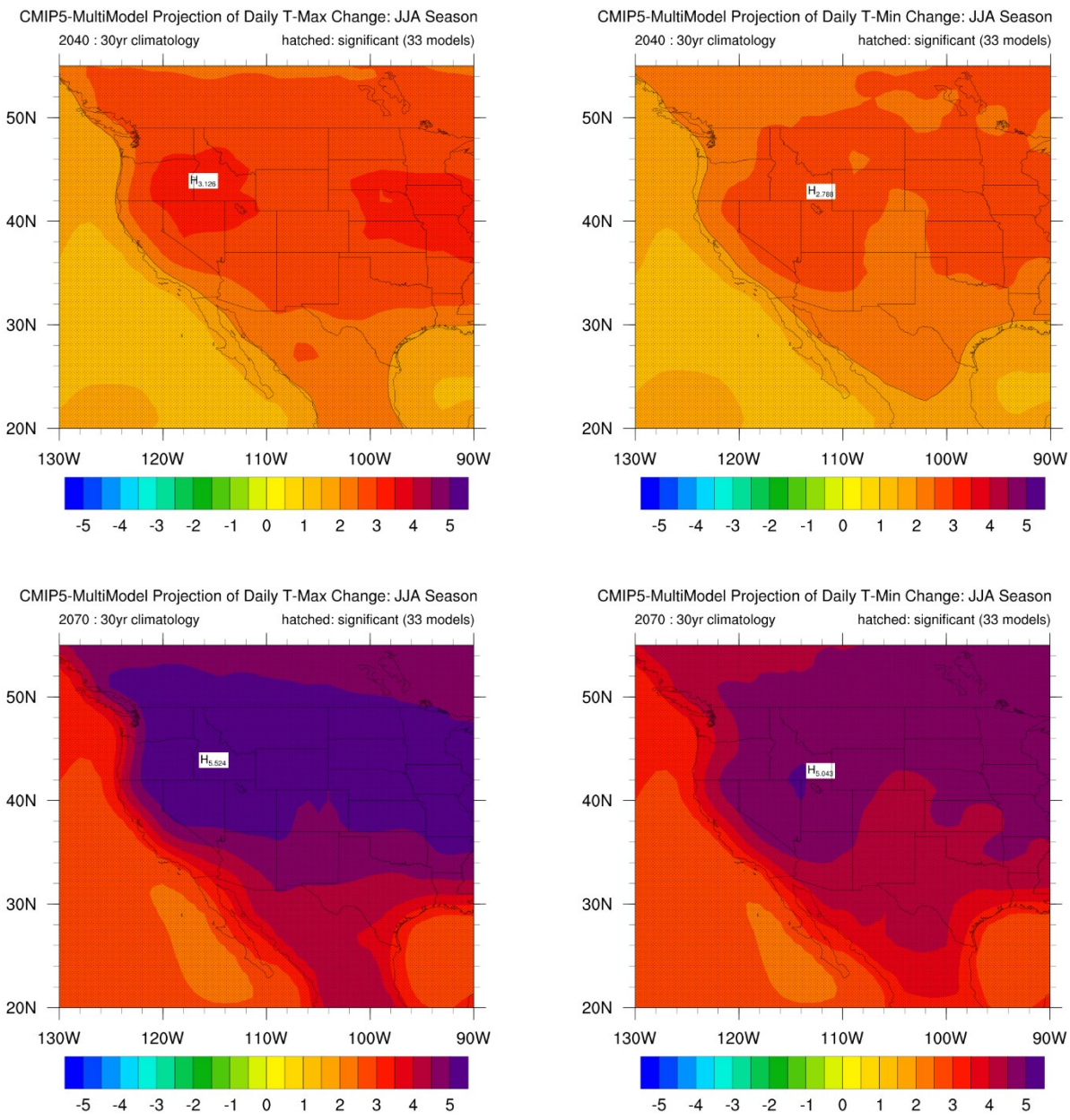


Figure 2-10. Same as Figure 2-9, for the Summer Season (JJA).

Figure 2-11 shows the monthly mean absolute temperature climatologies for the three 30-year averaging periods (2010, 2040, and 2070) for a location in the Southern Sierra Nevada near Hetch Hetchy as represented in the bias-corrected BCSD[4] dataset. This includes projections from 20 unique GCMs, many with multiple ensembles members resulting in 70 unique GCM projections. The plot suggests a robust warming signal across all months of the year and an increase of warming as time progresses. January's mean temperature for the 2010, 2040, and 2070 averaging period was about -3°C, -2°C, and 0°C respectively; while the July mean temperature for the same averaging periods were, 10°C,

[4] The Bias Correction and Spatial Downscaling (BCSD) is a trend-preserving statistical downscaling algorithm, which has been widely used to generate accurate and high-resolution data set. BCSD technique is used to statistically downscale projected weather variables such as temperature and precipitation.

11.5°C, and 13°C respectively. Note, the larger the projected mean changes, the broader the model spread. The key message of this figure is that the changes are systematic across models, despite internal climate variability (in contrast to precipitation, see Section 2.2.2.2).

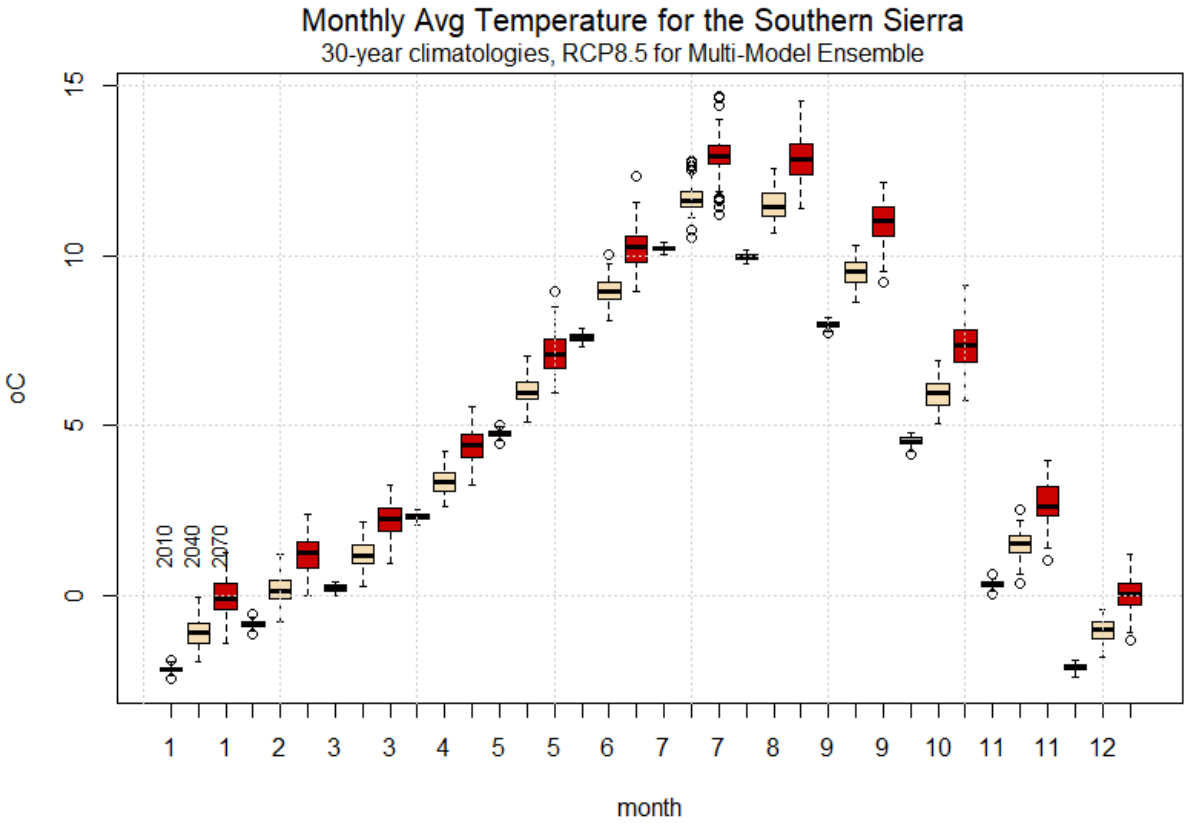


Figure 2-11. The Projected Changes in Mean Monthly Temperature Averaged across the BCSD CMIP5 Multi-Model Ensemble for the Averaging Intervals 1996-2025 (Grey), 2026-2055 (Yellow), and 2056-2085 (Red) as Bias-Corrected in the BCSD Archive for More Accurate Absolute Values.

The boxes represent the upper and lower quartile, while the whiskers are the 10th and 90th percentile range of the projections and open circles are outliers.

On inter-annual time scales, the variability of temperature also appears to be increasing slightly as time progresses, although the change is not significant. Looking at the CMIP5 RCP8.5 projections, Figure 2-12 shows the trend of warming over 3 periods of 30 years centered around 2010, 2040 and 2070. The figure indicates that the range across the models increases as the different models exhibit slight differences in the mean warming trajectory (due to different climate sensitivities as well as differences in regional dynamical changes). When examining the inter-quartile range, the distribution of annual mean temperatures in the CMIP5 models is slightly larger for future periods than for the evolving present-day window, and the overall rate of the warming trend over the period 2056-85 is greater than over the "near-current' 1996-2025 averaging interval.

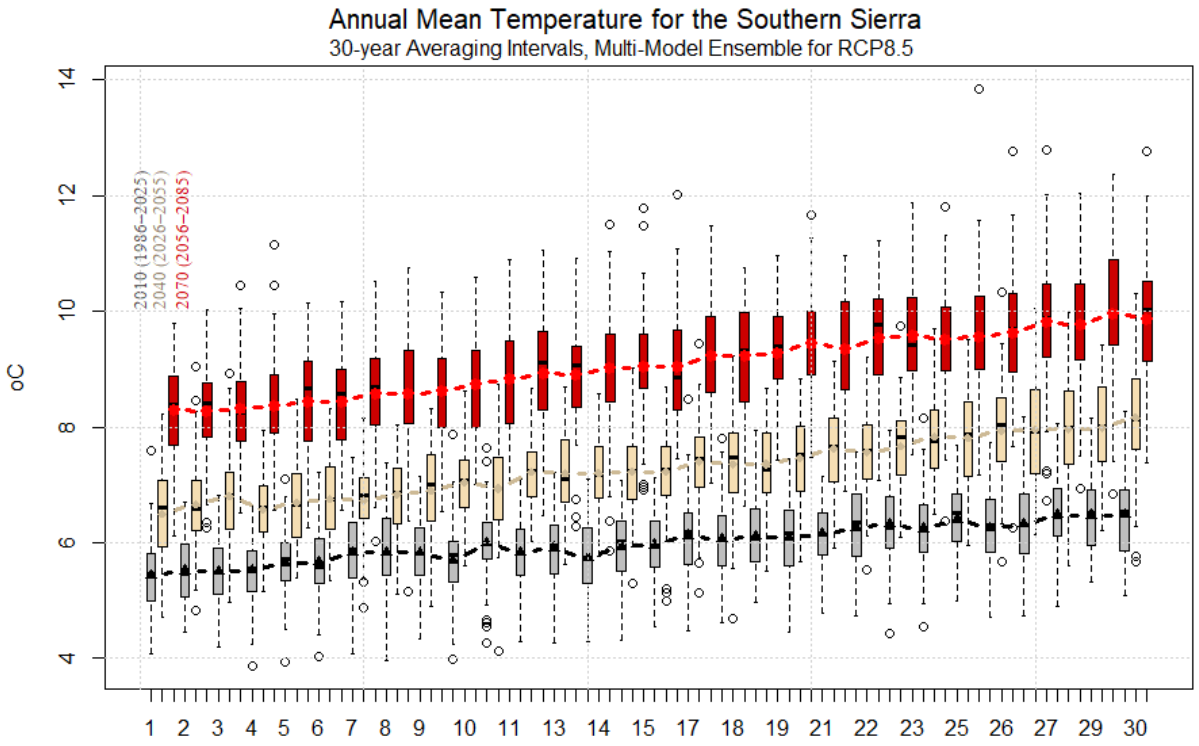


Figure 2-12. The Projected Mean Annual Temperature across a CMIP5 Multi-Model Ensemble for the Averaging Intervals (2010, 2040, and 2070) Including 1996-2025 (Grey), 2026-2055 (Yellow), and 2056-2085 (Red). Each year in the 30-year sequence is represented by the distribution across the 70 GCM RCP8.5 simulations from the BCSD archive. The boxes represent the upper and lower quartile, while the whiskers are the 10th and 90th percentile range of the projections and open circles are outliers, and the line is the mean of the ensemble. The segments connect seamlessly if assembled in sequence.

As discussed above, the changes seen in temperature are robust and showing a systematic warming across time in all consulted datasets. The mean of the 2010 averaging period is 6.0°C, while the 2040 ensemble mean is 7.3°C, while the 2070 ensemble mean is 9.1°C, with a clear trend over each of the 30-year averaging intervals. Figure 2-12 offers insight into the evolving nature of these changes (note, the segments connect seamlessly if assembled in sequence).

In summary, temperatures are clearly projected to increase across the SFPUC watersheds. The changes by the 2050s are expected to be around 2 to 2.5 °C, with the Upcountry region possibly slightly higher. Towards the end of the century, temperature changes could increase as much as 4 to 5°C under RCP8.5. The seasonal distribution of these changes is to a first order uniform, though locally in the high Sierra where snow feedback can operate larger values can be expected during the shoulder seasons and in the winter.

2.2.2.2 Precipitation Projections

The NCAR staff analyzed the precipitation projections keeping in mind the focus of the analysis being water resource management, thus this summary keeps an eye on key aspects such as seasonal changes in precipitation amount and the timing of that precipitation that could affect mean annual water volumes and, by proxy, water availability to the RWS. In addition, the NCAR staff looked for trends in the climate projections that exhibit more systematic signals such as seasonal changes. It important to note that many of the projected changes are not significant with respect to the range of variability experienced within the observational record. This type of “non-result” is actually useful in the context of

this climate change assessment because it indicates that, based on what we know today, precipitation regimes may not shift significantly, and this is useful information for surface water system planning. However, it may have other significant impacts on watershed land management and wildfire which are not in the scope of this study.

Figure 2-13 and Figure 2-14 show the projected precipitation anomalies (departure from long-term average) relative to the present-day baseline for the direct GCM output for the 2040 and 2070 30-year averaging intervals relative to the 2010 historic period, respectively. These average climatologies derived from the CMIP-5 archive for the 30-year averaging intervals of 2040 and 2070 relative to the historic 2010 period for the Upcountry region and East Bay and Peninsula regions.

While most individual months do not show significant precipitation volume change, when taken together across the seasonal cycle a more systematic story emerges: a clear seasonal evolution of the mean and median anomalies with relatively wetter winters and drier summers is seen. The two future periods show the GCM-derived projected changes in monthly precipitation, which suggest increasing winter-time precipitation as reflected in the multi-model median as well as the inter-quartile range, as shown in Figure 2-12. The other period of increase is during summer, though the absolute values are diminishingly small (the large percent change in summer is an artifact of the very small absolute values of precipitation.). However, these two periods are separated by shoulder seasons that show a tendency for drying, namely for April, May, and June, as well as October and November. Changes, both increases and decreases, are greater in 2070 than in 2040 for all months in both the Upcountry region and East Bay and Peninsula regions.

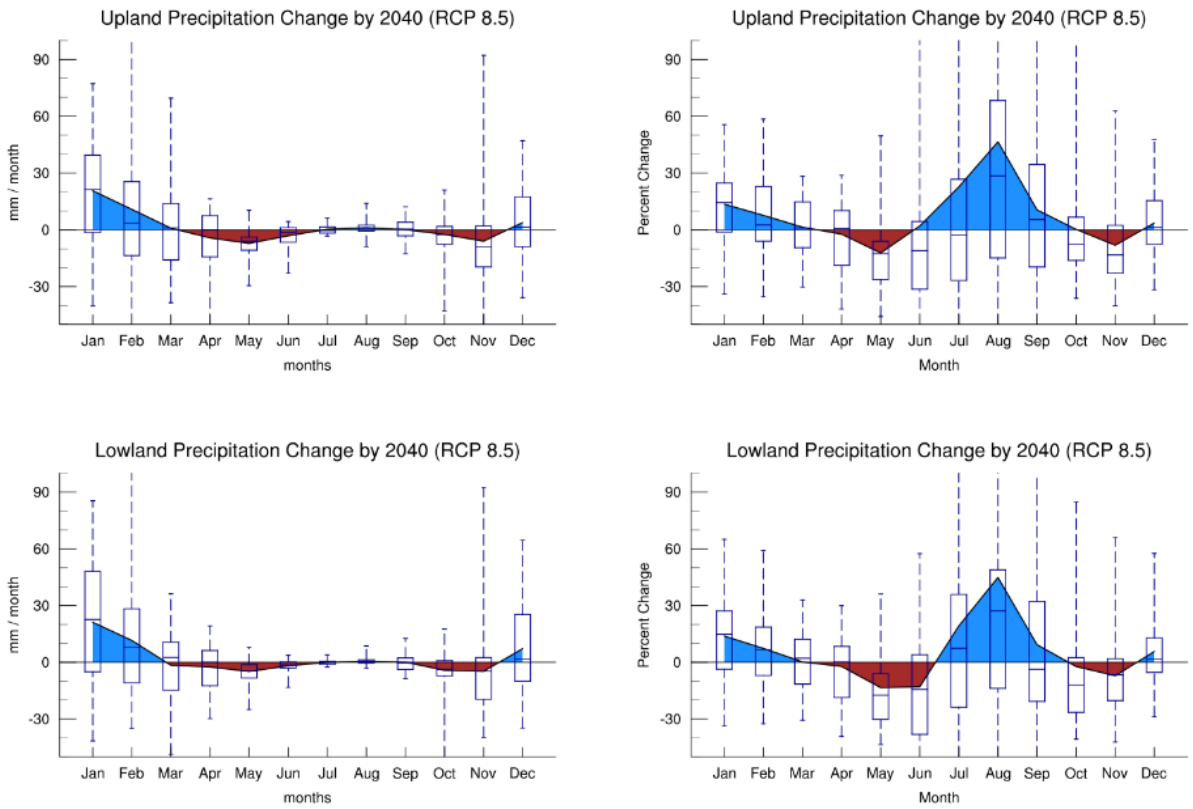


Figure 2-13. GCM Projected Changes in mm/Month (Left Panel) and Percent Changes (Right Panel) Compared to the Present-Day Baseline for the 2040 Averaging Interval of 2026-2055.

Upland refers to the Upcountry region and Lowland to the East Bay and Peninsula regions. The colored shapes are the means of the distribution for each month, and the color simply indicates positive (blue) or negative (red) changes, where the values are linearly interpolated between the months. The box-and-whiskers are the inter-quartile range (box), the 10th and 90th percentile (whiskers), and the median (bar). Clearly seen are the increases during the core winter and summer season and decreases in the shoulder season.

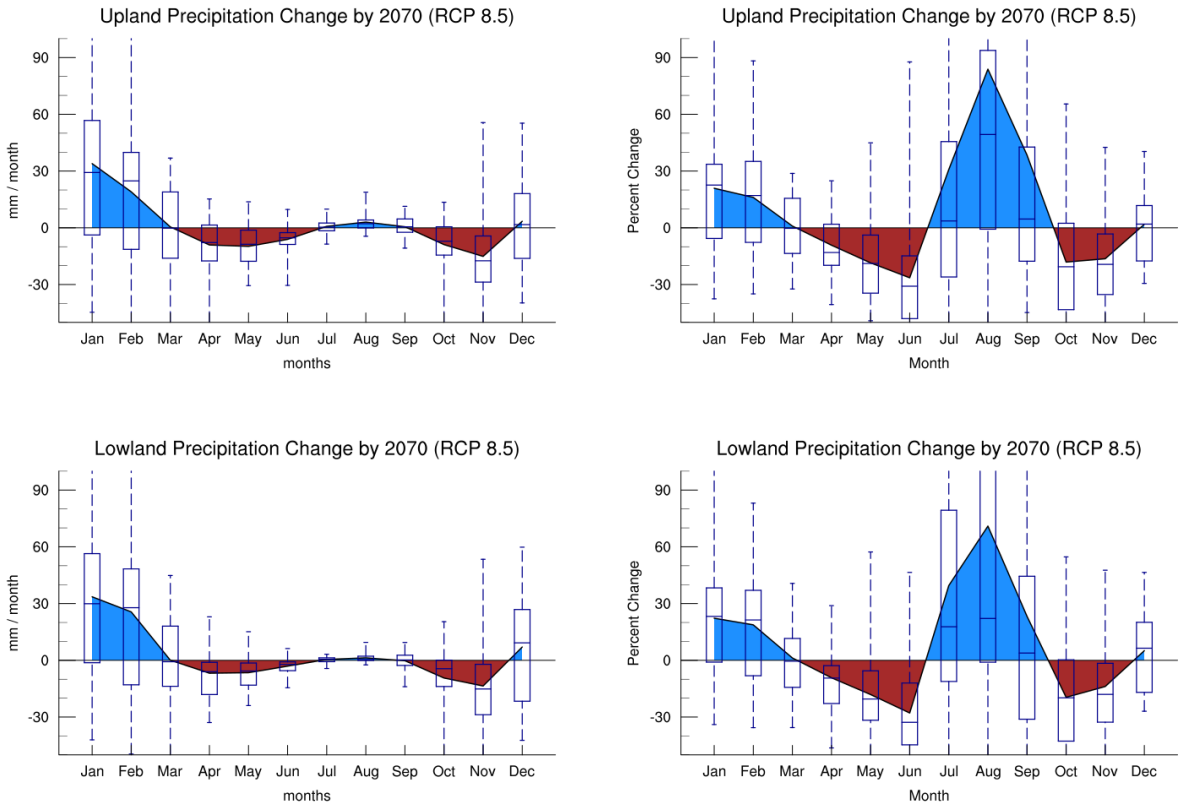


Figure 2-14. Same as Figure 2-13 , Except for the 2056-2085 (2070) Averaging Period.

Figure 2-15 shows the absence of trend in mean annual precipitation over 3 periods of 30 years centered around 2010, 2040 and 2070 using these same CMIP5 projections for RCP8.5. The figure shows that the inter-annual spread of the running average across the different GCMs are fully overlapping, suggesting that natural variability, not GHG forcing, dominates precipitation variability. Had the atmospheric forcing induced by GHG been more dominating in terms of its impact on the regional precipitation in California, then one would have expected some departure of the mean state for the 2070 averaging period when compared with the other averaging periods (2010 and 2040). This demonstrates that natural inter-annual variability in the Upcountry region masks the signal from climate change, as is the case in other regions (Deser et al. 2012). Although the mean precipitation change over the broader Upcountry region shows a slight increase over the 2070 averaging interval, the coherence of this signal across models is much lower than in regions further to the north (and much further to the south (Figure 2-16). Therefore, it is important to ask the question if the changes are in any form significant for the annual precipitation chance.

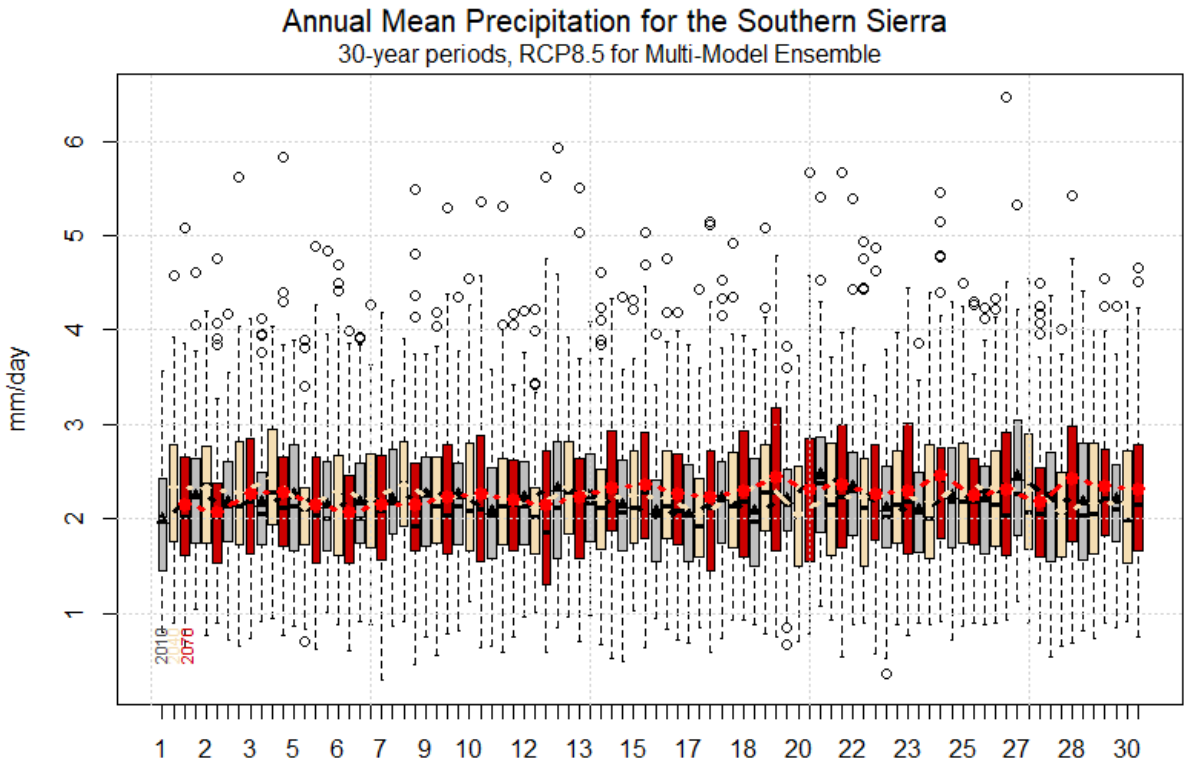


Figure 2-15. Projection of Annual Mean Precipitation for the Three, 30-Year Climatological Averaging Periods, 2010, 2040, and 2070 (1996-2025 Light Grey, 2026-2055 Tan, and 2056-2085 Brown) for the Upcountry Region. Each box represents 70 samples from the CMIP-5 data archive for RCP8.5. The running mean values across the 30-year intervals are shown as lines. The segments connect seamlessly if assembled in sequence.

Figure 2-16 illustrates significance of change in seasonality of precipitation using the approach applied in the IPCC report (Stocker et al. 2013, building on Tebaldi et al. 2011), where color is applied only in areas where at least two thirds of the models agree on the sign of change (the Upcountry region falls just outside of that area), and when additionally requiring a statistically significant deviation from the present day distribution using a 90% confidence, then only areas in Nevada and off the coast of Oregon fulfill these criteria for the DJF period in 2070 along the West coast of North America (hatched areas in Figure 2-16).

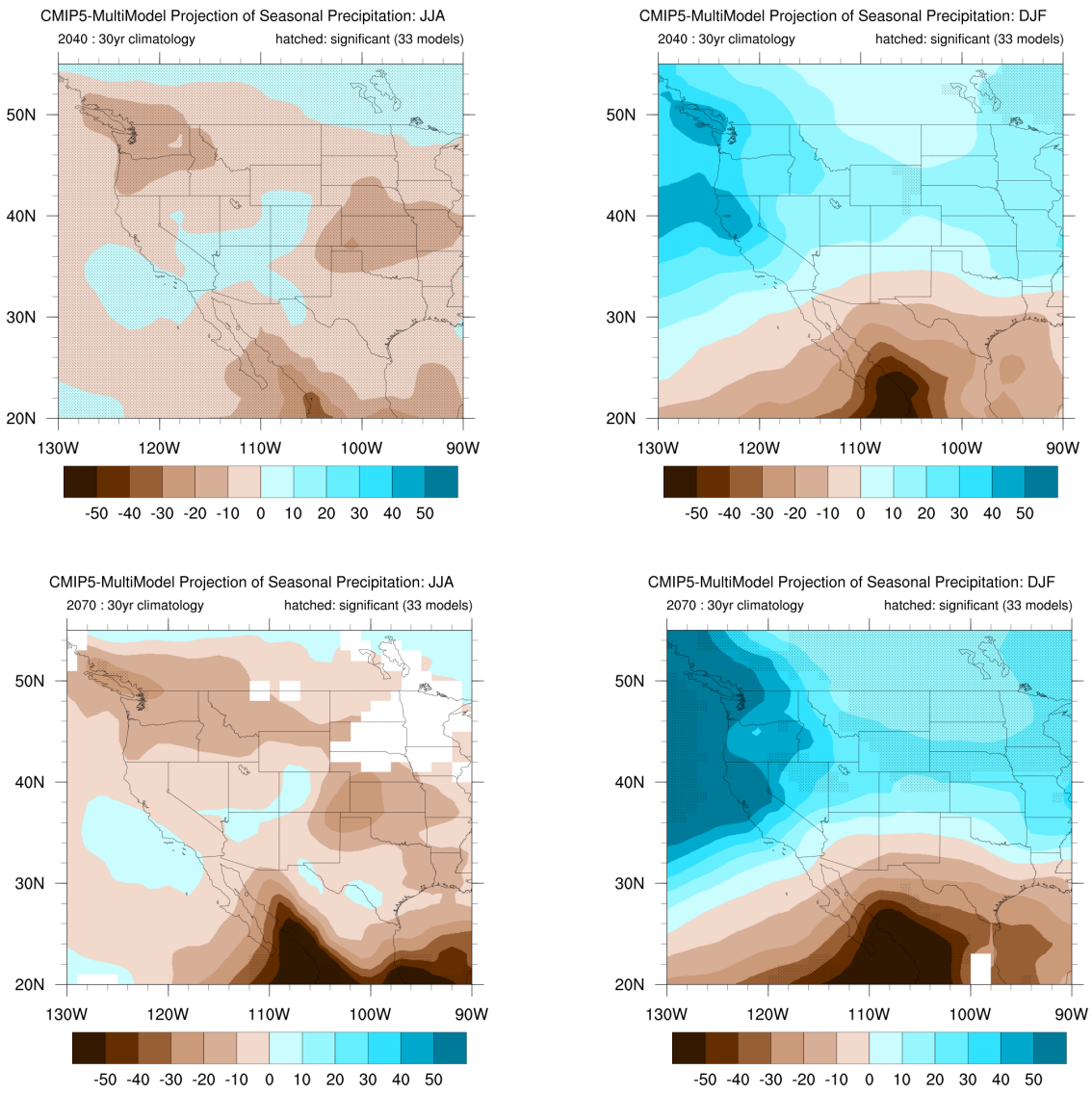


Figure 2-16. Projection of Precipitation Change (in mm) for the 2040 (Top) and 2070 (Bottom) Averaging Periods in Summer (JJA, Left) and Winter (DJF, Right) over the North American West.
Colored areas indicate at least two-thirds of models agree on the sign, and hatched areas additionally fulfill significance criteria at the 90% confidence level.

Figure 2-16 broadly outlines the dynamical feature of expected changes in the hydrologic cycle. Higher latitudes will very likely see an increase in precipitation due to an increase in transport of moisture in warmer air. The subtropical areas are more likely to see a decrease in precipitation due to enhanced downward motion in the descending branch of the Hadley cell. Based on the different climate model's depiction of this constellation, the intersect between these two large-scale domains falls somewhere over Southern California. The Upcountry region and East Bay and Peninsula regions are in the positive precipitation domain for the winter season in 2070, and thus are somewhat more likely to see positive than negative precipitation in the near future. This outlook appears consistent with the changes discussed above regarding the seasonal cycle. Positive trends during winter (Dec - February) align with the more robust trends to the north of the SFPUC regions. But because of the high variability, the significance of these changes in the means remains low.

When considering only the maximum precipitation events, however, then a trend towards higher values is apparent in the CMIP5 multi-model ensemble for RCP8.5. Figure 2-17 illustrates this, where the maximum daily precipitation in each month of each year is shown for the Upcountry region. This figure suggests an increasing trend in the maximum precipitation events over the coming decades in both the expected value (median and mean maxima) as well as an increase in the spread reflecting the heavy tail of the distribution of these maxima. This is consistent with the general idea that a warmer atmosphere can hold more water, and an intensified hydrologic cycle result in heavier precipitating events.

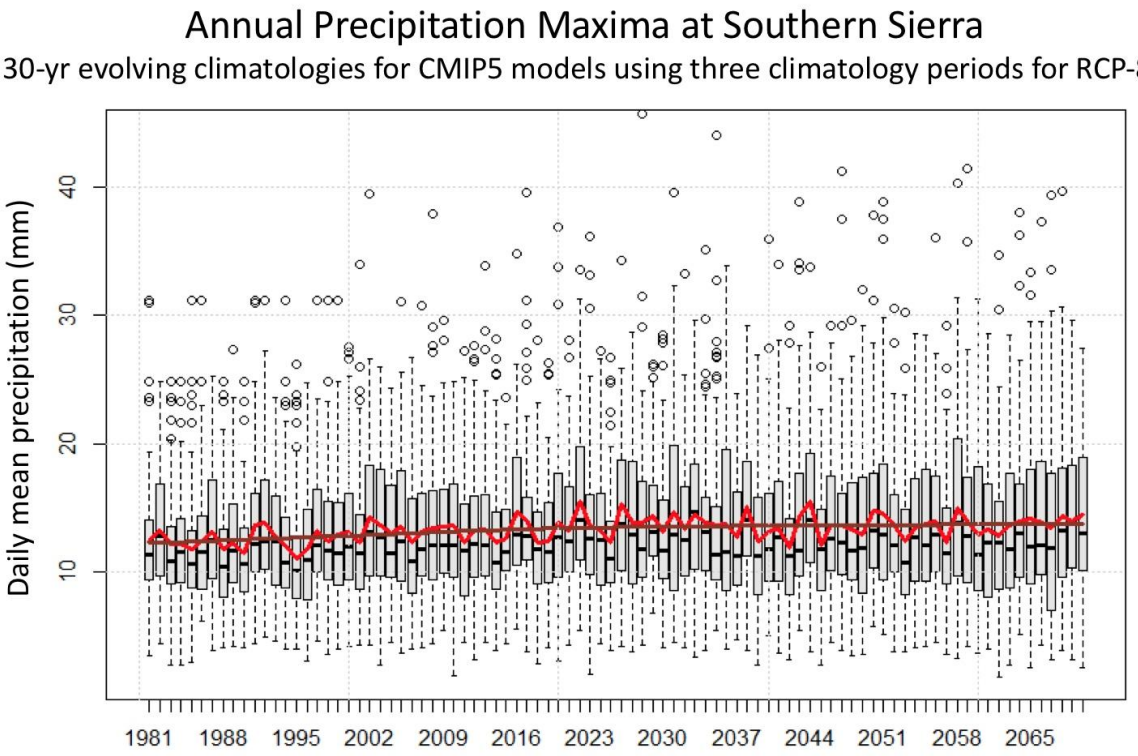


Figure 2-17. Projection of Change in Annual Maximum Precipitation (in mm/Day) for the Southern Sierra Showing an Increase from 12 to 14 mm.

2.2.3 Climate Information Elicitation Workshop

2.2.3.1 Elicitation Process

The SFPUC held a Climate Information Elicitation Workshop on March 21-22, 2019 (Figure 2-18). The workshop was organized by Robert Lampert, Sandy Berry, and Stephanie Tanverakul of RAND. The aim was to add depth to the information gathered from the CMIP5 model archive to inform the findings of the SFPUC vulnerability assessment. In particular, the elicited climate change information helps inform the extent to which the climate model projections are capturing the full uncertainty of future climate. The workshop sought expert opinion to help supplement the projections of mean annual temperature and precipitation changes from the CMIP5 archive for use in the decision-scaling analysis. The workshop focused on eliciting, where possible, most likely, best, and worst changes for the 2040 and 2070 planning horizons for climate change phenomena of greatest concern to the RWS. This included elicitation of opinions about mean changes in temperature and precipitation, but also about changes of extreme events identified by SFPUC staff and the consultant team as of significance to the SFPUC, but which don't emerge easily from climate model data. The workshop results are summarized here and are expected to appear in a future separate report.

Figure 2-18. Climate Workshop Venue in SFPUC Offices, March 2019.
Source: Courtesy of SFPUC.

The workshop employed a structured expert elicitation using a Delphi-based approach. The workshop convened nine climate scientists representing a variety of perspectives and disciplines to provide climate information relevant to SFPUC's vulnerability analysis. This workshop generated important information for the SFPUC and demonstrated a process that can be used by utilities and other adaptation planners more broadly. This work was led by the RAND Corporation and both process and outcomes are described in detail in 'Climate Information Workshop Goals, Process and Results' (Lempert et al. 2019). The experts' professional affiliations included universities (UCLA, University of Colorado), federal agencies (NASA, US Geological Survey), a federal research lab (Lawrence Berkeley National Laboratory), and research institutions (Scripps Institution of Oceanography and the National Center for Atmospheric Research). In addition, five SFPUC staff, three consultants from RAND, and three consultants preparing the vulnerability assessment gave presentations and were involved in facilitation and note-taking.

Participants were asked multiple times (2-3) for their expert judgement of the climate phenomena of interest, first a week before convening, and subsequently during convening. This allowed for clarification of objectives in asking the question and reconsideration of responses in light of group discussion. For the mean annual temperature and precipitation changes, scientists were asked for three estimates (most likely change, worst possible change and best possible change) for the horizon 2040 and the horizon 2070 for RCP8.5 and 4.5. These RCP scenarios relate to different possible levels of sensitivity of climate to any given GHG concentration in the atmosphere, an important source of uncertainty. All RCPs result in similar climate projections to mid-century (2040), but then projections diverge to the end-of-century (2070).

The questions were asked for changes in mean annual temperature and precipitation assuming 30 years centered on 2040 and 2070, baseline for estimating change was 1986-2005:

- What is the most likely change in mean annual temperature and precipitation?
- What is the worst possible case (hottest and driest)?
- What is the best possible case (coolest and wettest)?

For each question, degree of confidence was requested on a five-point scale from very low to very high based on the quality of scientific evidence. Workshop participants were also asked to evaluate other

climate parameters (OCPs) such as meteorological drought, the seasonal cycle of rainfall, frequency of atmospheric rivers and extreme precipitation, and changes in interannual variability. Participants were provided with description of the phenomena under consideration and a range of possible futures on either side of current conditions for selection – these phenomena, focused just on a 2070 future under RCP8.5. The OCPs are not discussed in this report but more information can be found in workshop report (Lempert et al. 2019).

Importantly, expert judgement can vary from climate model output for various reasons and in various ways. In general, in the literature and in this elicitation, extreme climate outcomes (in the "tail," or extreme end, of the distribution of outcomes) can be muted when available climate model output is considered due to limitations inherent in models seeking to project future climate outcomes under concentrations of GHGs that have not existed in the past. This statement applies to both extreme changes in climate as well as extreme weather events. As the IPCC AR5 stated (Flato et al. 2013), "Extreme events are realizations of the tail of the probability distribution of weather and climate variability. They are higher-order statistics and thus generally more difficult to realistically represent in climate models."

Models calibrate their projections against observed data, but atmospheric conditions that will exist in the future were never observed, nor do the extreme or high-end outcomes that might occur under those conditions (e.g., very hot temperatures, catastrophic melting of ice sheets, large storm events). Without observational data to train the models, they are thought to do a relatively poor job of projecting extreme outcomes. In addition, our understanding of how the climate will respond to increased GHG concentrations ("sensitivity") remains imperfect and it is therefore less likely that models will produce high quality projections for extreme climate change and extreme events. Coupled together, these uncertainties lead to a tendency for climate models to do a relatively poor job projecting extreme outcomes. Expert judgement, on the other hand, can supplement model output to incorporate general understanding of climate physics, data from outside the observational record (e.g., paleoclimatology), and other non-model judgements to project high-end outcomes. In recent years, these expert judgements have tended to suggest high end outcomes that are greater than those projected by models (e.g., Bamber et al. 2019). Finally, as discussed in the workshop report (Lempert et al. 2019), the convened climate science experts thought expansively when considering high end possibilities and felt unconstrained from rendering judgements that disagreed with the CMIP5 projection data archive. Precisely due to model limitations, expert judgement is increasingly being used to frame risk in adaptation planning.

2.2.3.2 Outcomes of Interest: Change in Mean Annual Temperature and Precipitation

The elicitation results for changes in mean annual temperature and precipitation are presented for the Upcountry, East Bay and Peninsula regions in Figure 2-19 for 2040 and 2070. The elicitation results are compared with CMIP5 model projections. Generally, the convened scientists believed warming and changes in precipitation could be more extreme than the CMIP5 model projections for all three regions. The envelope of uncertainty of change for these two climate variables increased compared to CMIP5 model projections. Statistics on the elicitation responses are presented in Table 2-1 for temperature change and in Table 2-2 for precipitation change. Elicitation results are discussed for each climate variable below.

Temperature

Workshop participants believe warming in the Upcountry region could exceed, under all scenarios considered, the picture provided by the CMIP5 model archive. For example, in a hot and dry scenario under RCP8.5 in 2070, many participants suggested warming could reach 6-7°C, while just one model

sampled reached this level of warming. For 2040, the elicitation produced a large cohort of opinions that warming would exceed 3°C, including several suggestions that warming could exceed 5°C; the CMIP5 archive included only a single estimate of temperatures above 3°C.

For the East Bay region, a similar level of increased warming in the elicitation emerged – as with Upcountry, the upper end for T under the hot and dry scenario for 2070 under RCP8.5 exceeded 5°C for most experts and for several exceeded 6°C, while model output projected under 5°C change in all but one realization. For the 2040 period, the elicitation skewed higher as well, though not as convincingly: the elicitation produced eight estimates of warming above 3°C for all scenarios, while the CMIP5 archive produced three.

For the Peninsula region, as would be expected in a coastal location, projected warming was significantly lower for both the elicitation and the models, with the elicitation skewing slightly higher for both 2040 and 2070. In 2040, elicitation experts thought warming at the high end would be in the 2.5-3.5°C range, while models at the upper end were around 2-3°C. For 2070, hot and dry conditions under RCP8.5 were thought to bring 4-5°C warming from virtually all the convened experts, while the CMIP5 archive fell largely around 3-4°C.

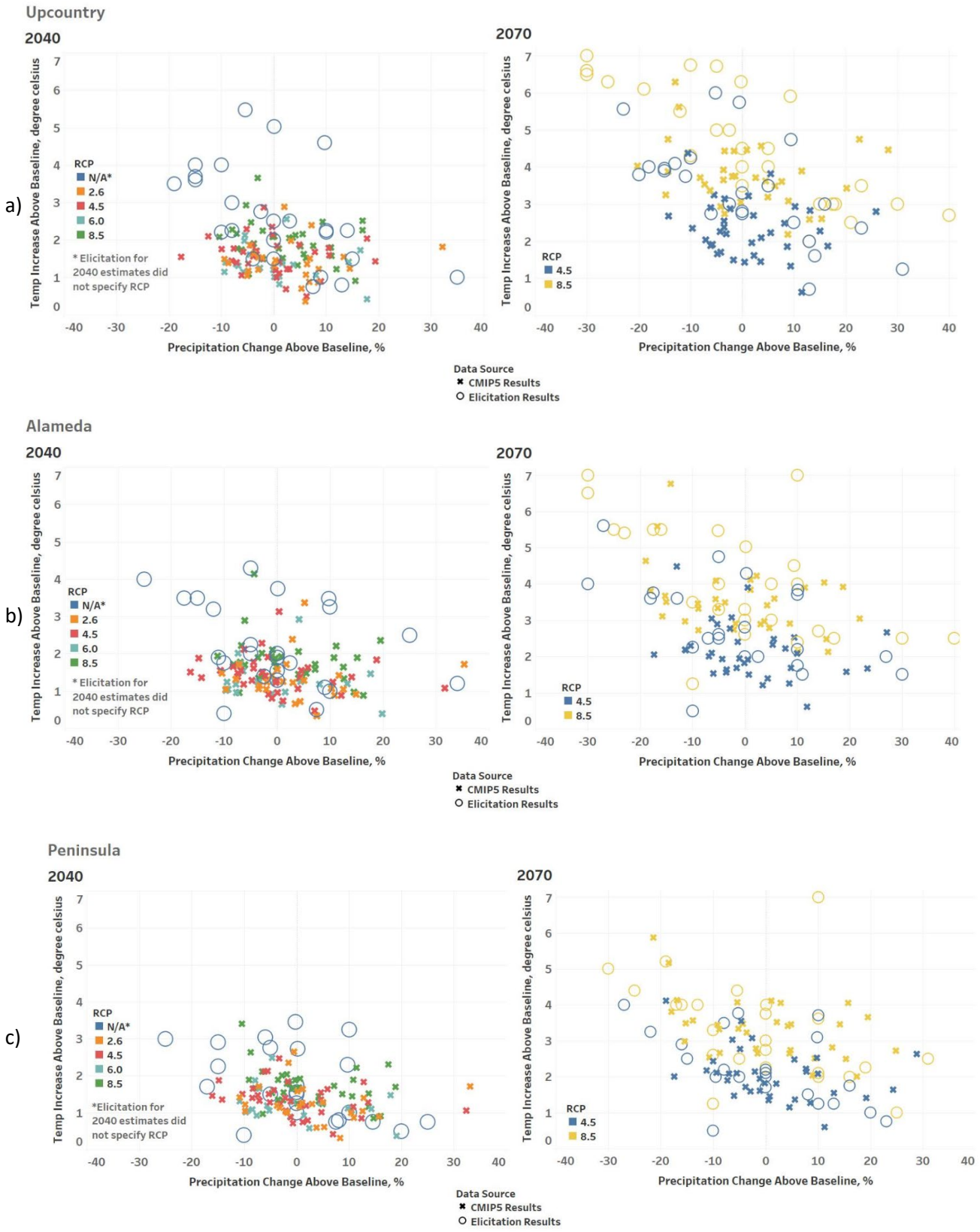


Figure 2-19. Elicitation Results for Changes in Mean Temperature and Precipitation Compared with CMIP5 Model Projections for Upcountry (a), East Bay (b), and Peninsula (c) Regions for 2040 and 2070.

In contrast, when asked about the "most likely" temperature outcomes across the three regions, elicitation expert judgements tended to mirror outputs from the CMIP5 archive. As indicated in Table 2-1 the elicitation responses to the "most likely" changes are:

- In the Upcountry region, for 2040, the "most likely" estimates ranged from 1.5 to 5°C with a median of 2.5°C. By 2070 RCP8.5, the range increased to 3.4 to 6.3°C and the median increased to 4.3°C.
- In the East Bay region, for 2040, the "most likely" estimates ranged from 1.3 to 3.8°C with a median of 1.9°C. By 2070 RCP8.5, the range increased to 2.6 to 5°C and the median increased to 3.3°C.
- In the Peninsula region, for 2040, the "most likely" estimates ranged from 1 to 2.7°C with a median of 1.5°C. By 2070 RCP8.5, the range increased to 2.3 to 4.2°C and median increased to 3°C.

Table 2-1. Descriptive Statistics of Elicitation Responses on Change in Mean Annual Temperature for Upcountry, East Bay, and Peninsula Regions for 2040 and 2070.

Mean Temperature Increase above Baseline (1986-2005, in °C)									
	2040			2070 Under RCP4.5			2070 Under RCP8.5		
	"Most Likely" Estimate	Coolest and Wettest	Hottest and Driest	"Most Likely" Estimate	Coolest and Wettest	Hottest and Driest	"Most Likely" Estimate	Coolest and Wettest	Hottest and Driest
Upcountry Region									
Mean	2.4	1.7	3.4	3.2	2.2	4.3	4.4	3.1	6.4
Standard Deviation	1.1	1.1	5.5	1.0	1.2	0.9	0.9	1.1	0.4
Maximum	5.0	4.6	5.5	5.7	4.7	6.0	6.3	5.9	7.0
Minimum	1.5	0.8	2.2	2.4	0.7	3.5	3.4	2.0	5.5
Median	2.5	1.0	3.5	2.8	2.30	4.0	4.3	3.0	6.5
East Bay Region									
Mean	2.0	1.4	3.1	2.6	1.9	3.9	3.6	2.6	5.7
Standard Deviation	0.7	1.0	0.9	0.7	0.9	0.9	0.7	0.8	1.0
Maximum	3.8	3.5	4.3	4.3	3.8	5.6	5.0	4.5	7.0
Minimum	1.3	0.4	1.8	2.0	0.5	2.3	2.6	1.3	4.0
Median	1.9	1.1	3.3	2.5	1.8	3.7	3.3	2.5	5.5
Peninsula Region									
Mean	1.6	1.0	2.6	2.2	1.5	3.3	3.2	2.2	4.7
Standard Deviation	0.5	0.6	0.7	0.6	0.8	0.7	0.7	0.8	1.2
Maximum	2.7	2.3	3.3	3.5	3.1	4.0	4.2	3.6	7.0
Minimum	1.0	0.4	1.3	1.7	0.5	2.0	2.3	1.0	2.6
Median	1.5	0.8	2.9	2.1	1.3	3.5	3.0	2.1	4.4

Precipitation

The convened scientists identified a range of possible precipitation futures on either side of no change for all futures and all regions – an outcome that largely mirrors output from the CMIP5 model

projections. Again, projections of extreme change – for both increased and decreased precipitation – tended to be more dramatic in the elicitation than the CMIP5 models. Excursions beyond the CMIP5 archive were more pronounced for the 2070 period, and most so for RCP8.5 scenario. For example, for the Upcountry watershed, model projections for 2070 RCP8.5 fell between -20% and +25% (all but one), while elicitation responses fell between -30% and +40%. For 2040, precipitation projections clustered for the most part in the -20% to +20% range for both experts and models, showing rare consensus between expert opinion and computer model ensemble. These results appeared for the East Bay and Peninsula regions as well, though in the Peninsula the effect was more muted.

For precipitation in East Bay region for 2070 RCP8.5, results were similar to Upcountry, with CMIP5 models projecting between -20% to +30% precipitation change while elicitation participants estimated between -30% and +40%. For 2040, expert opinion both projected precipitation would fall between -25 and +35%, though the preponderance of estimates from both sources fell in the -20% to +20% range.

For the Peninsula region, as with temperature, precipitation change – and the distinction between elicitation and models -- was more muted, with the elicitation finding a range of -30% to +30% by 2070 under RCP8.5, while the models suggested a range between -20% and +30%. For 2040, the result was somewhat anomalously mixed at the tails, with experts estimating a range of -30% to +25% while the models showed -25% to +35% for precipitation change.

As indicated in Table 2-2 the elicitation responses to the “most likely” changes are:

- In the Upcountry region, for 2040, the “most likely” estimates ranged from -4% to 3% with a median of 0%. By 2070 RCP8.5, the range increased to -10% to 5% and the median increased to 0%.
- In the East Bay region, for 2040, the “most likely” estimates ranged from -5% to 0% with a median of 0%. By 2070 RCP8.5, the range increased to -10% to 5% and the median increased to 0%.
- In the Peninsula region, for 2040, the “most likely” estimates ranged from -5% to 0% with a median of 0%. By 2070 RCP8.5, the range increased to -10% to 2% and median increased to 0%.

Table 2-2. Descriptive Statistics of Elicitation Responses on Change in Mean Annual Precipitation for Upcountry, East Bay, and Peninsula Regions for 2040 and 2070.

Mean Precipitation Change from Baseline (1986-2005, in %)									
	2040			2070 Under RCP4.5			2070 Under RCP8.5		
	Most Likely Estimate	Coolest and Wettest	Hottest and Driest	Most Likely Estimate	Coolest and Wettest	Hottest and Driest	Most Likely Estimate	Coolest and Wettest	Hottest and Driest
Upcountry Region									
Mean	-0.4	14.1	-11.5	-0.3	16.3	-14.6	-0.9	21.1	-20.2
Standard Deviation	1.9	8.3	4.3	2.9	6.9	5.5	4.7	9.3	9.5
Maximum	3.0	35.0	-5.4	5.0	31.0	-5.2	5.0	40.0	-4.9
Minimum	-4.0	7.5	-19.0	-6.0	9.5	-23.0	-10.0	9.3	-30.0
Median	0.0	13.0	-10.0	-0.0	14.0	-15.0	0.0	21.0	-19.5
East Bay Region									
Mean	-1.4	11.3	-11.0	-1.7	12.0	-14.1	-1.4	15.2	-14.3
Standard Deviation	2.2	12.6	9.6	3.1	12.1	11.9	4.5	14.5	15.7
Maximum	0.1	34.0	10.0	1.5	30.0	10.0	5.0	40.0	10.0
Minimum	-5.0	-10.0	-25.0	-7.0	-10.0	-30.0	-10.0	-10.0	-30.0
Median	0.0	9.8	-12.0	0.0	10.1	-16.5	0.0	14.0	-20
Peninsula Region									
Mean	-1.1	10.9	-10.5	-1.3	11.9	-12.2	-1.4	14.8	-14.4
Standard Deviation	2.2	9.7	9.9	3.1	9.6	10.8	3.7	11.8	11.8
Maximum	0.2	25.0	10.0	1.5	23.0	10.0	2.0	31.0	10.0
Minimum	-5.0	-10.0	-25.0	-8.0	-10.0	-27.0	-10.0	-10.0	-30.0
Median	0.0	10.0	-14.0	0.0	13.0	-15	0.0	16.0	-17.0

2.3 Setting Bounds for the Stress Test

Based on the findings of the CMIP5 models and the expert elicitation, it was decided to limit the climate stress test to changes to mean annual temperature and precipitation applied uniformly across seasons and regions. The central tendency for both CMIP5 projections and elicitations agreed to warming roughly around 2°C and minimal precipitation change. However, both datasets also suggest deep uncertainty about the potential climate change due to the large range of plausible changes for both temperature and precipitation as shown in Figure 2-20. The figure presents the CMIP5 projections and elicitation results for 2040 and 2070 for the Upcountry region. This study follows the principle of decision making under deep uncertainty which is to utilize an approach that incorporates the full plausible range of climate change projections, without regard to likelihood, to explore a comprehensive set of possible outcomes and find system vulnerabilities. Therefore, for the stress test, the warming in mean annual temperature will vary from 0 to 7°C and the change in mean annual precipitation will vary between -40% and +40% of historical conditions. The historical baseline is from 1986 to 2005.

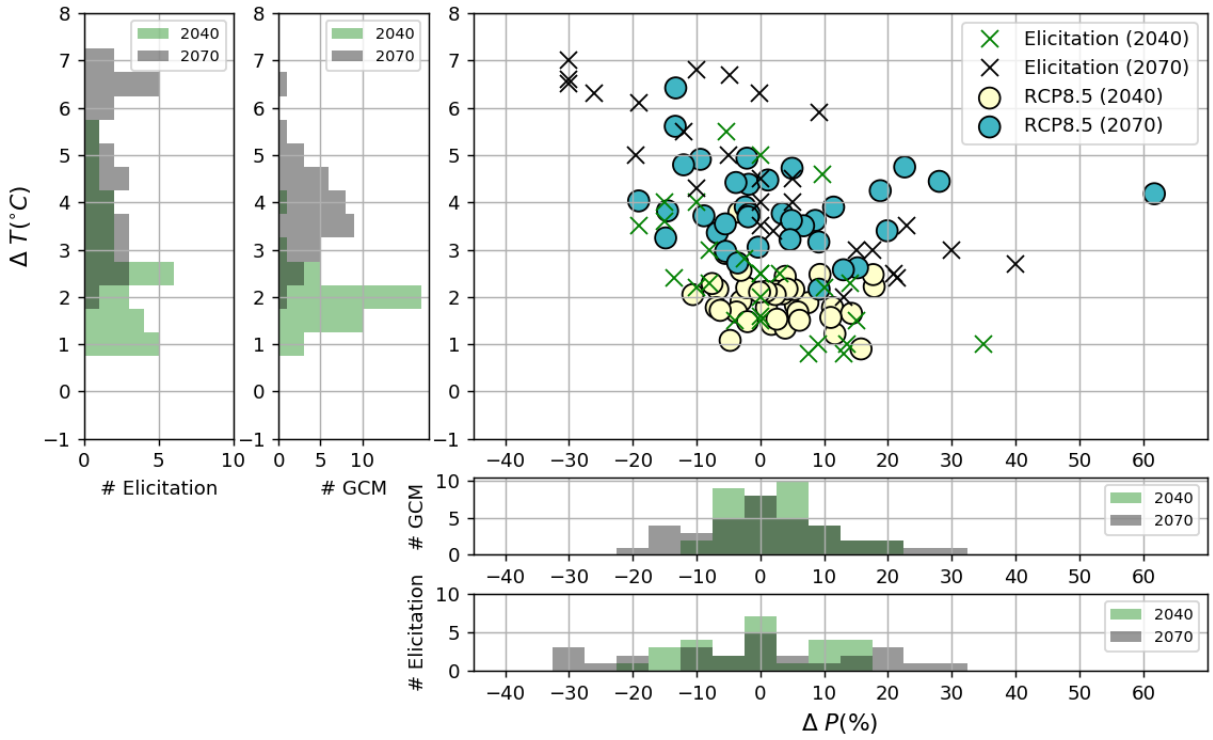


Figure 2-20. Comparison of the Results from the CMIP5 Projections and the Expert Elicitation for the Upcountry Region.

The scatterplot shows the average relative change in precipitation (x-axis in %) and the average absolute change in temperature (y-axis in ^{o}C). CMIP5 projections are shown with circles and expert elicitations are shown with cross symbols. Results are given for the Upcountry region, two future periods: 2040 (2026-2055; yellow color) and 2070 (2056-2085; blue color) and for the RCP8.5. Baseline is 1986-2005. The histogram plots show the distribution of CMIP5 projections and expert elicitations.

CHAPTER 3

Modeling Methods for the Hetch Hetchy RWS

The modeling system, including inputs, model relationships and connections, and outputs, are shown in Figure 3-1. The main inputs and outputs to each model, and their spatial and temporal scopes, are summarized in Table 3-1. This section describes each component of the modeling system so that one can understand the assumptions and the results of the LTVA. However, a more comprehensive description of each model and method is available in associated supplementary materials (HRG 2018, 2020a, 2020b, 2021a, 2021b, and 2021c).

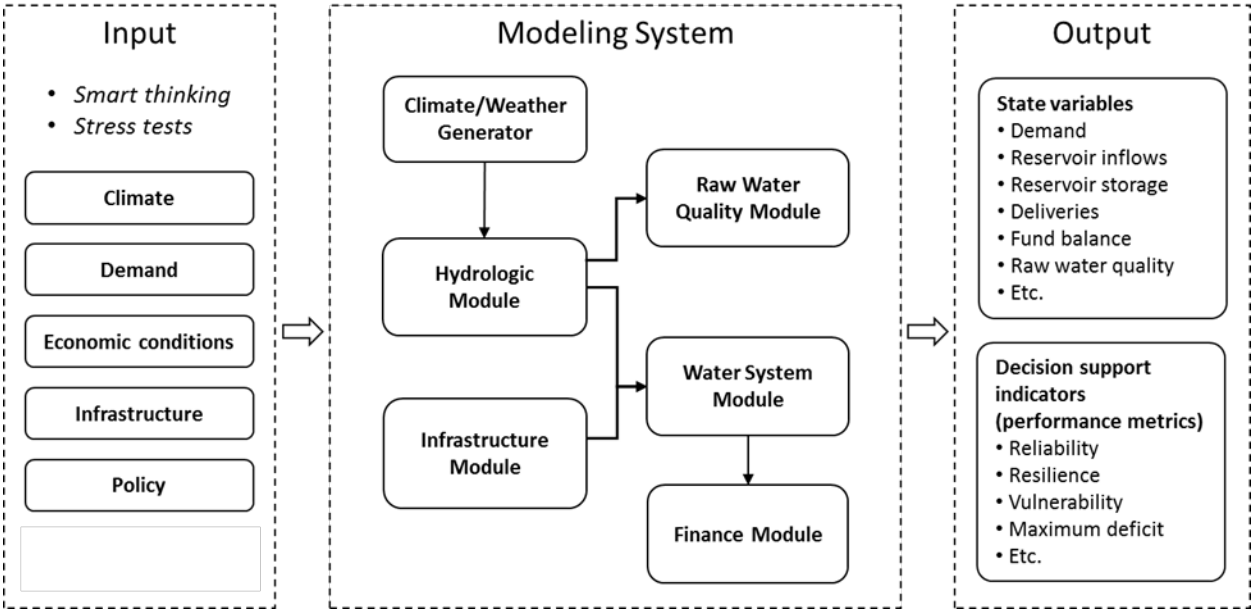


Figure 3-1. Conceptual Diagram of the Modeling System Developed for the LTVA.
including a summary of inputs and outputs.

Table 3-1. Summary of Specific Models in the Modeling System Developed for the LTVA.
All physical models use the daily time step. The only non-physical model—the finance model—is on an annual time step.

Module	Inputs	Outputs	Time step
Weather Generator	Observed temperature and precipitation	Simulated time series of temperature and precipitation; temperature and precipitation scenarios	Daily
Hydrology	Temperature, precipitation, topography, misc. physical-based parameters, river channel routing parameters	Simulated time series of streamflow into each reservoir	Daily
Water System	Physical system characteristics, inflow into reservoirs, operating rules, water demand, instream flow requirements, water rights, and others	A wide range of system state variables (storage in reservoirs, flow through main conveyances, deliveries to customers, …)	Daily
Demand	Retail and wholesale water price, climate conditions, population, policy and management objectives	Water demand for retail and wholesale service areas	Daily
Water Quality	Hydrometeorological conditions + system state variables (e.g., Hetch Hetchy storage, flow through San Joaquin pipelines …).	Total Organic Carbon (TOC) and Turbidity	Daily
Finance	Deliveries, OPEX, CAPEX, interest rate, key financial management policies	Price of water (wholesale and retail)	Annual

The weather generator is a key component of the climate stress test. It creates time series of temperature and precipitation that systematically sample plausible climate conditions that i) might have occurred over the historical period and ii) may occur in the future. It is specifically designed to create equally probable time series that represent realizations of natural climate variability. Weather generator outputs provide input to the hydrology models to model daily river flows across the three regions of interest: the Upper Tuolumne River watershed in the Central Sierra Nevada (“Upcountry”); the Alameda Creek watershed in the East Bay (“East Bay,” “Sunol Valley” or “Alameda”); and the San Francisco Peninsula (“Peninsula”). Weather generator outputs are also input to the water demand model that simulates daily demand for 27 wholesale customers, San Francisco city and sub-urban customers. Simulated time series of streamflow variable at various key locations of the Hetch Hetchy RWS and water demand at SFPUC customer locations are the main inputs for the water system model (hereafter called San Francisco Water System Model, or, for the short, SFWSM). Other inputs to SFWSM include major infrastructure (e.g., reservoirs and storage capacity) and of their operations/regulations (e.g., minimum and maximum preferred storage levels), precipitation and evaporation rate over the reservoirs, plus operational policies that reflect the management of the system, particularly during drought. SFWSM outputs are then input to the finance and water quality models; the latter also uses

outputs from the weather generator and the Upcountry hydrology model. Outputs from hydrology models, SFWSM, water quality and finance models are used to calculate various performance metrics (Section 4.4) that are used to assess system performance and/or vulnerability.

This section is organized as follows. The stochastic weather generator developed by the HRG is described in Section 3.1. This section also includes the presentation of, and comparison with, the weather generator developed by NCAR. The hydrology models built for the three regions of the Hetch Hetchy RWS are presented in Section 3.2. This section also describes the frequency analysis conducted for drought severity and duration at La Grange. The reason why the Tuolumne River flow at La Grange is important is because SFPUC is the junior water right holder on the Tuolumne River at Don Pedro Reservoir, while the Modesto and Turlock Irrigation Districts (MID and TID) are senior water right holders. The amount of water available for SFPUC to serve their customers (aka WAC, which stands for Water Available to the City) is determined based on these water rights, which are calculated based on flow at La Grange. The water demand model is described in Section 3.3. The water system model is presented in Section 3.4. Raw water quality models are described in Section 3.5 while the finance modeling approach is presented in Section 3.6.

3.1 Stochastic Weather Generator

3.1.1 Modeling Goals

Stochastic weather generators are mathematical algorithms that produce time series of synthetic weather data at desired spatial and temporal resolution. The parameters of the model are conditioned on existing meteorological records to ensure that the characteristics of historical weather emerge in the daily stochastic process. Stochastic weather generators can also be used to perform exhaustive assessments of a system's vulnerability to climate conditions across multiple temporal scales, including changes in mean climate and variability (Steinschneider and Brown 2013). Stochastic weather generators can be used to produce new realizations of a time series of weather variables that exhibit similar statistics as the historical record, thus producing an ensemble of time series that provides a sample of the historical or "natural" variability. By incrementally manipulating one or more parameters in a weather generator, one can simulate many climate scenarios that exhaustively explore potential futures that exhibit slight differences in nuanced climate characteristics, such as the intensity and frequency of daily precipitation, the serial correlation of extreme heat days, or the recurrence of long-term droughts.

The weather generator developed for the LTVA was used to produce time series of temperature and precipitation at locations of weather stations for use in the hydrologic (Bay Area and Sierra Nevada) and demand modeling (retail and wholesale service area). These output time series help explore the effects of variations similar to observed historical conditions, as well as climate variability beyond the historical record due to changes in future temperature and precipitation. Outputs from this module are used to evaluate effects of warming temperatures and changing precipitation on the streamflow to the RWS and on the resulting reservoir operations. This module enables evaluation of the RWS performance over a range of possible drought sequences.

3.1.2 Modeling Method

The CliWxGen weather generator developed for the LTVA is a simplified version of the weather generator proposed by Steinschneider and Brown (2013). This section provides a technical overview of the CliWxGen. Figure 3-2 illustrates the four major phases of the CliWxGen. Note that the description below summarizes the main steps of the methodology, rather than describing in detail every single step of the flow chart shown in Figure 3-2. A more complete and technical description of the different steps

involved in the generation of the climate realization via the CliWxGen stochastic weather generator is available in the Technical Report 1 (HRG 2018).

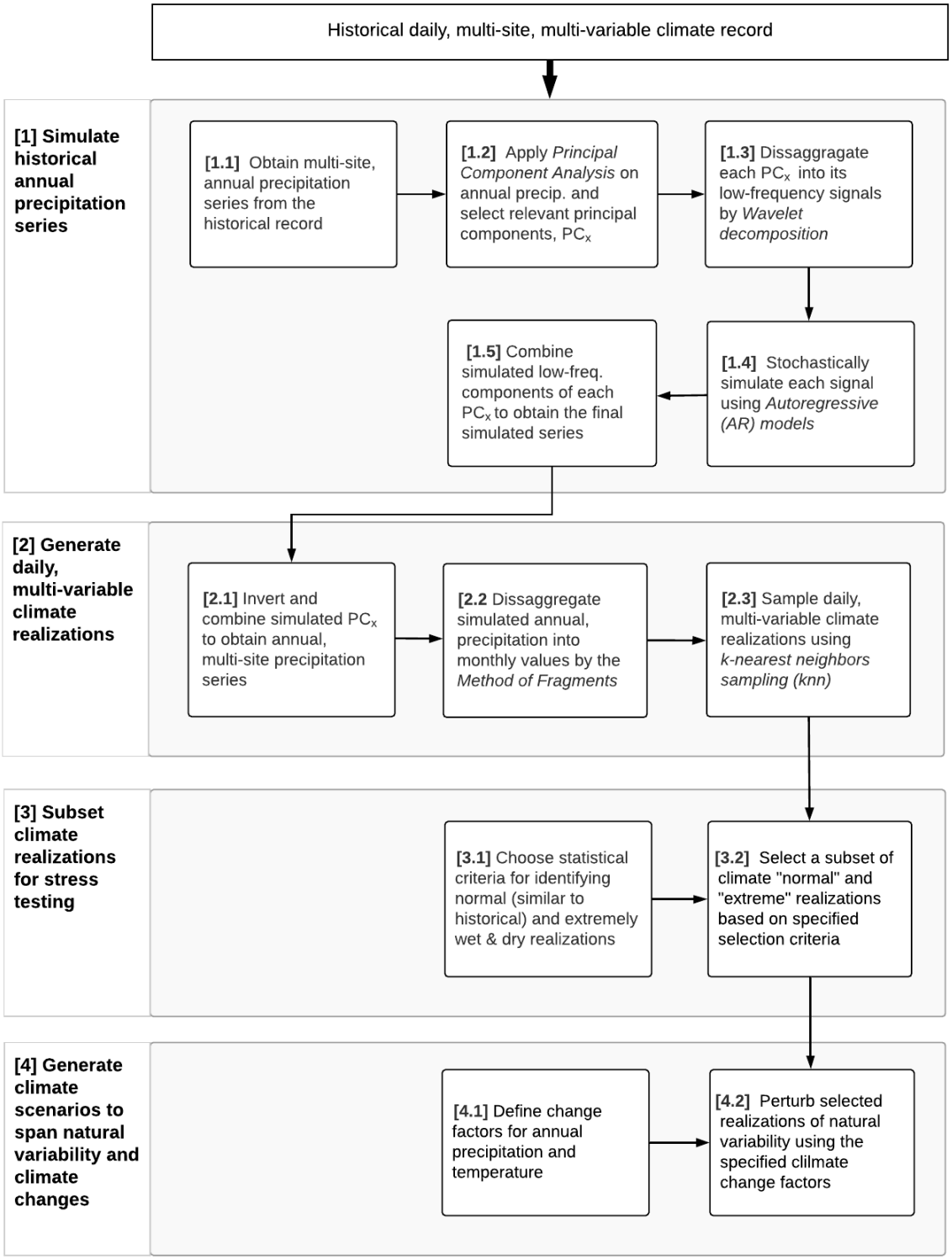


Figure 3-2. Flow Chart Describing the Four Main Phases for Development of the CliWxGen Stochastic Weather Generator for the Hetch Hetchy RWS.

In Phase 1, the goal is to produce stochastic simulations of historical annual precipitation over the study region. If the climatology is homogenous within the study area, location-specific annual precipitation data can be area-averaged to obtain a single, representative time-series. However, if there are significant differences in climatology within the region such as in this study, simple-averaging is not appropriate. Instead, principal component analysis (PCA) can be applied to preserve the variability between different locations. The stochastic time-series generation is applied using a wavelet autoregressive model (WARM) on the represented annual precipitation series (Kwon et al. 2007). The WARM procedure first decomposes the annual series into significant low-frequency signals and the residual error term (noise). Each low-frequency component and the residual error is then simulated stochastically using best-fit linear autoregressive (AR) models. Finally, the simulated low frequency and noise component(s) are aggregated to obtain the simulated representative series of annual precipitation.

In Phase 2, simulated annual series from Phase 1 is disaggregated in time and space to obtain daily climate variables (precipitation, temperature) at all locations. This is done in three subsequent steps. First, the simulated annual series is inverted to its original scale and disaggregated spatially into annual precipitation series for all gauge stations (Phase 2.1). Next, the inverted series is disaggregated into monthly values using the method of fragments, a non-parametric resampling approach widely applied in climate simulation and disaggregation (Silva and Portela 2012) (Phase 2.2). The result of this stage is a monthly precipitation time series at each of the considered rain-gauges location. Finally, a k-nearest neighbors (KNN) resampling algorithm is used to further disaggregate the monthly precipitation series into daily values. To preserve the spatial and seasonal consistency between precipitation and temperature, daily minimum, and maximum temperature time series are sampled from the closest k-nearest neighbor together with the precipitation variable. (In Phase 2.3). At the end of Phase 2, a complete set of climate realizations are obtained, which sample historical climate variability.

In Phase 3, the generated dataset of daily, multi-variable, multi-site climate realizations are reduced to a smaller set of realizations. This step is not mandatory in the stochastic weather generation process but is desired to reduce computational challenges (i.e., due to the need to store and simulate a large number of climate realizations in hydrosystem models). In this phase, a small set of realizations is carefully selected that can span the initial variability range. This is done by first selecting criteria for ranking of the climate realizations (e.g., drought severity, extreme precipitation), and then subset from the initial set of realizations using an appropriate statistical method such as a sequent peak algorithm (Whateley et al. 2016).

In Phase 4, daily, multi-site, multi-variable climate realizations are perturbed to simulate a wide range of future climate changes. This is the final phase of the stochastic weather generator, where the underlying statistics of climate realizations are perturbed to obtain a range of scenarios that represent both historical (natural) climate variability and a possible range of climate changes. This is done through different statistical procedures for temperature and precipitation variables. For temperature variables (i.e., minimum, average, and maximum daily temperature), additive change factors are used to impose an increase over historical conditions (e.g., 4°C). For precipitation, multiplicative change factors are used to perturb the historical mean values (e.g., 10%).

3.1.3 Spatial and Temporal Scopes of the Weather Generator

When developing a stochastic weather generator, it is desirable to have a weather record that is long enough (e.g., 50 years or more) to provide a good representation of the in-situ climatological features such as low-frequency variability. Furthermore, the resampling techniques used in the stochastic weather generator (Phase 2.3 in Figure 3-2) provides more diverse outcomes when the underlying input

climate series is sufficiently long. In most locations, long, continuous records of weather series may not be available, which poses a limitation for weather generator development. To circumvent this issue, weather generators often make use of gridded climate products, where corresponding time series of temperature of precipitation are available for large study regions.

As discussed extensively in the Technical Report 1 (HRG 2018), the length of the observed daily temperature and precipitation records varies across the available stations in the three regions of interest for this study. Because only a limited number of stations provide long, continuous data, the suitability of using gridded climate data was investigated by comparison with local observations. The suitability of the CONUS dataset, a gridded climate product covering a period of more than 50 years at $1/16^{o}$ spatial resolution (6 km^2) (Livneh et al. 2015), was evaluated for the study regions of interest. The CONUS gridded precipitation data was found to be not suitable for conditioning the weather generator (HRG 2018). As such, the thirteen precipitation gauges that have the longest records and are representative of precipitation across the three regions (Table 3-2) were used as inputs for the development of the weather generator (Phase 1.1 in Figure 3-2). Conversely, the CONUS gridded temperature data was found to be suitable for use in weather generator development. Thus, the weather generator makes use of 55 years (from 1956 to 2011) of data by combining the availability of precipitation gauge data and the CONUS gridded temperature dataset.

Table 3-2. Precipitation Gauges Used for Development of the CliWxGen Weather Generator.

Station	Short name	Region	Period of record (full range)	Period of record (used)
Upper Crystal Springs	UCS	Peninsula	1908 – 2017	1956 - 2011
Lower Crystal Springs	LCS	Peninsula	1915 - 2017	1956 - 2011
San Andreas	SA	Peninsula	1908 - 2017	1956 - 2011
Pilarcitos	PLD	Peninsula	1909 - 2017	1956 - 2011
Sunol	SUNO	East Bay	1907 - 2017	1956 - 2011
Calevaras	CAL	East Bay	1915 - 2017	1956 - 2011
Mt. Hamilton	HML	East Bay	1948 - 2017	1956 - 2011
Hetch Hetchy	HTH	Upcountry	1930 - 2017	1956 - 2011
Early Intake	IN	Upcountry	1930 - 2017	1956 - 2011
Moccasin	MCN	Upcountry	1930 - 2017	1956 - 2011
Cherry Valley	CVM	Upcountry	1952 - 2018	1956 - 2011
Yosemite	YOS	Upcountry	1956 - 2017	1956 - 2011
Sonora	SON	Upcountry	1956 - 2017	1956 - 2011

3.1.4 Climate Realizations

3.1.4.1 Climate Realizations from the Weather Generator

The climate stress test described in Section 4.2 is conducted by perturbing a set of weather realizations of daily precipitation and temperature over a 50 year time horizon that were developed using the CliWxGen weather generator above (Section 3.1.2). The reason for using multiple realizations of climate is to account for internal climate variability in addition to climate change. In this study, internal climate variability is deemed to be represented by the stochastic processes within the generation of the annual precipitation time series using WARM (Phase 1) and the spatial and temporal downscaling using the

method of Fragments (Phase 2.2) and KNN (Phase 2.3). For the study region, wavelet analysis conducted on annual precipitation in Section 2.1.3 identified a low-frequency signal of about 15 years, which is considered to be a relatively long wavelength (see the Technical Report 1 for more discussion about the wavelet analysis (HRG 2018).

The CliWxGen modeling framework described in Figure 3-2 resulted in the simulation of 1,000 precipitation and temperature realizations of 50-year long. To maintain a manageable volume of simulations for the climate stress test, a subset of nine climate realizations were selected from the initial ensemble of 1,000 realizations. This subset of nine realizations was selected to cover a broad and balanced range of possible future climates. As such, the subset includes three representative of "normal" realizations, which are similar to the historical record, and a total of six realizations that include realistic but extremely wet and dry years or sequence of years. It is important to emphasize that all nine realizations have approximately the same average annual precipitation but differ in terms of extreme wet and dry years.

The selection of the three "normal" realizations among the 1,000 realizations available relies on four criteria; all based on the monthly precipitation at the rain gauge near Hetch Hetchy Reservoir. The criteria are: i) the average, ii) the standard deviation, iii) the maximum and iv) the maximum cumulative deficit. The considered deficit metric used in iv) is based on the deviation to the mean. Candidate realizations for the selection of the "normal" realizations must not have an average rainfall that differs by more than 0.5% from the historic record, while all the other considered criteria must not differ by more than 2% from the historic. The three selected "normal" realizations are randomly sampled from the subset of realizations that satisfy the above criteria.

For illustration, Figure 3-3 displays the annual precipitation time-series at Hetch Hetchy gauge for the three selected "normal" stochastic realizations in comparison to the historical time series. Hereafter, the three 'normal' stochastic realizations are indexed from 1 to 3.

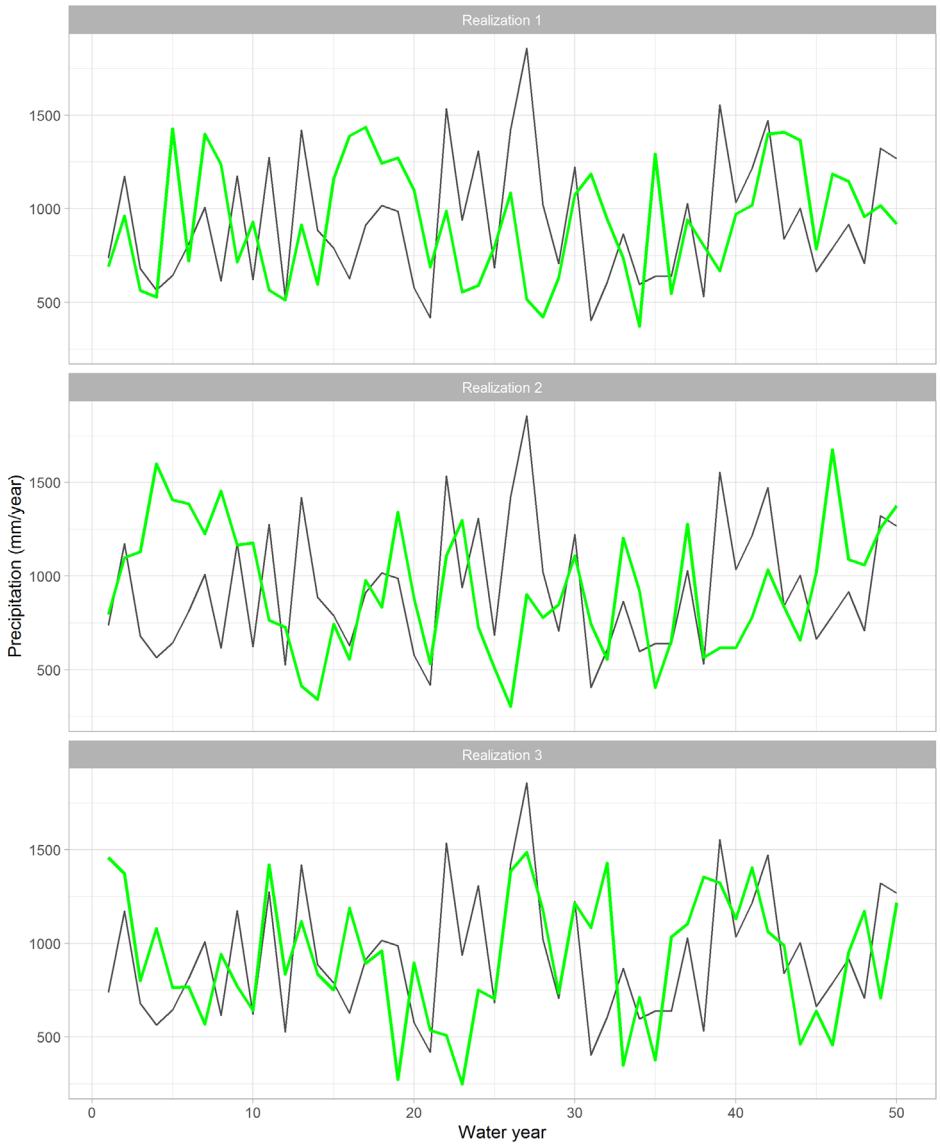


Figure 3-3. Annual Precipitation Series from Three Selected Climate Realizations (in Green) Compared to the Historical Time Series from 1956 to 2005 (in Grey).
All values are for the Hetch Hetchy gauge station.

To select the realizations with extreme wet/dry years or sequence of years, extremely wet and dry years were first identified, based on an analysis of historical annual precipitation series from the Hetch Hetchy gauge (from the year 1956 to 2011). Extremely wet and dry years were defined based on the 95th and 5th percentiles of the empirical distribution of historical annual precipitation, respectively. These threshold values correspond to greater than 1,524 mm/year (60 inches/year) and less than 465 mm/year (18.3 inches/year), for wet and dry years respectively. Using these thresholds, several criteria have been estimated to characterize the extreme years or sequence of years within each realization. These criteria are: i) frequency of extremes (i.e., count of values below or above the threshold), ii) maximum duration of extremes (i.e., longest sequence of consecutive extreme values), and iii) severity of extremes (i.e., maximum magnitude of extreme values). This ranking resulted in selection of three extremely wet and three extremely dry realizations used in the LTVA, which are shown in Figure 3-4 and Figure 3-5, respectively. Hereafter, the three 'wet' stochastic realizations are indexed from 4 to 6, while the three 'dry' realizations are indexed from 7 to 9.

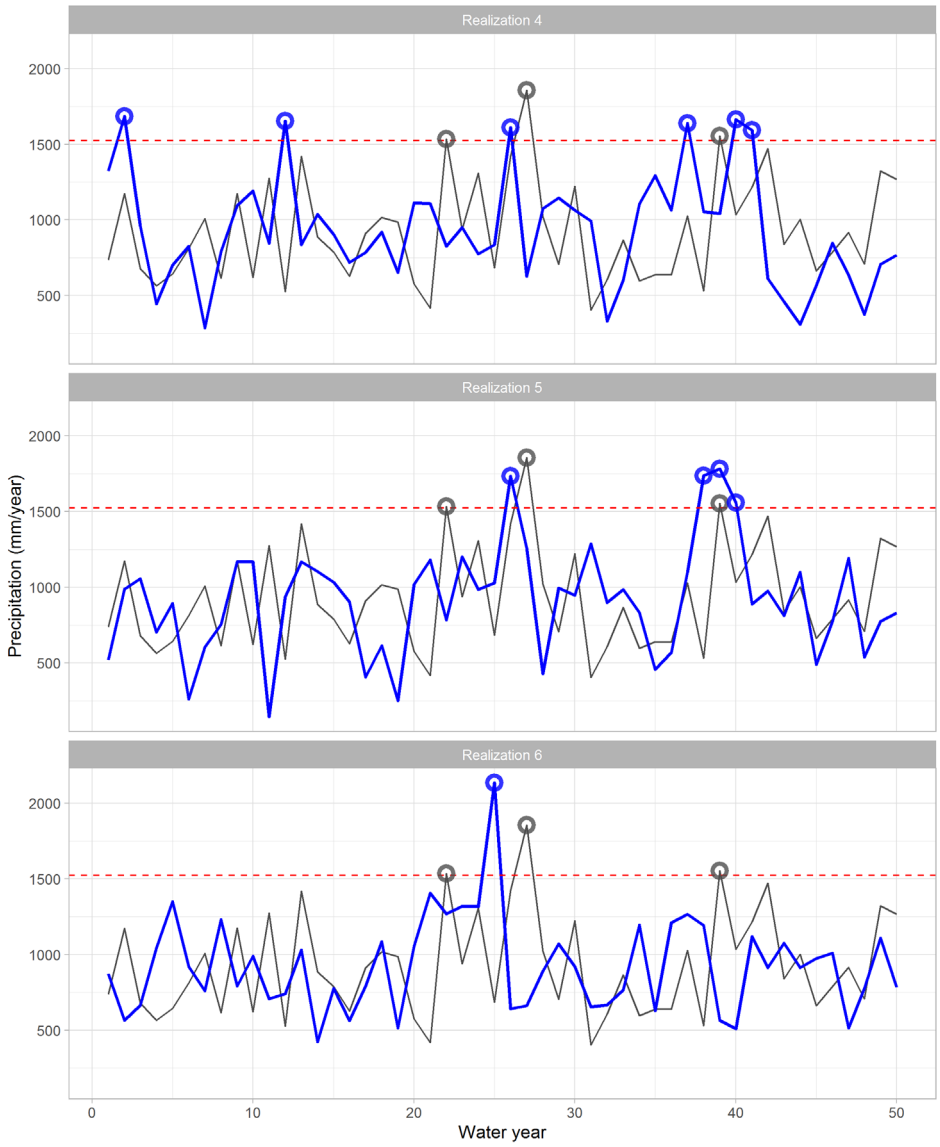


Figure 3-4. Annual Precipitation Series from Three Selected Wet Climate Realizations (in Blue) Compared to the Historical Time Series from 1956 to 2005 (in Grey).

All values are for the Hetch Hetchy gauge station. The red dashed line marks the threshold of 60 inches/year. The bullets show annual precipitation values that exceed this threshold.

Realization 4 is the realization with the largest number of extremely wet years (i.e., 6), which is twice the count of extremes found in the historical record. Realization 5 has the longest sequence of extremely wet years (i.e., 3) (water years 36, 37 and 38). Finally, Realization 6 has the wettest simulated year with total precipitation of about 2,133 mm/year (84 inches/year).

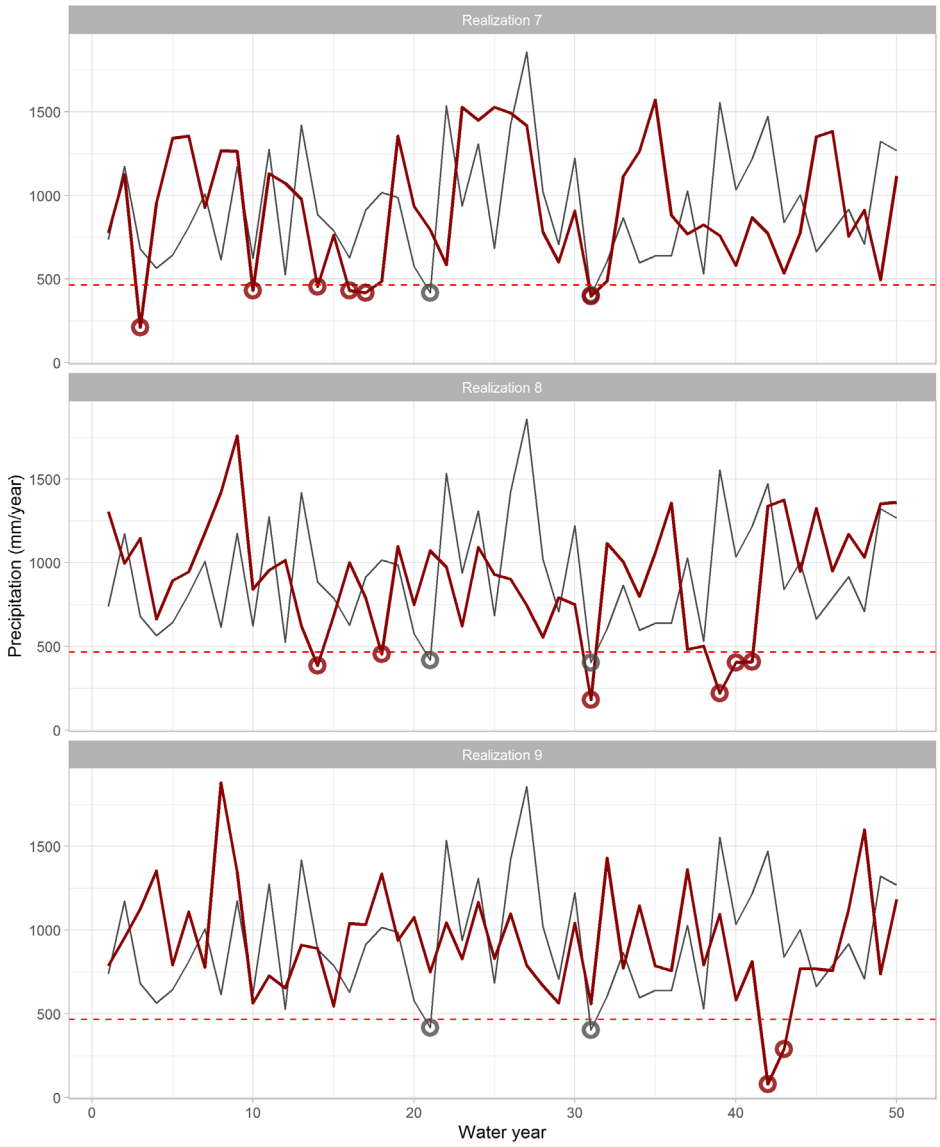


Figure 3-5. Annual Precipitation Time Series from Three Selected Dry Climate Realizations (in Red) Compared to the Historical Time Series from 1956 to 2005 (in Grey).

All values are for the Hetch Hetchy gauge station. The red dashed line marks the threshold of 18.3 inches/year. The bullets show annual precipitation values below this threshold.

The selection of the 'dry' realizations was based on the criteria mentioned above, although the most extreme realizations have not been selected. HRG selected three balanced yet challenging realizations. Realization 7 has six extremely dry years (<465 mm/year), which is two times that found in the historical time series. Note that 3 dry years occur within a window of 4 years. Realization 8 has a 3-year prolonged drought (years 39-41) and six extremely dry years. Note that the two years preceding the 3-year long sequence of extremely dry years are both barely above the extremely dry threshold of 465 mm/year (20 inches/year). Even a minor reduction in precipitation could turn the 3-year long sequence of extremely dry years into a 5-year long sequence. Finally, Realization 9 has a severe two-year extreme drought (years 42-43), in which the total precipitation during the first year is barely above 76 mm (3 inches).

Overall, these six extremely wet and dry traces provide a rich representation of stochastic (natural) variability of the climate in addition to the three representative traces previously shown.

The precipitation for the 9 selected realizations should be representing the yearly and monthly statistics of the historical observed data. Annual precipitation variability is well represented at Upcountry stations with extreme dry years having less precipitation than the observed data (Figure 3-6). The same observation is true for Bay Area stations (not shown).

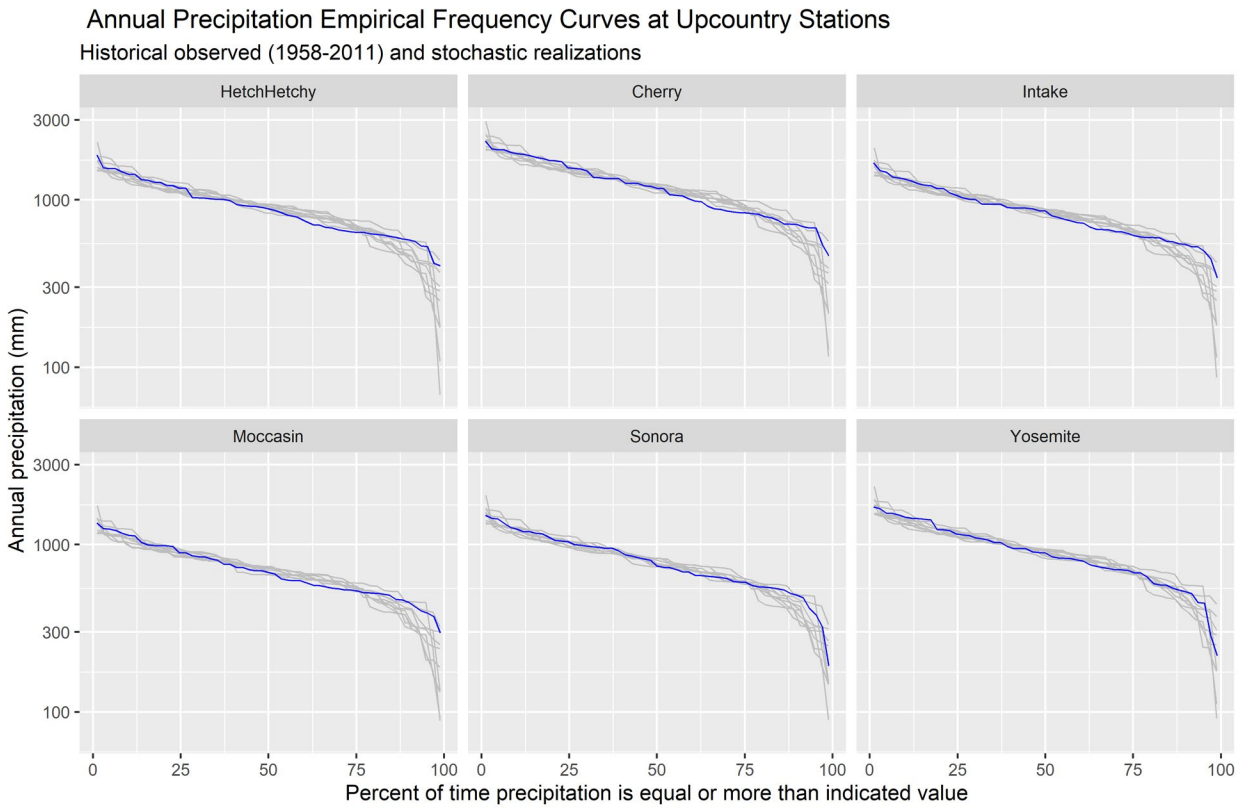


Figure 3-6. Empirical Frequency Curves of Annual Precipitation at Upcountry Stations for Historical Observed (Blue) and the 9 Selected Realizations (Grey).

Figure 3-7 and Figure 3-8 show mean and standard deviation, respectively, for each calendar month and for both historical observed and the 9 selected realizations. As expected by design of the selection of realizations, individual realizations represent a range of mean and standard deviation above and below the historical value. However, the mean and standard deviation of the 9 realizations together is a good match to the historical value (results are only shown for Upcountry stations).

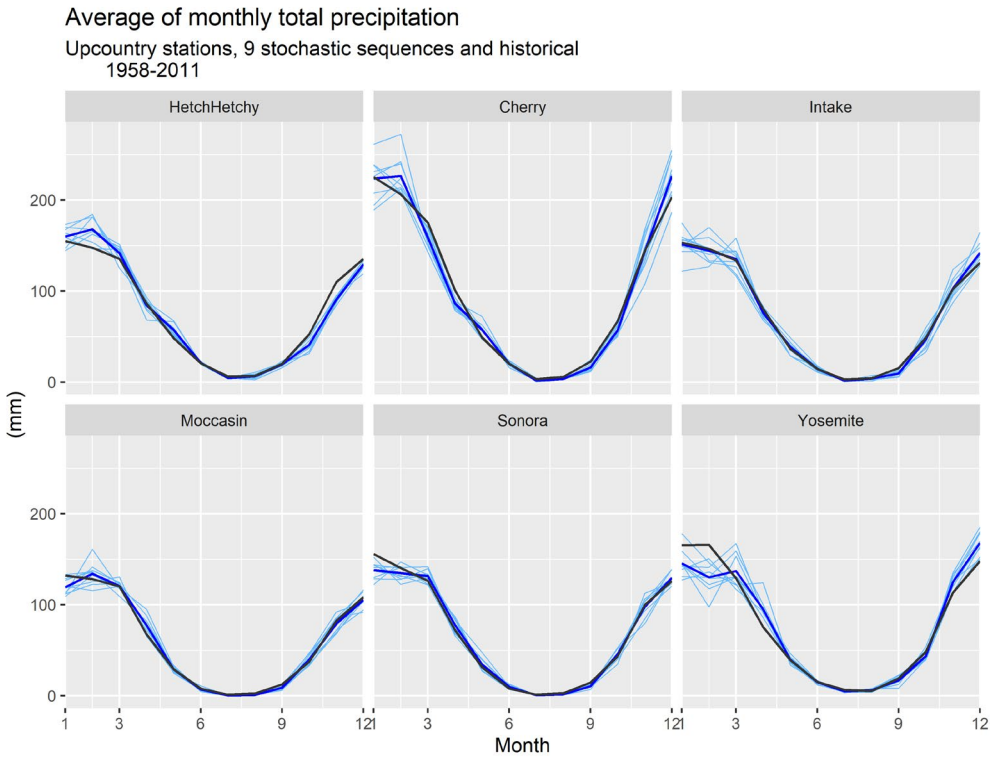


Figure 3-7. Average of Monthly Precipitation Totals at Upcountry Stations for Historical Observed (Black), the 9 Selected Realizations (Light Blue), and Their Average (Blue).

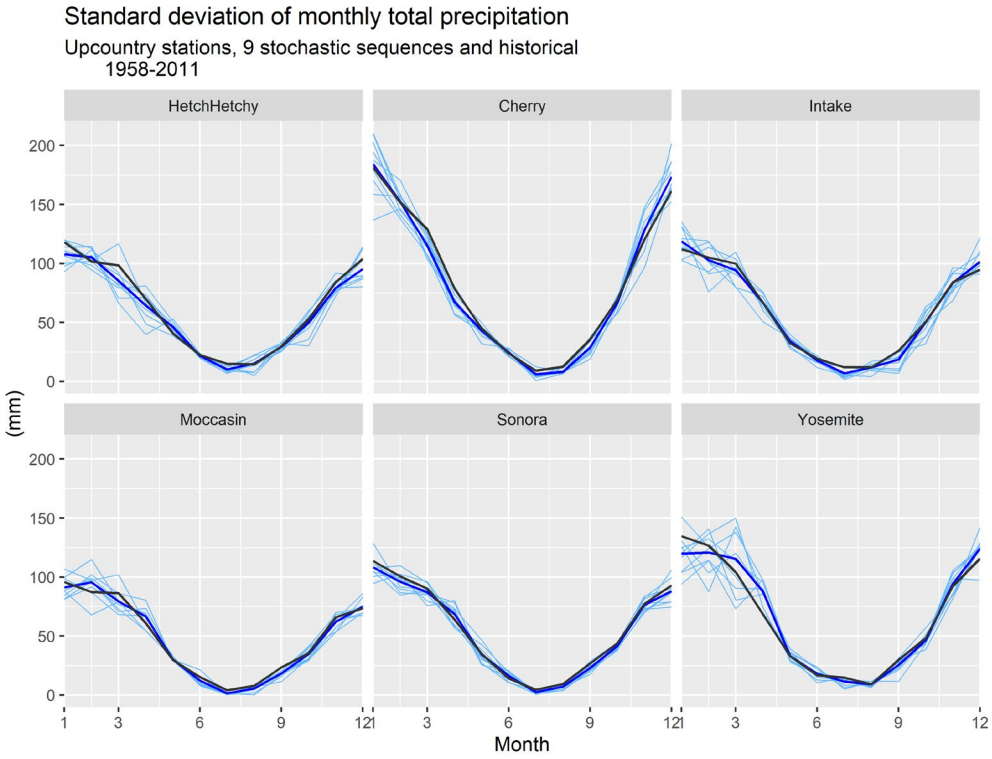


Figure 3-8. Standard Deviation of Monthly Precipitation Totals at Upcountry Stations for Historical Observed (Black), the 9 Selected Realizations (Light Blue), and Their Average (Blue).

3.1.4.2 Historical Climate Realization

In addition to the nine climate realizations described above, a tenth realization deemed 'historical' is considered. The creation of the historical realization used as much as possible historical weather data to force the various models used in the LTVA (e.g., hydrology and demand models). To align with the other 9 realizations used in the LTVA, the length of the historical realization (noted 'Realization 10', or 'R10') is 50 years. The selection of the 50-year long period from the historic was done considering the data availability across the RWS regions and the objective to include the most recent period. Given the data availability, the historical realization spans from 1961-10-01 to 2011-09-30. Most of the available weather data on hand at time of the work was not available after 2011, which does not allow to account for the most recent drought (2012-2016). However, the droughts of 1976-1977 and 1982-1987 are included in the historical realization. Note that to remain consistent with the other nine realizations, the dates assigned to the realization 10 spans from 2020-10-1 to 2070-9-30. As such, the droughts 1976-1977 and 1982-1987 are set as 2035-2036 and 2041-2046.

The creation of the realization 10 thus requires precipitation and temperature data from 1961-10-01 to 2011-09-30. Observed weather data for this period is available for East Bay and Peninsula regions. In Upcountry, precipitation data prior 1969 was not available for four rain gauges that are needed to simulate the streamflow across the region. Precipitation at the gauges were estimated from 1961-10-01 to 1969-09-30 from precipitation data from five other gauges that are also used to simulate the streamflow in Upcountry region. The detail about the interpolation, including estimation of the parameters and validation, are presented in the Technical Report 2: Hydrology Modeling Module (HRG 2021a). To keep consistency with the other nine stochastic realizations built by CliWxGen, the temperature data used to create the historical realization is form the Livneh et al. (2015) dataset.

3.1.5 Validation of the Climate Realization Sample Size

Although 1,000 stochastic realizations were generated by the climate generator, only nine realizations in addition to the historical realization were used for the climate stress test. In addition to the considerations outlined above, one must ensure that the ensemble of realizations provides a large enough sample size from which to assess the effect of internal variability on the RWS. This is especially pertinent in relation to the estimation of performance metrics that require several hundred years of data to converge, for example the water supply reliability metric (Hashimoto et al. 1982; Kjeldsen and Rosbjerg 2004). The reliability metric is used to report the frequency with which system performance is above a chosen threshold (more details about this metric are provided in Section 4.4).

Figure 3-9 illustrates the convergence of water supply reliability across different sample sizes. The y-axis shows the number of years used to calculate the reliability metric. Years are randomly sampled with replacement from the 10 realizations considered for the climate stress test. For each sample size, the boxplots show the distribution of the reliability metrics that were obtained from 100 random. For instance, the results show that when using a sample of only 50 years to calculate the reliability metric, the estimate falls in a wide range of values (between 88 and 100%). As the sample size increases, the variability across the random samples decreases and the median estimate of reliability converges toward a value similar to the ones obtained when using the selected 10 realizations. This justifies the selection of 10 realizations (or 490 years) to produce estimates of reliability and other statistics. Note that calculation of the water delivery reliability requires simulation of the RWS through the SFWSM (Section 3.4). A similar conclusion is assumed to hold for any other performance metrics used in the LTVA.

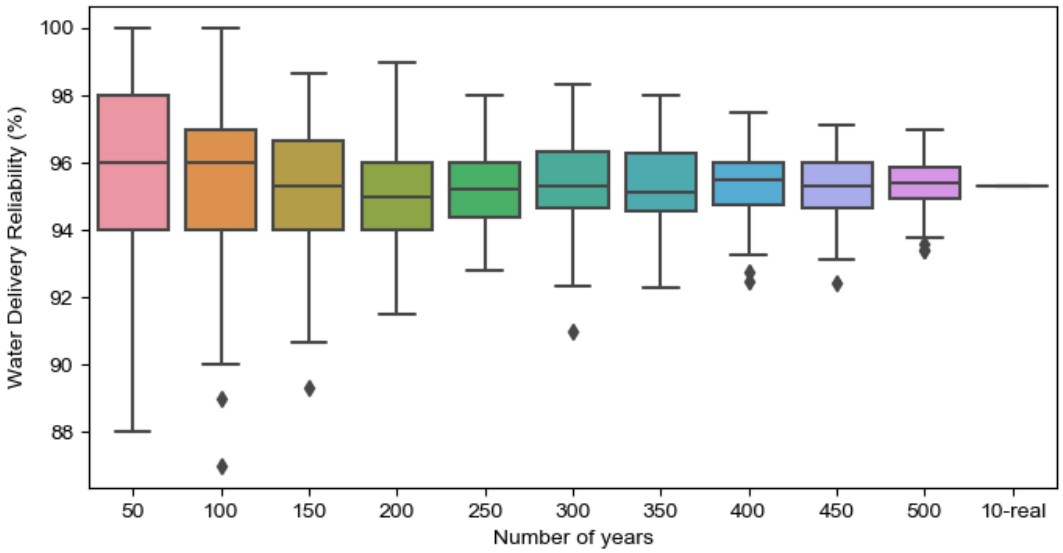


Figure 3-9. Estimated Water Delivery Reliability as a Function of the Number of Years Considered.
The x-axis shows the number of years used for assessing the RWS water delivery reliability. The y-axis shows the water delivery reliability. For each sample size, the boxplots show the distribution of 100 reliability values obtained by randomly sampling with replacement from the 490 years (10 realizations x 49 years) available from the outputs of the San Francisco Water System Model (SFWSM). Note that SFWSM output has a time horizon of 49 years because the first year of the 50-year time horizon is used to initiate the model. The last value indexed by '10-real' shows the water delivery reliability obtained when using all ten realizations (i.e., 490 years).

3.1.6 Model Limitations

Weather generators are paramount tools for planning water infrastructure (Albano et al. 2021). This section summarizes the main limitation of the CliWxGen model used to generate stochastic time series of weather for the LTVA. Overall, CliWxGen performance is very good at annual scale as it shows small if not no bias in annual precipitation and a rather well reproduction of the inter-annual variability. However, the seasonal variability is underestimated.

3.2 Hydrology Models

3.2.1 Modeling Goals

Water availability for SFPUC in the Bay Area and Sierra Nevada regions varies significantly within the year, from year to year, and there is large uncertainty on its change over the long term with climate changes. Understanding and representing ranges of potential future hydrologic conditions is a core component to the LTVA. Representations of hydrologic processes is needed specifically to develop time series of data for a range of system modeling purposes, including reservoir inflows, actual water entitlement, reservoir evaporation, and other hydrologic parameters. The hydrologic models are driven by the weather generator outputs to simulate streamflow and evaporation. Time series of streamflow and potential evapotranspiration are inputs to SFWSM (Section 3.4) which allows us to look at the effects of climate change and are also used for the drought frequency analysis (Section 3.2.4).

3.2.2 Modeling Methods

Hydrologic models were developed for each of the three RWS regions (Upcountry, East Bay, and Peninsula; Figure 3-10). The RWS watersheds in the Upcountry region were modeled using the Precipitation Runoff Modeling System (PRMS) (Markstrom et al. 2015), a physically-based distributed-parameter hydrologic modeling system developed by the USGS. Within the Upcountry region, hydrologic models were developed for the watershed above O'Shaughnessy Dam, the combined watersheds above

Cherry Valley Dam and Lake Eleanor Dam, and the watershed between New Don Pedro Dam and Cherry, Eleanor and O'Shaughnessy Dams (referred to as Don Pedro accretion or labeled as Don Pedro in figures).

The East Bay and Peninsula watersheds were modeled using the Sacramento Soil Moisture Accounting Model (SAC-SMA; Burnash et al. 1973), a conceptual and lumped hydrologic model developed by NOAA. A distributed version of SAC-SMA (noted SAC-SMA-DS) was used to model the watersheds in East Bay regions to better account for the variability of the precipitation across the region. Modeled East Bay watersheds include the Arroyo Hondo watershed above Calaveras Reservoir, the San Antonio Reservoir watershed, and the watershed upstream of the Alameda Creek Diversion Dam tunnel. Modeled Peninsula watersheds include the Crystal Springs Reservoir watershed, Pilarcitos Creek above Pilarcitos Dam, San Andreas Reservoir watershed, Stone Dam and Mud Dam watersheds below Pilarcitos Dam and San Andreas Reservoir, respectively.

PRMS and SAC-SMA are well-established models in the hydrology community. As such, a comprehensive description is not given here; rather, calibration results are shown and discussed below for each region. More details about PRMS and SAC-SMA modeling tools are available in the Technical Report 2 (HRG 2021a) and references within.

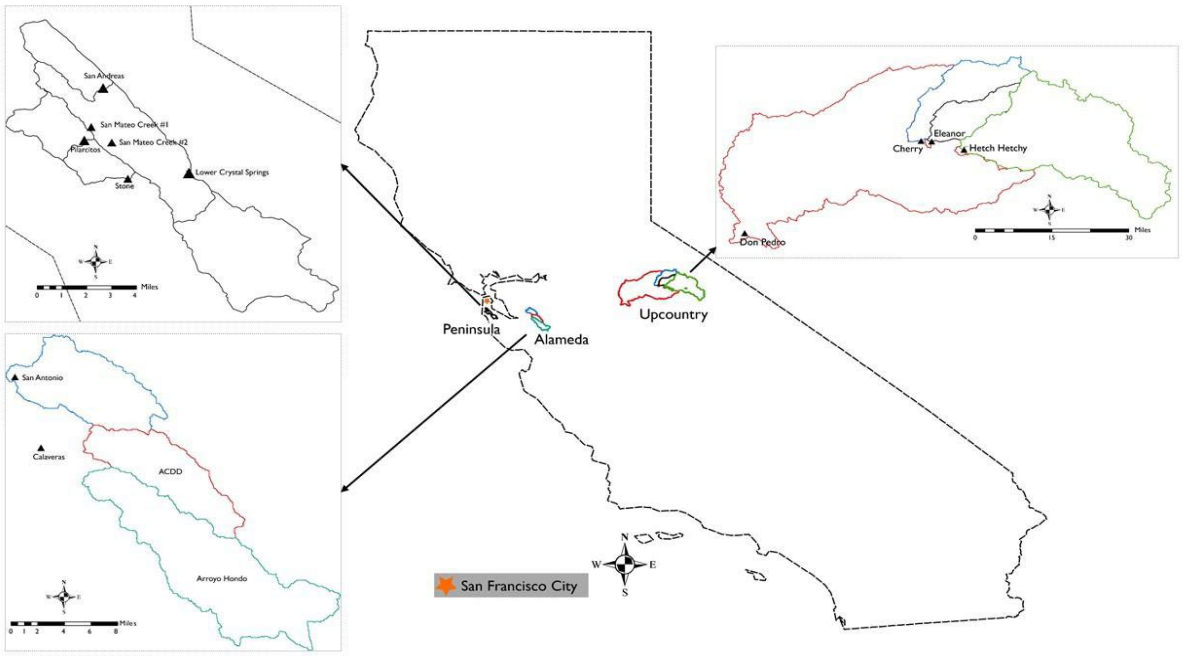


Figure 3-10. Map Showing Watersheds That Were Included in the Hydrologic Models Developed for the LTVA.

3.2.3 Calibration Results across the Upcountry, East Bay, and Peninsula Regions

Calibration results for Upcountry (PRMS), East Bay and Peninsula (SAC-SMA) regions are presented and discussed in this section. All models were calibrated using the Kling-Gupta efficiency (KGE; Gupta et al. 2009) as the objective function for the optimization algorithm. The Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe 1970) is also reported below for the various watersheds together with the percent bias. Kling-Gupta and Nash Sutcliffe efficiencies are positively oriented with an upper bound of 1 (i.e., a larger value means a better model). A negative bias indicates an underestimation of the observed streamflow, while a positive values indicates an overestimation.

While the calibration and validation periods vary for each basin (Table 3-3), the performance of the models are here discussed for the whole available period (i.e., calibration + validation periods). More details regarding the input data, models, and calibration and validation results are included in the Technical Report 2 (HRG 2021a).

Table 3-3. Calibration and Validation Period for the Watersheds in Each Basin.
Each region had different time-periods for calibration and validation based on data availability.

Watershed	Calibration Period	Validation Period
Upcountry region		
Hetch Hetchy Reservoir inflow	1970-10-01 → 2015-09-30	NA
Cherry Reservoir/Lake Eleanor inflow	1970-10-01 → 2015-09-30	NA
Don Pedro accretion flow[1]	1970-10-01 → 2015-09-30	NA
East Bay region		
San Antonio Reservoir inflow	1996-10-01 → 2006-09-30	2006-10-01 → 2011-09-30
Alameda Creek Diversion Dam inflow	1996-10-01 → 2006-09-30	2006-10-01 → 2011-09-30
Arroyo Hondo above Calaveras Reservoir	1969-10-01 → 1982-09-30	1995-10-01 → 2011-09-30
Peninsula region		
Pilarcitos Reservoir Inflow	1999-10-01 → 2010-09-30	2010-10-01 → 2011-09-30
Stone Dam Inflow[2]	1999-10-01 → 2010-09-30	2010-10-01 → 2011-09-30
San Andreas Reservoir Inflow	1999-10-01 → 2010-09-30	2010-10-01 → 2011-09-30
Crystal Springs Reservoir Inflow	2007-10-01 → 2010-09-30	2010-10-01 → 2011-09-30
Mud Dam Inflow (San Mateo creek)[2]	1999-10-01 → 2010-09-30	2010-10-01 → 2011-09-30

[1] Calibration performed across the full time-series.
[2] Stone Dam and Mud Dam inflows calculated for longer periods using basin-area adjusted flows from Pilarcitos Reservoir inflow values.

3.2.3.1 Upcountry Region

In the Upcountry region, PRMS models for Hetch Hetchy, Cherry Reservoir/Lake Eleanor watersheds were calibrated by SFPUC staff through previous work and were given to HRG. Note that due to Eleanor-Cherry Diversion Tunnel, which connects Cherry Reservoir and Lake Eleanor, the PRMS model representing the upstream watersheds was built as a single combined watershed. The PRMS model for the Don Pedro accretion watershed was calibrated by HRG. For doing so, the time series of Don Pedro accretion flows, used as calibration/validation target, was back-calculated from the mass balance equation defined in Equation 3-1:

$$Q_{LG} = Q_{HH} + Q_{CHEL} + Q_{DP,} \quad (3\text{-}1)$$

where the natural flow at La Grange(Q_{LG}) is defined as the sum of the flows upstream of Don Pedro Reservoir, where Q_{HH}, Q_{CHEL}, and Q_{DP} are Hetch Hetchy reservoir inflow, Cherry Reservoir/Lake Eleanor inflow and Don Pedro accretion flows, respectively. The time series of natural flow at La Grange (Q_{LG}) was provided by SFPUC at daily time step from 1969 to 2015. On average, Hetch Hetchy Reservoir watershed contributes 45% of Tuolumne flow at La Grange while Cherry/Eleanor contributes 25%. The accretion flow at Don Pedro contributes 30% of the total flow.

PRMS performance at these locations is shown in Table 3-4. Simulated and observed streamflow time series are illustrated at Hetch Hetchy Reservoir (Figure 3-11) and at La Grange (Figure 3-12) at annual, monthly, and daily temporal scales. In addition, Figure 3-13 illustrates the model performance at all locations using scatterplots for annual and monthly temporal scales.

Table 3-4. PRMS Streamflow Calibration Results for the Upcountry Region.
Performance criteria are estimated over the period 1970-2015. Performance criteria are given for daily, monthly, and annual water year temporal scales to highlight model performance at capturing the different processes involved at these scales (e.g., seasonal variability vs. inter-annual variability).

	Kling-Gupta Efficiency	Nash Sutcliffe	Percent Bias (%)
Hetch Hetchy Reservoir inflow			
Daily	0.85	0.70	3.9
Monthly	0.92	0.86	
Water Year	0.76	0.91	
Cherry Reservoir/Lake Eleanor inflow			
Daily	0.78	0.56	3.0
Monthly	0.91	0.83	
Water Year	0.67	0.86	
Don Pedro accretions			
Daily	0.65	0.56	2.2
Monthly	0.70	0.70	
Water Year	0.56	0.78	
Naturalized Tuolumne flow at La Grange			
Daily	0.79	0.68	3.2
Monthly	0.84	0.83	
Water Year	0.66	0.86	

PRMS model at Hetch Hetchy Reservoir performs well at daily and monthly scales; as demonstrated by the good KGE and NSE efficiency values and low bias (Table 3-4). The performance of the PRMS at Cherry Reservoir/Lake Eleanor at daily and monthly time scale is average, although the bias is low. The performance of the PRMS models at these gauges at annual is less straightforward, especially for Cherry Reservoir/Lake Eleanor model for which the KGE decreases significantly at this time scale. The time series of annual inflow to Hetch Hetchy Reservoir (Figure 3-11) and the scatterplots of annual flow at Hetch Hetchy Reservoir and Cherry Reservoir/Lake Eleanor (Figure 3-13) demonstrate a bias in variance at this temporal scale. This bias leads to an underestimation of the annual volume during the wet years and an overestimation of the annual volume during the dry years.

The PRMS model for the accretion flow into Don Pedro reservoir is reported as 'Don Pedro accretion' in Table 3-4. Despite the uncertainty stemming from the reconstruction of the Don Pedro accretion flows from the mass balance (Equation 3-1), PRMS model shows average performance at daily and monthly time step, with KGE values equal to 0.65 and 0.70, respectively. The performance of the Don Pedro accretion model at annual is more questionable. With a low bias (i.e., 2.2%) and rather good NSE value (i.e., 0.78), the low KGE (i.e., 0.56) might indicate a bias in variance. The scatterplots in Figure 3-13 confirms a similar bias in variance to the ones observed for Hetch Hetchy and Cherry/Eleanor models. The bias in variance is larger for the Don Pedro accretion model, leading to larger overestimation of low flow years and a larger underestimation of high flow years.

The whole watershed, which encompasses Don Pedro accretion, Hetch Hetchy, Cherry, and Eleanor watersheds, is reported as 'La Grange' in Table 3-4, in Figure 3-12 and Figure 3-13. A good reproduction of the unimpaired Tuolumne flow at La Grange is key for the RWS because it is the location where the Modesto and Turlock Irrigation Districts (MID and TID) entitlements and the Water Available to the City (WAC) are defined (see next paragraph for more detail about WAC). The performance of the combined PRMS model simulations through the mass balance equation somewhat averages out the performance obtained for the three calibrated PRMS models. Similar to its upstream components, the simulated flow at La Grange presents a bias in variance at annual time scale (Figure 3-12 and Figure 3-13), leading to overestimation of dry years and underestimation of wet years.

TID and MID have senior water rights on the water of the Tuolumne River. WAC represents the water from the unimpaired flow in the Tuolumne River at La Grange that belongs to SFPUC. The Raker Act specifies how much and when Tuolumne river flows must be available to MID and TID (district entitlements). On any given day, the actual district entitlements consist of the lesser of i) 4,792 AF/day (2,416 cfs) from June 14th through April 14th or 8,064 AF/day (4,066 cfs) from April 15th through June 13th (aka: maximum district entitlement), and ii) unimpaired flow in the Tuolumne River at La Grange. The implication is that water available to SFPUC from the Tuolumne River (i.e., WAC) is any unimpaired flow at La Grange (below Don Pedro) that is greater than actual district entitlements. Figure 3-14 illustrates the WAC for an extremely wet water year (1983) and an extremely dry water year (1987), which highlights the large variability of both the unimpaired flow of the Tuolumne River at La Grange and the WAC. Regarding the calculated WAC from the PRMS simulations, it is noted that the underestimation of the inter-annual variability of the unimpaired flow at La Grange leads to similar bias for WAC (Figure 3-13). The overestimation of the WAC during dry years is a concern because it might prevent a good representation of the droughts.

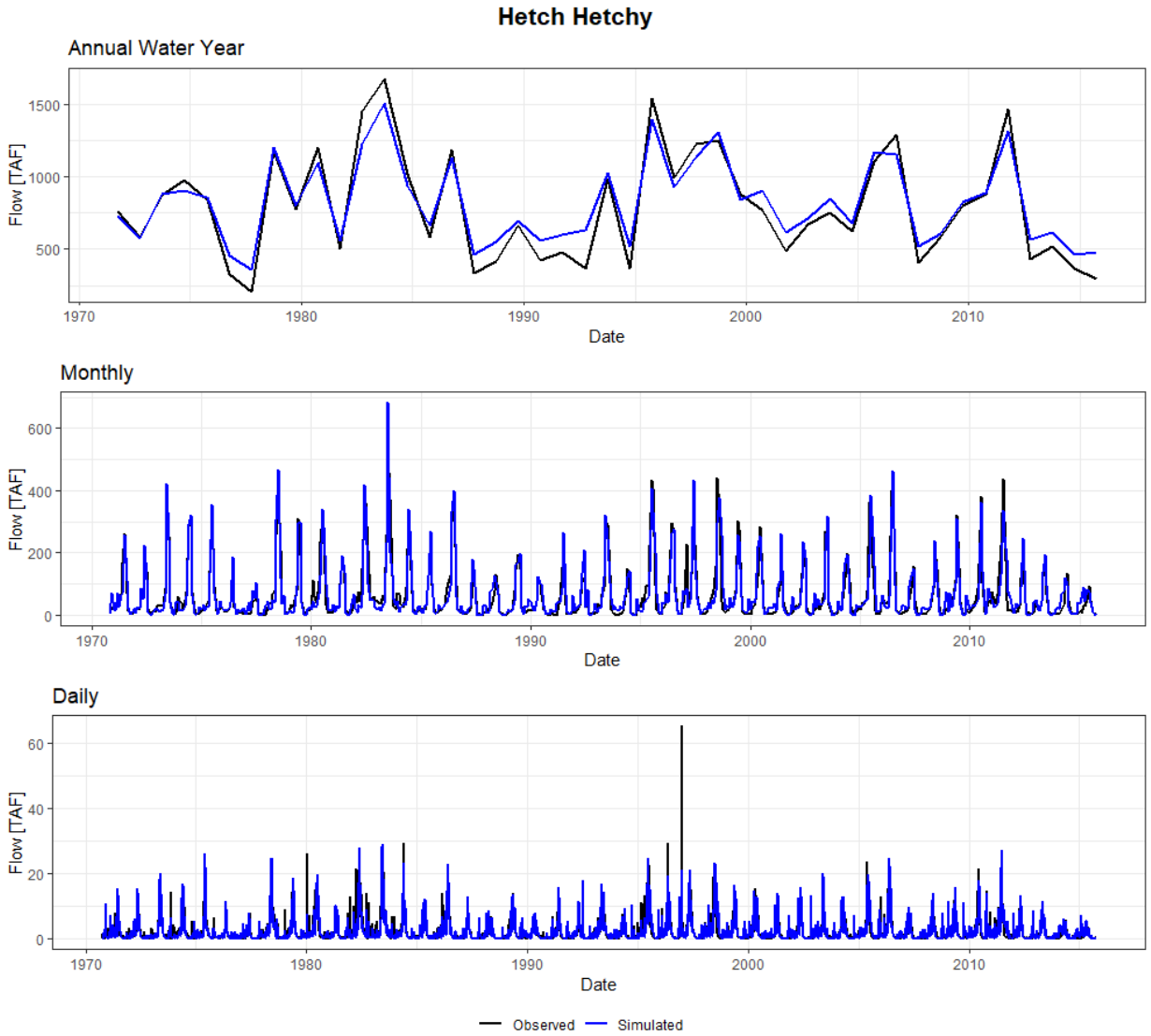


Figure 3-11. Comparison of the PRMS Simulated and Observed Streamflow at Hetch Hetchy Reservoir. Simulated (blue) and observed (black) inflows to Hetch Hetchy Reservoir at annual (top), monthly (middle) and daily (bottom) time step. Annual values are given for Water Year (October through September)

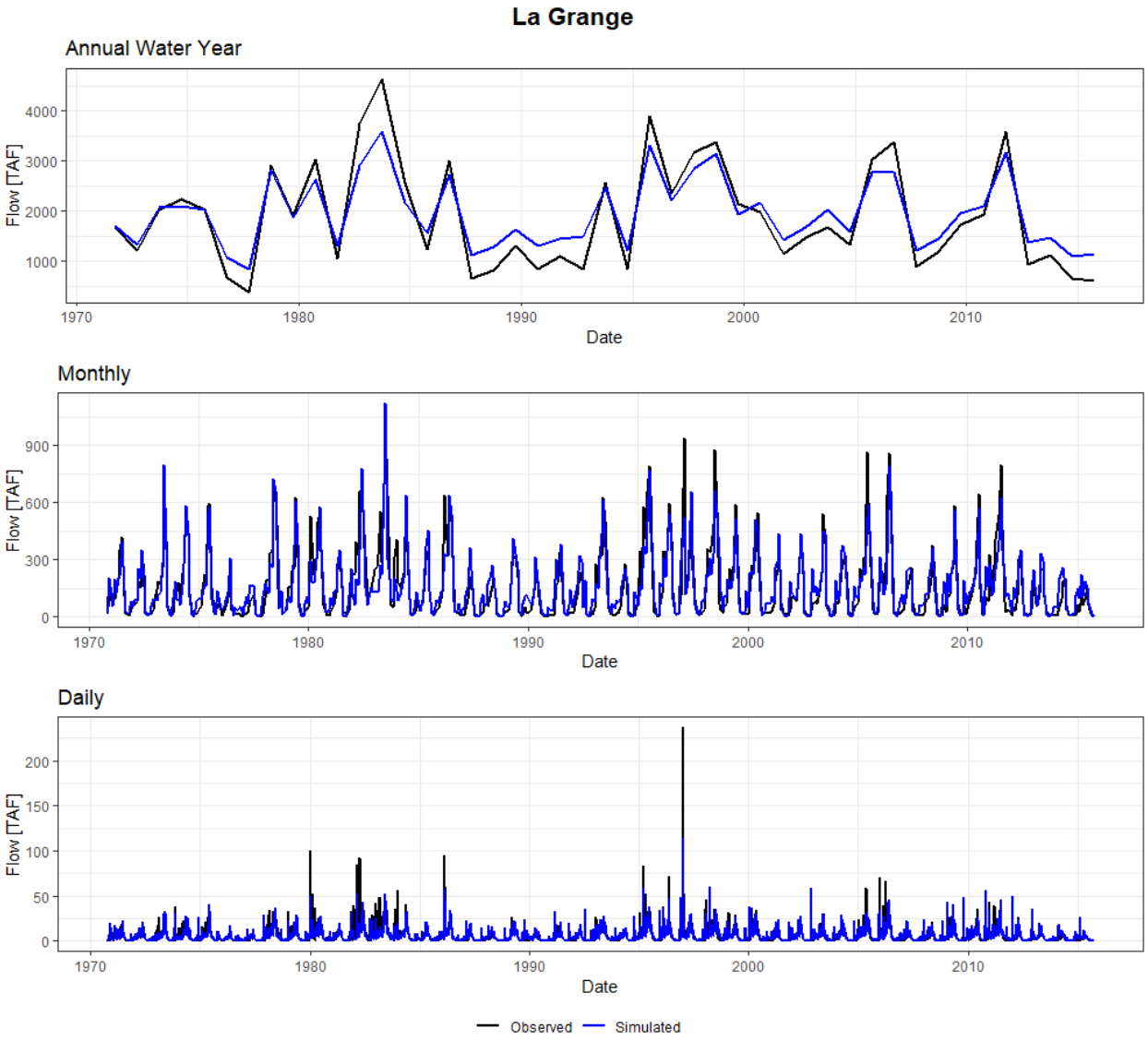


Figure 3-12. Comparison of the PRMS Simulated and Observed Streamflow at La Grange.
Simulated (blue) and observed (black) at annual (top), monthly (middle) and daily (bottom) time step. Annual values are given for Water Year (October through September)

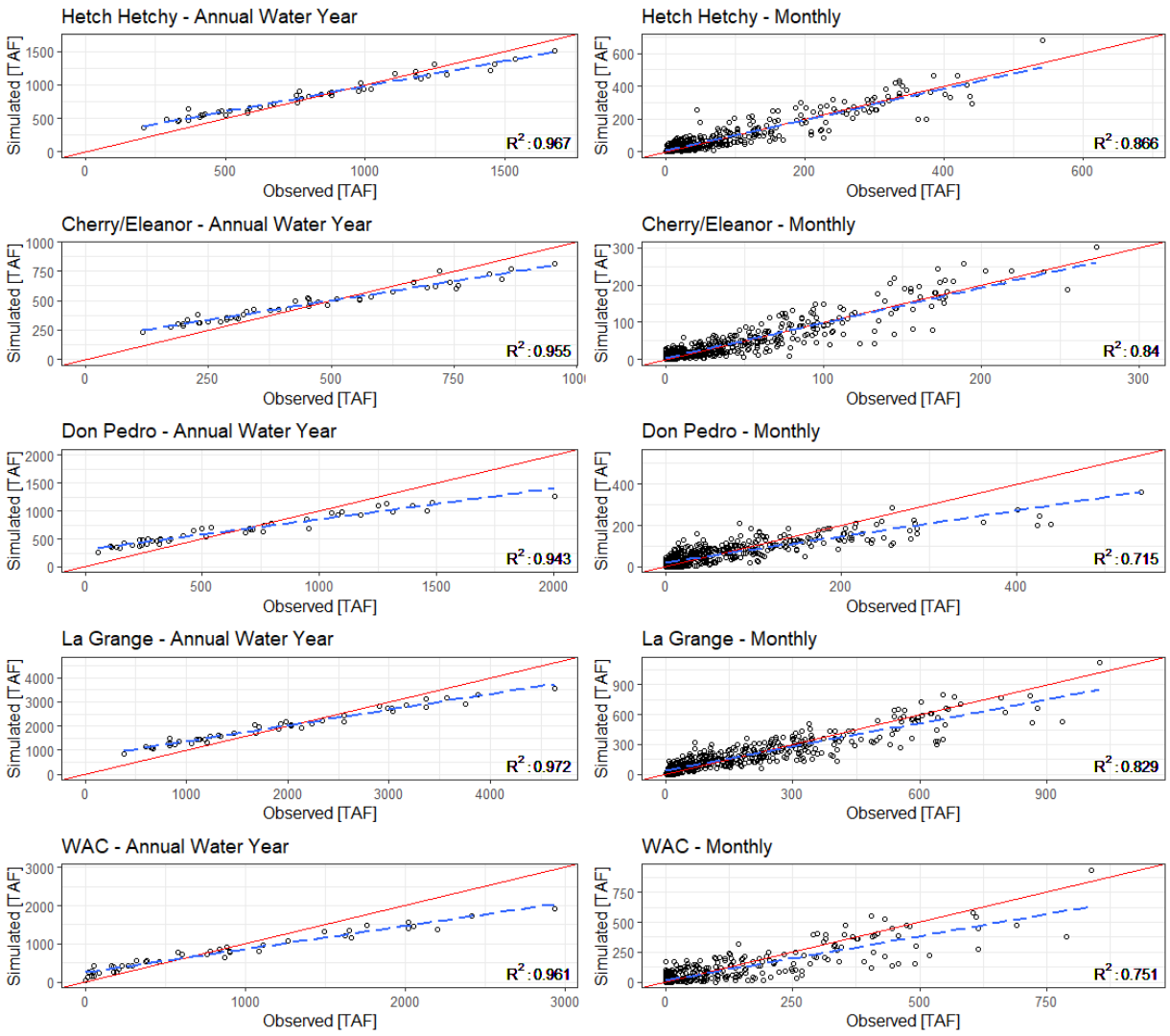


Figure 3-13. Scatterplots between PRMS Simulated and Observed Streamflow at La Grange (Top), Hetch Hetchy Reservoir Inflow (Second from the Top), Cherry Reservoir/Lake Eleanor Inflow (Third from the Top), and the Don Pedro Accretion Flows (Bottom) at Annual (Left) and Monthly (Right) Temporal Scales.

The red line is the 1-to-1 line with the dashed blue line being a fitted line. The R-squared value is calculated for the fitted blue line.

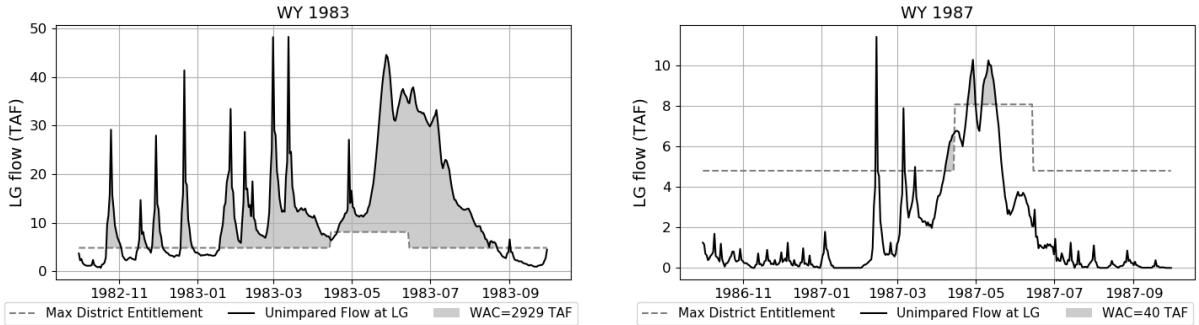


Figure 3-14. Water Available to the City (WAC) for an Extremely Wet Year (Left, 1983) and a Dry Year (Right, 1987).

The year 1987 is the first year of the historical drought 1987-1992. The shaded area shows the WAC, and the dash-grey line shows the maximum district entitlement. Note that the y-axis differs for the two subplots.

Given the importance of a good reproduction of the WAC for the RWS, UMASS HRG collaborated with SFPUC to correct the PRMS simulations across the Upcountry region. A post-processing model was developed by SFPUC to correct the PRMS daily streamflow. The considered method attends to correct the residual model errors using meteorological indices. For precipitation variable, an index called "Current Precipitation Index" is used to account for the basin wetness. For temperature variable, a heat index similar to a degree-day is used to account from the basin warmness. Note that the considered post-processed correction is a parametric method (i.e., it requires some parameters to be calibrated). The calibration of these parameters was done with the main objective to better reproduce the dry years. More details about the post-processing model, including calibration and validation, are provided in the Technical Report 2 (HRG 2021a). The comparison of the raw PRMS simulations (i.e., prior correction) with the post-processed simulations (i.e., after correction) is shown in Table 3-5 and in Figure 3-15 and Figure 3-16. The results clearly indicate a significant improvement of the inter-annual variability of the simulated streamflow, with significant improvement of the KGE and NSE scores at annual time scale, especially for Cherry Reservoir/Lake Eleanor watershed and Don Pedro accretion watershed. It also results in a much better reproduction of the inter-annual variability of the annual Tuolumne flow at La Grange and of the WAC. The flow duration curve (Figure 3-15) and the scatterplots (Figure 3-16) figures illustrate the significant improvement of the simulations during the low flow years/months. The downside of the post-processed model is that it worsens the bias across the region. At La Grange, the overall bias shifted from +3.2% prior correction to -6.8% after correction. Regarding WAC, the reduction of the bias in inter-annual variability (and a significant improvement of the low WAC years) comes with a significant improvement of the high WAC years (Figure 3-15 and Figure 3-16). Table 3-5 also shows the improvement of the KGE and NSE scores at all temporal scales, with a major improvement at annual scale, which, however, leads to a slight increase in the negative bias (from -8.2% to -8.8%).

Table 3-5. PRMS Streamflow Calibration Metrics for the Upcountry Region before and after Correction.
Performance criteria are estimated over the period 1972-2015.

	Raw PRMS simulations (prior correction)			**Post-processed PRMS simulations (after correction)**		
	Kling-Gupta Efficiency	Nash Sutcliffe	Percent Bias (%)	Kling-Gupta Efficiency	Nash Sutcliffe	Percent Bias (%)
Hetch Hetchy Reservoir inflow						
Daily	0.85	0.70	3.9	0.87	0.74	-3.8
Monthly	0.92	0.86		0.92	0.89	
Water Year	0.76	0.91		0.83	0.91	
Cherry Reservoir/Lake Eleanor inflow						
Daily	0.78	0.56	3.0	0.85	0.73	7.3
Monthly	0.91	0.83		0.90	0.88	
Water Year	0.67	0.86		0.91	0.92	
Accretion flow at Don Pedro Reservoir						
Daily	0.65	0.56	2.2	0.67	0.63	-20.8
Monthly	0.70	0.70		0.76	0.77	
Water Year	0.56	0.78		0.77	0.86	
Naturalized Tuolumne flow at La Grange						
Daily	0.79	0.68	3.2	0.87	0.79	-6.8
Monthly	0.84	0.83		0.91	0.90	
Water Year	0.66	0.86		0.87	0.94	
Water Available to the City (WAC)						
Daily	0.71	0.56	-8.2	0.81	0.70	-8.8
Monthly	0.78	0.75		0.88	0.85	
Water Year	0.61	0.82		0.84	0.93	

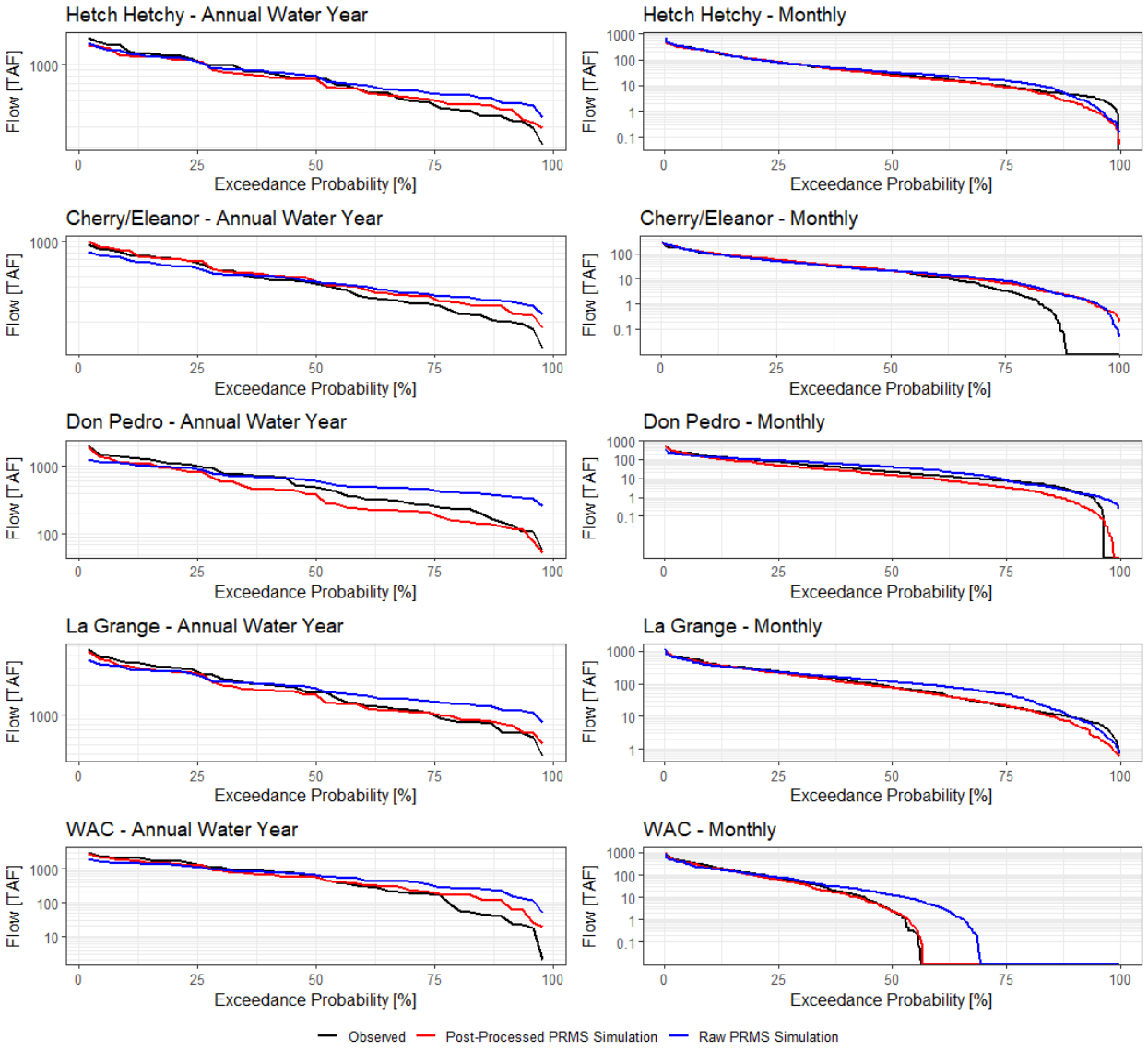


Figure 3-15. Flow Duration Curve for Tuolumne Flow at La Grange, Hetch Hetchy Reservoir Inflow, Cherry Reservoir and Lake Eleanor Inflows, Accretion Flow to Don Pedro Reservoir, and Water Available to the City (WAC) for Annual Water Year and Monthly Temporal Scales.

The black, blue, and red lines show the flow duration curve for the historic observed flow, the raw PRMS simulation (i.e., prior correction), and the post-processed PRMS simulation (i.e., after correction). A logarithm scale is used for the streamflow to ease visualization of the low flow years/months.

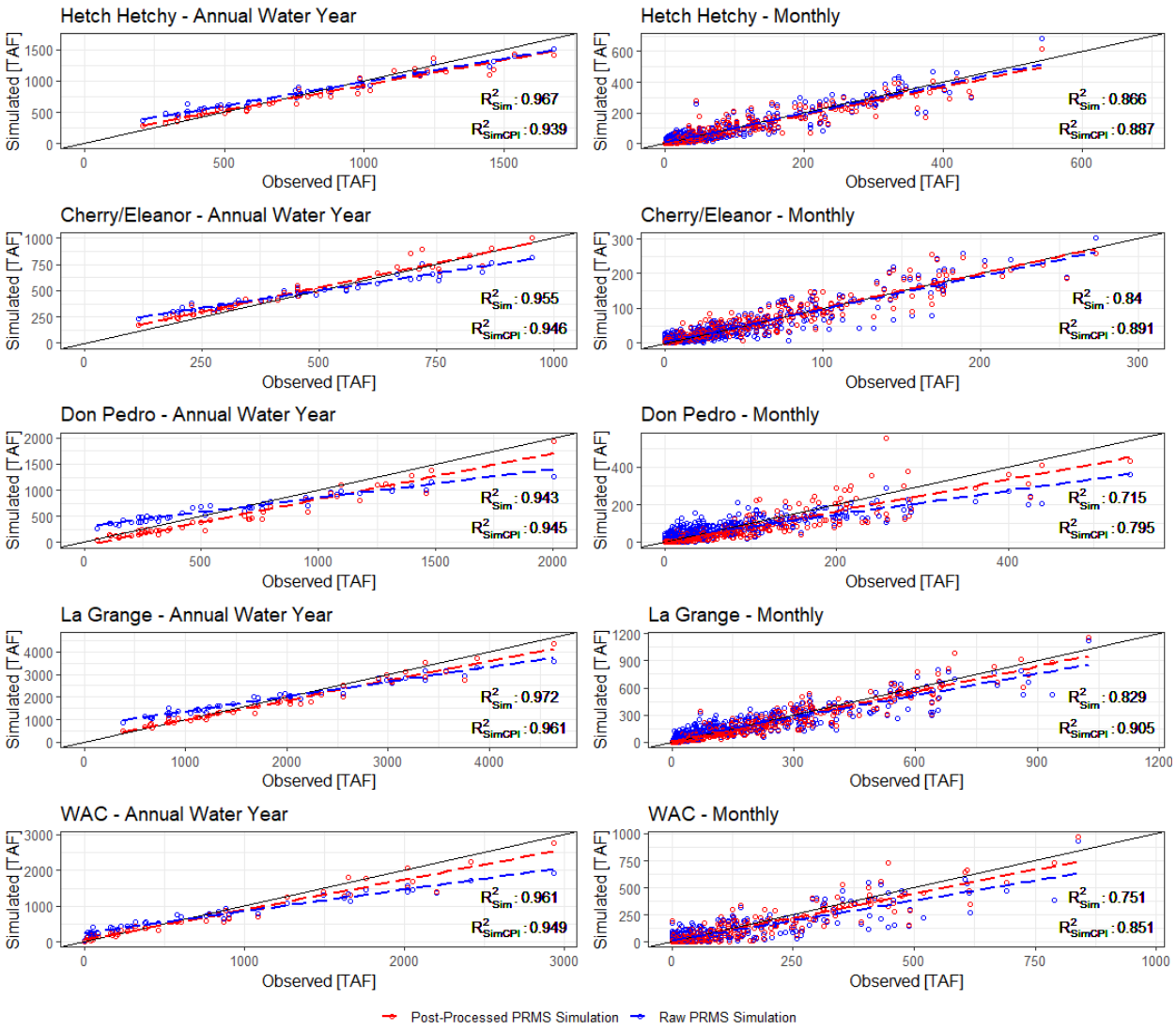


Figure 3-16. Scatterplot for Tuolumne Flow at La Grange, Hetch Hetchy Reservoir Inflow, Cherry Reservoir and Lake Eleanor Inflows, and Accretion Flow to Don Pedro Reservoir for Annual Water Year and Monthly Temporal Scales.

The black line represents the identity line whereas blue and red lines represent the linear regression between the observed flow and the raw and post-processed PRMS simulations, respectively.

Notably, while the post-processing model improves the results additional work needs to be performed to reduce the bias. The hydrologic model used to simulate the streamflow on the Tuolumne watershed in response to precipitation and temperature overestimates streamflow during dry years. For example, the flow computed at the Tuolumne River at La Grange is overestimated, and therefore WAC is also overestimated by about 482,000 acre-feet during the drought sequence 1987-1992 (observed is 813,000 acre-feet versus simulated is 1,295,000 acre-feet) (Table 3-6).

Table 3-6. Comparison of Observed and Simulated Water Available to the City over the 1987-1992 Drought Using Results from PRMS Hydrologic Model.

	Water available to the City (acre-feet)		
Water Year	Observed	Simulated	Error
1987	40,439	174,868	134,428
1988	23,521	112,425	88,905
1989	379,305	383,159	3,854
1990	46,632	114,957	68,325
1991	269,733	302,719	32,986
1992	53,525	207,221	153,697
1987-92 totals	813,155	1,295,349	482,194

3.2.3.2 East Bay Region

In the East Bay region, HRG has calibrated a distributed version of the SACramento-Soil Moisture Accounting model (noted SAC-SMA-DS) for Arroyo Hondo watershed above Calaveras Reservoir, Alameda Creek Diversion Dam (ACDD) watershed and San Antonio Reservoir watershed (respectively green, red, and blue color in Figure 3-10 bottom-left). These simulated streamflow time-series are three out of four (five if SFWSM is configured for Alameda Creek Recapture Project) time series that are used as inputs to the SFWSM. The two watersheds are the intermediate watershed between Arroyo Hondo USGS gauge (id: 11173200) and Calaveras Reservoir in addition to the watershed in-between Calaveras reservoir and the quarry pit F2 along Alameda Creek. After discussions with SFPUC personnel, streamflow time-series are obtained by scaling the simulated streamflow of Arroyo Hondo and San Antonio reservoir inflows, respectively. More details about the input data and model calibration and validation are provided in the Technical Report 2 (HRG 2021a).

Performance criteria obtained for the period 1996-2011 (for ACDD and San Antonio) and 1969-2011 (for Arroyo Hondo) are reported for the three basins in Table 3-7. Figure 3-17 through Figure 3-19 illustrate the calibration and validation results across the three modeled watersheds. Model performance is high in East Bay region. KGE and NSE scores for each temporal scale are above 0.8, with the exception of Arroyo Hondo NSE that equals 0.75 at daily time scale, which is already a fairly good score, especially for watersheds that have flashy response to precipitation events, as it is the case across East Bay region. Bias is low across all watersheds, with the largest bias being obtained for the San Antonio Reservoir inflow (i.e., 4.55%). An important results that is worth mentioning is the good performance of the SAC-SMA-DS model to model the range of water years as there is no systematic error for either dry or wet years (Figure 3-17b through Figure 3-19b). The streamflow seasonality is fairly well reproduced (Figure 3-17c through Figure 3-19c), while the distribution of daily streamflow is also fairly well simulated, although peak flow values tends to be underestimated, which is a well-known issue for any hydrological models (Figure 3-17d through Figure 3-19d).

Table 3-7. Same as Table 3-4 for East Bay Watersheds.
Performance criteria are given for the period 1996-2011 (for ACDD and San Antonio) and 1969-2011 (for Arroyo Hondo).

	Kling-Gupta Efficiency	Nash Sutcliffe	Percent Bias (%)
San Antonio Reservoir inflow, 1996-2011			
Daily	0.90	0.82	4.55
Monthly	0.92	0.96	
Water Year	0.90	0.93	
Alameda Creek Diversion Dam inflow, 1996-2011			
Daily	0.89	0.80	3.85
Monthly	0.90	0.95	
Water Year	0.89	0.93	
Arroyo Hondo above Calaveras Reservoir, 1969-2011			
Daily	0.87	0.75	1.70
Monthly	0.83	0.87	
Water Year	0.93	0.95	

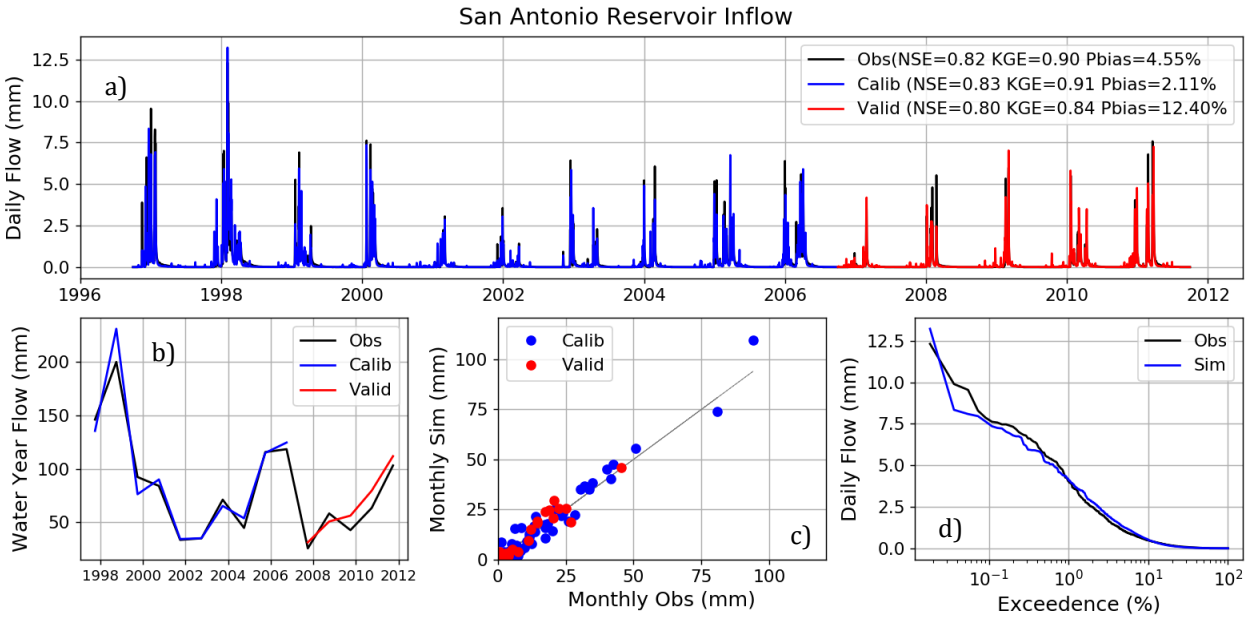


Figure 3-17. Calibration Results for San Antonio Sub-watershed.
a) Comparison of the daily observed (black) and simulated time series during calibration (blue) and validation (red) period. Nash-Sutcliffe (NSE) and Kling-Gupta (KGE) efficiency and the percent bias (Pbias, %) are given for the calibration, validation, and the entire periods; b) Same as a) but for the annual scale (i.e., water year from 10/01 to 09-30). The data point for the WY XXXX is located at the date 09/30/XXXX; c) Scatter plot showing the monthly simulated vs. observed discharges. Calibration and validation periods are illustrated in blue and red color respectively. d) Simulated (blue) and observed (black) flow duration curve. Here, the simulated flow duration curve include both calibration and validation periods.

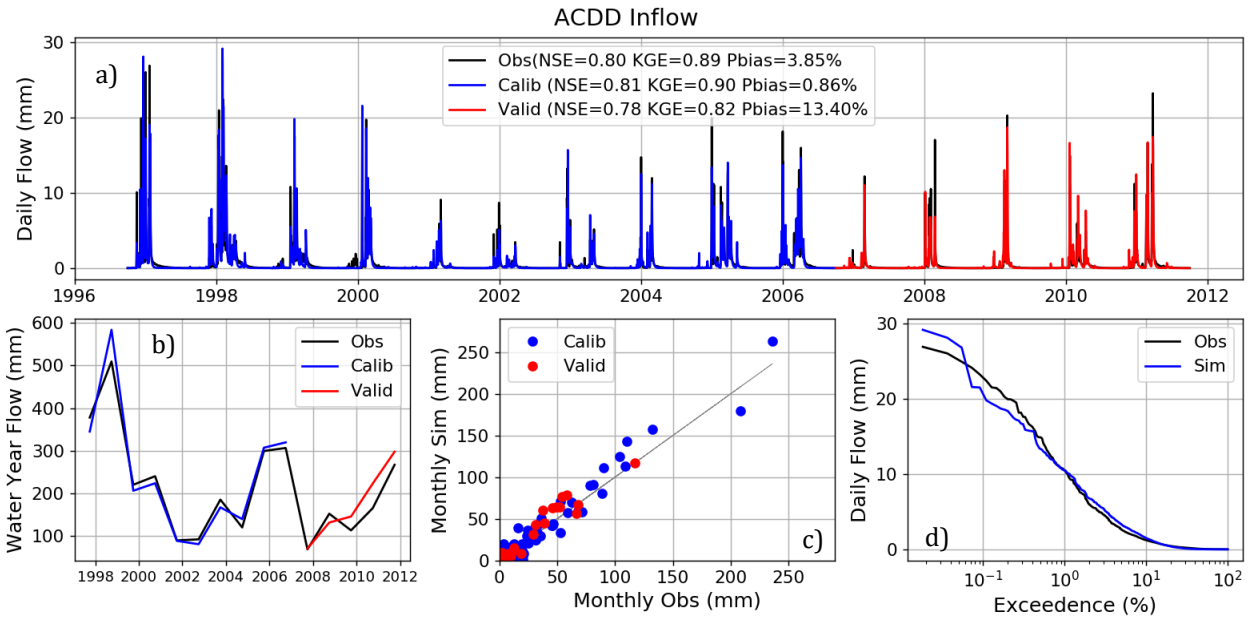


Figure 3-18. Same as Figure 3-17 but for ACDD Sub-watershed.

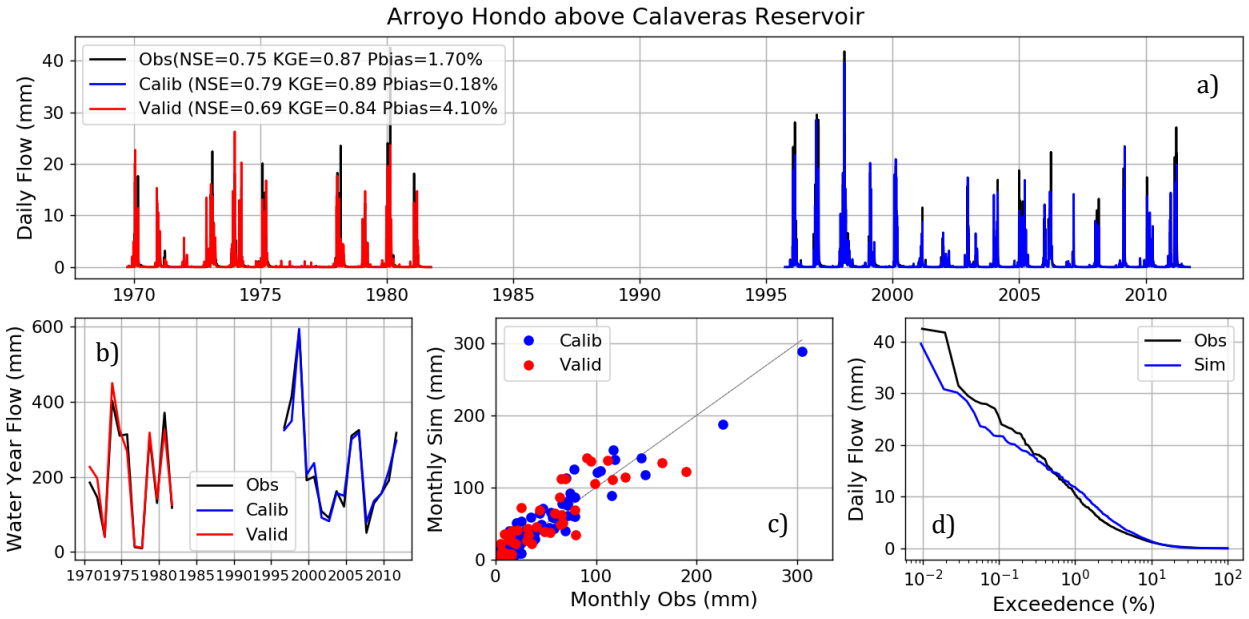


Figure 3-19. Same as Figure 3-17 but for Arroyo Hondo Watershed above Calaveras Reservoir.

3.2.3.3 Peninsula Region

In the Peninsula region, the HRG has calibrated the SACramento-Soil Moisture Accounting model (noted SAC-SMA) for five watersheds; the watershed above Pilarcitos Dam, the watershed in-between Pilarcitos Dam and Stone Dam accretion, the watersheds upstream of San Andreas Reservoir and upstream of Crystal Springs Reservoirs, and the watershed upstream Mud Dam 1 (Figure 3-10 top-right). Streamflow time series at San Mateo Creek upstream Mud Dam 2, an input to SFWSM, is obtained by scaling the time series obtained for San Mateo Creek upstream Mud Dam 1; the scaling factor is the ratio of watershed areas.

Time series of observed streamflow are scarce across the Peninsula region. Stone Dam inflow and Mud Dam inflow watersheds has observed streamflow data starting in 2011, which is outside of the temperature data ranges. Therefore, a correlation was calculated between the available streamflow time-series across the Peninsula region, and the time series of inflow to Pilarcitos Reservoir was found to have the highest correlation to the Stone Dam and Mud Dam observed inflows. The calibration target for these two watershed models were calculated on a basin-area ratio adjustment to the Pilarcitos Reservoir inflow.

Performance criteria for each watershed are reported in Table 3-8. Figure 3-20 through Figure 3-24 illustrate the calibration and validation results across the five modeled watersheds. KGE scores for each temporal scale are all good (e.g., KGE obtained for daily time scales is always above 0.8, but for San Andreas Reservoir for which it equals 0.79). At annual scale, the performance at Crystal Springs Reservoir appears low (i.e., KGE equals 0.62 and NSE equals 0.19); these low scores can be explained by the very limited number of years used for the assessment of the annual metrics (i.e., 4 years). Biases are generally low across all watersheds, with the largest bias being 6.5% at Crystal Springs Reservoir, which could also be explained by the limited availability of streamflow data. The streamflow for most watersheds are well produced for either wet or dry years and also shows a good visual fit (Figure 3-20 through Figure 3-24). Figure 3-22c and Figure 3-22d show a systematic slight underestimation of streamflow (which is significantly emphasized by the log-scale). Figure 3-23a-c illustrates good reproduction of Crystal Springs Reservoir inflow, although Figure 3-23d shows a small overestimation during low flow. Given the scarcity of observed inflow and meteorological data across the Peninsula region, these calibration results were deemed acceptable for the purpose of the LTVA. More details about the input data and model calibration and validation are provided in the Technical Report 2 (HRG 2021a).

Table 3-8. Same as Table 3-4 for Peninsula Watersheds.
The periods for performance metrics of each watershed are given in the table.

	Kling-Gupta Efficiency	Nash Sutcliffe	Percent Bias (%)
Pilarcitos reservoir inflow, 1999-2011			
Daily	0.86	0.77	-2.4
Monthly	0.92	0.88	
Water Year	0.86	0.80	
Stone Dam inflow (between Pilarcitos and Stone Dam), 1999-2011			
Daily	0.86	0.73	0.1
Monthly	0.90	0.83	
Water Year	0.82	0.77	
San Andreas reservoir inflow, 1999-2011			
Daily	0.79	0.57	0.3
Monthly	0.89	0.86	
Water Year	0.92	0.93	
Crystal Springs Reservoir inflow, 2007-2011			
Daily	0.86	0.74	6.5
Monthly	0.91	0.91	
Water Year	0.62	0.19	
Mud Dam inflow (San Mateo creek), 1999-2011			
Daily	0.87	0.76	-3.0
Monthly	0.91	0.87	
Water Year	0.90	0.86	

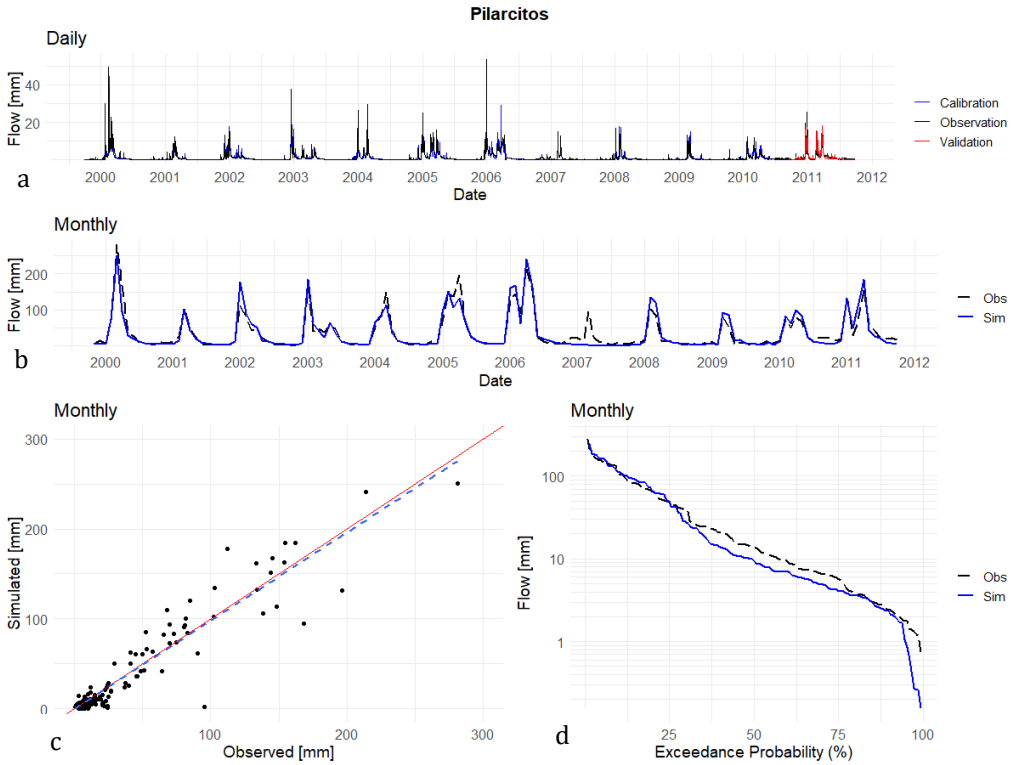


Figure 3-20. Observed and Simulated Streamflow for Pilarcitos Reservoir Inflow.

a) Comparison of the daily observed (black) and simulated time series during calibration (blue) and validation (red) period. b) Same as a) but for monthly time scale. Note: Notice the difference between the observed and simulated streamflow for the water year of 2007. Precipitation data distribution shows a single precipitation gauge (Pilarcitos gauge) covering majority of Pilarcitos watershed and all of Stone Dam watershed. For 2007, Pilarcitos gauge observed significantly less precipitation than the following years (WY 2008, 2009, and 2010), but the observed streamflow data shows noticeable flows. Further detail is available in the Technical Report 2 (HRG 2021a). c) Scatterplot showing the monthly simulated vs observed discharges. d) Flow duration curve for monthly simulated (blue) and observed (black) flows.

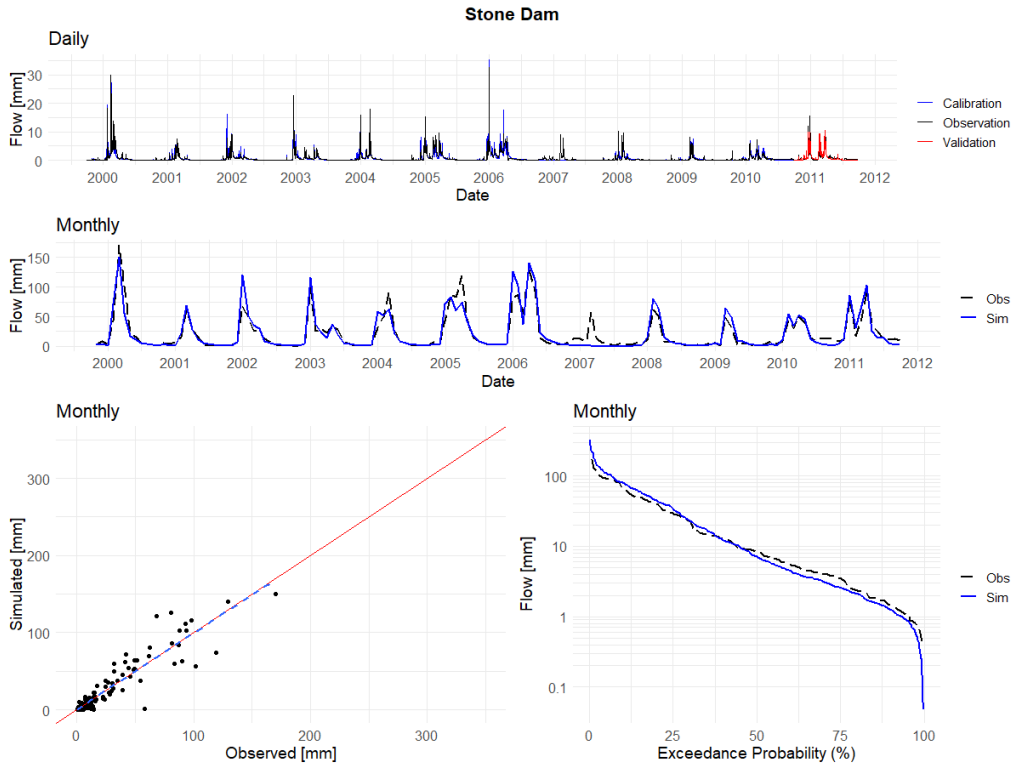


Figure 3-21. Same as Figure 3-20 but for Stone Dam Inflow.

The Stone Dam inflow is a basin-area ratio adjusted observed Pilarcitos Reservoir inflow.

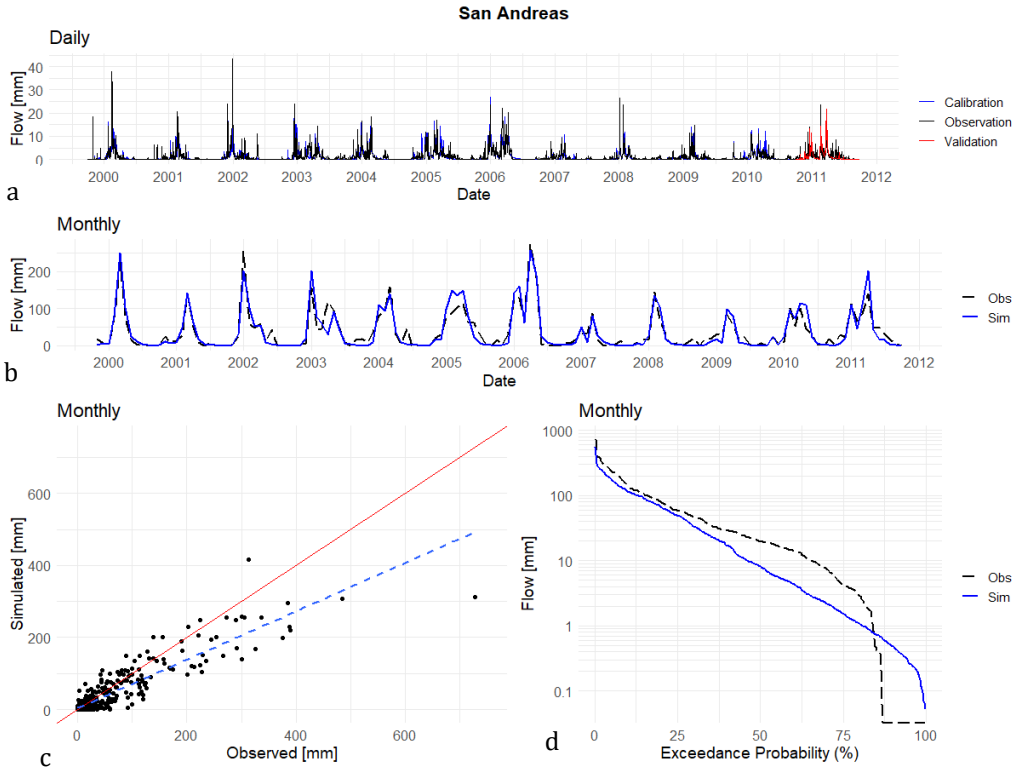


Figure 3-22. Same as Figure 3-20 but for San Andreas Reservoir Inflow.

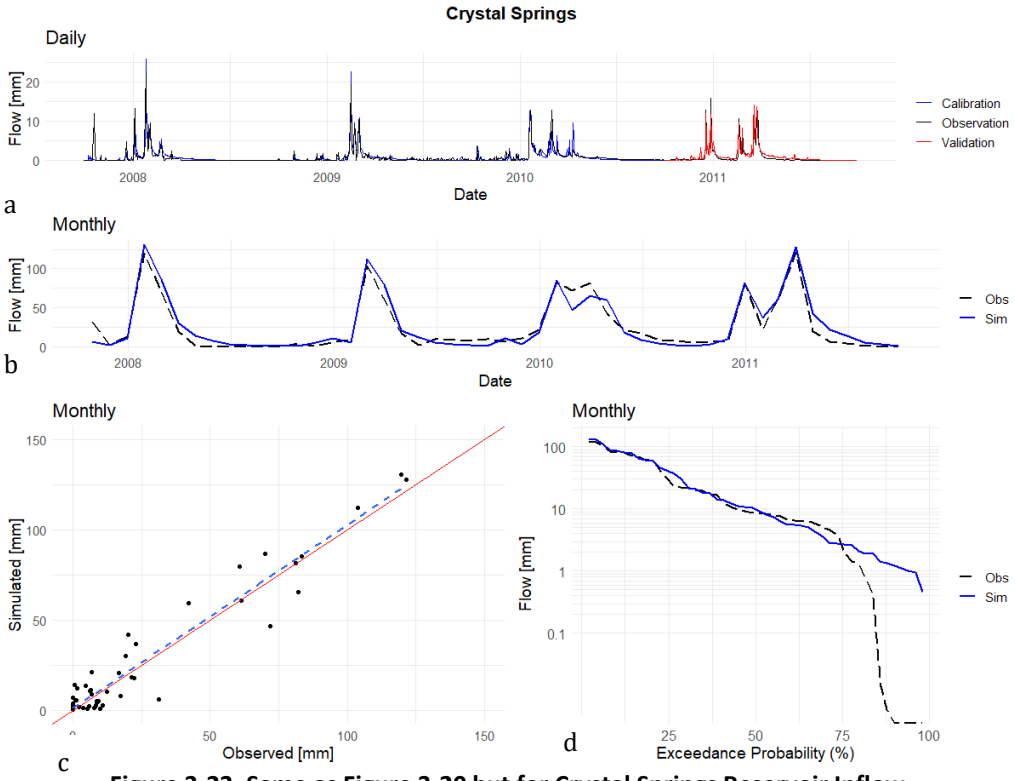


Figure 3-23. Same as Figure 3-20 but for Crystal Springs Reservoir Inflow.

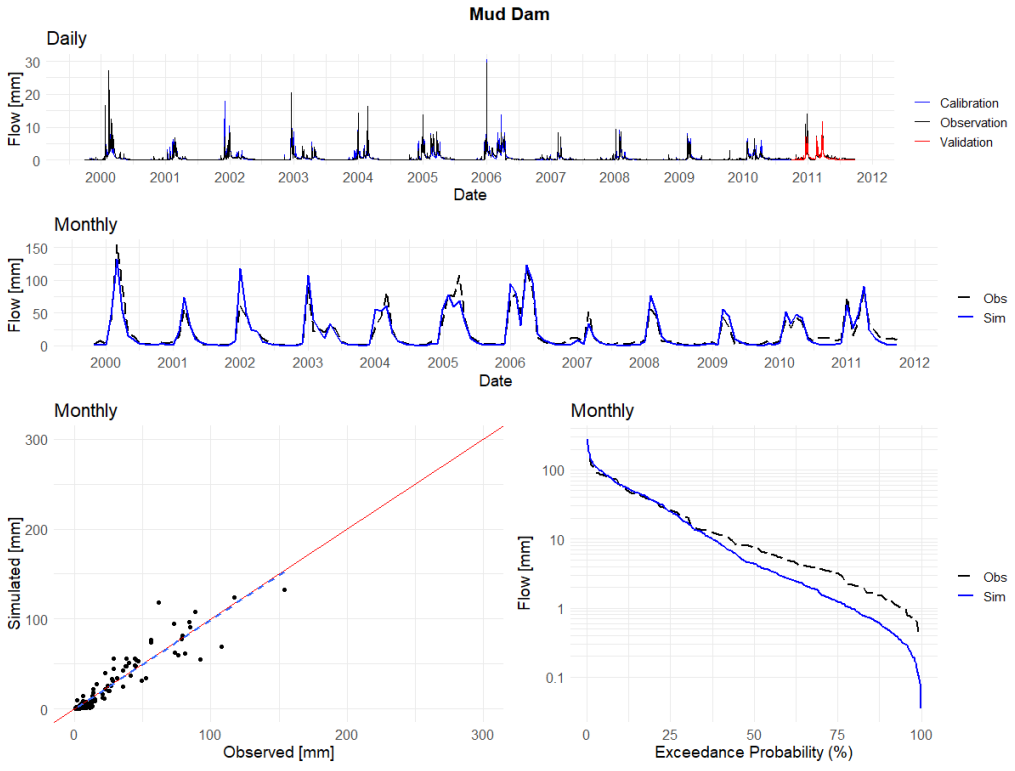


Figure 3-24. Same as Figure 3-20 but for Mud Dam Inflow (San Mateo Creek).
The Mud Dam inflow is basin-area ratio adjusted observed Pilarcitos Reservoir inflow.

3.2.4 Frequency Analysis of Drought Severity and Duration

The aim of this analysis is to estimate how sensitive the frequency of droughts in the Tuolumne River is to changes in climate, and so it will focus on hydrologic drought which is the deficit in streamflow below a specified threshold. The analysis is limited in scope to the Upcountry region because it represents roughly 85% of yield from the RWS. This analysis is strictly looking at the availability of water for SFPUC on the Tuolumne River in sequences of dry years using multiple datasets of streamflow: historical, paleo-reconstruction and simulated with PRMS hydrologic model using stochastic weather generated inputs[5]. The analysis of the impacts of hydrologic droughts on the ability of the RWS to meet demand is accomplished through other methods described in Sections 3.4 and 5.2. Figure 3-25 presents historical deviations of annual flow from the mean on the Tuolumne River at La Grange. Consecutive years with negative deviation can be considered droughts. In recent history, three droughts are especially remarkable for SFPUC: 1976-1977, 1987-1992 and the most recent 2012-2016.

This analysis will focus on the frequency of two characteristics of droughts: the severity – cumulative flow deficit – and the duration. The frequency of occurrence of events such droughts or floods is often expressed as a return period, which gives the estimated time interval between events of similar characteristics (same severity, same duration or same severity and duration). The return period between occurrences is the inverse of the average frequency of occurrence. For example, a 100-year flood has a 1/100 = 0.01 or 1% chance of being exceeded in any one year. This does not mean that if a 100-year flood occurs, then the next will occur in about one hundred years' time - instead, it means that, in any given year, there is a 1% chance that it will happen, regardless of when the last similar event was. The same reasoning is applicable to drought severity and duration.

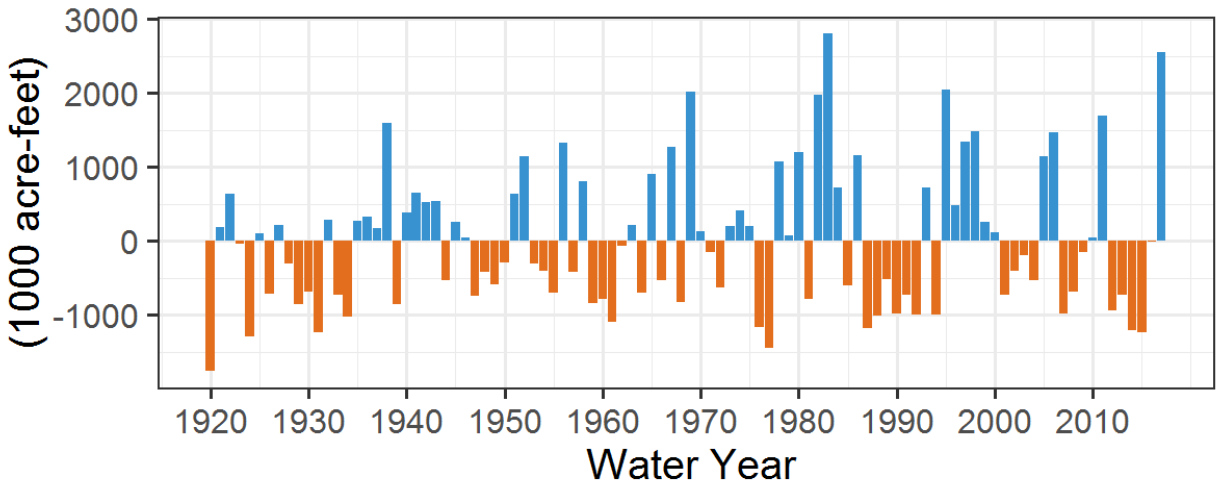


Figure 3-25. Unimpaired Annual Flow of the Tuolumne River at La Grange.
Deviations from historical mean (1924-2017).

First, the analysis estimates the distribution of severity and duration, separately. However, there is a strong dependence between severity and duration (really severe droughts tend to be longer). The separate analysis of severity and duration does not reveal the significant correlation relationship between them. Therefore, a better approach for describing drought characteristics of severity and duration is to derive the joint distribution of severity and duration. This analysis will allow to evaluate a

[5] The PRMS hydrologic model used to simulate the streamflow on the Tuolumne watershed in response to precipitation and temperature overestimates streamflow during dry years. The combination of weather generator and PRMS hydrologic model tend to underestimate the frequency of droughts (both severity and duration), as compared to the observed record.

change in frequency of drought severity and duration with changes in mean annual temperature and precipitation. The following sub-sections present the drought definition for the analysis (Section 3.2.4.1), the datasets (Section 3.2.4.2) and the fitting of statistical distributions to model the frequency of the drought severity and duration (Section 3.2.4.3). The last section (Section 3.2.4.4) presents the multivariate analysis where the joint distribution of drought severity and duration is modeled using copula functions.

3.2.4.1 Definition of Hydrologic Drought

The scientific literature has proposed several indices to characterize drought episodes (Stahl et al. 2020). In this study, drought characteristics are derived using the theory of run (Yevjevich 1967) by calculating the cumulative deficit and duration of shortfall of flow below a certain threshold. The considered 'flow' variable to be compared with the drought threshold is the Water Available to the City, or WAC. WAC is defined as the difference between the unimpaired Tuolumne River flow at La Grange $\mathrm{Q_{LG}}$, and the actual irrigation district entitlement $\mathrm{Q_{DE}}$:

$$WAC(t) = \mathrm{Q_{LG}}(t) - \mathrm{Q_{DE}}(t). \tag{3-2}$$

$\mathrm{Q_{DE}}$ is the minimum between $\mathrm{Q_{LG}}$ and the maximum irrigation district entitlements Q_{DE}^{max}:

$$Q_{DE}(t) = \min\left(Q_{LG}(t), Q_{DE}^{max}(t)\right), \tag{3-3}$$

where the maximum district entitlement is 4,800 AF/day (2,416 cfs) from June 13th to April 15th and 8,065 AF/day (4,066 cfs) from April 15th through June 13th. The cumulative deficit D during the water year Y is defined as:

$$D(Y) = \min(D(Y-1) + \mathbf{Threshold} - \mathrm{WAC}(Y), 0)\,, \text{where } D(Y = Y_i) = 0, \tag{3-4}$$

where Y_i is the first year of the time series. It is noted that Equations 3-2 and 3-3 are solved at daily time step so that the calculation of WAC can account for the variations in the maximum district entitlement Q_{DE}^{max}. An illustration of the calculation of WAC is given in Figure 3-14 for the water year 1983 (wet) and 1987 (dry). When streamflow data are not available at daily resolution, a regression is used to estimate the annual actual district entitlement $\mathrm{Q_{DE}}$ (Section 3.2.4.2) and Equation 3-4 is solved at annual time scale. Two drought thresholds are considered below. The first threshold is 269 TAF, which is the largest annual transfer in the past decade from the Upcountry region to the East Bay and Peninsula regions. It occurred during the fiscal year 2012-2013. The second threshold is 365 TAF, which is the maximum annual transfer capacity from the Upcountry region based on the capacity of the San Joaquin Pipelines. This second threshold is used to assess the upper bound, given the current transfer capacity from the Upcountry region, of what the drought distribution may look like in a hypothetical reality in where the San Joaquin Pipelines were used at full capacity.

Figure 3-26 illustrates the simulated cumulative deficit for the two considered thresholds (middle and bottom panels). The unimpaired annual streamflow time series at La Grange used in Figure 3-26 (top panel) is a reconstructed time series that was provided by SFPUC. Drought events start when the cumulative deficit value gets negative and end when they reach their local maximum (hereafter denoted as severity). The effect of setting a larger threshold is clear on this figure: larger drought threshold leads to more frequent, longer, and more severe droughts, despite some nonlinearities (cf. the one year long drought the happens right after the 1987-92 drought for 269 TAF threshold but does not exist with 365 TAF threshold). Table 3-9 lists the drought events recorded using both thresholds together with their

severity and duration. Note that all historical droughts highlighted in Figure 3-26 are correctly identified; and their duration and severity depends on the considered threshold.

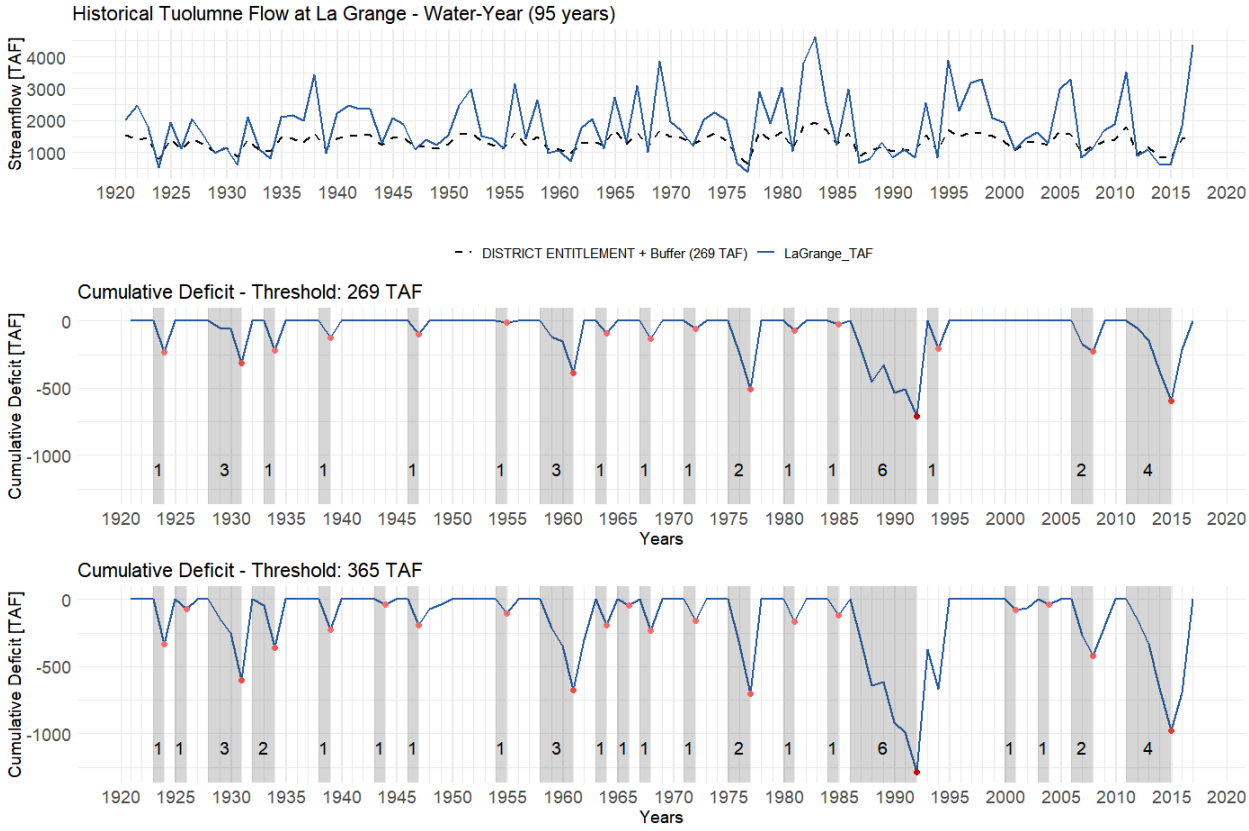


Figure 3-26. Historical Tuolumne Flow at La Grange with District Entitlement and Threshold at 269 TAF (Top) with the Corresponding Cumulative Deficit Plot (Middle) and Cumulative Deficit Plot for Threshold at 365 TAF (Bottom).

The grey shaded areas show the identified droughts, and the numbers tell their duration in years. The red dots show the associated severity to each drought event.

Table 3-9. Extracted Drought Events from Historical Tuolumne Flow at La Grange for Two Different Thresholds.
For each threshold, the drought events are sorted by decreasing severity.

Threshold: 269 TAF			Threshold: 365 TAF		
Year Drought ends	Severity [TAF]	Duration of Deficit [Years]	Year Drought Ends	Severity [TAF]	Duration of Deficit [Years]
1992	707.39	6	1992	1283.39	6
2015	594.35	4	2015	978.35	4
1977	510.18	2	1977	702.18	2
1961	389.44	3	1961	677.44	3
1931	312.14	3	1931	600.14	3
1924	233.66	1	2008	418.98	2
2008	226.98	2	1934	357.10	2
1934	218.34	1	1924	329.66	1
1994	204.77	1	1968	229.06	1
1968	133.06	1	1939	223.20	1
1939	127.20	1	1947	190.42	1
1947	94.42	1	1964	189.19	1
1964	93.19	1	1981	165.90	1
1981	69.90	1	1972	154.99	1
1972	58.99	1	1985	118.42	1
1985	22.42	1	1955	104.96	1
1955	8.96	1	2001	75.15	1
			1926	72.70	1
			1966	45.69	1
			1944	37.45	1
			2004	37.09	1

3.2.4.2 Streamflow Datasets

Following the methodology described in the previous section, the detection of drought events and estimation of their severity and duration only requires a time series of unimpaired flow at La Grange. Three different data sources are used to provide such time series.

The first dataset is an annual time series of reconstructed unimpaired streamflow at La Grange from 1921 to 2011. This time series was provided by SFPUC and illustrated in Figure 3-27. The temporal resolution of this time series is annual, which prevents from calculating the actual district entitlements using the above equations. However, an annual time series of actual district entitlements was provided by SFPUC. This dataset will be used as reference for the severity and duration of the historical droughts.

The second dataset considered for the drought analysis is a collection of simulated streamflow time series at La Grange obtained from PRMS model when forced by climate realizations generated with CliWxGen. In addition to the 10 realizations presented in Section 3.1.4, 500 other realizations are considered. The objective of using such a large number of stochastic realizations is to create a large collection of droughts to increase the robustness of the statistical inference of the distribution parameters for the severity and duration models.

The third dataset used for the drought analysis is a paleo record for the Tuolumne River at La Grange. This record is a tree-ring based reconstruction of the annual streamflow spanning from the year 900 to 2012 (Meko et al. 2014). The drought events extracted from the paleo dataset is of interest because it provides insights on the effect of the natural climate variability on the frequency, severity, and duration of droughts. Both paleo and reconstructed records are compared over the period 1921-2012 in Figure 3-27, allowing for an evaluation of the paleo streamflow. Over the period 1921-2012, the annual average of the paleo streamflow is 1,918.5 TAF, while the observed is 1,858.5 TAF. The result suggests a slight overestimation of paleo streamflow (+3.3%) with lower variability as outlined in Table 3-10.

Therefore, quantile mapping bias correction was applied to the full paleo dataset, which significantly reduced the bias in average and variability over the period 1921-2012 as indicated in Table 3-10. More details about the bias correction of the paleo record in the Technical Report 2 (HRG 2021a).

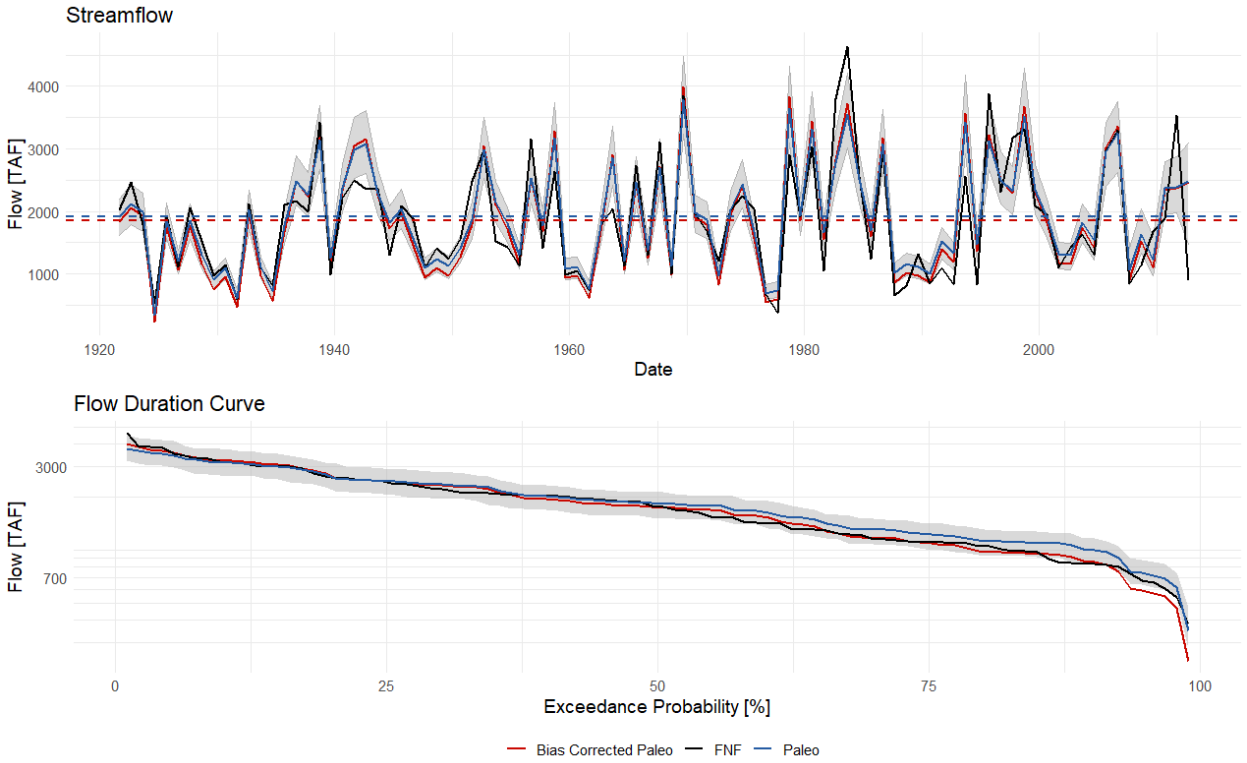


Figure 3-27. Comparison of Available Paleo Record Data with Observed Historical Full-Naturalized Flow of Tuolumne River at La Grange.

These annual water-year time-series are shown from 1921 through 2012. The grey shaded region is the 50% confidence interval for the Paleo record. The dashed lines represent the long-term average for full naturalized flow at La Grange (black), Paleo dataset (blue) and bias-corrected Paleo (red).

Table 3-10. Summary Statistics for Historic, Paleo, and Bias-Corrected Paleo Time-Series.

Dataset (1921-2012)	Mean (TAF)	Standard Deviation (TAF)
Historic	1858.5	907.9
Paleo	1918.5	810.8
Bias Corrected Paleo	1859.4	901.9

The paleo record at La Grange is available at annual temporal scale and, contrary to the historical record reconstructed by SFPUC, it does not come with a reconstruction of the actual district entitlements. As such, the actual district entitlements for the paleo records were estimated using the regression presented in Figure 3-28. This regression has been estimated using the historical reconstruction of the flow at La Grange and actual district entitlements provided by SFPUC. It uses annual unimpaired flow at La Grange as predictor to predict the annual actual district entitlement. The fit of the regression model was deemed satisfying for the drought analysis.

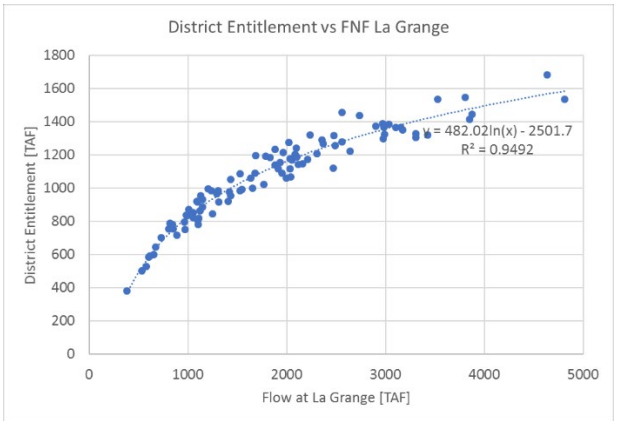


Figure 3-28. Regression between Actual District Entitlements and Unimpaired Flow at La Grange.

3.2.4.3 Fitting of Distributions for Severity and Duration

This section presents first the identified drought events using the methodology described in Section 3.2.4.1 and the datasets presented in Section 3.2.4.2. Then, statistical distributions are selected, and their parameters calibrated for modeling the frequency of drought severity and duration. Return periods for historical droughts are provided for both severity and duration. Last, copula functions are investigated to model the relationship between severity and duration and provide return periods accounting for the joint distribution of severity and duration of droughts.

Drought events were extracted from 1,113 years of paleo streamflow records (900 – 2012), 100 years of historic Tuolumne Flow at La Grange (1921 – 2020), 49 years from each of the 10 climate realizations from the CliWxGen-WG, and 49 years from the 500 drought realizations selected from the remaining realizations in CliWxGen-WG (only 49 years from the 50-year long simulations were used because the first year serves as warm-up period for PRMS). When put together, the different datasets accumulate to a total of 26,110 years. This latter dataset is further denoted as 'combined' dataset. Note that when combined, the years in the paleo records after 1920 were disregarded because this period is available from the reconstructed historical streamflow provided by SFPUC.

Figure 3-29 shows the distribution of drought severity and duration across the identified events using the combined dataset (numbers of identified events for each dataset are given in Table 3-11). Using the combined dataset and a drought threshold set to 269 TAF, a total 4,318 drought events have been extracted (average duration is 1.61 years and average severity is 217 TAF). When using a threshold equal to 365 TAF, this number grows to 4,351 (average duration is 2.04 years and average severity is 346 TAF). It is interesting to note that the number of events does not increase significantly with a larger threshold (i.e., +33 events). However, the average severity increases significantly by +129 TAF, which corresponds to an increase by almost 60%. The severity of the identified drought events ranges from roughly 0 to 1,218 TAF or 1,985 TAF, whether the 269 or 365 TAF is used. Similar to severity, the duration of the simulated drought events is significantly influenced by the chosen thresholds. Using a 269 TAF threshold, the duration ranges from 1 to 9 years. Only 4 droughts have a duration of 8 years or more (difficult to see on the figure due to scale and resolution). Using the 365 TAF threshold, the duration ranges from 1 to 14 years.

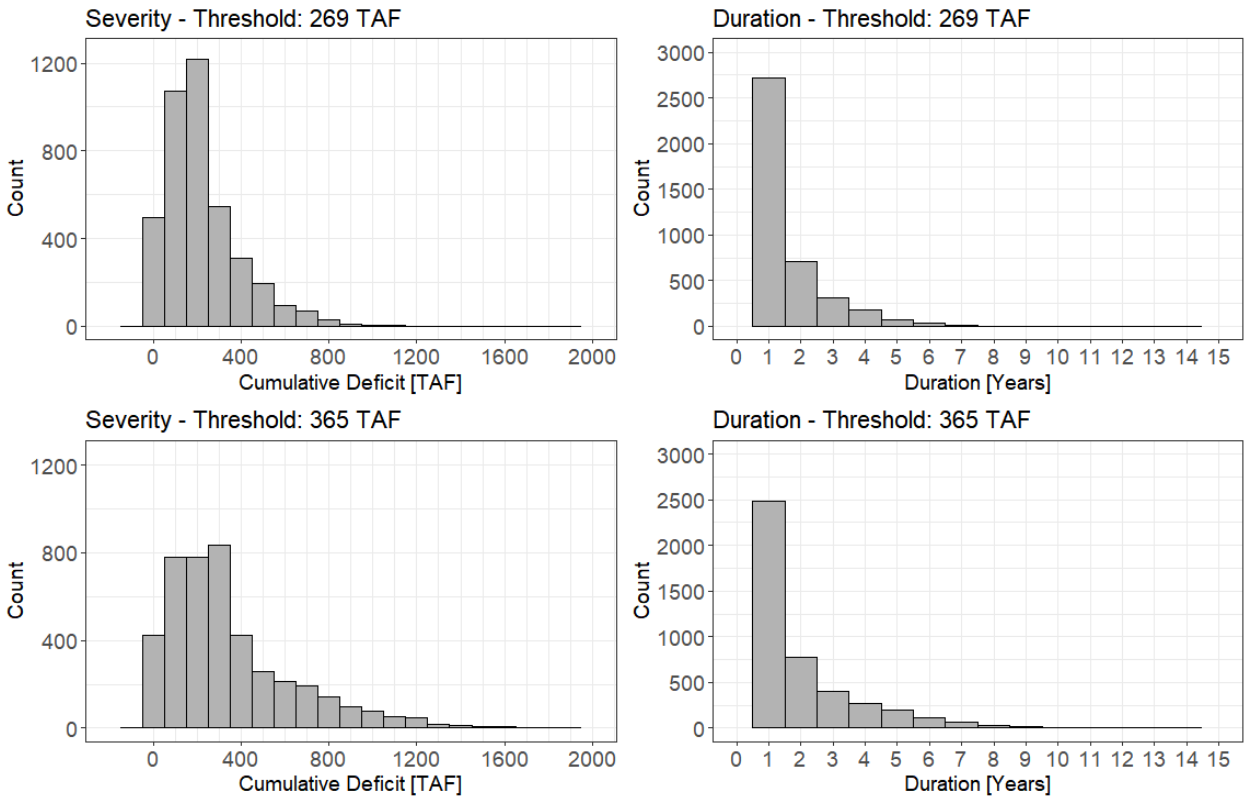


Figure 3-29. Distribution of the Severity (Left) and Duration (Right) of the Identified Drought Events Using the 269 TAF (Top) and 365 TAF (Right) Thresholds with the Combined Dataset.

Table 3-11. Number of Identified Drought Events Using Theory of Run for Each Dataset.
The combined dataset is a combined dataset with drought events from historical (1921-2020), Paleo (900-1920), and 510 stochastic realizations.

Dataset	Number of years	Average Severity (TAF)		Number of Drought Events	
		Threshold: 269 TAF	Threshold: 365 TAF	Threshold: 269 TAF	Threshold: 365 TAF
Historical (1921 – 2020)	100	236	333	17	21
Paleo (900 – 2012)	1,113	147	222	135	173
Bias Corrected Paleo (900 – 2012)	1,113	188	294	171	190
Stochastic Realizations	24,990	220	351	3,902	4,172
Combined	26,110	217	346	4,318	4,351

Several statistical distributions have been investigated to model the frequency of the severity and duration of the identified droughts. The considered distributions to model severity and duration are given in Table 3-12 and Table 3-13, respectively. Continuous distributions were tried for severity and both continuous and discrete distributions were used for duration. The method to calibrate the model parameters is the Maximum Likelihood Estimation (MLE). The fitted distributions were evaluated considering both the Akaike and Bayesian information criteria (AIC and BIC, respectively), in addition to the visual inspection of the goodness of fit of the empiric and simulated cumulative distributions for severity and duration. AIC and BIC criteria are commonly used to compare the goodness-of-fit across multiple distributions. They are both negatively oriented (low values are better than high values). For the combined dataset, Table 3-12 shows that a Weibull distribution provides the best fit to model the

frequency of the severity of the identified drought events. Still for the combined dataset, Table 3-13 shows that the Log-Pearson Type-III distribution provides the minimum AIC score, although the visual fit for the drought duration cumulative distribution is bad (Figure 3-30). For this reason, the selection of the distribution to model the frequency of the drought duration was based on the visual fit of the data. As such, the Generalized Pareto Distribution (GPD) was selected to model the frequency of the drought duration (Figure 3-30). Weibull (severity) and GPD (duration) were also found to be suitable distributions for the other datasets. Note that the choice of either 269 or 365 TAF for threshold does not influence the model selection. The calibrated parameters of the fitted distribution are given in Table 3-14 Further details are available in the Technical Report 2 (HRG 2021a).

Table 3-12. Cumulative Drought Deficit Distribution Fit Criterions for Thresholds 259 TAF and 365 TAF.
AIC and BIC are obtained for the combined dataset. Distributions are sorted by increasing AIC. The * symbol indicates the selected model for the drought severity.

	Threshold: 269 TAF		Threshold: 365 TAF	
Distribution	**AIC**	**BIC**	**AIC**	**BIC**
Weibull*	49689	49702	59611	59624
Gamma	49792	49804	59665	59662
Log-Pearson Type III	49868	49887	59759	59778
Generalized Pareto (GPD)	49872	49884	59650	59662
Generalized Extreme Value (GEV)	49898	49917	60061	60080
Exponential	50147	50153	59789	59795

Table 3-13. Drought Duration Distribution Fit Criterions for Thresholds 259 TAF and 365 TAF.
AIC and BIC are obtained for the combined dataset. Distribution are sorted by increasing AIC. The * symbol indicates the selected model for the drought duration. As indicted in the text, the selection of the model for drought duration was based on the visual fit of the empirical distribution rather than the AIC score. '*Inf*' stands for Infinity.

	Threshold: 269 TAF		Threshold: 365 TAF	
Distribution	**AIC**	**BIC**	**AIC**	**BIC**
Log-Pearson Type III	5959	5978	10139	10158
Gamma	9320	9333	13521	13534
Weibull	10099	10112	14044	14057
Poisson	11290	11297	15103	15109
Negative Binomial	11292	11305	15102	15109
Exponential	11635	11641	14964	14970
Generalized Pareto (GPD)*	12618	12624	16716	16722
Geometric	13666	13673	16820	16826
Generalized Extreme Value (GEV)	*Inf*	*Inf*	12463	12482
Binomial	*Inf*	*Inf*	*Inf*	*Inf*
Hypergeometric	*Inf*	*Inf*	*Inf*	*Inf*

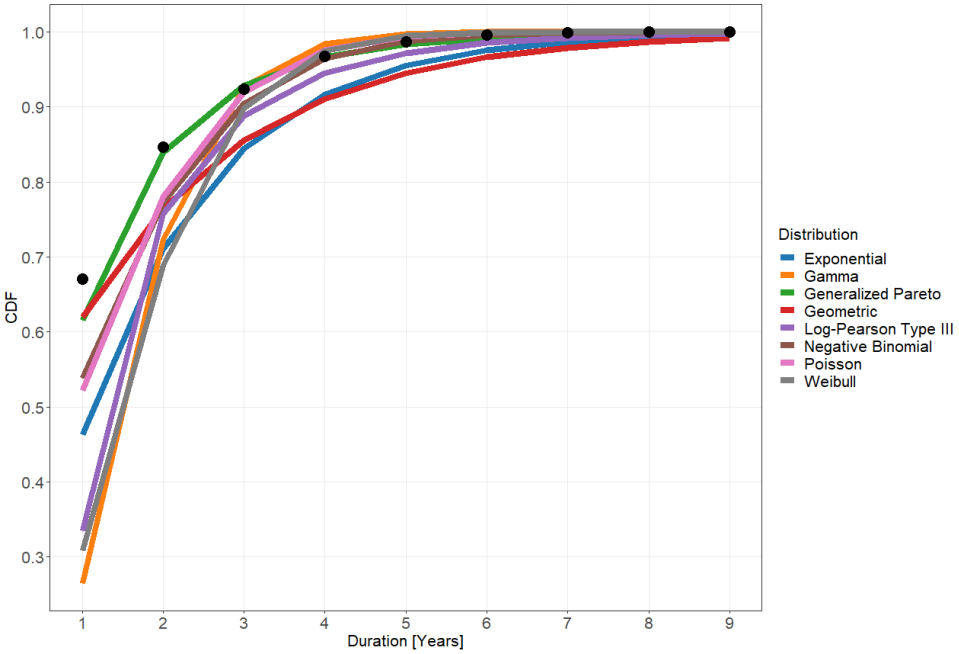


Figure 3-30. Empirical (Black Dots) and Theoretical (Color Lines) Cumulative Distribution Functions for Drought Duration.

Table 3-14. Distribution Parameters for Fitted Baseline Scenario Distributions for Severity and Duration.

Variable	Distribution	Dataset	Threshold: 269 TAF		Threshold: 365 TAF	
			Shape	Scale	Shape	Scale
Severity	Weibull	Historical	1.14	246.94	1.06	341.59
		Paleo	1.23	156.90	1.31	240.39
		Bias Corrected Paleo	1.39	205.59	1.29	316.43
		Stochastic Realizations	1.31	237.76	1.18	370.83
		Combined	1.30	234.85	1.18	365.58
Duration	GPD	Historical	0.22	0	0.17	0
		Paleo	-0.14	0	0.023	0
		Bias Corrected Paleo	-0.07	0	0.17	0
		Stochastic Realizations	0.10	0	0.33	0
		Combined	0.063	0	0.32	0

The return period T for a specific severity or duration value x is calculated based on:

$$T = \frac{\mu}{1 - F_x(x)}, \tag{3-5}$$

where $F_x(x)$ is the cumulative distribution function value for the variable x, and μ is the mean inter-arrival time (Gräler et al. 2013). The mean inter-arrival time is estimated as the ratio between the total

numbers of years and the number of drought events. As such, it varies depending on the considered threshold and the considered dataset. The mean inter-arrival time values for each dataset are given for each dataset and threshold in Table 3-15.

Table 3-15. Mean Inter-arrival Time Obtained for the Considered Datasets and the Two Drought Thresholds.

Dataset	Mean inter-arrival time μ (years)	
	Threshold: 269 TAF	Threshold: 365 TAF
Historic	4.76	5.88
Paleo	8.24	6.43
Bias Corrected Paleo	6.51	5.86
Stochastic Realizations	6.40	5.99
Combined	6.05	6.00

Figure 3-31 illustrates the drought severity and duration frequency curves obtained with the considered datasets and the 269 TAF threshold. Figure 3-32 shows similar frequency curves but obtained using 365 TAF threshold. The sampling uncertainty is shown using 95% confidence intervals obtained with bootstrap method (5,000 random samples were used). The three most recent major historical droughts are indicated on these figures. The estimated return periods for the historical droughts estimated from each dataset and using the two considered thresholds are given in Table 3-16.

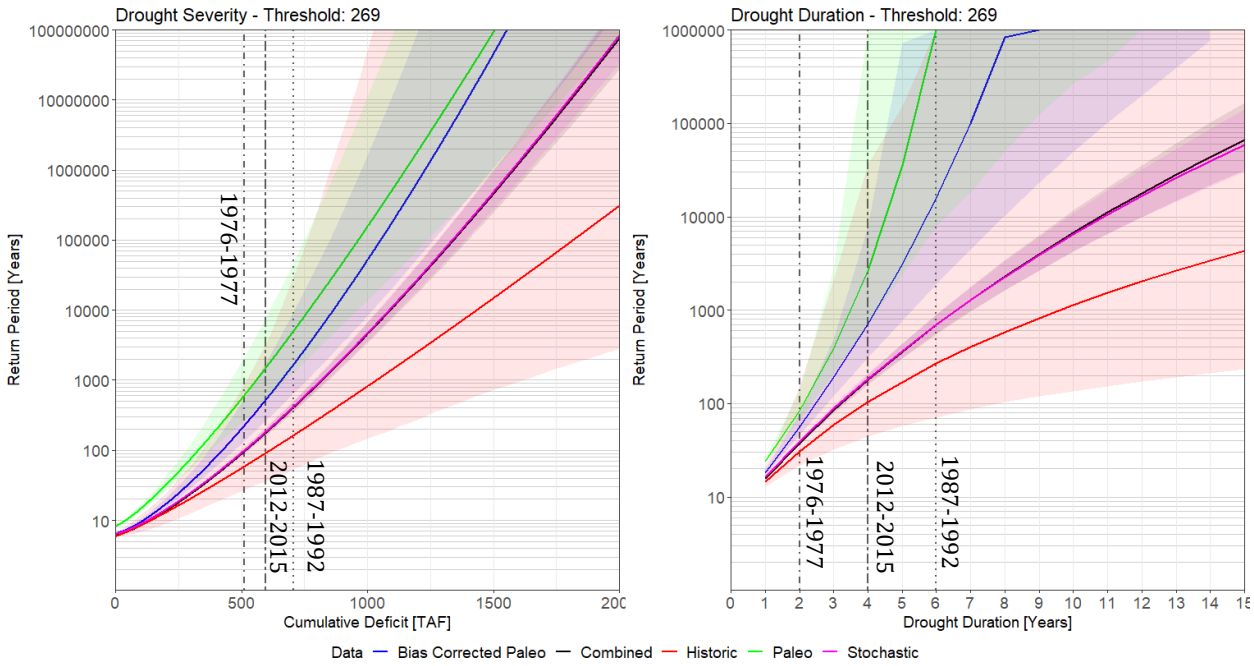


Figure 3-31. Drought Frequency Curves for Severity (Left) and Duration (Right).
Results are obtained for the 269 TAF threshold. The color lines show the frequency curves obtained for the considered datasets: bias corrected paleo (blue), paleo (green), historic (red), 510 stochastic realizations (purple) and the combined dataset (black). The shaded areas show the 95% confidence intervals obtained using the bootstrap method for each model. Severity and duration for the historic drought events are shown and labeled as vertical lines.

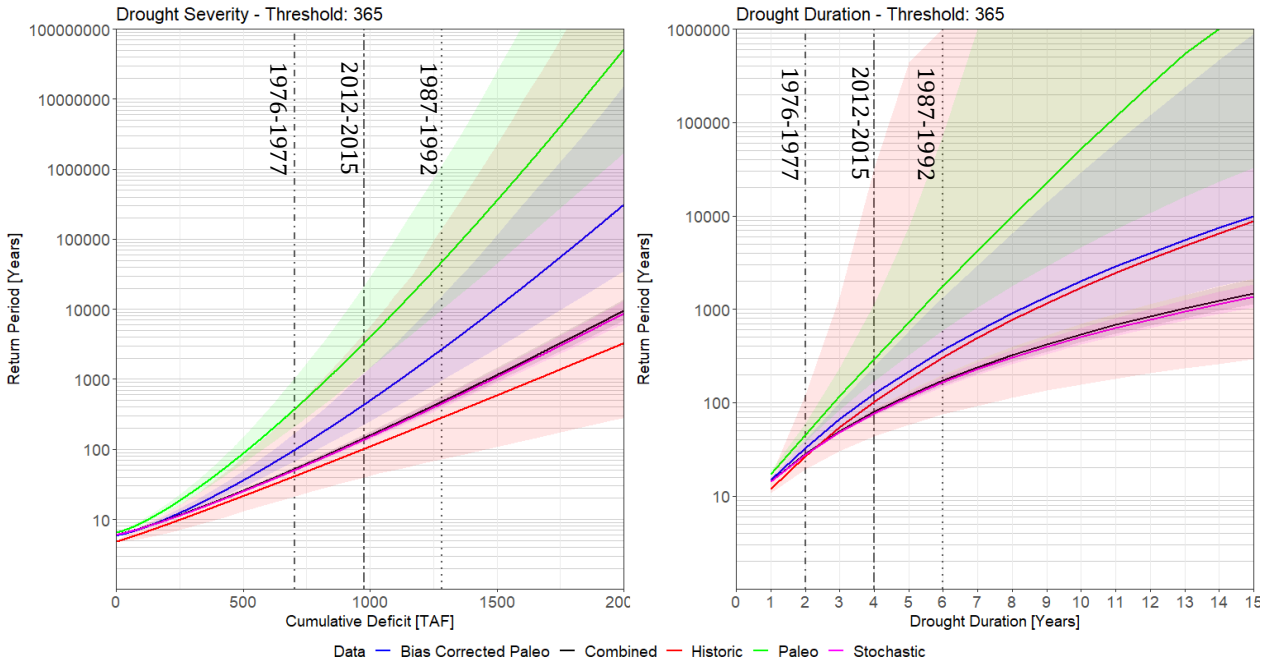


Figure 3-32. Drought Frequency Curves for Severity (Left) and Duration (Right).
Results are obtained for the 269 TAF threshold. Caption details are similar than Figure 3-31.

Table 3-16. Estimated Return Periods of Drought Severity and Duration for the Historic Drought Events.

Drought Event	Dataset	Threshold: 269 TAF Return Period [Years]		Threshold: 365 TAF Return Period [Years]	
		Severity	Duration	Severity	Duration
1976-1977	Historic	58	31	41	27
	Paleo	599	85	370	45
	Bias Corrected Paleo	221	56	97	33
	Stochastic Realizations	98	40	50	28
	Combined	95	38	52	28
1987-1992	Historic	163	267	280	305
	Paleo	5,033	*Inf*	47,385	1,769
	Bias Corrected Paleo	1,673	15,838	2,672	364
	Stochastic Realizations	420	698	447	165
	Combined	407	694	475	172
2012-2015	Historic	89	103	101	102
	Paleo	1,455	2,644	3,324	295
	Bias Corrected Paleo	508	714	436	124
	Stochastic Realizations	178	184	137	77
	Combined	173	179	144	79

Estimated return periods for drought severity and duration for historic droughts are shown to be highly dependent on the underlying datasets used to fit the statistical models. For instance, the return period estimates for the severity of the 1976-77 and 1987-92 droughts range from 58 to 599 years, and from 163 to 5,033 years, respectively (Table 3-16). Similar comments can be made for the drought duration (Table 3-16). One important explanation for the large estimate range across datasets is the difference in number of years and identified drought events within each datasets (Table 3-11). Return periods calculated with a limited number of identified drought events corresponding are expected to be unreliable, especially for severity or duration significantly beyond the range observed in the dataset (Bonaccorso, et. al. 2003). The sampling uncertainty can be visualized in Figure 3-32 with the 95% confidence intervals surrounding the frequency curve obtained from each dataset (i.e., shaded areas in color). Confidence intervals were obtained via the bootstrap method using 5,000 random samplings. Confidence intervals are large for datasets with a limited number of events available to fit the distribution (e.g., historic and paleo), while they are small for datasets with a large number of droughts.

3.2.4.4 Joint Distribution of Severity and Duration

As aforementioned, the separate analysis of severity and duration does not reveal the significant correlation relationship between them. Therefore, a better approach for describing drought characteristics of severity and duration is to derive the joint distribution of severity and duration. This analysis will allow to evaluate a change in frequency of drought severity and duration with changes in mean annual temperature and precipitation.

Figure 3-33 illustrates the relationship between drought severity and duration as seen across the identified events and for the two considered thresholds. The 1976-77 and 1987-92 droughts are labeled on the plot. Several observations can be made from this figure.

First, it is noted that the severity tends to increase with the duration of the events, although the severity varies significantly for each duration. For example, the severity of 2-year long drought events ranges from about 0 to about 550 TAF or 750 TAF for drought thresholds set to 269 or 365 TAF, respectively. Note that the historical 1976-77 drought is on the very high-end of this range.

Another highlight shown in Figure 3-33 is that droughts with rather short duration (e.g., less than 3 years) can be as severe, if not more severe, than droughts with significantly longer duration. For example, considering a 269 TAF threshold, the severity of the most severe 2-year long droughts is similar to the median severity of the 4-year long droughts, which is also twice as much as the severity of the less severe 6-year long drought.

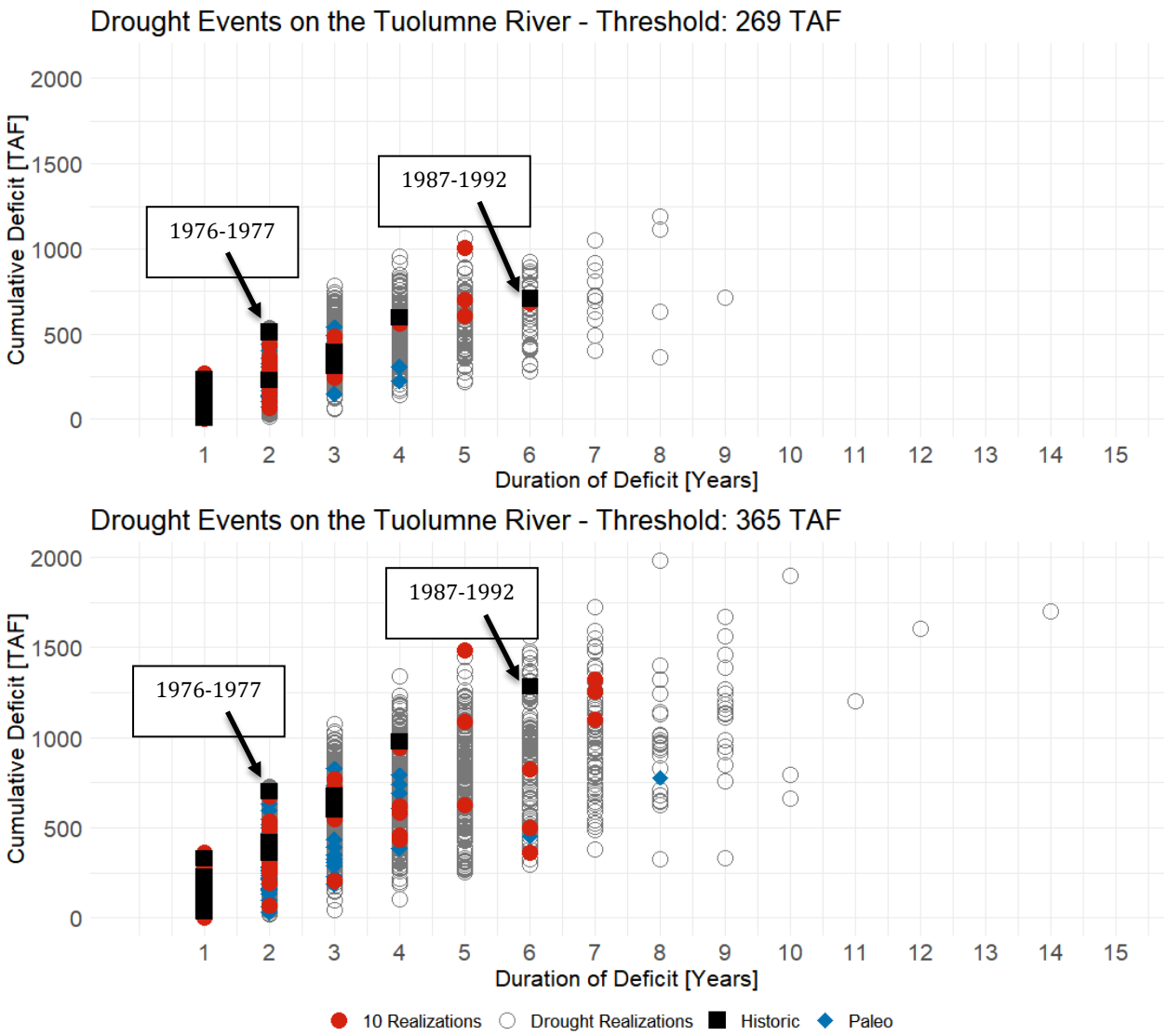


Figure 3-33. Extracted Drought Events across All 10 Climate Realizations (Red Circle), 500 Drought Realizations (Grey Circle), Historic Data (Black Square), and Paleo (Blue Diamond) Records for the Two Thresholds; 269 TAF (Top) and 365 TAF (Bottom).

The complex relationship between drought duration and severity illustrated above highlights the importance of accounting for the joint probability of the severity and duration for assessing drought risk. To provide insight into the relationship between drought severity and duration, Kendall's rank correlation coefficient, Spearman's rank correlation coefficient, and Pearson's linear correlation coefficient are shown in Table 3-17. All correlation coefficients are positive implying that longer droughts tend to be more severe. Pearson's correlation indicates an explained variance slightly above either 61% or 66% when using either 269 TAF or 365 TAF threshold, indicating a significant relationship between severity and duration (i.e., the explained variance is obtained by taking the squared value of the Pearson' correlation coefficient).

Table 3-17. Correlation Coefficients between Drought Severity and Duration.

Correlation	Threshold: 269 TAF	Threshold: 365 TAF
Kendall's Rank Correlation Coefficient	0.564	0.628
Spearman's Rank Correlation Coefficient	0.694	0.773
Pearson's Linear Correlation Coefficient	0.782	0.817

In statistics, the dependence between two random variables, say $\boldsymbol{Y}$ and $\boldsymbol{Z}$, can be modeled via their joint distribution using multivariate distributions, such as the bivariate normal distribution for instance. However, one major limitation of the multivariate distributions is that the marginal distributions must follow the same theoretical distribution. For instance, in the case of the bivariate normal distribution, the two marginal distributions are normal. This is a significant limitation for the drought analysis conducted for the LTVA because the drought severity and duration of the identified drought events are distributed following two different distributions; a Weibull distribution for the severity and the GPD for the duration (Section 3.2.4.3).

Following Sklar's theorem (Sklar 1959), however, any multivariate joint distribution $F(y,z)$ can be specified from the univariate marginal distributions for $\boldsymbol{Y}$ and $\boldsymbol{Z}$ and a copula function C that describes the dependence structure between the random variables $\boldsymbol{Y}$ and $\boldsymbol{Z}$ (e.g., Genest and Favre 2007):

$$F(y,z) = \Pr(Y < y, Z < z) = C(F_Y(y), F_Z(z), \vartheta) \tag{3-6}$$

where ϑ is a parameter that controls the degree of dependence between $\boldsymbol{Y}$ and $\boldsymbol{Z}$. In the case developed here, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are the duration and severity of the identified drought events. Various copula functions were fit using maximum likelihood estimation, and the selection was made based on the Akaike Information Criteria (AIC). The survival Clayton (Clayton 1978) and Gumbel (Gumbel 1960) copula functions were found to fit best the severity and duration joint distribution, respectively when using 269 TAF or 365 TAF thresholds (Table 3-18).

The structure of the Clayton copula is defined as:

$$C_\vartheta^{Cl}(u_1, \dots, u_n) = \left(\sum_i \left(u_i^{-\vartheta} - 1\right) + 1\right)^{1/\vartheta}, \text{where } \vartheta \geq 0 \text{ and } 0 \leq u_i \leq 1, i = 1, \dots, n \tag{3-7}$$

The survival Copula is a 180° rotation of the input pseudo-observation of the covariate u_n, which is shown in Equation 3-8 with survival copula $\check{C}$ and the copula $\bar{C}$.

$$\check{C}(u_1, \dots, u_n, \dots, u_N) = \bar{C}(1 - u_1, \dots, 1 - u_n, \dots, 1 - u_N) \tag{3-8}$$

The structure of the Gumbel copula is defined as:

$$C_\vartheta^{Gu}(u_1, \dots, u_n) = \exp\left(\left(\sum_i (-\log u_i)^\vartheta\right)^{1/\vartheta}\right), \text{where } \vartheta \geq 1 \text{ and } 0 \leq u_i \leq 1, i = 1, \dots, n \tag{3-9}$$

For the application of the above copula functions, n=2 as only 2 co-variates are considered (i.e., severity and duration).

Table 3-18. Copula Model Fit Performance for Various Copula Families.
Lower the value, the better the fit. Bold values show the minimum AIC values that indicate the copula function to use for each threshold.

	Threshold: 269 TAF		Threshold: 365 TAF	
Copula	**AIC**	**BIC**	**AIC**	**BIC**
Survival Clayton	**-2909.63**	-2903.33	-3622.42	-3616.04
Clayton	-2888.01	-2881.70	-3668.83	-3662.46
Joe	-2852.59	-2846.28	-3622.42	-3616.04
Gumbel	-2828.46	-2822.15	**-3746.15**	-3739.77
Gaussian	-2541.45	-2535.15	-3552.63	-3546.25
Student t-Copula	-2539.78	-2527.17	-3544.26	-3531.51
Frank	-2485.88	-2479.58	-3593.14	-3586.77

The fitted copula parameters are respectively $\vartheta = 2.54$ and $\vartheta = 2.54$ for the survival Clayton (269 TAF) and Gumbel (365 TAF) copula functions. The estimates of the Kendall's rank correlation coefficients from the fitted copulas respectively equal 0.537 and 0.606, which indicates that the fitted copula functions fairly well represent the observed rank correlation (0.564 and 0.628, respectively for 269 and 365 TAF threshold; Table 3-17).

The joint return period of an event with severity x and duration y is calculated as:

$$T_{joint} = \frac{\mu}{1 - F(x) - F(y) + C(x,y)}, \quad (3\text{-}10)$$

with $F(.)$ the cumulative distribution function for each margin and μ the average inter-arrival time.

Table 3-19. Calculated Joint Return Period for Thresholds 269 and 365 TAF.

	Threshold: 269 TAF	Threshold: 365 TAF
Drought Event	**Joint Return Period [Year]**	**Joint Return Period [Year]**
1976-1977	98	61
1987-1992	772	537
2012-2015	236	176

The estimates of the return period for the most important historical droughts are shown in Table 3-19. The 1987-92 drought is shown to be the historical drought with the largest return period (772 and 537 years for 269 TAF and 365 TAF thresholds). The return period of the 1976-77 drought was estimated to 98 years with 269 TAF threshold, and to 61 years for 365 TAF threshold. Finally, the return period of the most recent 2012-2015 drought was estimated to 236 and 176 years, respectively for 269 and 365 TAF threshold.

3.2.5 Model Limitations

Prediction of future streamflow under changing climate is subject to various uncertainty sources and assumptions that are important to acknowledge. A paramount assumption that is unavoidable is the one of the robustness of the model parameters to future climate. This assumption somewhat disregards any changes other than the modification of the weather inputs to the hydrology models. For instance, the modification in land-use and land-cover that could follow from warming conditions is not accounted for,

while the type of vegetation is expected to change should the temperature changes. The robustness assumption of the model parameters does not account either for any possible modification of the physical mechanisms that could follow from the modification of the phase of the precipitation, or from the modification of the precipitation regime (e.g., long rain events vs. convective storms).

Some limitations are tied with the work and models used in the LTVA. Regarding the Upcountry region, the PRMS models used to model streamflow have been calibrated using the whole available dataset and no model validation was considered. As such, the model performance out of sample is unknown. The considered implementation of potential evapotranspiration module in PRMS is lumped (i.e., no spatial heterogeneity in the evapotranspiration model parameter), which, given the observed spatial variability of the evapotranspiration using remote-sensed product (i.e., MODIS), a distributed model would have been more appropriate. More details are given in the Technical Report 2 (HRG 2021a). Some inputs to the PRMS had to be reconstructed from the outputs of the weather generator CliWxGen. A multivariate linear interpolation was used, which led to an underestimate of the inter-annual variability of the precipitation variable at 4 out of 9 locations across Upcountry region. The PRMS hydrologic model overestimates streamflow during dry years. For example, the flow computed at the Tuolumne River at La Grange is overestimated, and therefore WAC is also overestimated by about 482,000 acre-feet during the drought sequence 1987-1992 (observed is 813,000 acre-feet versus simulated is 1,295,000 acre-feet, Table 3-20).

Table 3-20. Comparison of Observed and Simulated Water Available to the City over the 1987-1992 Drought Using Results from PRMS Hydrologic Model.

	Water available to the City (acre-feet)		
Water Year	**Observed**	**Simulated**	**Error**
1987	40,439	174,868	134,428
1988	23,521	112,425	88,905
1989	379,305	383,159	3,854
1990	46,632	114,957	68,325
1991	269,733	302,719	32,986
1992	53,525	207,221	153,697
Sum	813,155	1,295,349	482,194

In the Bay area, the streamflow data available for the Peninsula region is limited and does not necessarily align well with the availability of the weather database. As such, only a limited validation of the SACSMA models in this region has been possible. In addition, while the role of the fog can be important during some period across the region, this phenomenon is not accounted for in the modeling.

Regarding the drought analysis, several limitations could alter the results of the frequency analysis. First, the drought analysis relies on the PRMS skill to reproduce the unimpaired flow of the Tuolumne River at La Grange. As discussed in Section 3.2.3.1, PRMS simulations were post-processed to improve the performance of the raw PRMS simulations at this location, although some bias remains after post-process, especially on the left-side of the annual flow distribution (i.e., years with low annual flow values). This bias then spreads to the WAC distribution, which is the one the drought analysis is based

on. Overall, WAC in dry years is overestimated. Therefore, one can assume an underestimation of the simulated frequency of drought duration and severity.

Droughts, even more than floods, are rare events. As such, long time series are really paramount for inferring robust model parameters to model drought frequency. For the flow at La Grange, less than 100 years were reconstructed by SFPUC, which is rather small. Other datasets are used (i.e., paleo records, stochastic simulations) but they all have their own biases and errors. The considered datasets actually lead to really different frequency curves and estimates for the return period for the historical drought events.

3.3 Water Demand Model

3.3.1 Modeling Goals

The purpose of the demand model is to generate a spatially disaggregated time series of water demand for the City of San Francisco and each suburban retail and wholesale customer at the daily time step for various scenarios of annual demand. It is important to make the distinction between water demand in the service area and demand on the RWS (SFPUC share of demand). Several wholesale customers in the service area have multiple sources of supply to meet their demand, therefore the demand on the RWS is a fraction of the total service area demand. The modeling effort first estimates annual demand projections to 2070 for each customer, then determines the share of this annual demand met by the RWS, and finally disaggregates annual demand on the RWS temporally on a daily time step. The baseline annual demand is constructed using sales data for fiscal year 2012-2013 to represent 'normal' conditions prior to the 2012-2016 drought. The daily time series of demand for retail and wholesale customers serve as input to the water system model described in Section 3.4. Details on the water demand model are provided in the Technical Report 3: Urban Water Demand (HRG 2020a).

3.3.2 Modeling Methods

Water demand on Hetch Hetchy RWS is split between retail demand in the City of San Francisco and wholesale demand in the service area. Retail customers include the residents, businesses, and industries located within San Francisco city limits, referred to as the in-city retail service area. Retail service is also provided to a patchwork of customers located outside the City, such as the Town of Sunol, San Francisco International Airport (SFO), Lawrence Livermore National Laboratory, and the Groveland Community Services District (GCSD) in Tuolumne County. These areas are not contiguous and are collectively referred to as the suburban retail service area. In addition, SFPUC serves key municipal, fire and irrigation accounts both within the City of San Francisco and the surrounding suburban areas.

In order to produce projections, it is necessary to establish a baseline for annual demand in the service area and also on the RWS. Baseline demand consists of both retail and wholesale demands. In consultation with SFPUC, the HRG decided to use pre-drought conditions of fiscal year (FY) 2012-2013 as the basis for baseline demand.

Baseline values of retail demand is provided in Table 3-21 and add up to 78.9 mgd. The components of retail demand are drawn from a mixture of sales data in FY2012-2013 and values provided in SFPUC's 2015 Urban Water Management Plan (SFPUC 2016). SFPUC delivers 100% of the retail customer demand.

Table 3-21. Baseline Demand for Suburban Retail Customers.

Customer/ Region	Average Daily Demand (mgd)
In-City Retail	
Single Family Residential[1]	16.60
Multi-Family Residential[1]	22.60
Commercial and Industrial[1]	18.80
Other retail demand[2]	9.90
Water loss[3]	6.00
Total in-City	**73.9**
Suburban Retail	
Golden Gate Bridge Cemetery[4]	0.25
Menlo Park Country Club[4]	0.22
NASA[4]	0.62
San Francisco International Airport[4]	1.16
Town of Sunol[4]	0.58
Groveland CSD[4]	0.40
Lawrence and Livermore National Laboratories[4]	0.80
General Electric[4]	0.03
Other suburban retail[5]	0.92
Other[6]	0.01
Total suburban retail[6]	**5.00**
RETAIL TOTAL	**78.90**

[1] Historical sales data for FY2012-13, which is used by SFPUC as a benchmark for 'normal' conditions prior to the 2012-16 drought.
[2] This category includes other CI, institutional /governmental and landscape irrigation (both potable and non-potable). The estimates follow the same procedure as for the 2015 UWMP (SFPUC 2016).
[3] Water loss is assumed constant at 6 mgd and follows the procedure of the 2015 UWMP (SFPUC 2016).
[4] Sales data from SFPUC Customer Bureau. This value is an average and is assumed constant over time.
[5] 2015 UWMP provides a suburban retail demand total of 5 mgd. Other suburban retail represents the difference between this total demand value and the sum of suburban retail customers listed here.
[6] Cordilleras MWC is a wholesale customer not part of BAWSCA. For the purpose of the system model, it was added to the suburban retail demand with a constant demand of 0.01 mgd.

Wholesale customers include 27 BAWSCA customers and Cordilleras Mutual Water Company. Table 3-22 shows the annual sales for FY2012-13, which is the baseline for this study, and the fraction of the total demand served by SFPUC (data from BAWSCA Annual Report; BAWSCA 2021). For the baseline year, total customer demand is established by dividing their RWS water use by the observed SFPUC share of total demand. The fraction of demand met by the RWS varies from year to year for customers with multiple sources of supply, in particular in dry years. For this study, this ratio held constant and set at that observed in FY 2012-2013. The baseline total demand for the wholesale customers is 226.6 mgd with 147.88 mgd supplied by the RWS.

Table 3-22. Baseline Total Demand and Baseline Demand on the RWS (SFPUC Share of Total Demand) for Wholesale Customers and the City of San Francisco (Values Are Drawn from FY2012-13 FY Sales).
Historical sales data for FY2012-13, which is used by SFPUC as a benchmark for 'normal' conditions prior to the 2012-16 drought.

Wholesale Customer	Total Demand in FY2012-13 (mgd)	Demand on the RWS in FY2012-13 (mgd)	SFPUC share of Total Demand (%)
Alameda County Water District	43.17	9.06	21
Brisbane/GVMID	0.32	0.32	100
Burlingame	4.48	4.16	93
CWS - Bear Gulch	12.71	12.08	95
CWS - Mid Peninsula	14.04	14.04	100
CWS - South San Francisco	7.41	6.89	93
Coastside County Water District	1.88	1.67	89
Daly City	7.38	4.13	56
East Palo Alto Water District	2.08	2.08	100
Estero MID	4.05	4.05	100
Hayward	15.48	15.48	100
Hillsborough	3.25	3.25	100
Menlo Park	3.24	3.24	100
Mid-Peninsula	3.00	3.00	100
Millbrae	2.30	2.28	99
Milpitas	10.21	6.63	65
Mountain View	10.83	9.09	84
North Coast County Water District	2.51	2.51	100
Palo Alto	11.80	11.33	96
Purissima Hills Water District	1.99	1.99	100
Redwood City	9.90	9.31	94
San Bruno	3.73	2.01	54
Santa Clara	21.77	2.18	10
San Jose	4.89	4.50	92
Stanford University	3.39	2.14	63
City of Sunnyvale	19.87	9.54	48
Westborough Water District	0.91	0.91	100
Cordilleras MWC	0.01	0.01	100
Total	**226.6**	**147.88**	

The SFPUC maintains both end-of-use and econometric models to develop water demand projections. Most recently, SFPUC has been using an econometric model (Brattle Group 2018) to generate a point forecast of annual demand for each customer based on projected increases in population growth, employment, income, and the price of water. This model was used to develop a range of projections of annual demand in the service area by 2070. Changes in income and price were coupled with estimations of consumer elasticities of demand to reflect the demand response of customers to these exogenous variables. Figure 3-34 depicts the overall approach to generating all daily demand scenarios used in the stress test.

Once annual demand projections of total demand in the service area and demand on the RWS are established, annual demands on the RWS are disaggregated to a daily time step using the influence of weather and climate through application of heat functions (Maidment and Miaou 1986) and nine synthetically generated sequences of temperature generated by the weather generator (HRG 2018) plus the historical realization. In their review of approaches to modelling the response of water demand to changes in climate and weather, Kiefer et al. (2013) established that the approach provided by

Maidment and Miaou (1986) provides greater flexibility in defining the effects of seasonality relative to other approaches, and thus was chosen by SFPUC as the approach for this work.

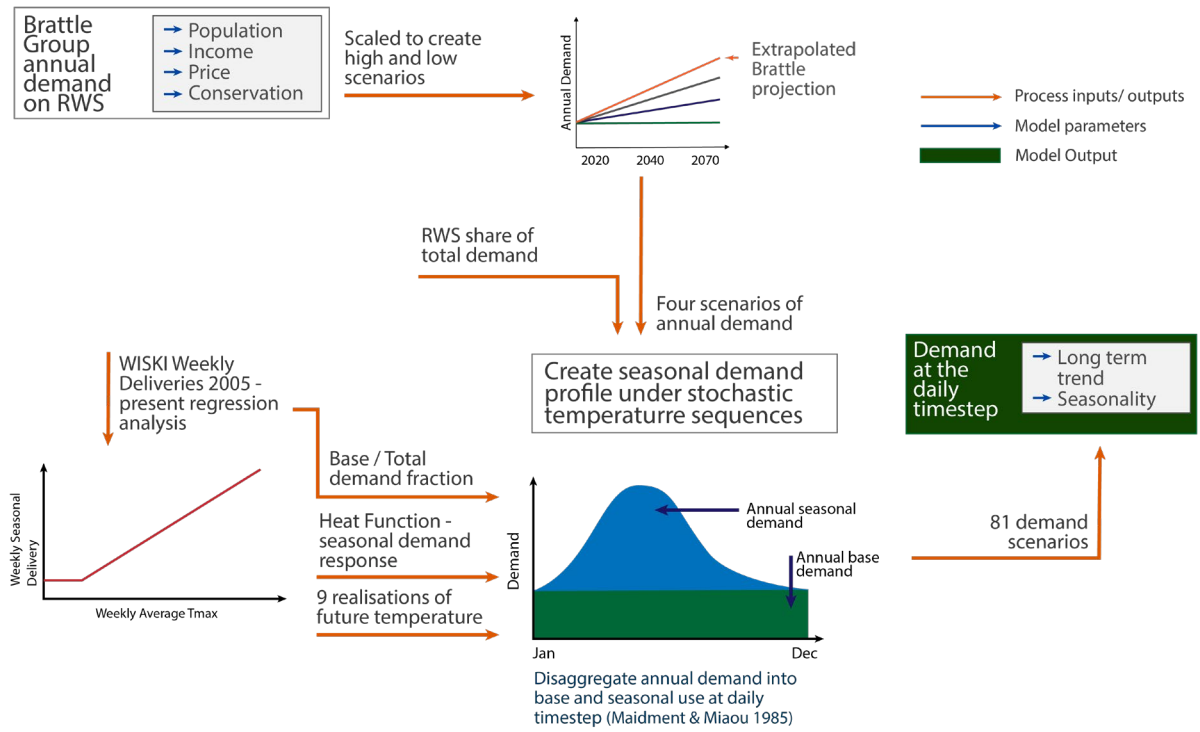


Figure 3-34. Schematic of the Water Demand Model for Generating Daily Demand Scenarios.

The use of heat functions relies on decomposing observed water use into base and seasonal components which represents the proportion that is insensitive to changes in temperature - the base fraction - and that which is sensitive to changes in temperature - the seasonal fraction. The base fraction is generally considered to represent essential uses such as for drinking consumption, washing, toilet flushing etc., whilst the seasonal portion generally represents non-essential uses such as for watering gardens, washing cars etc. The base fraction is assumed constant through the year, while the seasonal fraction varies throughout the year based on temperature. Base fractions for each customer were estimated by SFPUC using historical data from 2005 to 2018. The observed base fraction for each customer in FY2012-13 are shown in Figure 3-35. Higher values of the base fraction, such as in San Francisco, represent a demand with less seasonal fluctuations between winter and summer.

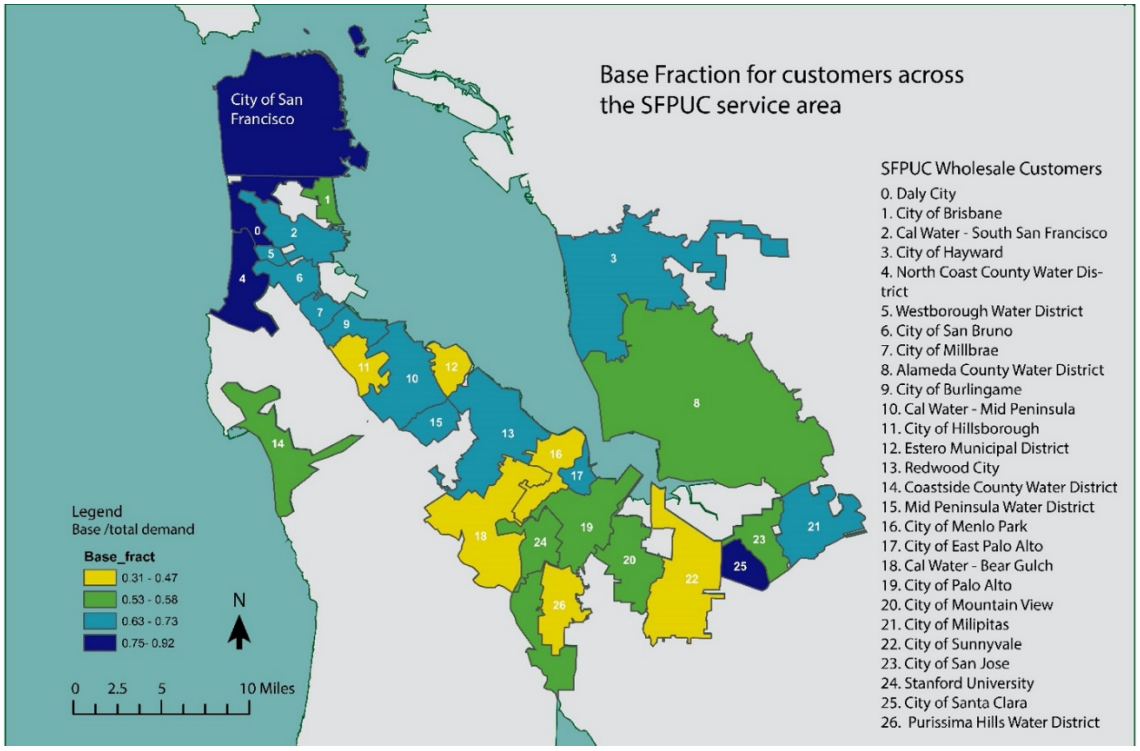


Figure 3-35. Range of Base Demand Fraction (Base Demand over Annual Demand) Values across the SFPUC Service Area for FY 2012-2013.

Observed fluctuations in the seasonal component of demand are related to observed maximum temperature in the same timestep. This observed relationship is used to establish a demand-temperature response function, called a "Heat Function" by Maidment et al. (1985). The change in demand due to temperature is called a Heat Function factor. Heat Function factors were developed based on historical data for each of the three SFPUC delivery centers: the City and County of San Francisco (CCSF), Peninsula (i.e., the Peninsula without CCSF), and South and East Bay. Figure 3-36 shows a slice of the Heat Function factor time series for the three delivery centers of SFPUC for an arbitrary year and illustrates how demand would change through the year. The Heat Function can be used with different climate realizations to establish seasonally varying daily demand downscaled from an annual demand value.

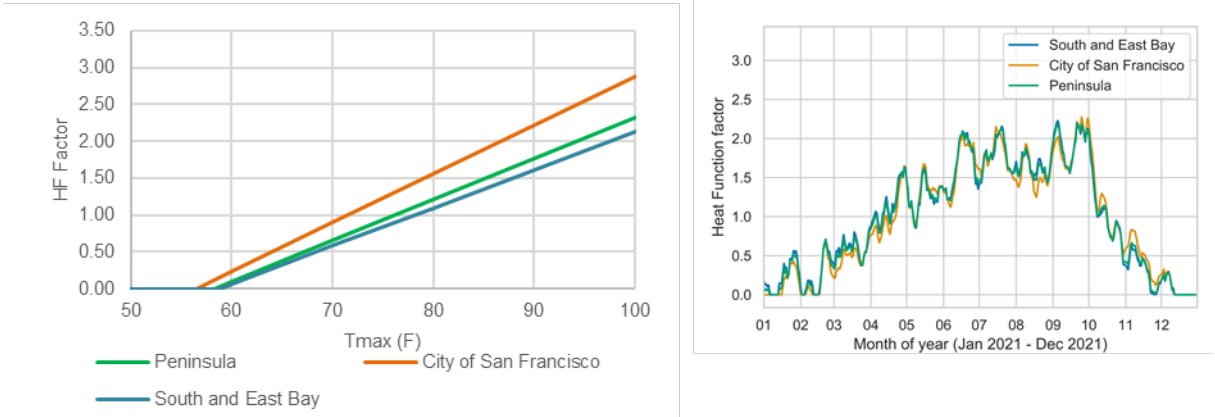


Figure 3-36. Heat Functions for the 3 Delivery Centers (Left), and a Slice of Heat Function Factor for Delivery Centers (Right).

3.3.3 Model Limitations

Figure 3-37 shows the SFPUC service area, the area covered by the weather generator, and the location of the San Jose (SJ) and San Francisco Airport (SFO) temperature gauges, which are the gauges of interest for modelling seasonal demand. As detailed in the Technical Report 3 (HRG 2020a), average daily seasonal demand is scaled using a heat function which relates daily maximum temperature at SJ and SFO temperature gauges to a 'Heat Function Factor' illustrated in Figure 3-36.

As detailed in Section 3.1.4, nine stochastic realizations of temperature were generated for the areas covering the three hydrological basins upon which the RWS relies. The relevant spatial domain for demand modeling is shown in Figure 3-37. While the weather generator outputs temperature data at the grid cell level (shaded area in Figure 3-37), the heat functions illustrated in Figure 3-36 use station data for modeling seasonal demand. For this reason, the temperature at the SJ and SFO gauges are estimated from the nearest available grid cell. For SFO, this grid cell is the one in which the SFO gauge lies; and for SJ gauge, that which is located outside the spatial domain included in the weather generator, this grid cell is located east from its location (i.e., grid cells bordered by the red dashed lines in Figure 3-37).

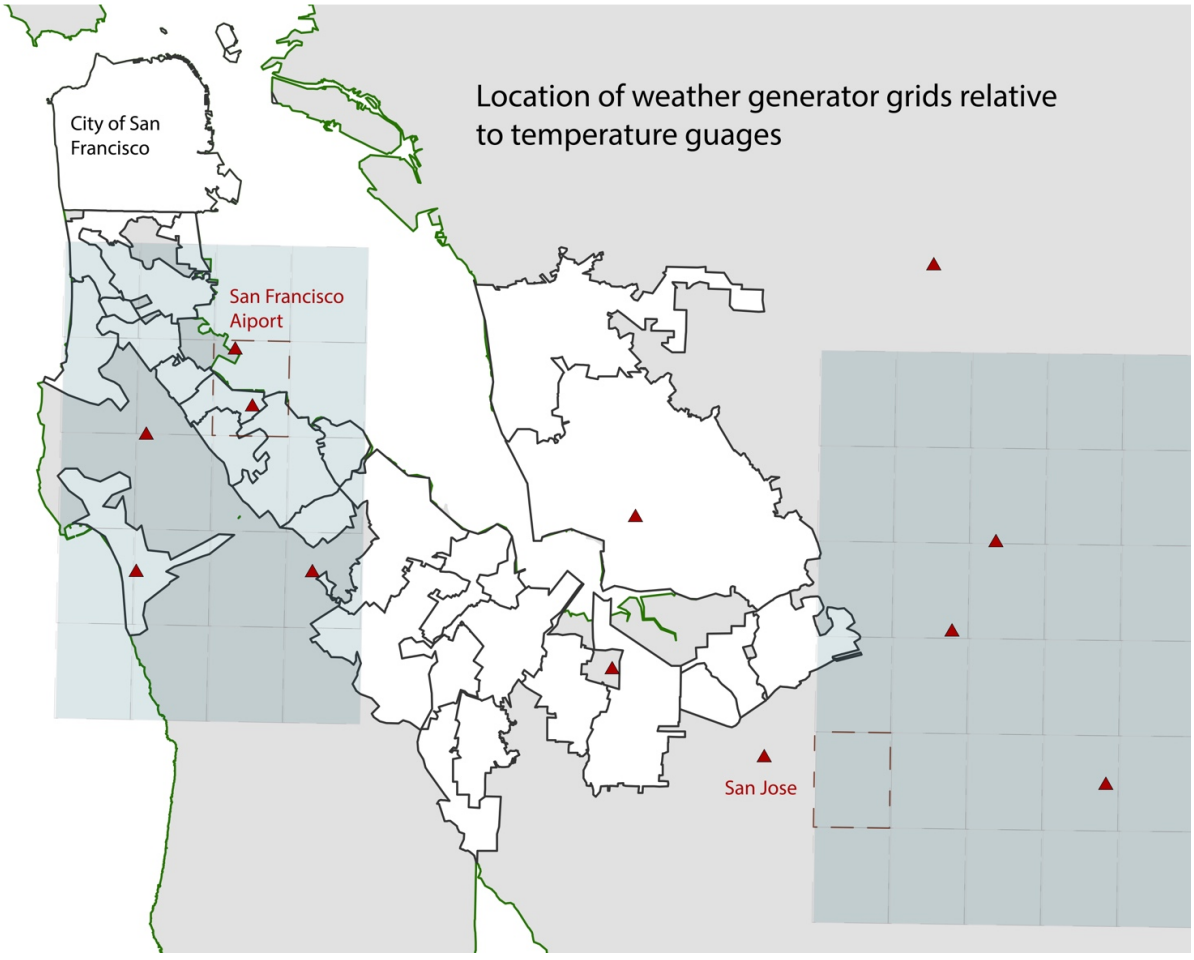


Figure 3-37. Map Showing the SFPUC Service Area.

The location of grid cells where the weather generator generates temperature data for South and East Bay and Peninsula regions (shaded areas). Red triangles indicate the location of temperature gauges across the region. San Francisco Airport and San Jose gauges are those used in the development of Heat Functions that are used to scale daily seasonal demand based on daily maximum temperature. The two grid cells bordered with a red dashed line are those used to establish stochastic realizations of climate from the weather generator.

The linear regressions illustrated in Figure 3-38 were developed to estimate the maximum daily temperature at SJ and SFO from the maximum daily temperature at the above mentioned grid cells and obtained from the stochastic weather generator. However, application of these linear regressions does not allow a good reproduction of the observed distribution at SFO and SJ, as shown in Figure 3-39. A quantile mapping correction was used to correct the identified bias. Figure 3-39 demonstrates that although the distribution of the corrected simulated temperature although (orange lines) is significantly closer to the observed one, the fit with the distribution of the observed temperature at the gauge is not perfect.

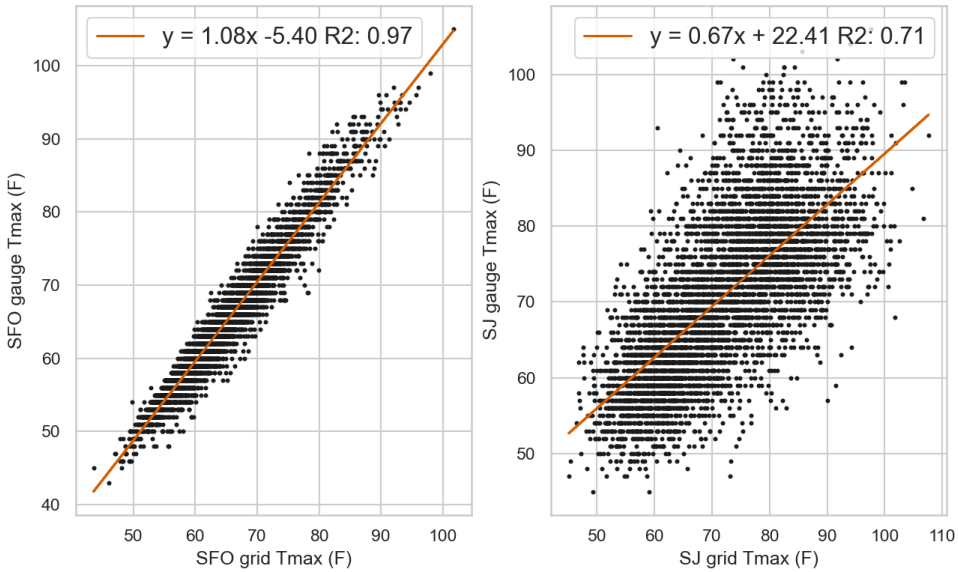


Figure 3-38. Linear Regression Showing the Relationship between the Historical Temperature at SFO and SJ Temperature Gauges vs Historical Temperature for the Nearest Grid Cell.

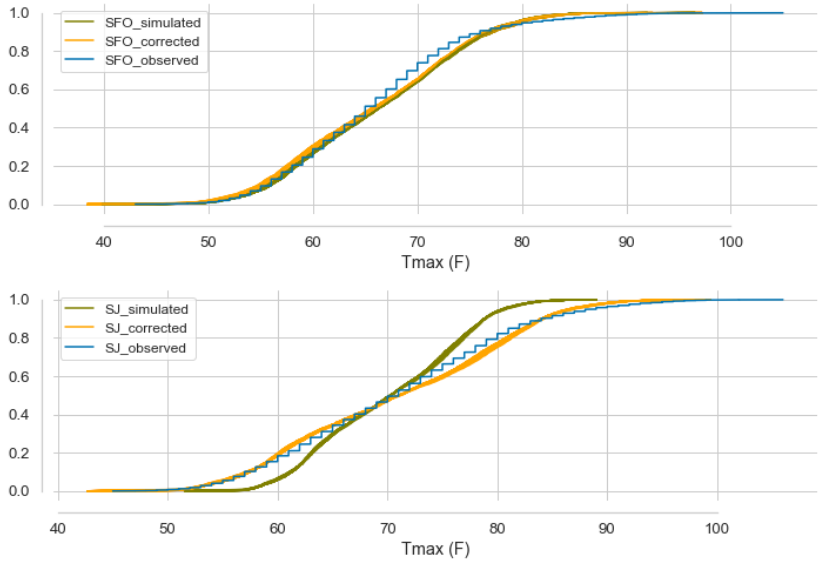


Figure 3-39. Comparison of the Cumulative Distribution Functions Obtained from the Observed Record at the Gauge Station (Blue), the Simulated Temperature at the Gauge Using the Weather Generator Output from the Nearest Grid Cell and the Linear Regression Shown in Figure 3-38 (Green) and the Bias Corrected Simulated Temperature (Orange).

The top and bottom panels show the distributions for SFO and SJ gauge locations, respectively.

This bias propagates through into the application of the heat functions that shape the seasonal portion of demand and ultimately into average daily demand that is fed into SFWSM. For example, the demand 227 mgd demand scenario is in fact 225 mgd on average. This bias is outlined in Table 3-23 which provides for each demand scenario the simulated daily average base, seasonal and total demand together with the error.

Table 3-23. Description of the Bias in the Demand Scenarios.

The average bias across the 10 realizations from the weather generator is shown. The variability across the stochastic realizations is not significant (standard deviation of the error across the 9 realizations is 0.02 mgd, 0.03 mgd, 0.03 mgd and 0.04 mgd, respectively for the demand scenarios 227 mgd, 265 mgd, 300 mgd and 334 mgd.

Demand Scenario for the RWS				Simulated Demand for the RWS after daily disaggregation via the heat functions			Model Error	
Name	Base Demand	Seasonal Demand	Total Demand	Simulated Base	Simulated Seasonal	Total	Absolute (mgd)	Relative (%)
227 mgd	157.37	69.42	226.79	157.37	67.84	225.21	-1.57	-0.69
265 mgd	181.61	83.40	265.01	181.61	81.43	263.04	-1.97	-0.74
300 mgd	204.78	94.80	299.58	204.78	92.56	297.34	-2.24	-0.75
334 mgd	227.96	106.18	334.14	227.96	103.68	331.64	-2.50	-0.75

3.4 San Francisco Water System Model (SFWSM)

3.4.1 Modeling Goals

The main purpose of the water system operations model is to evaluate the impact on water delivery reliability and other performance metrics related to system operations with variations in climate/hydrologic conditions, level of water demand, system configuration (shutdowns, outages, new projects, etc.) and various management policies. Currently, SFPUC uses an Excel VBA-based simulation model of the RWS for planning (HHLSM) that operates at the monthly time step. An objective of the LTVA water system model was to represent system operations in greater detail and accuracy than is possible with HHLSM, by using the daily time step and including a greater number of facilities and spatially disaggregating to all wholesale customers and retail service areas.

3.4.2 Modeling Methods

The core modeling approach of LTVA water system model—called the San Francisco Water System Model (SFWSM)—is to use a linear programming (LP) optimization model run anew each time step. The LP problem is formulated with Pywr, a Python programming language packaged for water system modeling (Tomlinson et al. 2020). The objective of the LP model is to minimize the overall costs associated with water that flows or is stored within the system in each time step, bound by a variety of constraints that represent infrastructure capacities (reservoir capacity, pipeline capacity, etc.) and initial and boundary conditions (initial reservoir storage, preferred storage levels for reservoirs, inflow to reservoirs, etc.), and operational objectives. Operational objectives are achieved by assigning minimum and maximum flow and storage to each flow point and reservoir, respectively, as well as related relative costs of flow/storage, using a variety of input methods and values. In SFWSM, this is implemented with

a set of Python scripts, JSON files, and CSV files that are read and interpreted by Pywr. More details regarding SFWSM are given in the Technical Report 4: San Francisco Water System Model (HRG 2021b).

During model development, three major model versions were developed: a "historical" version to simulate historical operations (from July 1, 2005, to June 30, 2014), a "HHLSM" version to include operations included in HHLSM but not necessarily in the historical record (e.g., Recapture project), and a "Planning" version to include updates to operations not included in HHLSM and to be used for the LTVA. The Planning version was modified to represent narratives (see Section 4.3). The differences among these three versions are summarized in Table 3-24 while the main differences between HHLSM and the HHLSM validation version of SFWSM are presented in Table 3-25.

Table 3-24. Key Differences among SFWSM Versions.

	Historical validation	HHLSM validation	Planning
Recapture Project (including Pit F2)	NO	YES	NO
ACDD Tunnel	NO (Calaveras construction period)	YES	YES
Historical shutdown of WTPs & SJPL for maintenance	YES	NO	NO
Max and min flows through the WTPs	Same for all three version		
Maintenance schedule for SJPL and BDPL	NO	YES	YES
Storage reduction at CA, CH, CS due to construction and/or maintenance	YES	NO	NO
CS storage reduction for Fountain thistle	YES	NO	YES
Post-WSIP UTREP Snowmelt Management Program below HH	YES	YES	YES
Tuolumne River Transfer to Water Bank	NO	YES	NO
San Francisco Groundwater Project (1 MGD to SF City)	YES	NO (HHLSM demand is reduced instead)	YES
Groundwater Storage and Recovery Project (7.2 MGD during dry year)	YES	YES	YES
San Francisco Westside Recycled Water Project (2 MGD of recycled water in San Francisco)	NO	NO	YES
IFR below O'Shaughnessy Dam	YES	YES	YES
IFR below Cherry Valley Dam	YES	YES	YES
IFR below Lake Eleanor Dam	YES	YES	YES
IFR at La Grange from SFPUC (i.e., below Don Pedro Dam)	NO	NO	Included as narrative only
IFR below San Antonio Dam	NO	NO	Included as narrative only
IFR below Calaveras Dam	NO	YES	YES
IFR below Crystal Springs Dam	YES	YES	YES
IFR below Pilarcitos Reservoir (Stone Dam)	YES	YES	YES
IFR San Andreas Dam	NO	NO	NO
Drought Rationing	NO (historical deliveries are used as demand)	YES	YES
TID/MID rationing as function of DP storage & inflow forecast	NO (historical deliveries are used as demand)	YES	YES
Modelled demand	NO (historical deliveries are used as demand)	YES	YES

Table 3-25. Key Differences between HHLSM and Both Planning and Validation Versions of SFWSM.

Characteristic/component	HHLSM	SFWSM (validation)
Model platform/language	Excel with VBA	Python with Pywr
Model method	Traditional simulation	Linear programming-based simulation
Time step	Monthly	Daily
SFPUC customers	Five demand centers	36 demand centers (26 wholesale customers; City & County of SF; 9 suburban retail customers)
Upper Tuolumne River Ecological Program	Not included	Included
Rafting flows	Not included	Included
Emergency storage in the Bay Area	Not included	Included (120 days)
5% minimum demand reduction for SFPUC during drought rationing	Not included	Included
Anticipate SJPL maintenance by filling San Antonio with HH water	Not included	Included
Lower Cherry Aqueduct	Included	Not included
SJPL to SVWTP	Included	Not included

Inputs to and outputs from SFWSM are summarized in Figure 3-40. Inputs to the system model include a range of parameters related to initial conditions (reservoir storage, groundwater storage, etc.), boundary conditions (inflow hydrology, water demand, etc.) and operating rules (drought policy, flow requirements, reservoir preferred storage levels, etc.). Outputs include reservoir state (storage for reservoirs, the water bank, and groundwater), system flows in conveyances (pipelines, etc.) and other facilities (e.g., water treatment plants and hydropower plants) and deliveries (to water customers, instream flow requirement locations, etc.). Technically, deliveries are a subclass of flows, since any delivery is recorded as a flow through a delivery point. Each of the outputs listed in Figure 3-40 are time series. Aggregated outputs, such as performance metrics and policy violation checks, can be calculated from these model outputs.

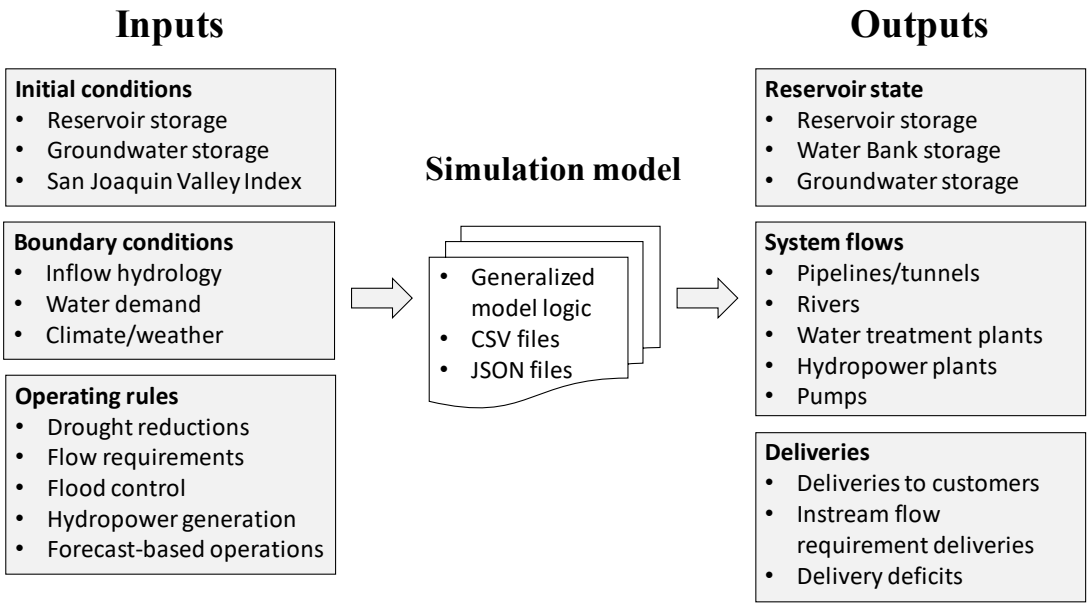


Figure 3-40. Inputs to and Outputs from the San Francisco Water System Model.

The SFWSM includes a detailed but still simplified representation of system facilities and operates at the daily time step. Most dominant system facilities are included (e.g., reservoirs, aqueducts, treatment plants, pump stations, etc.), with some aggregation/simplification as needed and able to help reduce computation time. The daily time step was selected based the need for greater detail. The following operational aspects are included in SFWSM:

- Urban water demand (retail and wholesale service areas)
- Agricultural water canal diversion for Modesto and Turlock Irrigation Districts
- Reservoir operations
- Instream flow requirements
- Drought policy (drought management, and more specifically demand restriction during drought is described below).
- Specific supporting programs and projects, including:
  - SFPUC Water Bank
  - Upper Tuolumne River Ecological Program (UTREP) snowmelt management flows
  - Alameda Creek Recapture Project (ACRP)
  - Regional Groundwater Storage and Recovery Project
  - San Francisco Groundwater Project
- Streamflow forecasts using perfect foresight
- Numerous miscellaneous operational objectives such the objectives of having 120 days of emergency storage in the South East Bay and Peninsula areas, and the water level constraint at Crystal Springs reservoir to protect the endangered fountain thistle species.
- "Virtual water" (see below)

Don Pedro Reservoir is included to model deliveries to Modesto and Turlock Irrigation Districts, who have higher priority rights than SFPUC for water from the Tuolumne River, and for compliance with flow requirements downstream of Don Pedro, for which SFPUC is partially responsible. Don Pedro Reservoir levels are also required for estimation of the lake evaporation that is needed for the calculation of the evaporation from the water bank.

The most important operational policies related to water quantity are included in SFWSM ("water first, power second," etc.), as appropriate for the spatial and temporal scope. Most operations from HHLSM are included.

Virtual water (the last item in the list above) is water that can enter the RWS in the system model if there are no other alternatives given the constraints of the model. Virtual water is often used as a modeling "trick" when working with LP models to prevent infeasibility conditions: without virtual water, LP models can completely fail under some conditions. However, virtual water can also provide insight into where, when, how frequently, and how much additional water may be needed to satisfy delivery shortfalls after accounting for demand rationing. Virtual water is included in the water system model at several locations in the system (i.e., entrance of the Mountain Tunnel, entrance of the Sunol Valley and Harry Tracy water treatment plants, and at existing interties with neighboring water agencies). Virtual water is tracked as an output to analyze deficit in water supply.

Allocation of shortages to retail and wholesale customers during droughts are determined by the Water Shortage Allocation Plan described in the Water Service Agreement (WSA) between SFPUC and wholesale customers (SFPUC 2009). The WSA defines how shortages may be imposed, including using either voluntary (preferable) or mandatory (if needed) reduction measures. It describes how shortages are allocated between SFPUC and the wholesale customers as a group (Tier 1 split) and among wholesale customers (Tier 2 split). Shortages result in percentage use reduction requirements on the

part of wholesale customers for shortages up to 20%. In SFWSM, allocation for SFPUC retail customers is split among CCSF and the sur-urban retail customers. As the WSA does not provide guidance for splitting the allocation between CCSF and sur-urban retail customers, the long-term average contribution from each customer to total SFPUC retail demand is used. This calculation makes CCSF take almost 95% of the allocation for SFPUC retails customers. The second largest allocation is for the San Francisco International Airport with around 1.5% (HRG 2021b).

Demand reductions are accounted for in the system model in three parts.

- Calculate reduction targets. This involves two parts:
  - Calculate projected water supply.
  - Use a lookup table to find demand reductions based on projected water supply value triggers.
- Calculate Tier 1 split based on the WSA.
- Calculate Tier 2 split based on the WSA.

This method for projecting water supply is not explicitly defined in the WSA, but is instead described generally as: "the determination of projected available water supply shall consider, among other things, stored water, projected runoff, water acquired by the SFPUC from non-SFPUC sources, inactive storage, reservoir losses, allowance for carryover storage, and water bank balances, if any…." (SFPUC 2009). Based on discussions with SFPUC, the number of years of water supply remaining in storage system-wide, including in the Water Bank, was selected as the projected water supply metric to use to trigger drought rationing. This metric, called "years of remaining supply" (YRS), includes July 1st reservoir storage and predicted demands for municipal supply and instream flows and evaporation losses. It provides an objective and easily understood measure of the supply of water. As new sources of supply or new demands occur, the metric remains appropriate. The YRS metric is calculated as:

$$\boldsymbol{YRS} = \frac{\boldsymbol{Total\ Active\ Storage}}{\boldsymbol{Total\ Demand}} \qquad (3\text{-}11)$$

Total active storage is total actual storage less the total of each reservoir's inactive zone. Total demand is calculated as:

$$\boldsymbol{Total\ Demand} = \sum \boldsymbol{Urban\ Demand} + \sum \boldsymbol{IFR} + \sum \boldsymbol{Reservoir\ Evap}, \qquad (3\text{-}12)$$

where IFR is instream flow requirements. The IFRs included in Equation 3-12 include the IFRs that are not collected by a dam downstream (i.e., below Calaveras, Crystal Springs and Stone Dam in current conditions but also San Antonio and New Don Pedro in plausible narratives [Section 4.3.1]). YRS is calculated once per year, on July 1st, based on the observed demand and reservoir evaporation from the previous year (from July 1 of the previous year through June 30 of the current year). Thus, YRS is an approximate value based on recent observations, rather than on demand and/or hydrologic projections.

The approach of using YRS is different than the one in HHLSM, which uses total system storage directly. However, total system storage does not account for changing demand when projecting supply, and was therefore deemed insufficient for modeling changes in long term demand, a critical component of the LTVA. Demand reduction factors and types (voluntary or mandatory) and their corresponding YRS trigger values are listed in Table 3-26, along with Tier 1 splits and established total system storage trigger values from HHLSM planning simulation of WSIP 2018[6] infrastructure program with an annual demand of 265

[6] Water System Improvement Program (WSIP); ONESF, n.d.

mgd. These WSIP 2018 storage triggers were used to calculate the YRS thresholds using HHLSM simulation results for urban demand, IFRs and reservoir evaporation. For example, when the YSR calculated in a simulation drops below 3.4 years system wide deliveries are reduced by 10%. Voluntary reductions are modeled as mandatory reductions.

Table 3-26. Demand Reduction Factors (DRF) with Corresponding Exceedance (Inclusive) Trigger Years of Remaining Supply (YRS) Thresholds and Historical Total System Storage.
"Voluntary" reductions are assumed as mandatory in SFWSM**.**

Trigger	YRS trigger (years)	Demand Reduction Factor	Reduction type	Tier 1 split (Retail/Wholesale, %)	Hist. total system storage trigger (TAF)
Trigger 1	3.6	0%	N/A (groundwater use)	N/A	1285
Trigger 2	3.4	10%	voluntary	37/63	1100
Trigger 3	3.1	20%	mandatory	37.5/62.5	900

The Tier 1 split for allocating reductions (Step 2) depends on YRS. If 3.4>YRS ≥ 3.1 (Demand Reduction Factor = 10%), then the Tier 1 split is 37% of available supply for SFPUC and retail customers, and 63% for BAWSCA customers. If YRS<3.1, the split is 37.5% for SFPUC and retail customers, and 62.5% for BAWSCA customers. When 3.6≥YRS >3.4, no reduction is implemented but the use of the Groundwater Storage & Recovery Project (South Westside Basin) is allowed. Note that a minimum reduction is imposed to SFPUC and its retail customers. Tier 2 splits are as listed in Table 3-27.

The application of demand reduction differs slightly between wholesale customers and in-city retail customers. As described in Section 3.3, daily demand is partitioned into base (non-variable) demand and seasonally variable demand. As a first approximation, demand reduction during droughts are applied to base demand for San Francisco city and to the seasonal demand for wholesale customers and suburban retail customers. If the reduction of the seasonal demand is not sufficient to match the requested reduction, the base demand is reduced accordingly.

Table 3-27. BAWSCA Tier 2 Splits for Distributing Tier 1 Drought Allocations.

BAWSCA Customer	Share
ACWD	7.6%
Brisbane GVMID	0.5%
Burlingame	2.9%
Coastside CWD	1.2%
CWS Bear Gulch	6.5%
CWS Mid-Peninsula District	8.1%
CWS South SF	4.6%
Daly City	2.7%
East Palo Alto	1.4%
Estero MID Foster City	3.2%
Hayward	13.2%
Hillsborough	1.7%
Menlo Park	1.9%
MID Peninsula WD	2.1%
Millbrae	1.6%
Milpitas	4.6%
Mountain View	6.2%
North Coast County WD	2.0%
Palo Alto	7.8%
Purissima Hills WD	0.8%
Redwood City	5.8%
San Bruno	1.5%
San Jose	2.1%
Santa Clara	2.1%
Stanford	1.5%
Sunnyvale	5.8%
Westborough WD	0.6%
TOTAL	*100%*

Source: Courtesy of SFPUC.

The result of the above-described drought management is illustrated in Figure 3-41. It is noted that delivery rationing (red) is triggered when the years of remaining supply gets lower than the thresholds given in Table 3-26. On this figure, which has been obtained using SFWSM (HHLSM validation) and the climate realization number 2, three droughts are present; one 4-yr long drought (204-2037) during which 10% rationing was used for three years and 20% rationing for the last year; another 2-yr long drought (2046-2047) during which 10% and then 20% rationing was trigged, and finally another 2-yr long drought (2060-2061) during which 10% rationing was applied.

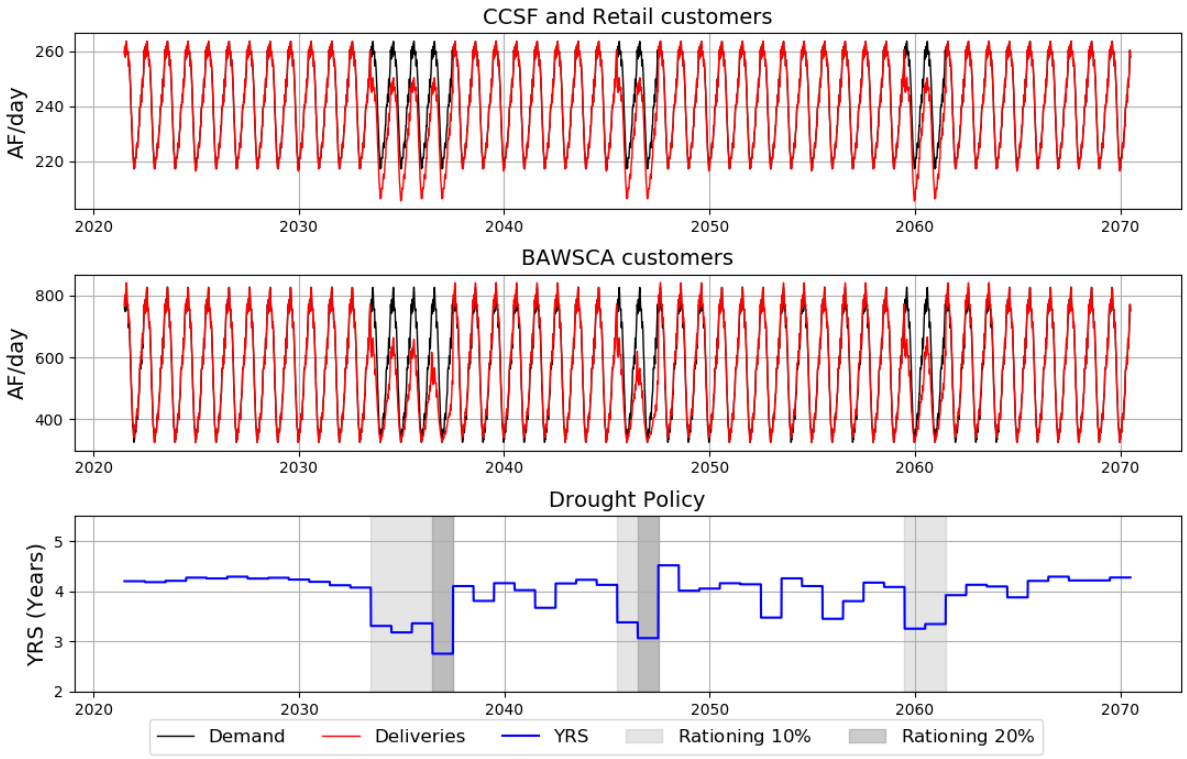


Figure 3-41. Illustration of the Demand Rationing during Drought.
Top and middle panels show the water demand (black) and deliveries (red) to San Francisco plus sub-urban retail customers (Top) and BAWSCA wholesale customers (Middle). The bottom panel shows the year of remaining supply (blue) and the rationing (grey bars). These simulation results are obtained using the SFWSM (HHLSM validation) model and with the climate realization 2. Dates assigned to the time series are for illustration only. No change in precipitation and/or temperature is included in this example.

Following what SFPUC does in practice, the rationing by either 10% or 20% is applied by SFWSM on the annual demand experienced the previous year. The reason is that at the time the decision to ration deliveries is made, on July 1st, the demand for the year to come is unknown, especially the seasonal component of the demand as it depends on the local temperatures for the year to come. As such, if the decision is made to trigger a 10% rationing at system scale, the actual reduction can actually be slightly more or less than the targeted reduction when compared with the demand that would have materialized without rationing.

Table 3-28 summarizes the efficiency of this rationing mechanism by showing the simulated deliveries, demand and actual delivery reduction for each rationing period simulated using the 10 realizations used for the climate stress test. Only baseline climate and demand conditions are considered for the illustration.

Table 3-28. Characteristics of the Rationing Applied by SFWSM under Baseline Climate and Demand Conditions.

Date end	Realization number	Simulated deliveries (TAF)	Demand without rationing (TAF)	Actual reduction (%)	Targeted reduction (%)
7/1/2049→6/30/2049	1	202	227	11.098	10
7/1/2035→6/30/2035	2	202	223	9.168	10
7/1/2037→6/30/2037	2	185	222	16.681	20
7/1/2047→6/30/2047	2	181	228	20.51	20
7/1/2061→6/30/2061	2	207	222	6.919	10
7/1/2043→6/30/2043	3	203	222	8.906	10
7/1/2044→6/30/2044	3	178	224	20.392	20
7/1/2046→6/30/2046	3	179	219	17.99	20
7/1/2028→6/30/2028	4	204	224	9.009	10
7/1/2065→6/30/2065	4	187	229	18.277	20
7/1/2066→6/30/2066	4	183	223	17.678	20
7/1/2068→6/30/2068	4	207	225	7.853	10
7/1/2069→6/30/2069	4	181	227	20.038	20
7/1/2040→6/30/2040	5	179	225	20.327	20
7/1/2038→6/30/2038	7	203	234	12.866	10
7/1/2039→6/30/2039	7	187	224	-0.16259	20
7/1/2052→6/30/2052	8	205	224	8.351	10
7/1/2060→6/30/2060	8	182	228	20.154	20
7/1/2061→6/30/2061	8	183	218	15.865	20
7/1/2062→6/30/2062	8	175	222	21.095	20
7/1/2063→6/30/2063	9	206	226	9.195	10
7/1/2064→6/30/2064	9	181	229	21.058	20
7/1/2037→6/30/2037	10	196	219	10.64	10

3.4.3 Comparison of SFWSM (Historical Validation) Against the Historical Operations of the Hetch Hetchy RWS

This section includes a series of representative validation figures. More discussion and qualitative assessments of model performance compared with historical operations are available in the Technical Report 4 (HRG 2021b). The assessment covers the most important operations across the region, system-wide operations, and selected operations in each region.

Comparisons of major operations system-wide are shown in Figure 3-42, including total system storage (Figure 3-42a), total storage in each region (Figure 3-42b-d), flow in San Joaquin Pipelines (Figure 3-42e), and outflow from each WTP (Figure 3-42f-g). Given the complexity of the RWS system and management, all system-wide operations are reasonably well simulated. Both total system storage and Upcountry storage include the Water Bank. Because the extra flood space available for the Water Bank (the bubble account) is not used in SFWSM, which results in peak storage less than observed. Storage otherwise generally follows observed historical trends.

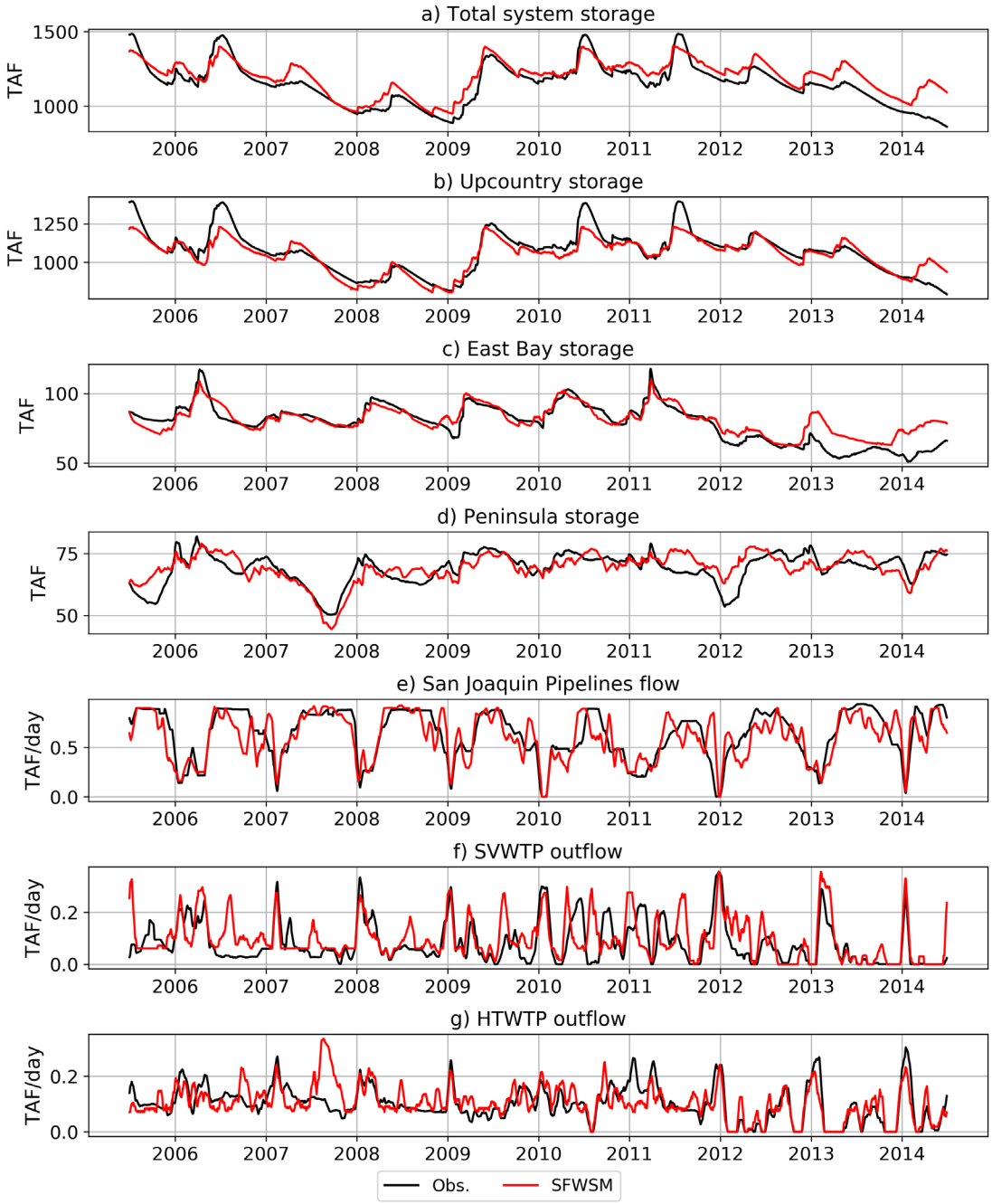


Figure 3-42. Observed and Simulated Major Operations System-Wide.
Operations with flow are with a 20-day moving average.

Generally, flows in SFWSM that are shown in Figure 3-42 are more variable than observed. This is particularly apparent in the San Joaquin Pipelines (SJPL). This follows from the inherent linear nature of linear programming, which results in discrete jumps from one state to another. Flow through the SJPL and outflow from the treatment plants are well represented. More details regarding the comparison are found in Technical Report 4: San Francisco Water System Model (HRG 2021b).

Figure 3-43 compares observed and simulated daily storage for each reservoir in SFWSM, including the Water Bank. Most reservoirs are reasonably well simulated with both seasonal and inter-annual

variability being well reproduced by the model. This is especially true for the Upcountry reservoirs and the Water Bank. Again, it is noted that the high storage portion (also known as the bubble account) of the Water Bank was not considered important to represent in SFWSM. For the purposes of the LTVA, reservoir behavior was deemed acceptable, as all important reservoir operations are well simulated by SFWSM. San Antonio and San Andreas Reservoirs are the least well represented. More detailed discussions are provided in Section 5.2 of the Technical Report 4 (HRG 2021b).

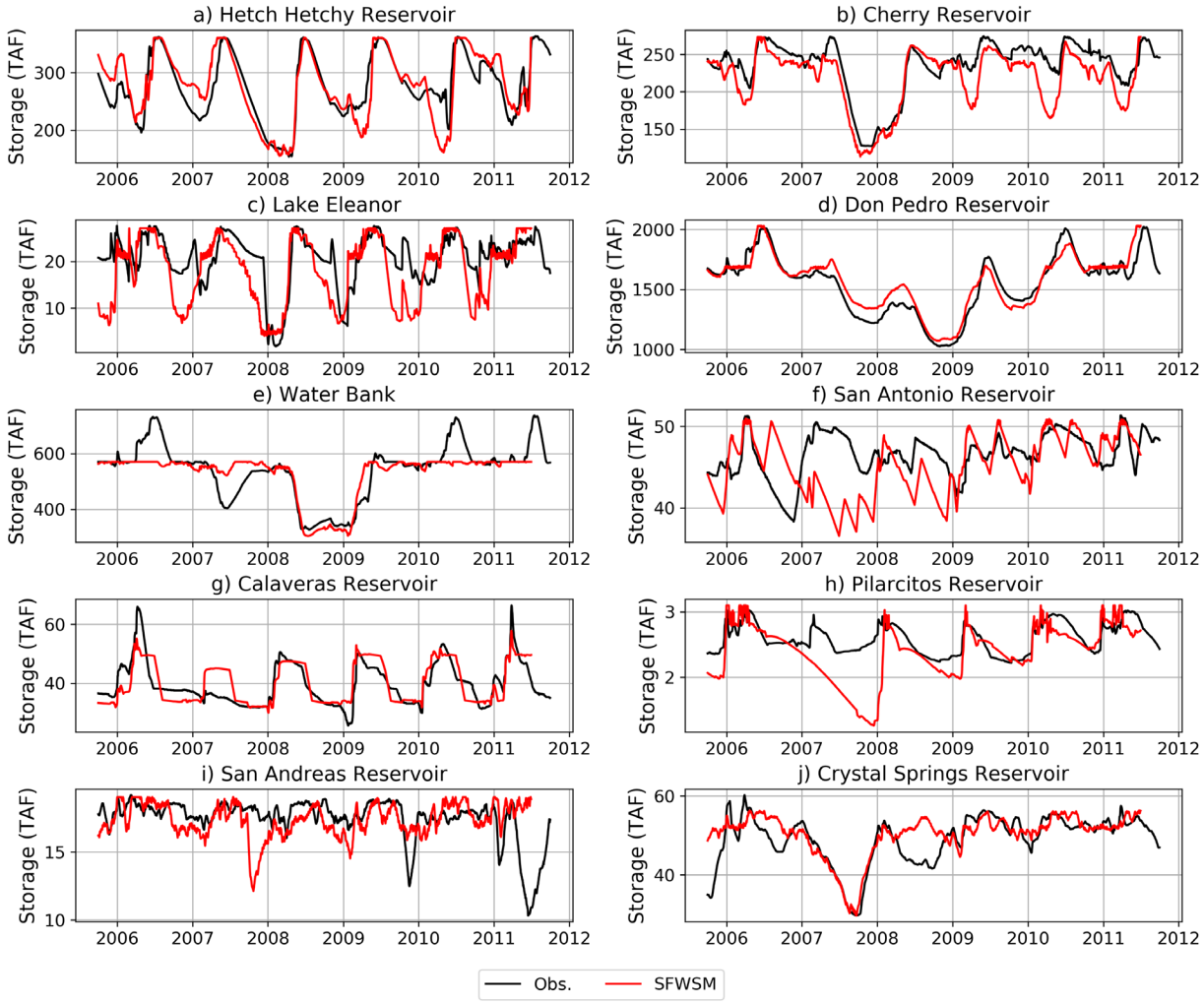


Figure 3-43. Observed and Simulated Daily Storage for All Reservoirs, Including the Water Bank.

3.4.4 Comparison of SFWSM (HHLSM Validation) and HHLSM

It is worth noting differences between HHLSM (i.e., version WSIP 2018 265) and the HHLSM version of SFWSM. The purpose of this comparison is to show that SFWSM operates generally in a similar manner to HHLSM. Note that SFWSM (HHLSM validation) does not use observed streamflow and water demand time series as input. While HHLSM uses reconstructed monthly streamflow from 1920 to 2011 and the same monthly pattern for demand, which repeats year on year. SFWSM uses simulated streamflow values obtained as output of the weather generator (Section 3.1), hydrological models (Section 3.2). Regarding urban water demand, the HHLSM version of SFWSM uses times series for each wholesale and retail customer that were developed so that SFWSM demand aligned with HHLSM demand at the delivery center scale (i.e., Peninsula, South East Bay and City of San Francisco; see Section 4.2 of Technical Report 4 (HRG 2021b). As such, comparison of SFWSM (HHLSM validation) and HHLSM models

cannot be carried out by comparing simulated time series of reservoir storage or flow through the main conveyances, as in Figure 3-42 and Figure 3-43, for example. Instead, results obtained from SFWSM (HHLSM validation) are averaged across the 9 historical realizations presented in Section 3.1.4. Figure 3-44 compares reservoir storage for all 9 reservoirs plus the Water Bank between HHLSM and SFWSM, with SFWSM results averaged across 9 historical (no change in precipitation or temperature) climate realizations. Figure 3-45 shows a similar comparison for flow in the San Joaquin Pipeline (SJPL). More details regarding the comparison are found in Technical Report 4: San Francisco Water System Model (HRG 2021b).

The most important and largest reservoirs—the upper six reservoirs in Figure 3-44, which includes the Water Bank—operate broadly consistently between both HHLSM and SFWSM. When comparing the variability of storage in each reservoir, as indicated by the light shaded areas that highlight the deviation between the minimum and maximum values, it is noted that SFWSM results in a greater range in storage compared to HHLSM in both smaller reservoirs and the Cherry/Eleanor Reservoirs and, to a lesser degree, Don Pedro Reservoir. Flows in the SJPL are also generally consistent between the two models (Figure 3-45), though, as with some reservoirs, SFWSM operates with a somewhat greater range than HHLSM. These differences generally—including the range differences—result from both slight differences in operations and differences in the hydrologic inputs into the respective models; SFWSM uses 9 generated realizations, which includes more extreme events compared to the historical record used in HHLSM (see Section 3.1.4).

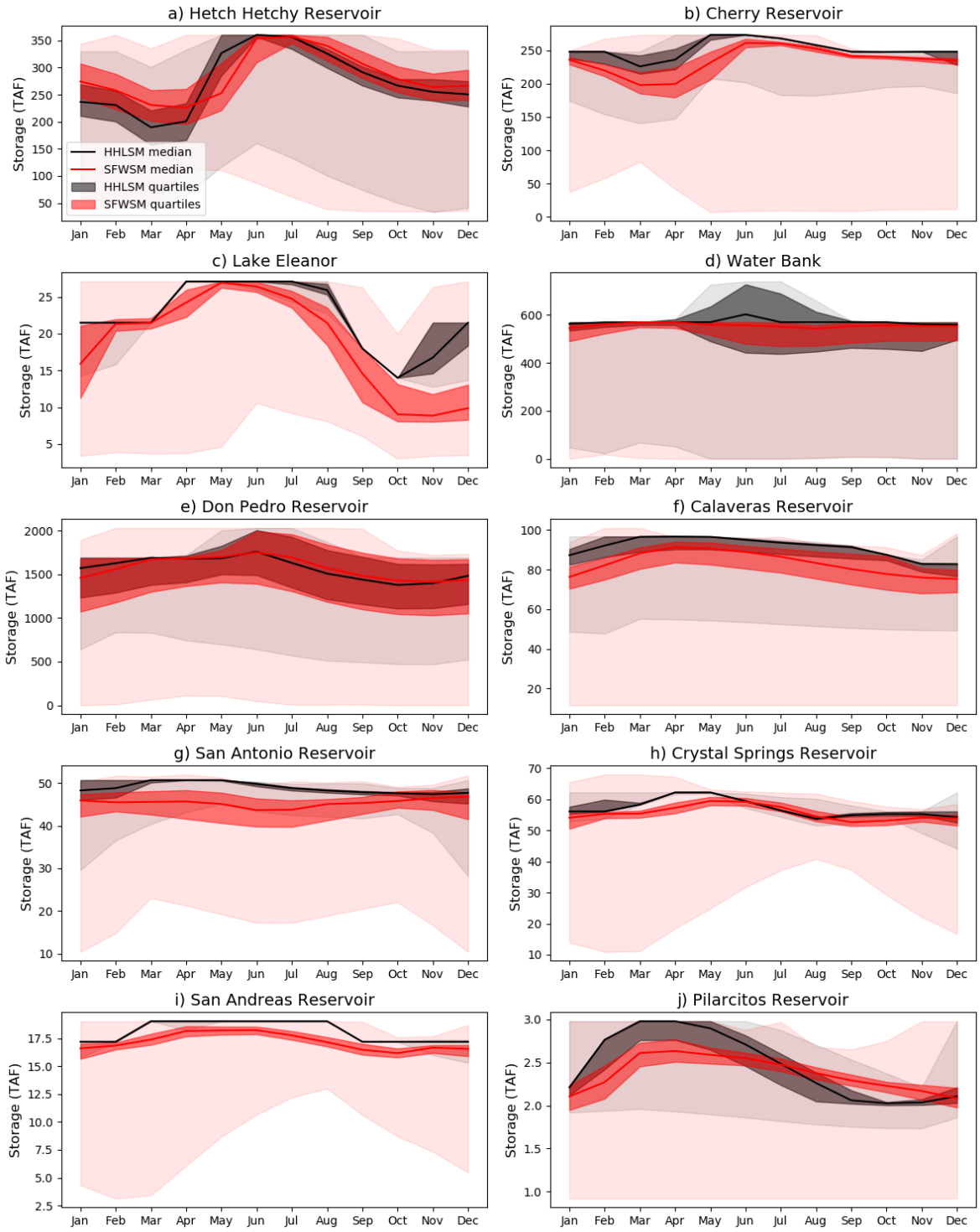


Figure 3-44. Comparison between Simulated Reservoir Storage for HHLSM and SFWSM.
HHLSM uses monthly time step. SFWSM simulations are aggregated to the monthly time step for comparison. The bold curves show the median values obtained for the period 1920-2011 for HHLSM and across the 9 50-year climate realizations presented in Section 3.1.4. Dark shaded areas show the deviation between the 25th and 75th percentiles. Light shaded areas show the deviation between the minimum and maximum values.

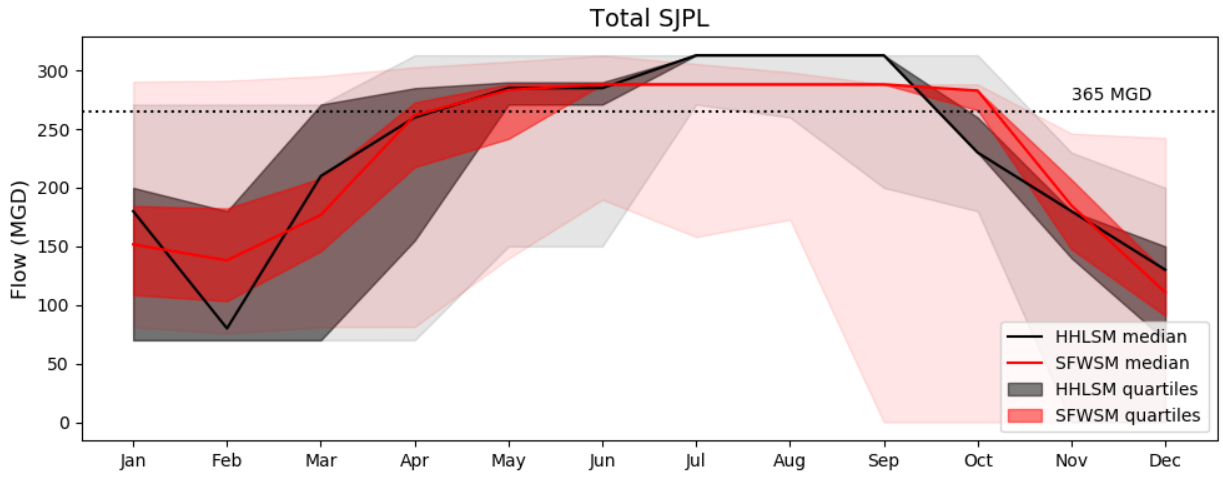


Figure 3-45. Comparison of Monthly San Joaquin Pipeline (SJPL) Flow for HHLSM and SFWSM.
HHLSM uses monthly time step. SFWSM simulations are aggregated to the monthly time step for comparison. The bold curves show the median values obtained for the period 1920-2011 for HHLSM and across the 9 50-year climate realizations presented in Section 3.1.4. Dark shaded areas show the deviation between the 25th and 75th percentiles. Light shaded areas show the deviation between the minimum and maximum values.

3.4.5 Comparison of SFWSM (HHLSM Validation) and SFWSM (Planning)

For the application of SFWSM to the LTVA (i.e., for the stress tests and narratives in this report), the Planning version of SFWSM was used. This version involved modifying several operational details found in the HHLSM version. While specific operational differences are described in detail in the Technical Report 4 (HRG 2021b), Figure 3-46 illustrates some of the impacts on reservoir storage operations for one arbitrarily selected climate realization (R4). Major operational differences include Cherry, San Antonio, and Crystal Springs Reservoirs, each of which operate at lower reservoir levels than in the HHLSM version during some or all years. Crystal Springs Reservoir is explicitly operated to a lower level to protect the fountain thistle along the reservoir banks. Cherry Reservoir is also operated at lower elevation during fall and winter seasons following the updated operations to limit chances of spillway flow. Finally, following discussion with SFPUC, the Recapture project was not included in the planning version SFWSM, which reduces the filling of San Antonio. More details regarding the comparison are found in Technical Report 4: San Francisco Water System Model (HRG 2021b).

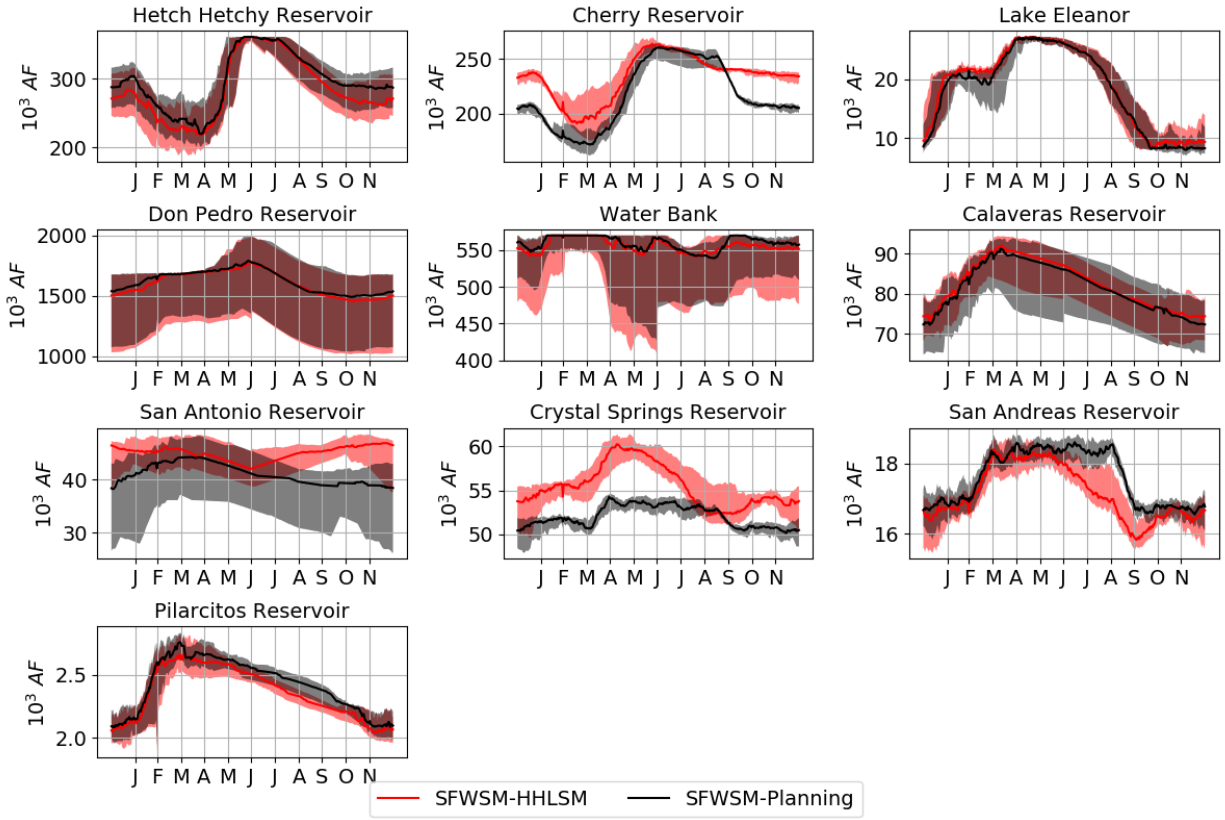


Figure 3-46. Median and Range of Simulated Reservoir Storage for the HHLSM (Red) and Planning (Black) Version of SFWSM.

Climate realization number 4 is used to force both models.

3.4.6 Model Limitations

SFWSM is a planning model. As such, several simplifications were made to represent the Hetch Hetchy RWS, such as the aggregation of several pipelines and the simplifications of some operations. Some infrastructures have not been accounted for, such as Priest Dam and the Cherry Lower Aqueduct.

Cordilleras Mutual Water Company is wholesale customer of SFPUC that is not part of BAWSCA. As such, Cordilleras is not part of the Tier 2 split (Table 3-27). In agreement with SFPUC, Cordilleras Mutual Water Company is deemed a sub-urban retail customer in SFWSM.

To reduce the computational burden, deterministic streamflow forecasts are used to manage operations of several reservoirs. This is especially important for operations of Upcountry reservoirs for which inflows are significantly driven by snowpack dynamic. Today, operations of Upcountry reservoirs are facilitated by frequent observations of snowpack, which can provide a reliable estimate of the expected volume during spring runoff. However, under warming conditions, the snowpack accumulation during winter months will be less, which will force the operators of the Upcountry reservoirs to rely more on seasonal precipitation forecasts that are less accurate than snowpack-based forecasts. In SFWSM, the use of deterministic forecasts does not account for the reduction of forecast skill the operators will have to cope with in a warmer future.

Regarding the rationing during droughts, the SFPUC Tier 1 allocation in SFWSM is 37% when 10% rationing is trigged at system scale (Table 3-26). However, following the Water Shortage Allocation Amendment, this number should have been 36%. Figure 3-47 illustrates the error in actual delivery

reduction caused by using 37% instead of 36% for the SFPUC Tier 1. With an average error roughly equal to 0.2% and a maximum error lower than 1%, the overall error is deemed insignificant and does not impact the conclusion of the LTVA.

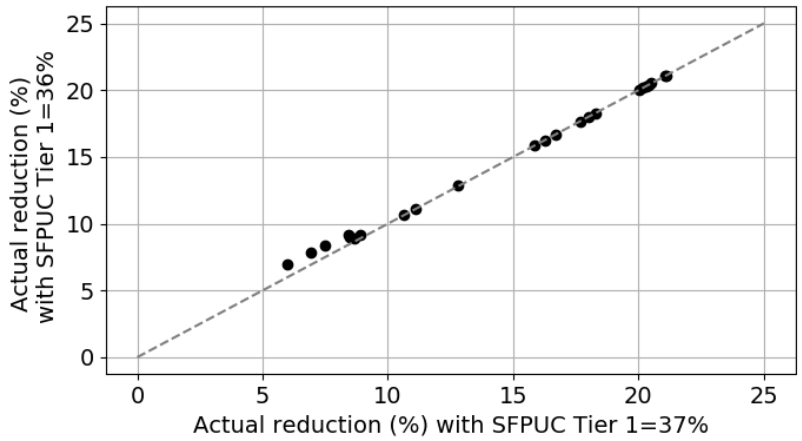


Figure 3-47. Comparison of the Actual Reduction in Deliveries When SFPUC Tier 1 Equals Either 37% (X-axis) or 36% (Y-axis).

3.5 Raw Water Quality in the Hetch Hetchy Supply

3.5.1 Modeling Goals

The RWS mainly relies on the Hetch Hetchy Reservoir supply for which SFPUC holds a filtration avoidance permit from the U.S. EPA. For planning purposes, SFPUC wants to understand drivers of raw water quality in order to assess change in frequency of water quality events due to changes in climate conditions and operating rules, as this can have significant impact on RWS operations.

Currently about 85% of SFPUC's water supply is unfiltered water originating from the Upper Tuolumne River captured in Hetch Hetchy Reservoir. To maintain the filtration avoidance permit, water needs to meet criteria stipulated in Title 22 of the California Code of Regulations 64652.5 (22CCR §§ 64652.5) relating to total coliform and turbidity. The compliance point in the system is Tesla Portal, where SFPUC operates a treatment plant for water disinfection. Losing the filtration avoidance permit is a vulnerability for the RWS as it has a high adaptation cost, necessitating the need to understand future water quality possibilities.

In addition, since filtration is not required prior to disinfection at Tesla Portal, the natural organic matter (or total organic carbon, TOC) in the unfiltered water can combine with the chlorine disinfectant, which can lead to elevated disinfection byproducts (DBPs) in the RWS downstream of Tesla Portal. While usually not a problem, elevated DBPs can constrain the operation of the RWS, by reducing the amount of Hetch Hetchy Reservoir water allowed into the system. Operationally, elevated TOC levels can trigger a reduction in SJPL flows to meet a specified blend followed by other chain-reaction effects (e.g., increase the production rate of Sunol Valley and Harry Tracy water treatment plants to compensate the reduced transfer from the SJPL). A separate strategy would be to increase SJPL flows so as to reduce detention time in the Coast Range Tunnel (CRT) to reduce DBP formation time. A last strategy is to filter the SJPL supply at SVWTP. The water quality model aims to help answer some key questions, including: What changes in climate/weather could influence frequency and severity of high turbidity and TOC events on Hetch Hetchy water (and, therefore, elevated DBPs levels)?

The goal of the raw water quality model is to simulate several indicators of water quality in time at various locations in the RWS. The model results are focused on Tesla Portal, the primary water quality compliance point for the RWS, and the O'Shaughnessy Dam at Hetch Hetchy Reservoir. The outputs of the raw water quality model can then be used to help assess the long-term vulnerability of the SFPUC Water Enterprise to long term changes in climate conditions.

Two indicators of water quality were considered:

- **Turbidity** – Turbidity is a measure of the light scattered by suspended particles in the water and, as such, is a tracer of the concentration of suspended sediments. It is an indicator of water pollution. The objective of SFPUC is to keep turbidity below 1 Nephelometric Turbidity Units (NTU). A threshold of 2 NTU is possible but may be detected by wholesale customers. The filtration avoidance threshold is 5 NTU (2 times in a 12-month rolling window).
- **Total Organic Carbon (TOC)** – TOC is the amount of carbon in organic compounds. It is a tracer of organic substances. Furthermore, TOC can combine with the disinfectant in RWS (free chlorine) to cause disinfection byproducts (DBPs). TOC should not exceed a level of 2 mg/L and DBP violations are likely when exceeding 3 mg/L.

The water quality model outputs will also be used to define water quality narratives to be used by SFWSM. Several other water quality indicators should be looked at to describe raw water quality in reservoirs. However, the scope and budget of the study did not allow for additional modeling. For example, coliforms, algae blooms, and stratification in reservoirs were left out since they would require a more important level of effort. The water quality indicators were selected in discussion with SFPUC.

3.5.2 Modeling Methods

Process-based models, such as CE-QUAL-W2 (Cole and Wells 2002), were briefly evaluated but the level of effort to setup such the model far exceeded the budget available. In addition, a process-based model requires large amount of inputs (i.e., an extensive data collection). The required inputs to a process-based model do not align with the outputs of SFWSM, making it impossible to stress test. For these reasons, a hybrid machine learning-statistical model was developed to characterize observed physical processes of raw water quality. A Composite Quantile Regression Neural Network (CQRNN) model was developed to predict water quality indicators at the following locations within the RWS: the O'Shaughnessy Dam (Turbidity) and the Tesla Portal (Turbidity and TOC). CQRNN is a machine-learning algorithm that combines artificial neural networks (ANN) with Quantile regression (Cannon 2018). Given a set of inputs (predictors), the output of CQRNN model is a set of pre-defined percentiles of the distribution of the considered water quality indicators (predictors). Inputs to the CQRNN model are limited to variables that will be simulated by the other models used for the LTVA. The input variables include hydro-climatic variables and RWS state variables. Hydro-climatic variables include: Hetch Hetchy (HH) inflow, air temperature and water temperature together with variables that describes the dryness of the basin (e.g., number of antecedent dry days). System state variables include San Joaquin Pipeline (SJPL) flow, HH storage, HH spill/release, Canyon Power Tunnel flow. The modeling time step for water quality is daily. Five percentiles of the distribution of the predicted water quality indicators are simulated (i.e., 2.5^{th}, 25^{th}. 50^{th}, 75^{th}, 97.5^{th}). The ability of CQRNN to simulate high percentiles of the predicted distribution makes the approach attractive when extreme events are of interest. More details about CQRNN are given in the Technical Report 5: Raw Water Quality Models (HRG 2021c).

3.5.3 Calibration and Validation Results

Calibration and validation results for the CQRNN approach are presented in this section. The model performance is evaluated using two goodness of fit criteria in addition to visual comparison of the

simulated and observed time series and distributions for the considered water quality indicators. The considered goodness of fit metrics are: i) the loss function used for training the models and, ii) the mean squared errors (MSE). The loss function used for model training is analog to an average error term that is averaged out across the considered percentiles (HRG 2021c). Similar than for the loss function, the MSE is calculated across the considered five quantiles (or percentiles). Lower the value of the loss function and of the MSE, better the model. The performance of CQRNN is also compared with commonly used data-driven approaches: quantile regression (QR), linear regression (LR), multivariate adaptive regression spline (MARS), and K-nearest neighbors (KNN).

CQRNN calibration and validation results for turbidity at the O'Shaughnessy Dam are presented in Section 3.5.3.1. Results for turbidity at the Telsa Portal are presented in Section 3.5.3.2, while results for TOC at the Tesla Portal are presented in Section 3.5.3.3. For each water quality indicator, results are given for the median of the distribution and for high and low extreme values of the distribution (i.e., 2.5^{th}, 25^{th}, 75^{th}, and 97.5^{th} percentiles). Since CQRNN can potentially reproduce very well any set of known events by finding complex relationships among predictors, caution must be taken to prevent overfitting of their parameters. Overfitting here means that the model parameters are calibrated to represent not only the underlying processes linking predictors and the target variable, but also the noise embedded within these values. In practice, it means that good performance is sought for both training and testing sets. To prevent overfitting QRNN model parameters, an approach called 'bootstrap-aggregating-bagging' was used (Cunningham et al. 2000). This approach is described in detail in Section 3.3 of the Technical Report 5 (HRG 2021c).

3.5.3.1 Turbidity at the O'Shaughnessy Dam

Figure 3-48 through Figure 3-50 show the performance of CQRNN in simulating turbidity levels at the O'Shaughnessy Dam. The calibration is considered satisfactory because i) no significant decrease in performance is observed when comparing model results obtained across the training and testing datasets, and ii) CQRNN outperforms the other considered approaches (Figure 3-48). Only KNN shows a better performance than CQRNN during training, but a large error for the validation sample highlights the lack of robustness of the approach in this context. Large MSE values are obtained for QR, LR and MARS during training. These large values are somewhat explained by the poor skill of these approaches to correctly simulate the turbidity peak that occurred during the largest ever observed turbidity event. This event, which took place early January of the year 1997, is illustrated in Figure 3-50 (a more detailed description of the event is given in the Technical Report 5 (HRG 2021c). Across all considered approaches, CQRNN and KNN are reproducing at best the turbidity peak during the 1997 event. In addition, the range of observed turbidity values is fairly well reproduced with the five considered quantiles (2.5^{th}, 25^{th}, 50^{th}, 75^{th}, and 97.5^{th} percentiles) as illustrated in Figure 3-50.

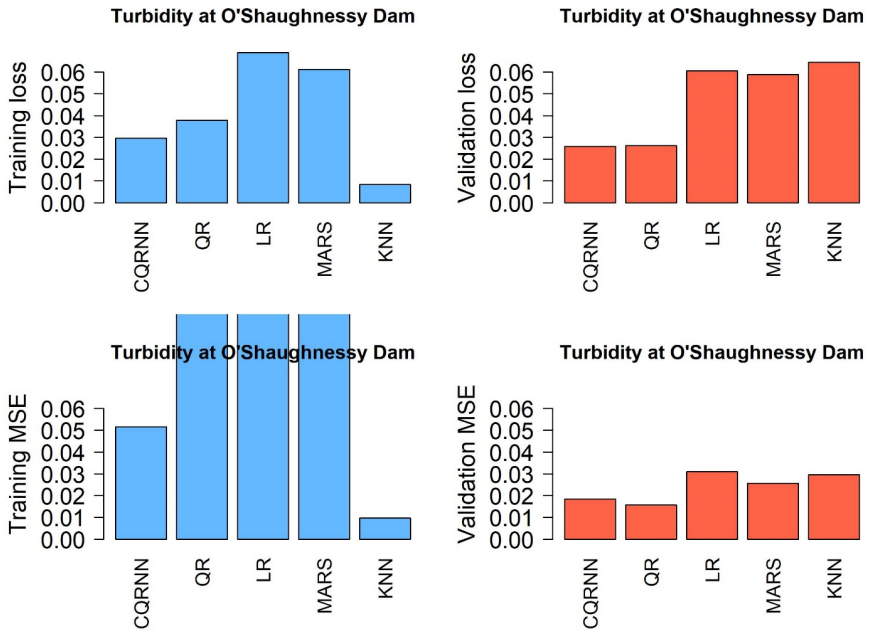


Figure 3-48. Comparison of the Training and Validation Results of the CQRNN Model (CQRNN) with Quantile Regression (QR), Linear Regression (LR), Multivariate Adaptive Regression Spline (MARS), and K-Nearest Neighbors (KNN).

Results shown are for turbidity levels at the O'Shaughnessy Dam. The loss function (top) and the MSE are shown for training (left) and validation (right) samples.

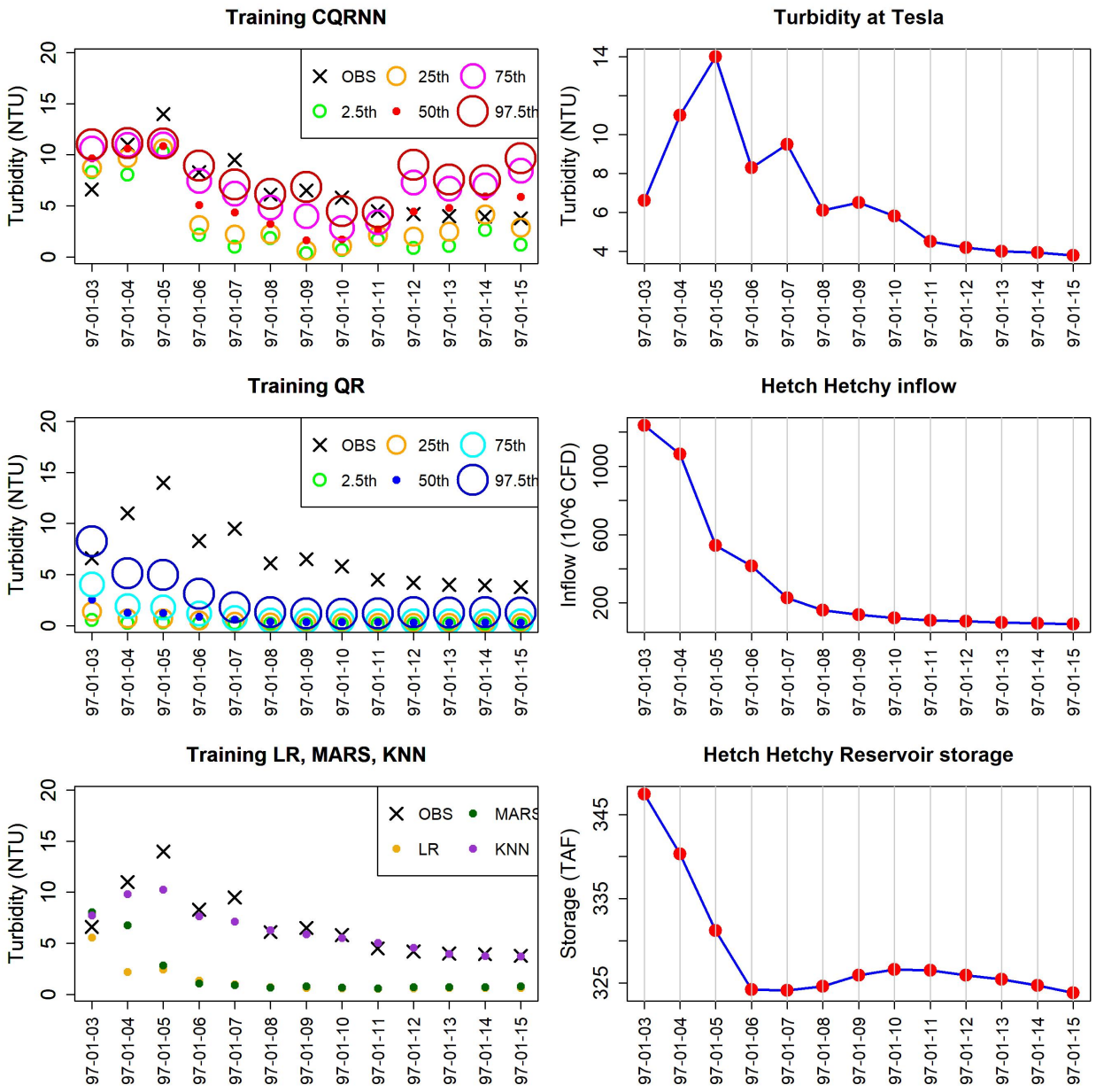


Figure 3-49. Comparison of the Model Performance during the 1997 Event at the O'Shaughnessy Dam. The left panel shows respectively model results for CQRNN (top), QR (middle), and LR, MARS, KNN (bottom). The right panel shows respectively observed measures for turbidity (top), the Hetch Hetchy inflow (middle), and the Hetch Hetchy Reservoir storage (bottom). Based on the left panel, CQRNN can capture turbidity peaks better than other methods (QR, LR, MARS, KNN).

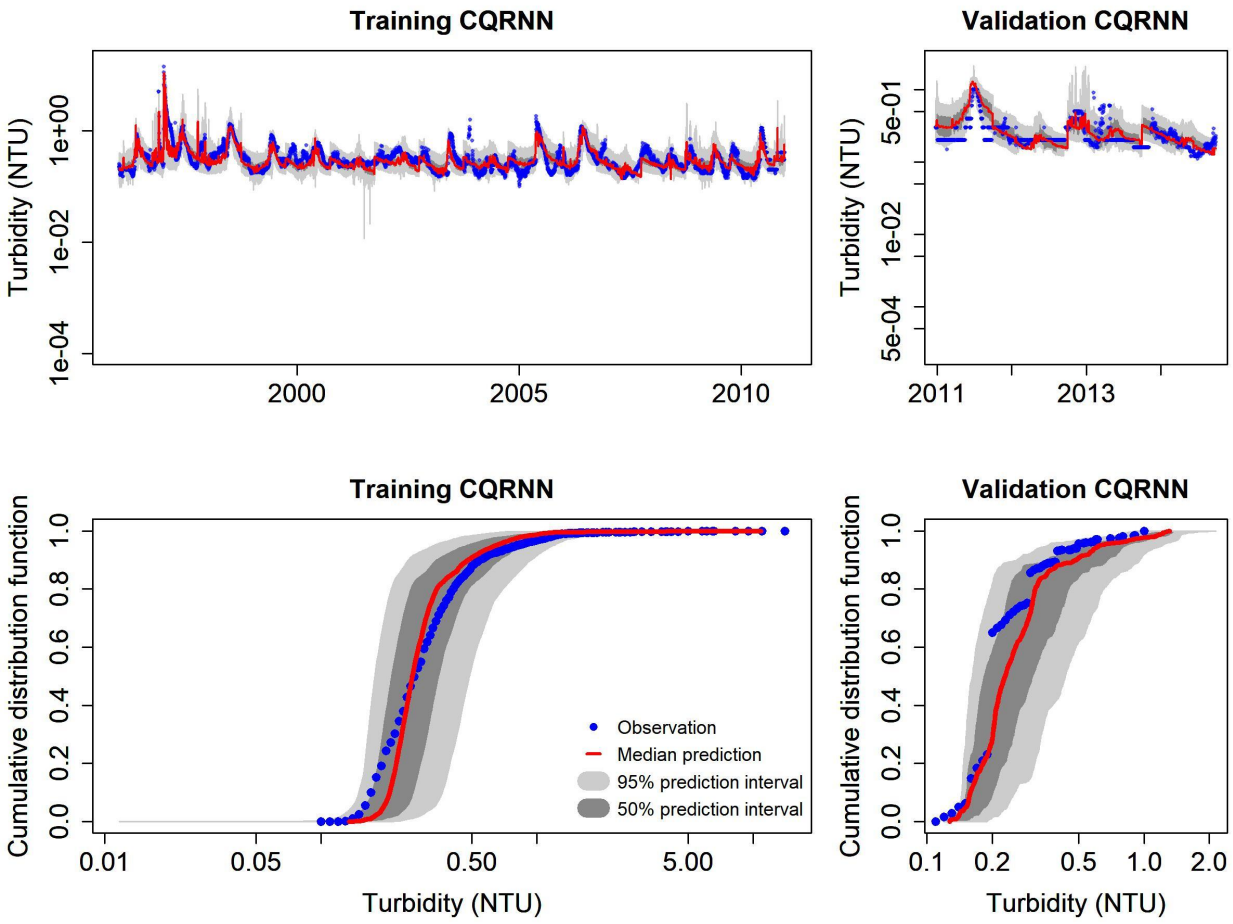


Figure 3-50. Illustration of the Performance of the Prediction Model (CQRNN) for Turbidity Levels at the O'Shaughnessy Dam Scaled in Natural Logarithm.

Performance obtained for the training (i.e., calibration) and for the testing (i.e., validation) are shown in left and right columns, respectively. Observed turbidity levels are shown in blue dots. The 50th percentile of the distribution is in red color. The deviation between the 25th and 75th percentiles (i.e., 50% prediction interval) is shown in dark grey color, while the deviation between the 2.5th and 97.5th percentiles (i.e., 95% prediction interval) is shown in light grey color. A logarithm scale is used to ease visualization of the turbidity distribution and time series.

3.5.3.2 Turbidity at the Tesla Portal

Figure 3-51 through Figure 3-53 show the performance of CQRNN in simulating turbidity levels at the Tesla Portal. The calibration is considered satisfactory because i) no significant decrease in performance is observed when going from the training to testing (or validation) sample and ii) CQRNN outperforms the other methods (Figure 3-51). Contrary to CQRNN, the performance of LR, MARS and KNN decreases significantly on the validation sample. The performance of QR is close to the one of CRQNN, although CQRNN performance is slightly better across the training sample. Another advantage of CQRNN is its ability to simulated larger turbidity values than the other methods. This is for instance illustrated in Figure 3-52 with the example of the 1997 event. Note that the 97.5th percentile simulated by CQRNN is larger than the observed turbidity value during the event. In addition, the range of observed turbidity values at the Tesla Portal is rather well reproduced by the five considered quantiles (2.5th, 25th, 50th, 75th, and 97.5th percentiles) (Figure 3-53).

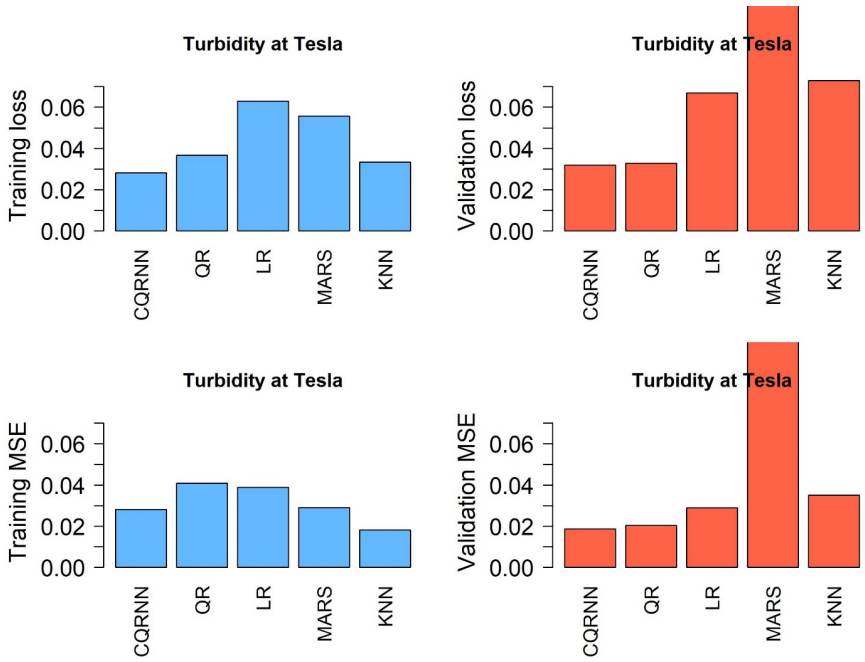


Figure 3-51. Comparison of the Training and Validation Results of the CQRNN Model (CQRNN) with Quantile Regression (QR), Linear Regression (LR), Multivariate Adaptive Regression Spline (MARS), and K-Nearest Neighbors (KNN).

Results shown are for turbidity levels at the Tesla Portal. The loss function (top) and the MSE are shown for training (left) and validation (right) samples.

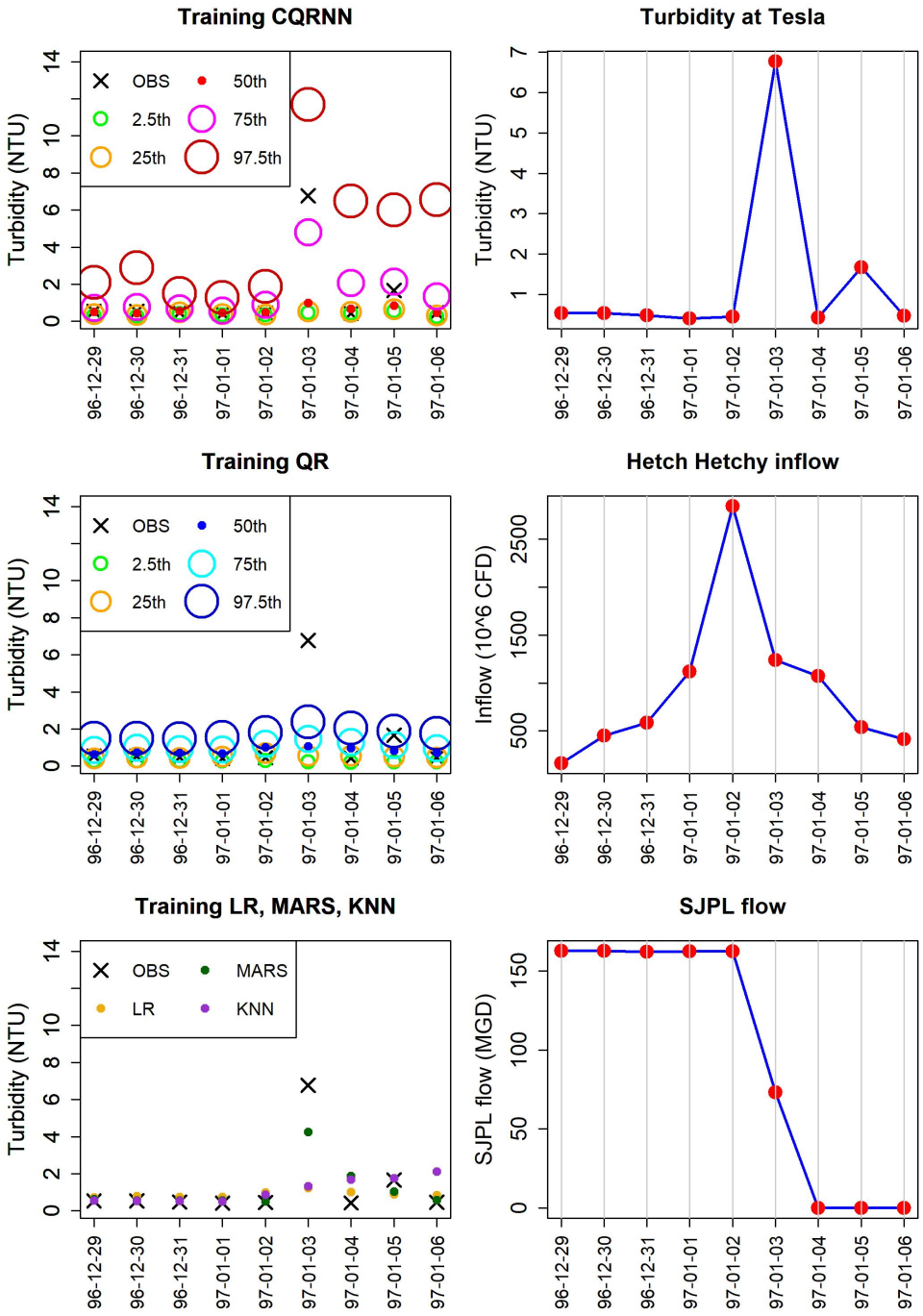


Figure 3-52. Comparison of the Model Performance during the 1997 Event at the Tesla Portal.
The left panel shows respectively model results for CQRNN (top), QR (middle), and LR, MARS, KNN (bottom). The right panel shows respectively observed measures for turbidity (top), the Hetch Hetchy inflow (middle), and the Hetch Hetchy Reservoir storage (bottom). Based on the left panel, CQRNN can capture turbidity peaks better than other methods (QR, LR, MARS, KNN).

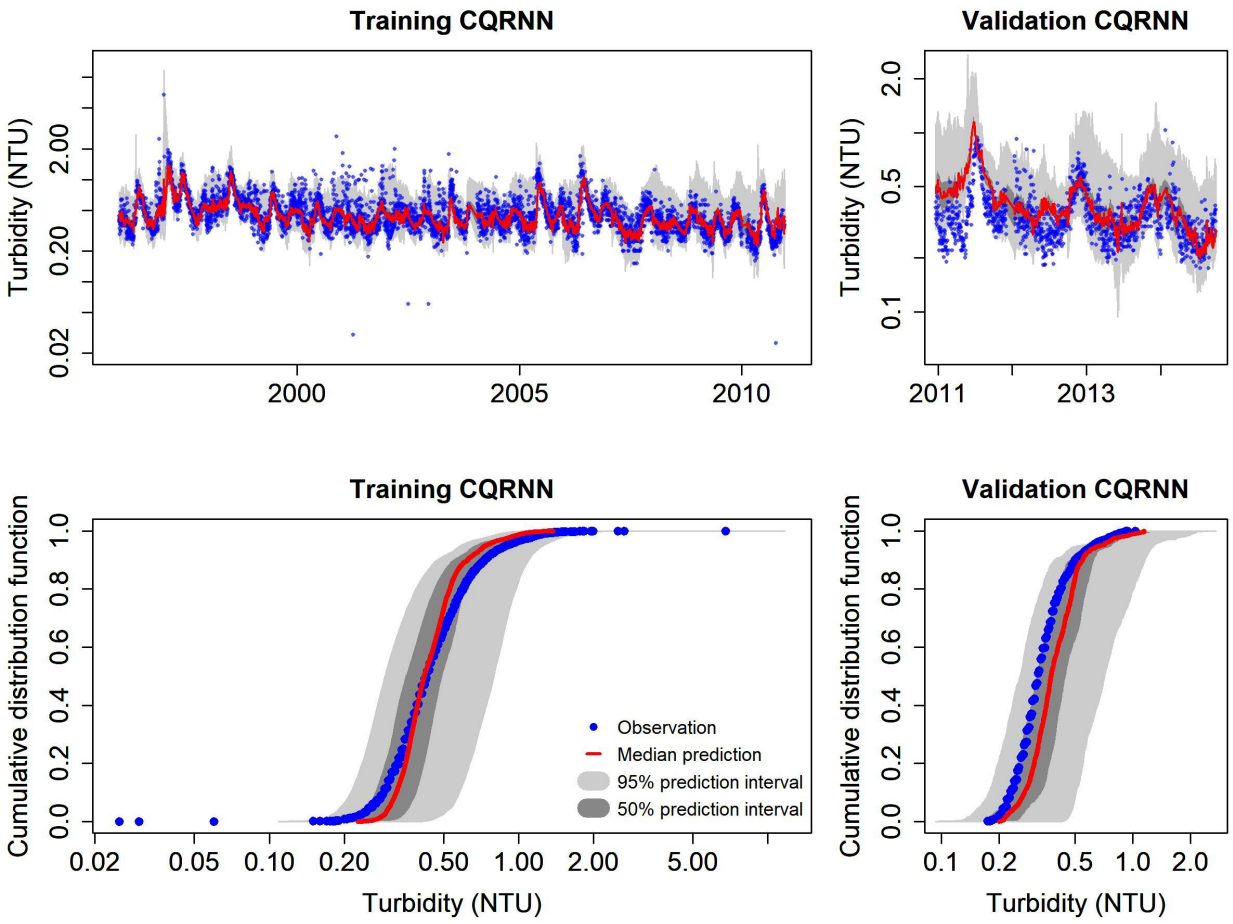


Figure 3-53. Illustration of the Performance of the Prediction Model (CQRNN) for Turbidity Levels at the Tesla Portal Scaled in Natural Logarithm.

Performance obtained for the training (i.e., calibration) and for the testing (i.e., validation) are shown in left and right column, respectively. Observed turbidity levels are shown in blue dots. The 50th percentile of the distribution is in red color. The deviation between the 25th and 75th percentiles (i.e., 50% prediction interval) is shown in dark grey color, while the deviation between the 2.5th and 97.5th percentiles (i.e., 95% prediction interval) is shown in light grey color. A logarithm scale is used to ease visualization of the turbidity distribution and time series.

3.5.3.3 TOC at the Tesla Portal

Figure 3-54 and Figure 3-55 show the performance of CQRNN in simulating TOC levels at the Tesla Portal. The calibration is considered satisfactory because i) no significant decrease in performance is observed when going from the training to testing dataset and ii) CQRNN outperforms the other considered approaches (Figure 3-53). Seasonal variations in TOC levels are fairly well reproduced across both training and testing sets, and most TOC peaks are caught by the high quantile prediction (Figure 3-54).

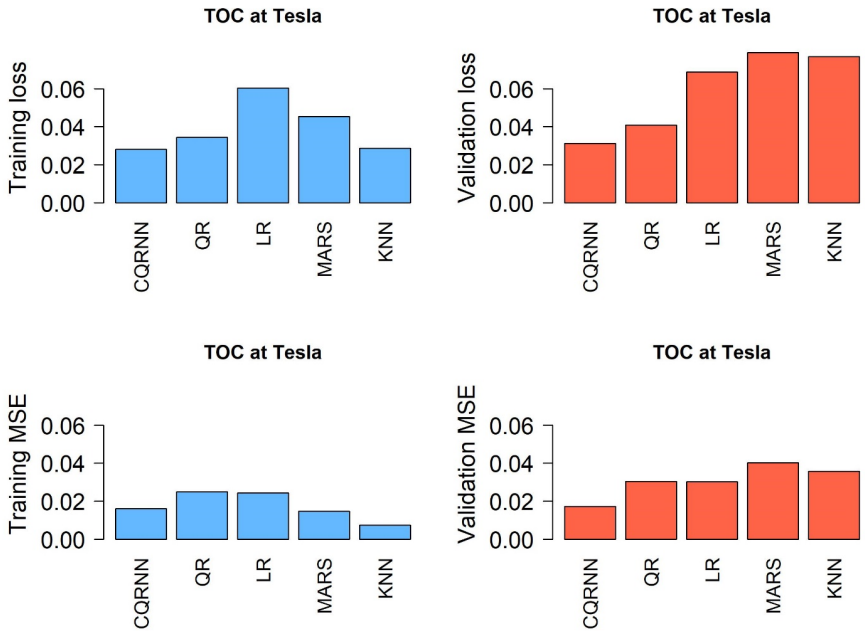


Figure 3-54. Comparison of the Training and Validation Results of the CQRNN Model (CQRNN) with Quantile Regression (QR), Linear Regression (LR), Multivariate Adaptive Regression Spline (MARS), and K-Nearest Neighbors (KNN).

Results shown are for TOC levels at the Tesla Portal. The loss function (top) and the MSE are shown for training (left) and validation (right) samples.

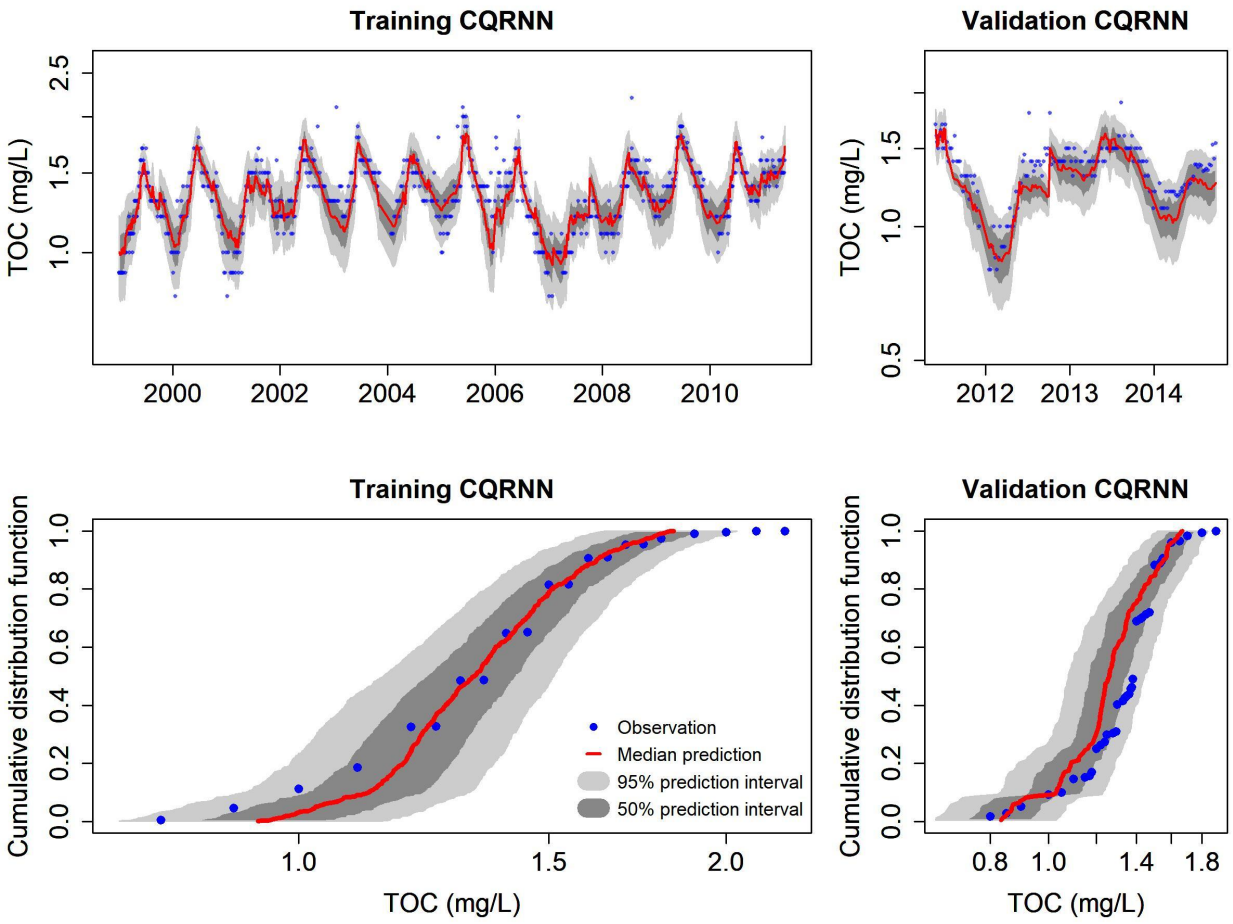


Figure 3-55. Illustration of the Performance of the Prediction Model (CQRNN) for TOC Levels at the Tesla Portal. Performance obtained for the training (i.e., calibration) and for the testing (i.e., validation) are shown in left and right column, respectively. Observed TOC levels are shown in blue dots. The 50th percentile of the distribution is in red color. The deviation between the 25th and 75th percentiles (i.e., 50% prediction interval) is shown in dark grey color, while the deviation between the 2.5th and 97.5th percentiles (i.e., 95% prediction interval) is shown in light grey color. A logarithm scale is used to ease visualization of the turbidity distribution and time series.

3.5.4 Model Limitations

A screening of the available approaches for modeling raw water quality (i.e., turbidity and TOC) was carried out at the beginning of the study, which resulted in disqualifying process-based approaches, mostly due to their data requirement. For instance, the well-known CE-QUAL-W2 model (Cole and Wells 2002) requires a large variety of input variables that are not available (e.g., bathymetry, solar radiation shading, wind, inflow constituent concentrations, etc.). In addition, it is not possible to predict these variables under climate change given the available modeling tool used for the LTVA.

Following discussion with SFPUC, HRG developed a data-driven approach using the available data (e.g., inflow, temperature, precipitation, reservoir storages, among others). Data driven approaches do require a significant volume of data to learn the relationship between the target variable (e.g., turbidity or TOC) and the considered predictors. The more data used for training/validation of the model, the more skillful the predictions. When extreme values are of interest, it is paramount for model training to include extreme events in the training sample. It is also important to include extreme event in the validation sample, so that the ability of the model to predict accurately the water quality indicators

during extreme events outside the training sample can be checked. As described below, this has not been possible given the available turbidity and TOC data.

For turbidity models at the O'Shaughnessy Dam and the Tesla Portal, the number of available data is reasonable, with respectively 6,767 and 6,496 days where both model inputs and turbidity observations are available. However, a single day in the historical record shows a high turbidity value larger than 5 at the Tesla Portal (i.e., during the 1997 turbidity event illustrated in Figure 3-49 and Figure 3-52). At the O'Shaughnessy dam, more values above that threshold are available, although all recorded during the same 1997 event. The lack of extreme turbidity events may limit the model skill at predicting accurately the absolute value of the turbidity peak.

For TOC model, the available TOC data is much less than for turbidity (855 data points). No recorded TOC values are above the threshold of 3 mg/L, while only five values are above 2 mg/L. The lack of high TOC events is a major limitation for training a skillful data-driven model to simulate accurately extremes. In addition, TOC historical record is available at weekly basis, which is an additional difficulty for model training because the dynamic of the hydro-meteorological event leading to high TOC values might be more difficult to learn.

Other important limitations are not necessarily tied to the models but more on the intended use of these models for the LTVA. First, most predictors used for the stress test (defined Section 4.2.1) have been obtained from model simulations whose assumptions do not necessarily align with the observed hydro-climatic and system state variables used to train and validate the water quality models. For example, the demand scenarios for the stress test are set to 227, 265, 300 and 334 mgd (baseline is 227 mgd), although, the average demand for the training/validation periods was estimated to 246 mgd. A larger demand to the SFPUC service area means larger transfer from Hetch Hetchy Reservoir to the Bay area through the San Joaquin Pipelines, which is a predictor for the water quality models at Tesla Portal. Another example is the one of the different predictors that are based on the 'number of antecedent dry days.' Following recent work on the topic (e.g., Cha et al. 2016; Allen et al. 2017), the dry days were assumed to be days with 0 precipitation. These days are used to calculate three different predictors that are meant to inform the model of the dryness of the basin (the first predictor is the length of the current sequence of dry days while the second and third predictors are the total number of dry days within the previous and current water year). As such, precipitation scenarios of either increase or decrease in precipitation obtained using the delta approach, as it is the case in this study (Section 4.2.2), do not alter the sequence of dry days defined under the baseline scenario with no change in precipitation. As such, the modification of TOC and turbidity that would result from changes in these predictors are not accounted for.

In addition, beyond the considered assumptions to create or simulate the predictors under various climate or demand conditions, any modeling errors or simplification of the modeled processes of the models (i.e., PRMS and SFWSM) used to create input time series for the predictors of the water quality models would potentially lead different predicted values than the one obtained for the training/validation periods. Figure 3-56 provides an example of the discrepancies that may exist between the predictors used for the stress test and the training/validation dataset. The figure shows the observed and simulated empirical cumulative distribution functions for the flows through the San Joaquin Pipelines (simulated values are shown for the demand scenarios considered in the LTVA and for baseline climate; no change in precipitation and temperature). It is shown that the observed and simulated distributions are quite different, which could lead to difference in predicted distributions for TOC and turbidity.

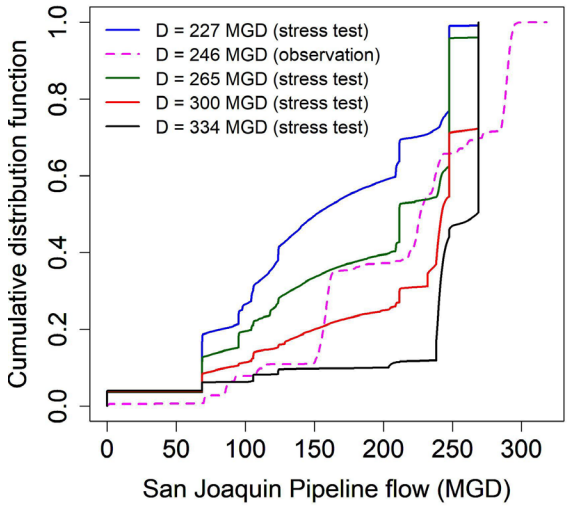


Figure 3-56. Empirical Cumulative Distribution Function (ECDF) for the San Joaquin Pipelines (SJPLs) Flow. Blue, green, red, and black curves show the ECDF of the simulated flows through the SJPLs for each demand scenario (227, 265, 300 and 334 mgd, respectively) and across the 10 realizations under baseline climate conditions (no changes in precipitation, temperature). The dashed pink curve shows the ECDF of the historical flow through the SJPL during the training/validation period (i.e., 1996-2010).

Wrapping-up the various limitations discussed above, the water quality models presented in this section have been challenging to train and validate for the various reasons mentioned above. The performance of the models for predicting out of sample extreme events is difficult to anticipate, especially given the limited or lack of extreme events in the training and validation samples. For instance, Figure 3-49 shows that the predicted 97.5th percentile overestimates significantly the turbidity peak at Tesla Portal during the 1997 event by nearly 100%, which suggests that 97.5th percentile will overestimate the turbidity peaks through the stress test. However, the model successfully predicts the day of the event. The recommendation is then to use the predicted 97.5th percentile for the turbidity at Tesla Portal to detect timing of turbidity values above the 5 NTU threshold; however, the actual turbidity values is likely to be largely overestimated. On the other hand, both 50th and 75th percentiles have been shown to underestimate the 1997 peak.

3.6 Finance Model

3.6.1 Modeling Goals

Along with natural hazards, financial risk has been identified by SFPUC Water Enterprise staff as the most important source of vulnerability and a key uncertainty affecting future performance of the SFPUC Water Enterprise. The LTVA Finance module is designed to simulate the potential impact of changing environmental and socioeconomic conditions on the price of water.

The general goals of the finance module are to 1) simulate the potential impact of a changing climate on the price of water and 2) allow future users to adjust key parameters driving uncertainty in the model, including the interest rate, debt repayment period, capital expenditures (CAPEX) and operating expenditures (OPEX) in particular.

SFPUC's finance department currently maintains a long-term finance planning model with a 10 year horizon and incorporates SFPUC's Capital Investment Program, tracks forecasted spending on

operations and debt service, and estimates the price of water and affordability of water bills for customers. Although this model does extend through to 2070, SFPUC only prepares detailed financial plans 10 years into the future. The model projects a decline in sales from 192 mgd in 2020 to 183.5 mgd 2030 and remains static at this level through to 2070 (FY20 Water 10 yr plan). The objective of the finance module in the LTVA is to compliment this existing model by providing insight into the potential impact of long term trends on key indicators of financial performance.

For this phase of work, the finance module is not tightly coupled to the SFWSM, but instead takes the output of the SFWSM model (i.e., deliveries) as its input in a loosely coupled approach. As such, feedback loops that take account of consumer response to price change are not taken into account. In subsequent phases of work, this can be revisited to explore the role of price in driving demand of consumers, both as an adaptation strategy, and as a risk associated with drought periods.

3.6.2 Modeling Methods

This section summarizes the description of the finance module while more details are given in the Technical Report 6: Finance Module (HRG 2020b). The finance module combines water deliveries output from the system model (SFWSM) with a long term average values drawn from SFPUC's financial planning model to evaluate the impact of these changing conditions on the health of SFPUC's finances. Inputs and outputs to the model are summarized in Figure 3-57. Inputs from the SFPUC financial model include fixed and variable CAPEX, fixed and volumetric OPEX, and average annual revenues from sources other than water sales (such as rental income, interest etc.). Input from the system model comes in the form of annual deliveries. The model is run at the annual time step, which aligns with key financial planning processes such as rate setting and submission of annual financial accounts.

The finance module includes two primary functions: 1) evaluating expenditure for a given year, and 2) setting retail and wholesale customer water rates to ensure revenues are sufficient to cover costs. Results of the finance model are presented for both step and trend model runs (see Section 4.2.4 for the explanation of step and trend runs). Inclusion of trend model runs allow one to explore how costs and specifically debt service, evolve over time in response to variables such as interest rate. Each of the variables and assumptions included in the finance module and depicted in Figure 3-57 are described in detail in the Technical Report 6 (HRG 2020b).

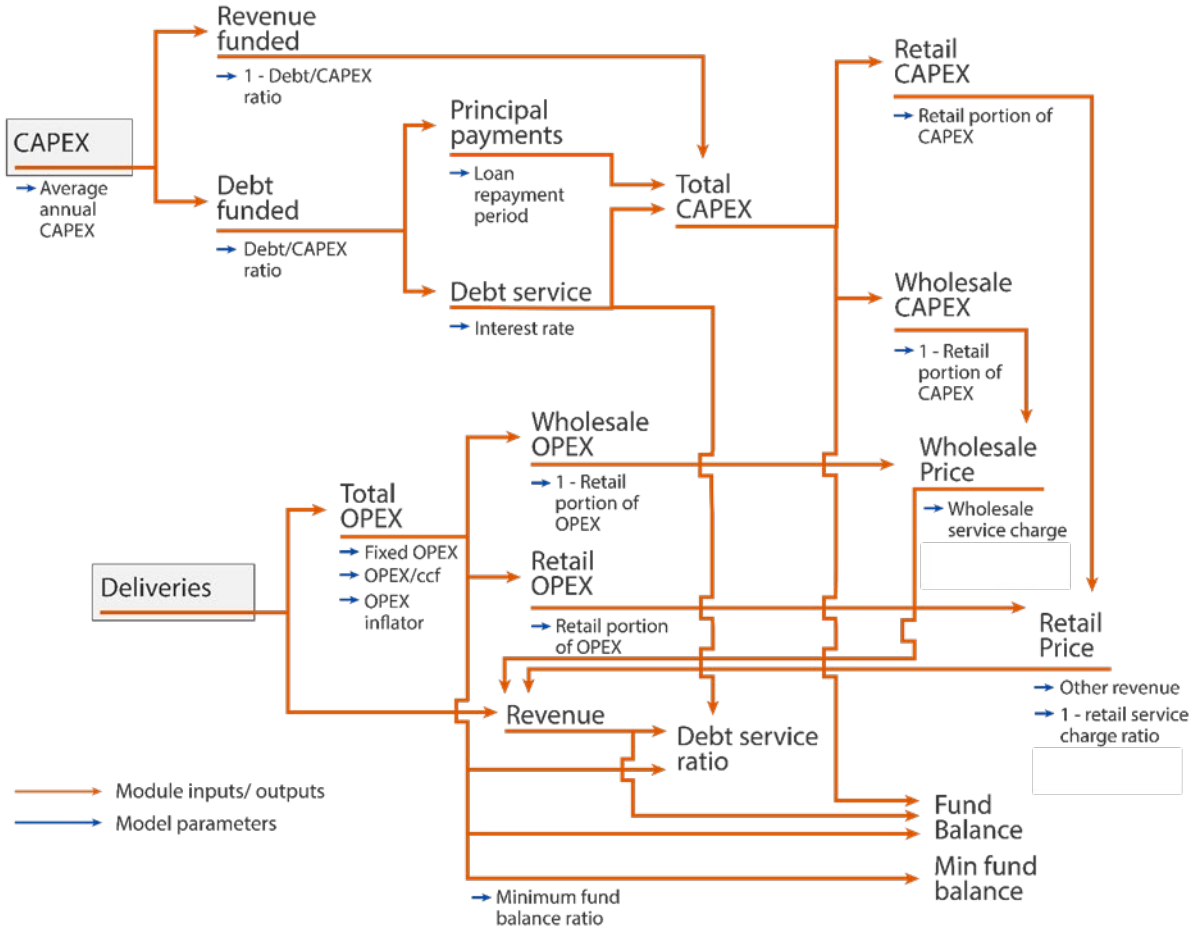


Figure 3-57. Schematic of the Finance Module Developed for the LTVA.

3.6.2.1 Expenditure

The finance module presented here does not itemize costs but instead draws on the detailed breakdown provided by the SFPUC 10-year plan. High level aggregate costs are drawn from the model to provide figures for OPEX, CAPEX, and debt as a proportion of CAPEX. Figure 3-58 provides a breakdown of CAPEX, OPEX and debt service in the 10-year plan versus that assumed in the LTVA Finance Model under both trend and step model runs.

In order to estimate system performance using step model runs, eight scenarios of CAPEX and two scenarios of OPEX are used to explore the sensitivity of the system to different levels of expenditure. CAPEX scenarios are placed between $0.35B and $1.75B at intervals of $0.175B. OPEX scenarios are centered on OPEX expenditure in 2020 - $0.29B – and OPEX expenditure in 2040 - $0.5B.

There is significant divergence between CAPEX spend in the 10-year plan and that assumed in the trend model runs for this study, particularly beyond 2030. This is due to the fact that detailed forecasts of CAPEX spend in the 10-year plan do not extend beyond 2030 and steadily decline over time. It was agreed with SFPUC that this study should assume a continued increase in CAPEX through time. An overview of the approach to estimating CAPEX and OPEX spend under trend model runs is provided below.

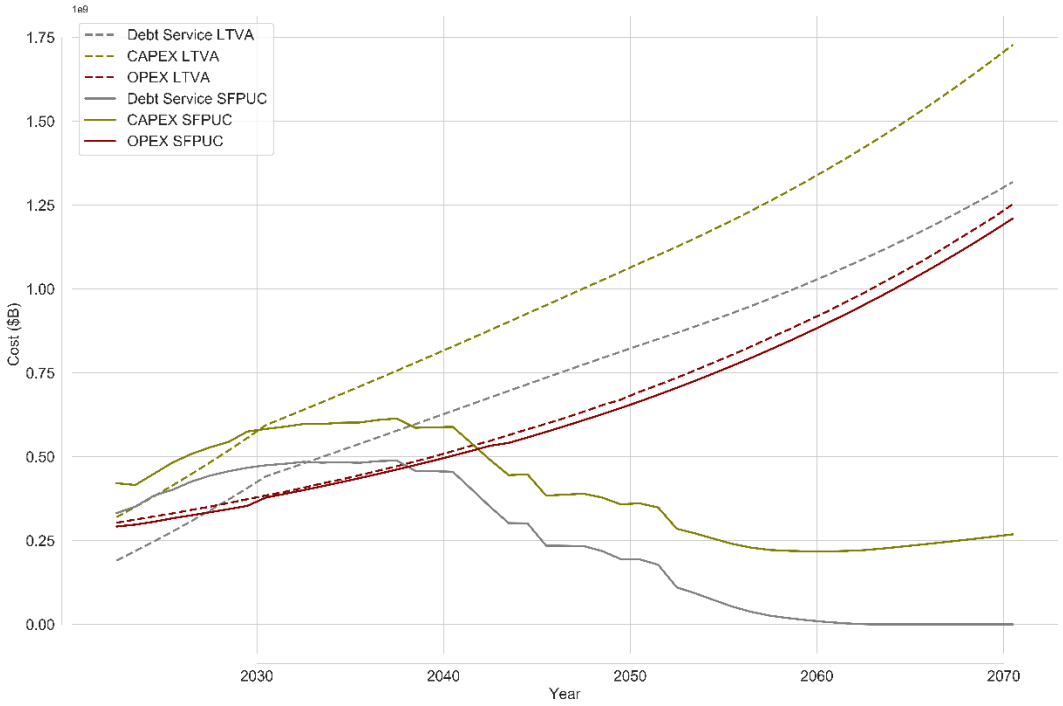


Figure 3-58. Comparison of CAPEX (Green), Debt Service (Grey), and OPEX (Red) between SFPUC 20-Year Plan (Solid Lines) and LTVA (Dashed).

Capital Expenditure

In order to establish a forecast of CAPEX across the 50-year time horizon utilized in the LTVA, the finance model generates a time series of CAPEX spend from 2000 through to 2070 using the logic provided in Figure 3-59. This time series initializes in 2000 to incorporate legacy debt that SFPUC holds in 2020, the LTVA's initial time step.

To avoid tracking individual CAPEX expenditures and loans (which is done in the SFPUC's existing 20-year plan), the Finance model considers CAPEX spend as an aggregate value for each year and is made up of two components:

- Fixed – representing the expected new capital investment in each year and makes up 45% of total CAPEX spend.
- Variable – representing ongoing repair and replacement (including significant modification) of existing infrastructure that is inflated by 3% year on year. Variable CAPEX makes up 55% of total CAPEX.

CAPEX spend is assumed to be 75% debt funded and 25% revenue funded where the debt funded portion is assumed to be a new loan taken out each year.

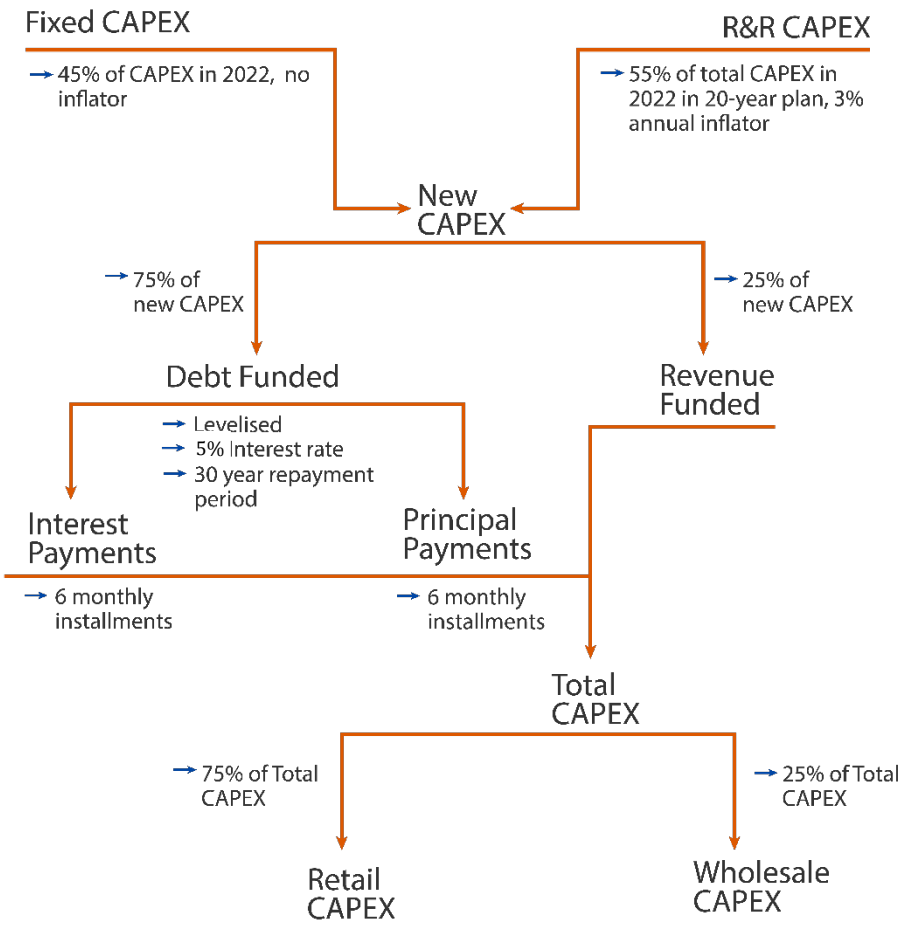


Figure 3-59. Schematic of CAPEX Logic in LTVA Finance Module.
Orange arrows indicate model inputs/outputs and blue arrows indicate model parameters.

Operational Expenditure

Figure 3-60 provides an overview of the logic utilized in establishing OPEX. OPEX spend is made up of two components:

- Volumetric: this OPEX is a function of the amount of water delivered and is inflated at a rate of 3% per year from a base value of $0.11 /ccf. This value is derived by taking 3.5% of total OPEX (SFPUC's finance department estimate the volumetric component of OPEX to be 3.5%) spend in 2020 from the 10 year plan and dividing by the estimated volume of water delivered for that year.
- Fixed: OPEX represents operational costs that do not depend on the volume of water delivered and makes up 96% of total OPEX.

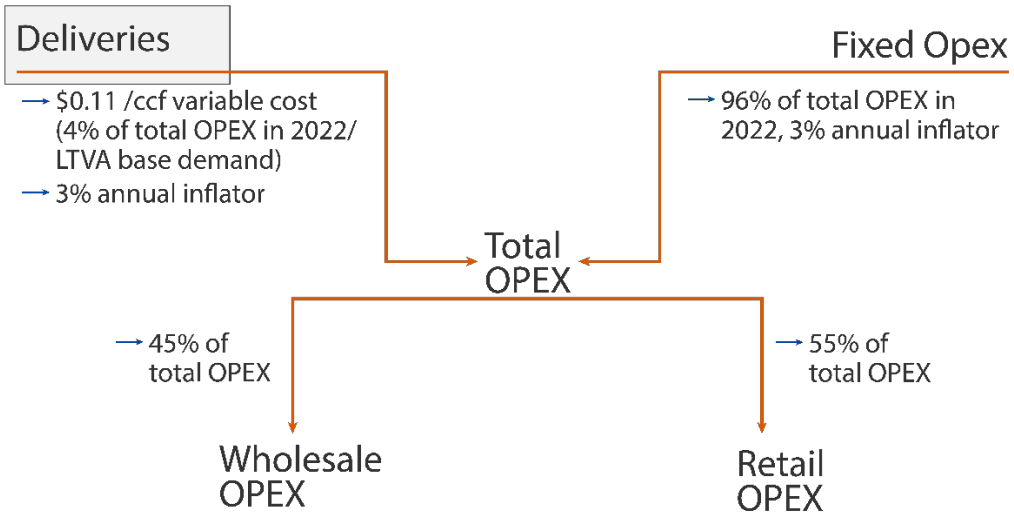


Figure 3-60. Schematic of OPEX Logic in LTVA Finance Module.
Orange arrows indicate model inputs/outputs and blue arrows indicate model parameters.

3.6.2.2 Price Setting

The logic for setting water rates is based simply on cost recovery. Each year rates are set by dividing the total expenditure for that year by the volume of water delivered. While the general logic for setting wholesale and retail rates is the same, the following describes how costs are apportioned to wholesale and retail customers and the specific logic for each.

- CAPEX: 25% to wholesale customers, 75% to retail customers
- OPEX: 45% to wholesale customers, 55% to retail customers
- When setting retail water rates, revenue from sources other than water sales (rental income, interest, etc.) is discounted from the retail portion of total costs and multiplied by a factor of 0.85. The remaining 0.15 is split amongst all retail customer accounts to make up the service charge portion of their water bill.
- A service charge of $4,277,000 is discounted from the wholesale portion of total costs.

3.6.3 Model Limitations

The demand, infrastructure and system models are currently loosely coupled: one model receives a 50-year time series as input to another model in the chain. However, in reality, dynamic feedbacks exist between these processes that are not captured in a loosely coupled approach. For example, one would expect to see a reduction in demand corresponding to an in increase in price resulting from drought event. These dynamics are potentially of interest in the context of a changing climate, where long term changes in precipitation and temperature have the potential to increase the depth and duration of drought events.

In order to undertake a stress test on the system, the finance model currently considers an average CAPEX and OPEX cost drawn from SFPUC's 10 year Finance Plan (HRG 2020b). In order to explore the impact of specific investment portfolios, a time series of CAPEX and OPEX would need to be provided as input to the model in order to understand the implications of these investments on the price of water.

Several assumptions are made regarding key parameters in the model - such as interest rate, inflation rate, split of costs between retail and wholesale customers etc. These assumptions are held static across

the 50 year time horizon. The impact of investment portfolios on the price of water is likely to be greatly influenced by changes in these parameters and so should be subject to sensitivity analysis when assessing the performance of different investment options.

CHAPTER 4

Vulnerability Assessment Design

4.1 General Approach

The general approach of the Long-Term Vulnerability Assessment is to develop and use a suite of interconnected computer models and supporting analytical tools representing important processes involved in the long-term water system planning of the RWS (Chapter 3). These models—which include a stochastic weather generator informed by climate assumptions (CliWxGen, Section 3.1), PRMS and SAC-SMA-DS hydrologic models (Section 3.2), a water demand model (Section 3.3), the San Francisco Water System Model (Section 3.4), raw water quality prediction models (Section 3.5), among others — use input based on assumptions about future climate and other conditions such as demographic, economic, political and legislative conditions, and result in quantitative estimates of corresponding future conditions, which are identified as vulnerable or robust. This approach was developed based on the sources of vulnerability identified by the SFPUC staff (Figure 1-2) and is depicted conceptually in Figure 4-1. The figure does not represent all interactions between these different modules. For example, feedback effects are missing and neglected at the modeling approach at this point.

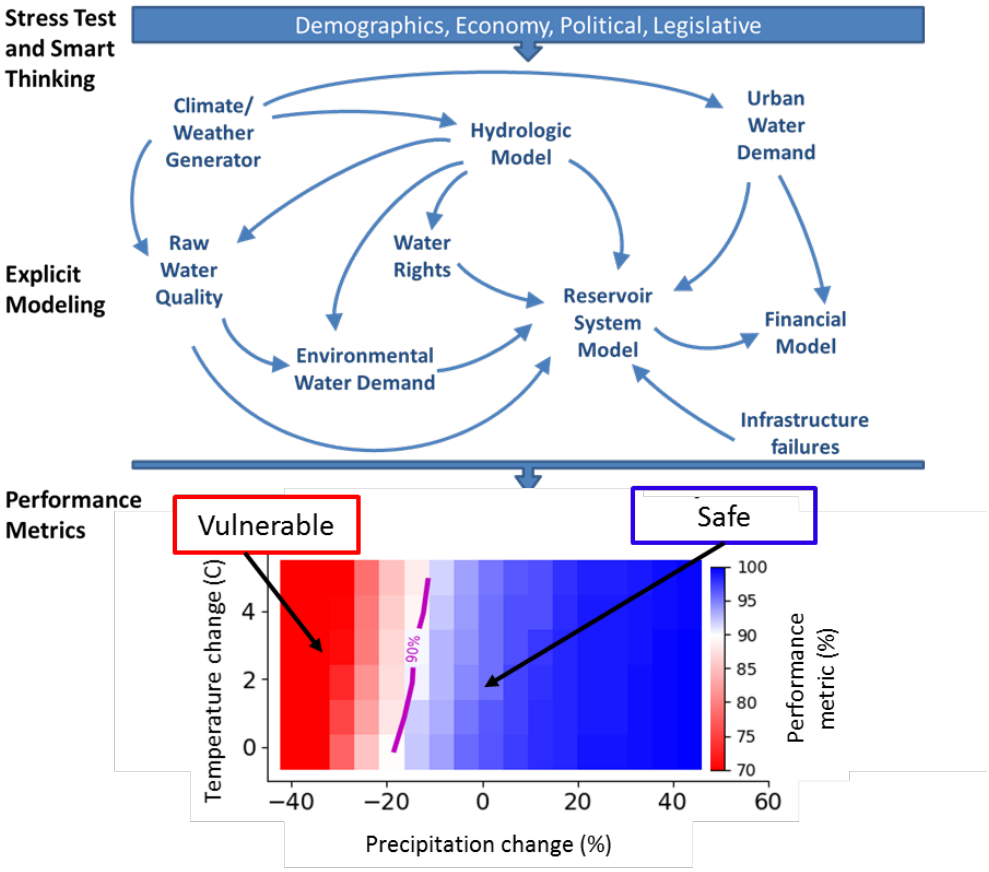


Figure 4-1. General Approach to Assessing Vulnerability.
Assumptions about the future, based on "smart thinking" narratives and stress tests developed by experts, can be used in a suite of models to assess vulnerability under a range of climate and other conditions to assess vulnerability and robustness.

The vulnerability assessment itself (Chapter 5) is conducted both with visual inspection and interpretation of results to identify conditions or combination of conditions, leading to system vulnerability (failure) under a large range of future scenarios. Two broad classes of approaches are used to generate future scenarios (Figure 4-2).

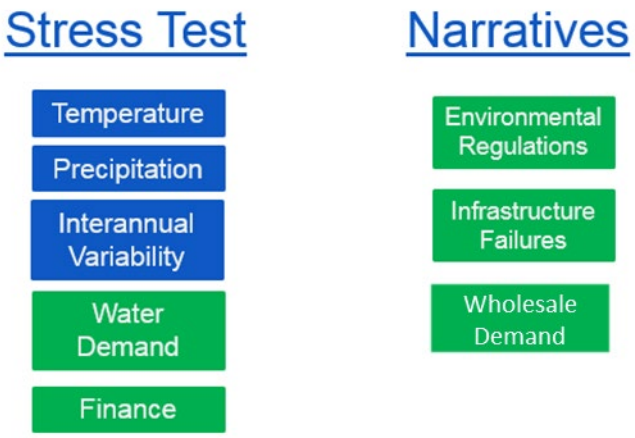


Figure 4-2. Approaches Used to Generate Future Scenarios.
Blue and green colors show natural and human-related drivers of change. 'Stress test' approach considers a large range of scenarios of change for each driver while 'Narrative-based Scenarios' are defined as a set of carefully selected scenarios to represent either the aftermath of short-term shocks (e.g., earthquake, wildfire) to the system or change in environmental regulations leading to change in instream flow requirement below the dams operated by SFPUC.

First, sources of vulnerability known to be important are systematically varied across a range of values to "stress test" various components of the water system and understand the consequent vulnerability associated with stressor variations. Stress tested system parameters include temperature, precipitation, climate variability, water demand and finance related variables (left column in Figure 4-2). For the 'inter-annual variability' driver of change, the stress test makes use of several stochastic realizations that sample the potential effect of climate low frequency variability on the sequence of dry and wet years (see Section 3.1.4). For the other parameters, change is represented using a large range of annual change factors. Given the combinations of changes across the considered system parameters, the stress test results in a very large number of possible, uncertain modeled futures. The assessment of the system performance across these futures can be used to map the vulnerability of the system as illustrated by the color map in Figure 4-1. This color map, hereafter termed as 'climate response function,' shows the range of parameter values (e.g., temperature and precipitation as in Figure 4-1) for which the performance of the system is acceptable (i.e., either above or below a chosen threshold/target). More details regarding the stress test are given in Section 4.2, including how to read climate response functions Section 4.2.4.1).

Second, a "smart thinking" approach is used, whereby qualitative narratives about future conditions are transformed into quantitative trajectories of specific system parameters. Future conditions span demographic, economic, political, and legislative domains, but may be extended to include other factors, such as technological advancements. For the study, narratives are limited to describing possible changes in instream flow requirements (i.e., environmental water demand in Figure 4-1) and some infrastructure failures (e.g., earthquake along the Calaveras fault, wildfires). Two demand narratives have also been included to represent the hypothetical situation for which SFPUC would deliver to its customers their individual supply guarantee only (with and without Santa Clara and San Jose as

permanent customers). A detailed description of the considered narratives is given in Section 4.3. The quantitative scenarios based on the narratives are called narrative-based scenarios to be consistent with the scientific literature.

Stress test and narrative-based scenarios are combined to further explore the response of the system under the trajectories of changes in climate and demand conditions.

4.2 Stress Tests

4.2.1 Baseline – Summary

In order to evaluate the effects of drivers of change on the RWS, the system baseline has to be established. This section summarizes information defining the baseline used for the climate stress test.

Regarding weather data, the nine stochastic realizations have been developed using the period 1956-2011. The historical realization covers 1961-2011. As such, averages and climate variability signals embedded within these ten realizations is representative of the internal variability identified for this period. Increase in temperature and changes in mean annual precipitation are considered above the historical mean of period 1956 – 2011.

Regarding demand, baseline total demand is 305.5 mgd, from which 226.8 mgd is demand for SFPUC RWS. This value is based on FY2012-2013. The spilt of total demand is given in Table 3-21 and Table 3-22.

The state of the SFPUC system infrastructure is summarized in Table 4-1 and Table 4-2.

Table 4-1. Summary of the Hetch Hetchy Infrastructure and Projects Considered for the Baseline.
For the sake of simplicity, all major infrastructure, such as reservoirs and pipelines, are not mentioned here. Further details are available in the Technical Report 4 (HRG 2021b).

System infrastructure or project	SFWSM planning version - baseline
Westside groundwater basin	Included
Recapture project	Not included
New Calaveras reservoir	Included
UTREP spill	Included
Fountain thistle constraint at Crystal Spring reservoirs	Included
Bubble account in the water bank	Not included
Lower Cherry Aqueduct	Not included
Transfer of Hetch Hetchy water from Alameda siphons to Sunol Valley water treatment plant	Not included (only in infrastructure narrative NAI1)
Scheduled maintenance for the San Joaquin pipelines and the Bay division pipelines	Included (details of the schedule are available in the Technical report 4 (HRG 2021b)

Table 4-2. Reservoir Storage Information Used for the Baseline of the Planning Version of SFWSM.
The maximum operating storage is the storage level from which SFWSM release water through the valves to prevent the reservoirs from spilling. Although the operating storage for each system reservoir vary within the year, only the maximum value is reported in this table. More details are given in the Technical Report 4 (HRG 2021b). Note that storage capacities in SFWSM for Calaveras and San Antonio reservoirs are significantly larger than their maximum operating levels. Storage capacity for these reservoirs in SFWSM corresponds to the capacity of the reservoir at dam crest. As detailed in the Technical Report 3: San Francisco Water System Model (HRG 2021b), this make possible to better describe the controlled spill through the valves when water storage gets larger than their maximum operating storage.

Reservoir	Storage capacity (AF)	Dead pool (AF)	Maximum operating storage (AF)
Hetch Hetchy	360,360	25,100	360,360
Cherry	273,300	1,000	273,000
Eleanor	27,100	0	27,100
Don Pedro	2,030,000	0	2,030,000
Water Bank	570,000	0	570,000
Calaveras	117,877	11,535	96,669
San Antonio	66,782	2,101	50,636
Crystal Springs	66,975	1,366	54,967
San Andreas	19,027	3,137	19,027
Pilarcitos	2,977	921	2,976
Westside groundwater storage	75,000	0	75,000

4.2.2 Climate Scenarios for the Stress Test

Stress tests for climate (precipitation and temperature) and demand (Section 4.2.3) were developed by analyzing the projected regional changes of climate and water demand through 2070.

The set of future climate scenarios were obtained by perturbing the selected stochastic climate realizations using the delta change approach (absolute change for temperature and relative change for precipitation). The considered range of temperature and precipitation change factors are based on the findings of climate projections (Section 2.3). Temperature scenarios range from +0°C to +7°C with 1°C increments, for a total of 8 temperature scenarios. Precipitation scenarios range from -40% to 40% with 5% increments, for a total of 17 precipitation scenarios. Overall, this results in 136 unique climate scenarios for each stochastic realization (i.e., 1,360 climate scenarios total) (Table 4-3).

Table 4-3. Summary of Climate Change Scenarios for the Climate Stress Test.
The range of scenarios is applied to both step and trend change (see Section 4.2.4).

Type of uncertainty	Sampling range	Sample size
Natural climate variability	Stochastic realizations	10 realizations
Changes in mean annual precipitation (%)	-40 % to 40 % with 5% increments	17 change factors
Changes in mean annual temperature (°C)	0 to 7°C with 1°C increments	8 change factors
TOTAL		1360 climate scenarios

The final set of annual precipitation series that result from the combinations of the ten realizations and 17 levels of precipitation changes and 8 temperature changes represent a wide range of challenging scenarios for the SFPUC water system, in terms of both wet and dry conditions as they consider natural climate variability and possible climate changes together. The change factors were applied uniformly to the precipitation and temperature time series. This means that minimum and maximum daily temperature have the same perturbation and the seasonality is not affected.

4.2.3 Demand Scenarios for the Stress Test

SFPUC uses several models to generate demand projections and continuously publish demand projections for ongoing planning processes. This study developed a baseline of total demand in the service area of 305.5 mgd (retail demand of 78.9 mgd and wholesale demand of 226.6 mgd) and a demand on the RWS of 226.6 mgd (retail demand of 78.9 mgd and wholesale demand of 147.9 mgd) (Section 3.3). In order to stress test the RWS, four demand scenarios were used to represent a range of demands to fully stress the system. The uncertainty in demand originates from uncertainty in projections for population, income, and other predictors as well as in estimates of demand elasticity. An uncertainty analysis was conducted under two cases: 1) uncertainty in projected exogenous driver of demand, and 2) adding the uncertainty associated with elasticity of demand estimates. Through this analysis a broader range of 9 demand scenarios was established. Detailed description of this sensitivity analysis can be found in the supporting document Technical Report 3 (HRG 2020a).

Since 1970, SFPUC has supplied approximately 65% of total Wholesale Customers' demand. The dependence of each Wholesale customer on the RWS varies, with some entirely reliant on the SFPUC for their supply (SFPUC 2016). In order to disaggregate the SFPUC portion of total demand (RWS demand), the observed share of total demand for FY2013 was used. FY2013 is used by SFPUC as a benchmark for 'normal' conditions prior to the 2014-17 drought. The RWS demand (established by applying the observed share of total demand in FY2013) under each demand scenario is provided in Table 4-4 and in Figure 4-3 (DB denotes baseline demand and, D45 denotes a 45% increase in total demand by 2070 relative to the baseline).

In developing demand scenarios, it is important to make distinction between total demand for all customers (retails and wholesales) and the demand to the RWS (aka RWS demand). Scenarios of total demand preserve the spatial variation in demand growth present in the SFPUC service area (more details are provided in the Technical Report 3 (HRG 2020a). Thus, an increase in total demand of 30% does not necessarily mean a 30% increase in RWS demand. This explains the differences in the RWS demand and total demand under each demand scenario (right hand column of Table 4-4).

Table 4-4. Water Demand Scenarios on the SFPUC RWS Accounted for the Stress Test.
From left to right: total annual demand (mgd), % change in total annual demand relative to baseline, the RWS share of total annual demand (mgd) and RWS demand as a % of total annual demand.

Scenario	Id	Total demand (mgd)	RWS demand (mgd)	Wholesale RWS demand (mgd)	Retail RWS demand (mgd)	RWS share of total demand (%)
Baseline	DB	305	227	148	79	74%
+15%	D15	351	265	181	84	76%
+30%	D30	397	300	207	93	75%
+45%	D45	444	334	233	102	75%

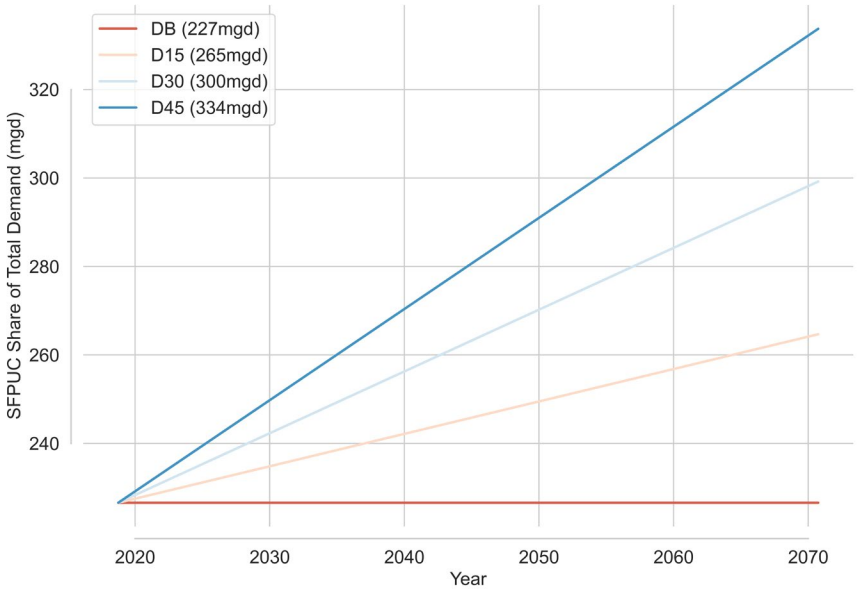


Figure 4-3. Annual Demand Scenarios on the SFPUC RWS (SFPUC's Share of Total Service Area Demand).
DB denotes baseline SFPUC demand and, D45 denotes a 45% increase in demand by 2070 relative to the baseline.

4.2.4 Stress Test Implementation: Step vs. Transient Changes

Each stress test composite scenario (precipitation + temperature + demand) is represented in two ways that offer different and complimentary perspectives on system performance: 1) *step*, where changes in climate and demand do not occur over time but instead occur over the entirety of the time series*, and 2) trend* (also called *transient* in the climate change literature), whereby changes in climate and demand occur linearly over time.

4.2.4.1 Step Change Scenarios

Due to the stochastic nature of the weather generator outputs, the calculation of several performance metrics requires many years of simulation over a stationary conditions (i.e., in terms of both demand and weather forcing) in order to ensure the estimated value of the metric converges to a reasonably representative value (cf. the discussion in Section 3.1.5 and illustrated in Figure 3-9). This is accounted with step change scenarios, which are used in the vulnerability assessment. Each step change scenario can be considered to represent the performance of the system as a snapshot in time under stationary conditions of climate and demand. Summary statistics across all ten realizations for each step change scenario are used to estimate performance metrics, and thus are derived from nearly 500 years of data.

Results of the climate stress test are commonly illustrated using a set of climate response functions showing the system response (e.g., change in performance). Two kinds of climate response functions are used in this report.

Response functions that are said to be 'univariate' show the system response (e.g., change in delivery reliability) that would result from change of a single factor of stress (e.g., change in temperature). When not indicated otherwise, it is assumed that all other factors of stress (e.g., precipitation) are baseline values. An example of univariate response function is illustrated in Figure 4-4.

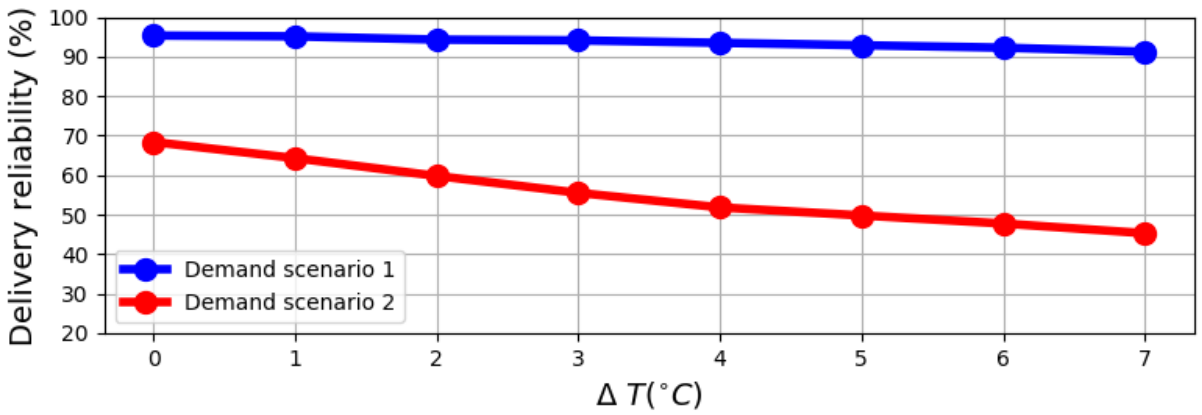


Figure 4-4. Example of Univariate Climate Response Function.

The x-axis shows the range of temperature changes that has been considered (with a step change of 1°C). The y-axis axis shows the estimated value for the considered performance metric (for this example, the water delivery reliability is used). Reliability values are also given for two scenarios of water demand (blue and red colors). Given the vulnerability assessment design considered in this study, each water delivery reliability for any given temperature scenario (i.e., represented as a bullet point) was obtained considering the 9 climate realizations, which means that its calculation used 500 years (or 490 years for metrics using SFWSM outputs). Note: delivery reliability values are here given for illustration purpose and do not correspond to the performance of the RWS.

Considering the example illustrated in Figure 4-4, one could for instance notice that under the demand scenario 1 (blue), the delivery reliability is close to 97% for the temperature baseline (e.g., ΔT=+0°C) and its sensitivity to temperature change is low. Conversely, under the demand scenario 2, the reliability obtained for the baseline temperature is significantly lower than the one obtained under demand scenario 1 (i.e., lower than 70%), and its sensitivity to temperature change is such that reliability decreases by almost 20% if temperature increases by 5°C (from 75% to 55%).

Bi-variate response functions are another typical decision-relevant output from the stress test approach. An example of bi-variate response function is illustrated for a representative performance metric in Figure 4-5. In this figure, delivery reliability is plotted in a surface plot with change in precipitation on the x-axis and change in temperature on the y-axis. Given a performance metric threshold above which performance is acceptable and below which performance is unacceptable, the surface plot is often colored to identify the combinations of precipitation and temperature change that result in future acceptable/unacceptable performance. Note that similar to Figure 4-4, delivery reliability metrics shown in Figure 4-5 results from the use of the ten climate realizations used for each point shown; however, they would be obtained for a single demand scenario.

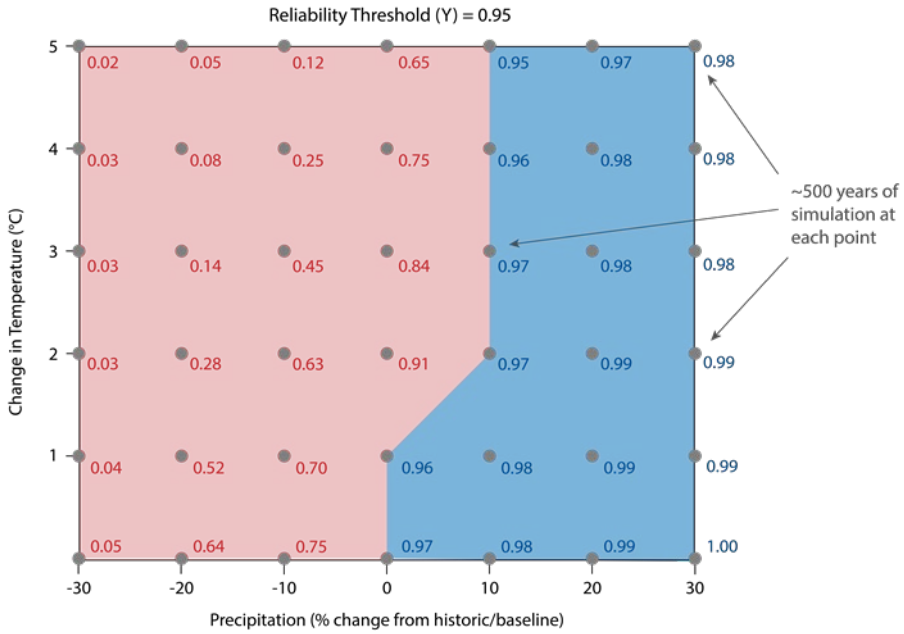


Figure 4-5. Demonstration of How a Climate Response Surface Is Constructed for a Reliability Metric with a Threshold of 0.95.

Red areas (left-hand side) indicate conditions under which system performance is not satisfactory, i.e., below 0.95. Blue areas (right-hand side) indicate conditions under which system performance is satisfactory, i.e., above 0.95. Note: delivery reliability values are here given for illustration purpose and do not correspond to the performance of the Hetch Hetchy RWS. Values not consistent with reliability values shown in Figure 4-4.

The bivariate response function illustrated in Figure 4-5 allows increased understanding of the system response to combined changes in two stress factors, in this case change in precipitation and change in temperature. For instance, under baseline condition (i.e., no change in precip and no change in temperature), the reliability is 0.97 or 97%. If temperature increases by two degrees the reliability gets to 0.91 or 91%, which is the example is below the target performance. However, if the two degree increase occurs while preciptiaton increases by 10% the reliability appears to be 0.97 or 97%.

The univariate (Figure 4-4) and bi-variate (Figure 4-5) reponse functions are useful vizualization tools that highlight the climate and demand conditions for which the system is vulnerable. However, i) the likelihood of these conditions are unknown at this stage and ii) how the change in system vulnerabilities and/or performance could manifest over time is not explicit either. To address these two points, one can map performance metrics against Global Circulation Model (GCM) projections of temperature and precipitation, as illustrated in Figure 4-6. The mean value across a 30-year moving window is taken for each GCM projection (in the case of Figure 4-6, temperature). Box plots show the range of performance across all the GCM projections at intervals across the time series.

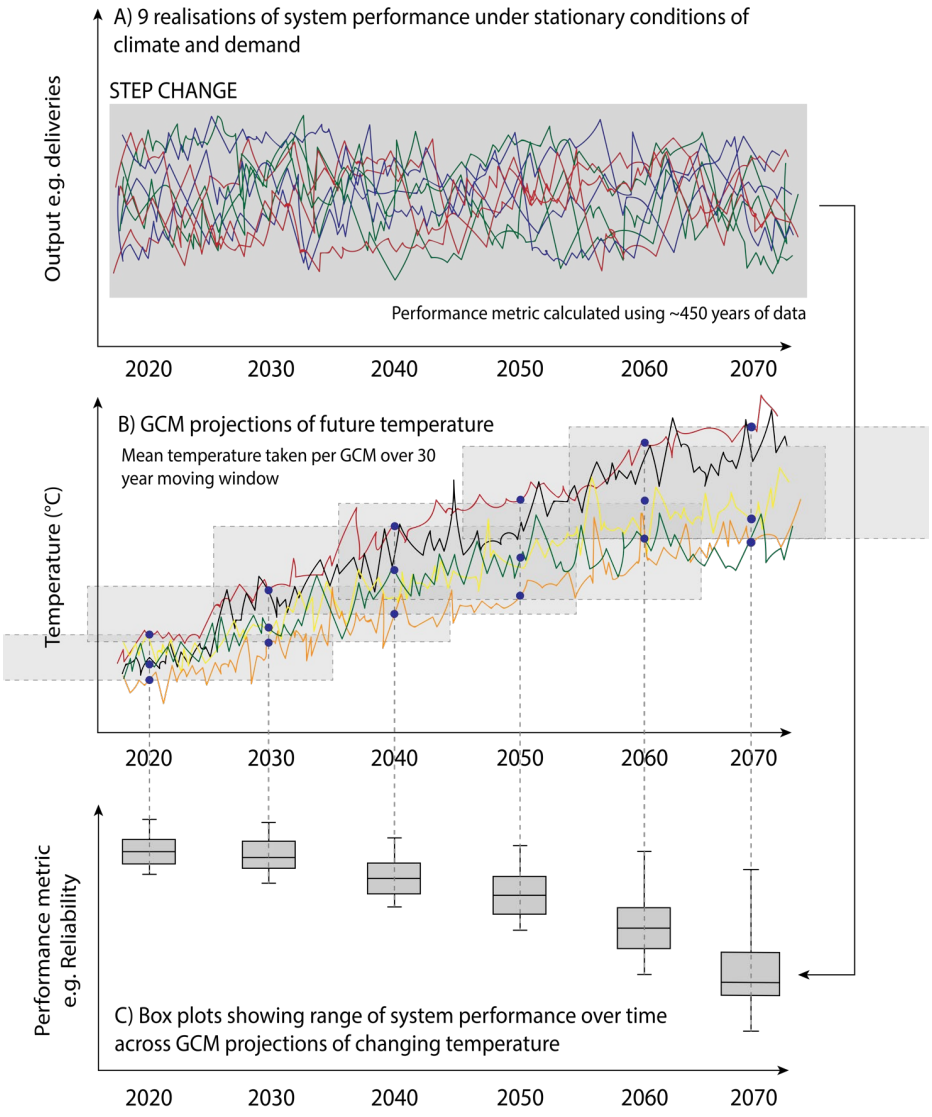


Figure 4-6. Example of How Results from Step Change Scenarios Can Be Mapped Against GCM Projections to Show How System Performance Is Likely to Change over Time.

4.2.4.2 Trend Change Scenarios

Trend scenarios, which are those depicted in Figure 4-3 for demand, define a representation of the reality of how system performance would manifest under continuous and linear changing conditions of climate and demand over time. The range of change in precipitation, temperature and demand considered for the Trend change scenarios is the same as the one used for step change (Table 4-3).

As far as the naming convention is concerned, the precipitation trend change scenario ΔP=-20% means that average precipitation decrease by 20% at the end of the time series. This also means that overall, the precipitation is less than the baseline by 10%.

Trend scenarios will be especially useful when implementing alternative supply options to reduce SFPUC vulnerabilities. However, some results obtained using the Trend scenarios are discussed in this report.

4.2.5 Limitations of the Stress Test

The perturbation of the weather (temperature and precipitation) and demand variables have been carried out following the incremental approach (or Delta Change approach). As such, potential changes in variability of these different variables are not investigated. This could be important to account for in future work as precipitation variability is expected to increase across California, even when and where precipitation is not expected to change on average (i.e., precipitation whiplash; Swain et al. 2018).

In addition, the considered stress test does not account for modification of the auto-correlation of the annual precipitation across the RWS. In other words, the stress test does not create longer sequences of meteorological droughts excepted for the ones created by reducing the whole precipitation time series by a specific multiplier.

It is also worth mentioning that CliWxGen samples all weather variables at the same time for the sake of consistency across variables. For instance, this means that through the KNN approach, if the daily precipitation for a specific month are selected, the minimum and maximum temperatures for this specific month will be sampled together with the precipitation variable. Although this is a very common practice in the literature, Diffenbaugh et al. (2015) shows that the likelihood of warm years being associated with dry years is likely to increase in the future. While the current implementation of the climate stress test somewhat allows such a change by reducing precipitation and increasing temperature, the change in co-occurrence of warm/dry years has not been explicitly accounted for.

4.3 Narrative-Based Scenarios

Several priority narratives were identified and developed through discussions between the HRG and SFPUC, and are described here along with the specific method for defining each narrative. The narratives are not exhaustive and represent a small number of potential futures or events. The intent of these narratives is to address important management issues. Specifically, they relate to new instream flow requirements (Section 4.3.1), infrastructure failure/outages (Section 4.3.2), and demand (Section 4.3.3). The narratives serve as the basis for narrative-based scenarios used in the analysis.

4.3.1 Environmental Narratives

SFPUC is interested in understanding the potential implications for water system performance of changes to the instream flow requirements (IFRs) under which it operates, as several new/expanded IFR schemes are likely to be imposed in the near future through regulatory processes. The objective of the IFR narratives is to answer the question: under what conditions do potential changes in IFRs cause the system to fail?

Several IFR scenarios were developed by SFPUC; these are summarized in Table 4-5 and described in detail below. No combination of IFRs were included in the stress test.

Table 4-5. Summary of Instream Flow Requirement Narratives.

The name (ID) convention for the environmental flow narrative is the following. 'NAE' stands for the Environmental nature of each Narrative. The first number gives the location (1: Hetch Hetchy, 2: Don Pedro water bank; 4: San Antonio reservoir and 5: Stone dam). The second number indicates the narrative index at the specific location.

ID	Name/location	Summary
NAE1.1	UTREP proposed base flows below O'Shaughnessy Dam	Proposed revised Tuolumne River base flow releases below O'Shaughnessy Dam under the Upper Tuolumne River Ecological Program.
NAE2.1	State-proposed New IFR below New Don Pedro Dam	Increase IFR to 40% of unimpaired Tuolumne River flow at La Grange from February 1st to June 30th, as proposed by the State Water Resources Control Board (SWRCB)
NAE2.2	Draft Tuolumne Voluntary Agreement New IFR below New Don Pedro Dam	Alternative to the State's 40% proposal.
NAE4.2	New IFR below Turner Dam (II)	New IFR in San Antonio Creek below Turner Dam (San Antonio Reservoir). No such IFR currently exists, nor is one planned, but one could be implemented to provide sufficient flow for frog species downstream of Turner Dam.
NAE5.1	New IFR below Stone Dam	New IFR in Pilarcitos Creek below Stone Dam. No such IFR currently exists, though SFPUC voluntarily releases 1.5 cfs below Stone Dam. New dam improvements would reopen environmental permits and an IFR will likely be required.

4.3.1.1 UTREP Baseflows below O'Shaughnessy Dam (NAE1.1)

This narrative includes the base flow instream flow requirements below O'Shaughnessy Dam proposed under UTREP. The UTREP base flows include both a refinement of the original water year type (WYT) classification scheme, which includes three types, and changes to both the magnitude and timing of flow volumes (Figure 4-7). The new WYT classification includes five types (instead of the original three) and emphasizes a balanced distribution around the 50th percentile of flows. Like the existing classification scheme, the future WYT depends on both cumulative precipitation (Jan-Jun) and inflow (Jul-Aug). The UTREP base flows not only are higher than the existing base flows, but also more closely align with the historical natural runoff regime, with flows shifted earlier and with some reduction in flows in the summer (see Figure 4-7). This new requirement would entirely replace the existing IFR, including the additional 64 cfs required when Canyon Power Tunnel flows exceed 920 cfs. A detailed description of the UTREP base flows, including their ecological motivation, is included in SFPUC (2014).

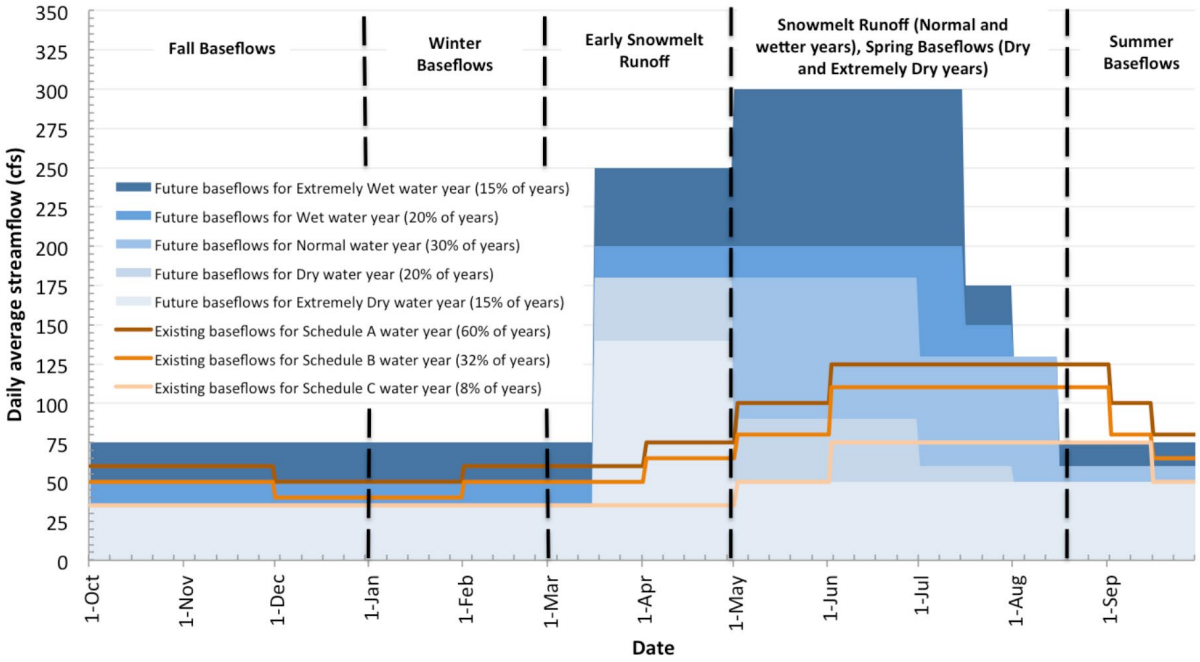


Figure 4-7. Existing and Future Base Flow Release Schedule in the Tuolumne River below O'Shaughnessy Dam under the Upper Tuolumne River Ecological Program.
Source: SFPUC 2014.

4.3.1.2 New IFR below New Don Pedro Dam (NAE2.1 and NAE2.2)

This narrative includes two related narratives, as follows.

SWRCB Proposal Adopted Water Quality Control Plan (NAE2.1)

The State Water Resources Control Board (SWRCB or State) adopted an amendment to the San Joaquin Bay-Delta Water Quality Control Plan in 2018 to increase flows in the San Joaquin River generally, which includes contributions from each contributing major tributary, including the Tuolumne River. The requirement for the Tuolumne River is 40% of unimpaired flow at La Grange from February 1st to June 30th. For the purposes of this study, any increase in flow requirements at La Grange are split 51.7 – 48.3% between SFPUC and the irrigation districts respectively as is possible under an agreement between the irrigation districts and SFPUC, known as the 4th Agreement. The SWRCB IFR narrative implements this. This narrative assumes a SFPUC share (51.7%) of additional flow requirement will be drawn from the Water Bank. Additional releases will be needed from Cherry Reservoir to keep the Water Bank full. The IFR will be proportional to unimpaired flow at La Grange, and the SFPUC share of the 40% of unimpaired flow requirement will be at the daily time step. When the existing flow regime is larger than the new flow schedule, the difference is credited to the Water Bank. This would supplement the existing FERC license requirements, which is the only IFR currently in place below Don Pedro (and serves as the baseline IFR in this study).

Draft Tuolumne Voluntary Agreement Proposal (NAE2.2)

The second Don Pedro IFR narrative is the draft Tuolumne Voluntary Agreement (March 1, 2019). The Tuolumne Voluntary Agreement is intended to replace the SWRCB adopted Water Quality Control Pan Amendment of December 2018.

4.3.1.3 New IFR below Turner Dam (San Antonio Reservoir) (NAE4.2)

Although there is currently no IFR in place below San Antonio Reservoir (Turner Dam), nor are any IFRs currently planned, it is expected that there is likely to be an IFR implemented within the time horizon

considered in this study. A potential impetus includes the need to provide sufficient flow for frog species downstream of Turner Dam. The IFR considered in this narrative includes adopting the IFR below Calaveras Dam for San Antonio Creek between December through June, scaled to the San Antonio watershed area, and no IFR from July through November.

4.3.1.4 New IFR below Stone Dam (NAE5.1)

There is currently no IFR on Pilarcitos Creek below SFPUC dams. However, SFPUC voluntary operates Pilarcitos Dam and Stone Dam to meet a target flow of 1.5 cfs year-round below Stone Dam. New dam improvements would re-open environmental permits and an IFR will likely be required below Stone Dam. This narrative includes and IFR below Stone Dam proportional to the IFR at Crystal Spring, in a similar fashion to the San Antonio IFR. In this case, the scaling factor is defined as the ratio between watershed areas (i.e., equal to 0.16). The new IFR would replace the voluntary 1.5 cfs releases currently in place and includes a minimum base flow of 2 cfs.

4.3.2 Infrastructure Narrative-Based Scenarios

Several infrastructure failure/outage narratives were developed in consultation with SFPUC staff to investigate the vulnerability of system operations to infrastructure failure, including subsequent recovery. There was no effort to identify an exhaustive list of potential failures. Instead, five narratives of importance were selected to demonstrate the approach. Narrative-based scenarios are summarized in Table 4-6 and described in more detail below. No combination of infrastructure narrative-based scenarios was included in the analysis.

Table 4-6. Summary of Infrastructure Narratives.

The name (ID) convention for the infrastructure narrative is the following. 'NAI' stands for the Infrastructure nature of the Narrative. The first number gives the typology of the narrative (1: water quality contamination, 2: dam safety; 3: Failure at the Moccasin switchyard; 4: Earthquake along Calaveras fault, and 5: Fire across Crystal Springs reservoirs watershed.

ID	Name	Summary of the aftermath for the system	Duration
NAI1	Water quality	High turbidity or TOC levels in Hetch Hetchy water leads to requiring filtration at the Sunol Valley Water Treatment Plant at a rate of 90 mgd (276 AF/day).	60 days
NAI2	New dam safety regulations combined with aging infrastructure (wear and tear)	Hetch Hetchy storage is reduced by 20%.	All years of the simulation
NAI3	Major failure at the Moccasin switchyard	San Joaquin pipelines are shut down.	365 days
NAI4	Earthquake along the Calaveras fault	The Sunol Valley Water Treatment Plant (SVWTP) is offline. The San Joaquin pipelines capacity is reduced to 120 mgd (368 AF/day) during the first 30 days and then 160 mgd (491AF/day) during the last 30 days.	60 days
NAI5	Major Fire across Crystal Springs reservoirs watersheds	Harry Tracy Water Treatment Plant is shut down.	365 days

Infrastructure narratives NAI1, NAI3, NAI4 and NAI5 are based on specific outages. For instance, the SFWSM simulation under NAI4 accounts for the aftermath of an earthquake along the Calaveras fault. The time of year at which the outage might occur may have a significant impact on the system's ability to cope with the additional stress caused by the failure. For instance, the consequences of an earthquake along the Calaveras fault could be different if it occurs during the summer season, when

demand is high, compared to the configuration for which it would occur during the winter season, when demand is low. To investigate potential issues resulting from the timing of the failure, four dates have been considered for each event. These dates have been selected to represent conditions for which the system could be either in a good or bad position when the outage occurs. Examples of such conditions include high storage or very low storage in the Bay Area reservoirs. Both storage levels and demand volume are obtained from the reference simulations (i.e., without narrative under baseline climate and demand conditions).

Although it is reasonable to think that an emergency rationing could take place at the onset of some outage events, no rationing other than the ones resulting for the drought is applied (no modification in the drought rationing was made). Rationing the deliveries is one possible response to minimize deficit throughout the duration of an outage, but others exist such as importing water from neighbor systems. Finding the optimal operational response to these outages requires a more comprehensive analysis that is beyond the scope of this vulnerability assessment.

4.3.2.1 Water Quality (NAI1)

Degradation of the quality of Hetch Hetchy reservoir water could follow from either a rain-on snow runoff (e.g., the 1997 flood) or a large snowmelt flood after a wildfire in the previous fall resulting in a high turbidity event. It could also occur after a drought, during which organic materials would accumulate across the watershed and become available for transport to the riverbed during a storm, and ultimately lead to a peak in TOC. Under this narrative, filtration treatment of Hetch Hetchy water at SVWTP prior to delivery is required for a duration of 2 months.

Hetch Hetchy water is treated at a rate of 90 mgd (i.e., 276AF/day, sustained capacity of the San Antonio pump station) at SVWTP. When Hetch Hetchy water is treated, San Antonio Reservoir water cannot be sent to SVWTP. Treating Hetch Hetchy water is a priority unless this results in spill at San Antonio Reservoir. Water from Calaveras Reservoir can be treated while water from Hetch Hetchy Reservoir is sent to SVWTP.

4.3.2.2 New Dam Safety Regulations Combined with Aging Infrastructure (Wear and Tear) (NAI2)

This narrative considers new dam safety regulations for O'Shaughnessy Dam combined with aging infrastructure (wear and tear). Specifically, the maximum Hetch Hetchy Reservoir operating level is reduced, resulting in storage capacity decreased by 20% in all years of the simulation. The guide curve for Hetch Hetchy Reservoir that defines the preferred level of storage is also reduced by 20%.

4.3.2.3 Major Failure at the Moccasin Switchyard (NAI3)

Under a Moccasin Switchyard failure, Foothill Tunnel flow (and, therefore, SJPL flow) is completely curtailed for a duration of 1 year. This narrative imagines a catastrophic failure at Moccasin Switchyard. If equipment begins to fail within the switchyard, HHWP will have to rely on the Moccasin Powerhouse Generator Bypass. The lead time for repair in the switchyard is about 13 months. Assuming the Moccasin Powerhouse Generator Bypass could be used for one month, this could result in a SJPL delivery outage for an extended period of time.

4.3.2.4 Earthquake along the Calaveras Fault (NAI4)

Under this narrative, an earthquake on the Calaveras Fault results in the following facilities to be out of service: Alameda Siphons No. 1, 2 and 3, partial outage of Bay Division Pipelines No. 1 – 4, partial outage of Calaveras Pipeline, San Antonio Pipeline, SVWTP and Pulgas Pump Station. This outage scenario was defined in the WSIP System Assessment for Levels of Service Objectives (Parsons 2006). The estimated repair times are 60 days for treatment plant and pump station major structural damages and 14-30 days for pipeline breaks. It is assumed that the cumulative repair time would be about 2 months with only

AS4 in service (AS4 has been specifically designed to resist a major seismic event). AS4 is only connected to Hetch Hetchy water (i.e., AS4 water cannot be sent to SVWTP for treatment). AS4 capacity during the event is assumed to be 368 AF/day (120 mgd) during the first 30 days, and to 491 AF/day (160 mgd) for the last 30 days. The result is that any water beyond these capacities cannot be transferred from Hetch Hetchy to the Bay area. To represent this, the Alameda siphons flow is reduced to 368 AF/day for 30 days and then 491 AF/day for the last 30 days; and SVWTP is offline.

4.3.2.5 Major Fire across Crystal Springs Reservoir Watershed (NAE5)

In this narrative, a large wildfire in the Crystal Springs and San Andreas watersheds leads to deposition of ashes on the reservoirs and potential contamination (heavy metals, TOC, PCB) of reservoir water. The following winter has heavy rainfall events generating large runoff from the burned areas resulting in turbidity events. A combination of raw water quality in Crystal Springs and San Andreas (turbidity, ashes, heavy metals, TOC, PCB) and public perception of raw water quality results in an outage of HTWTP for 12 months.

4.3.3 Wholesale Demand Narrative-Based Scenarios

Wholesale customers have a perpetual supply assurance of 184 mgd from the RWS going back to a 1984 Agreement (Table 4-7). The supply assurance is shared by 25 permanent customers (not including the cities of Santa Clara and San Jose). SFPUC has to supply the totality of demand for the City and County of San Francisco in addition of the obligation to the wholesale 184 mgd supply assurance.

Table 4-7. Individual Supply Guarantees for Wholesale SFPUC Customers.

Wholesale Customer	Individual Supply Guarantees (mgd)	Demand on the RWS in FY2012-13 (mgd)	Ratio between ISG and FY2013 demand
Alameda County Water District	13.76	9.06	0.66
Brisbane/GVMID	0.98	0.32	0.33
Burlingame	5.23	4.16	0.80
CWS - Bear Gulch	13.28	12.08	0.91
CWS - Mid Peninsula	14.66	14.04	0.96
CWS - South San Francisco	7.74	6.89	0.89
Coastside County Water District	2.18	1.67	0.77
Daly City	4.29	4.13	0.96
East Palo Alto Water District	3.46	2.08	0.60
Estero MID	5.90	4.05	0.69
Hayward	22.08	15.48	0.70
Hillsborough	4.09	3.25	0.79
Menlo Park	4.46	3.24	0.73
Mid-Peninsula	3.89	3.00	0.77
Millbrae	3.15	2.28	0.72
Milpitas	9.23	6.63	0.72
Mountain View	12.46	9.09	0.73
North Coast County Water District	3.84	2.51	0.65
Palo Alto	16.58	11.33	0.68
Purissima Hills Water District	1.63	1.99	1.23
Redwood City	10.93	9.31	0.85
San Bruno	3.25	2.01	0.62
Santa Clara	-	2.18	-
San Jose	-	4.50	-
Stanford University	3.03	2.14	0.70
City of Sunnyvale	12.58	9.54	0.76
Westborough Water District	1.32	0.91	0.69
Cordilleras MWC		0.01	

Figure 4-8 provides a summary of the year in which each permanent customer would exceed their ISG specified volume under the different demand scenarios considered in this report (under trend model runs). The year in which total system demand exceeds 184 mgd is also provided. For example, considering total system demand, supply assurance of 184 mgd is surpassed in 2043 under 300 mgd demand scenario (+30% relative to baseline) and in 2036 under a 334 mgd demand scenario (+45%).

The cities of Santa Clara and San Jose are not part of the 184 mgd supply assurance and thus are not permanent customers. As a result, if Wholesale Customers' projected demand exceeds 184 MGD, the SFPUC may issue a conditional five-year notice of interruption or reduction in supply of water to San Jose and Santa Clara. SFPUC is interested in understanding what the performance implications would be of making San Jose and Santa Clara permanent customers and therefore increase the supply assurance above 184 mgd. To explore this question, two narrative-based scenarios are considered as summarized in Table 4-8 and described in more detailed below.

Table 4-8. Summary of the Demand Narratives.

ID	Summary of the narratives	SFPUC wholesale customer demand
NAD1	Wholesale customers supply assurance is limited to 184 mgd. Santa Clara and San Jose are not permanent SFPUC customers. SFPUC retail demand continues to grow.	184 mgd
NAD2	Wholesale customers are delivered their ISG only. Santa Clara and San Jose are permanent SFPUC customers. SFPUC retail demand continues to grow.	199.5 mgd

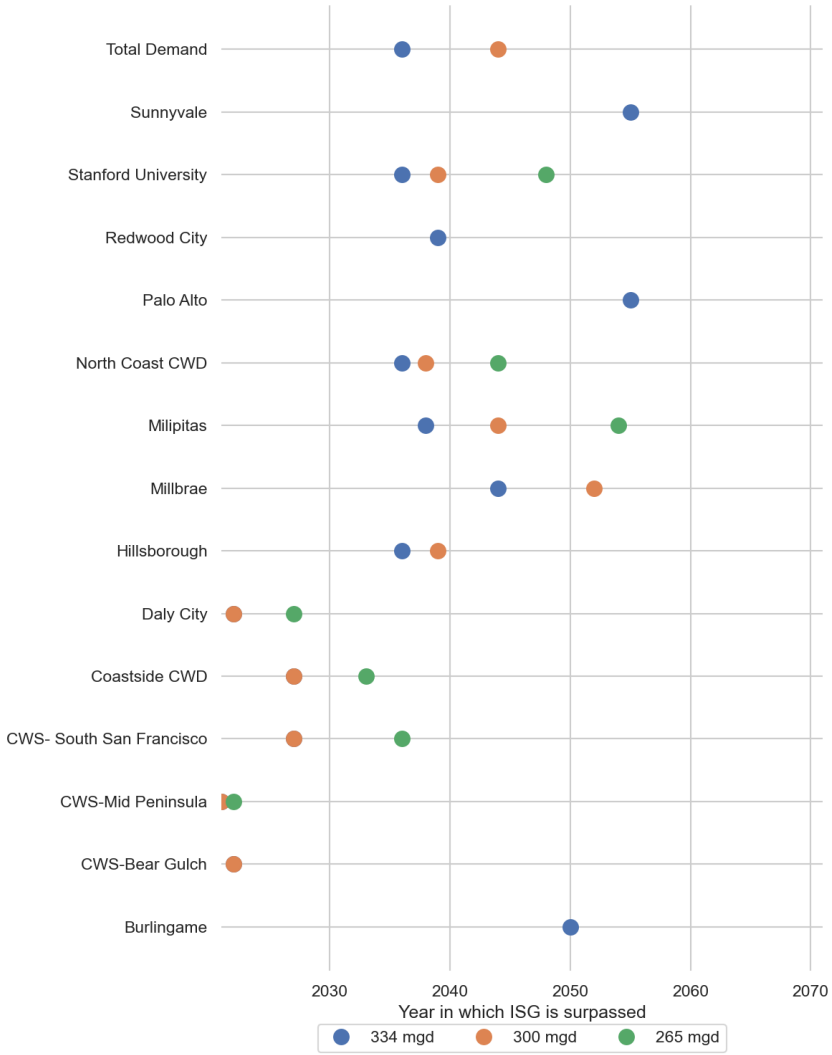


Figure 4-8. Plot Showing the Year in Which Customer's Demand Surpasses Their Agreed ISG Value under the 265 mgd (Green Dots), 300 mgd (Orange Dots), and 334 mgd (Blue Dots) Demand Scenarios.
Customers that do not appear in this plot do not surpass the ISG value under any of the demand scenarios explored here. No customers surpass their ISG value in the 227 mgd demand scenario.

4.3.3.1 Wholesale Customers' Supply Assurance Limited to 184 mgd (NAD1.1)

In this narrative, existing permanent customers are assigned a constant demand at their ISG level across all demand scenarios. In accordance with the 2009 WSA, the City of Hayward is assigned a value of demand that represents the difference between the total supply assurance (184 mgd) and the sum of all permanent customer ISGs (22.8 mgd). Non-permanent customers are assigned a demand of zero (San Jose and Santa Clara). San Francisco demand on the other hand is varied across demand scenarios in

accordance with the baseline demand narrative. Other retail customers have a constant demand across all demand scenarios in accordance with the base narrative.

4.3.3.2 Add Santa Clara and San Jose as Permanent Customers (NAD1.2)

SFPUC is interested in the performance implications of making Santa Clara and San Jose permanent customers. This narrative differs from demand narrative 1 only with regard to Santa Clara and San Jose being included as permanent customers with an ISG value of 6.5 mgd and 9 mgd, respectively.

4.3.4 Integration of the Narrative-Based Scenarios within the Climate Stress Test

Environmental, demand and infrastructure (NAI2 only) narrative-based scenarios were applied using a coarse-resolution stress test approach (i.e., with varying combinations of precipitation, temperature, and/or demand) (Table 4-9).

Table 4-9. Summary of Climate Change Scenarios for the Coarse Climate Stress Test Combined with Environmental, Demand, and Infrastructure (NAI2) Narratives.

Type of uncertainty	Sampling range	Sample size
Natural climate variability	Stochastic realizations	10 realizations
Changes in mean annual precipitation (%)	-30 % to 30 % with 15% increments	5 change factors
Changes in mean annual temperature (°C)	0 to 6°C with 2°C increments	4 change factors
TOTAL		200 climate scenarios

Infrastructure narratives (other than NAI2) were also applied using a coarse-resolution stress test approach with the difference that not all ten realizations were considered (Table 4-10). Dates for each specific infrastructure narratives are given in the relevant sub-sections of Section 5.6.

Table 4-10. Summary of Climate Change Scenarios for the Coarse Climate Stress Test Combined with Infrastructure Narratives (Other Than NAI2).

Type of uncertainty	Sampling range	Sample size
Timing	Starting date for the infrastructure failure/outage	4 dates
Changes in mean annual precipitation (%)	-30 % to 30 % with 15% increments	5 change factors
Changes in mean annual temperature (°C)	0 to 6°C with 1°C increments	4 change factors
TOTAL (for each narrative)		80 climate scenarios

4.4 Performance Metrics

For the LTVA, performance metrics are necessary to help identify key system vulnerabilities and subsequently inform decisions about infrastructure and policy interventions. In other words, they help answer the original questions posed in the LTVA: “under what conditions and when will the water system no longer meet system performance criteria?” The LTVA seeks to provide SFPUC with sufficient information to make informed diagnoses of system vulnerabilities and, later, to make informed choices given these tradeoffs.

Performance metrics allow decision makers to assess the performance of the system against some defined failure thresholds for any given system variable of interest. The magnitude of failure is the difference between the desired (target) value of a variable and the actual value of the variable in any given time step. Magnitude of failure in any given time step is defined for calculating performance metrics as:

$$X_t = \begin{cases} X_t - X_{target}, \textit{if } X_{target} > X_t \\ 0 \textit{ otherwise} \end{cases} \quad (4\text{-}1)$$

where X is the variable of interest, X_t is the actual value of X at time t, and X_{target} is the desired value of X. For some outputs, $X_{target} < X_t$ in Equation 4-1 (e.g., water quality targets).

4.4.1 Water Supply

Several key performance metrics related to water supply were calculated. These include water delivery reliability, frequency of 20% rationing, duration of rationing, system wide storage reliability, maximum deficit, and cumulative deficit. For the duration of rationing and storage reliability metrics, the performance of the RWS under WSIP 2018 at 265 mgd demand over the Design Drought is used as a point of comparison. The Design Drought is a synthetic planning sequence of 1987-92 appended with 1976-77 creating an 8.5 year drought event. The key performance metrics are defined as:

- **Water delivery reliability** – The frequency of years in which the system delivers full demand (i.e., no rationing is applied) across the planning horizon. Mathematically, reliability is defined as the probability that the system will meet given a criteria (Hashimoto et al. 1982):

$$Reliability = Prob\,[X_t \in S] = \frac{\text{No. of times } X_t = 0}{\text{No. of time steps}} \text{x}100 \quad (4\text{-}2)$$

  where S is the set of all times when water delivery is greater than or equal to system demand (failure is when deliveries are less than demand), after accounting for rationing. Reliability ranges from 0 to 1 (higher is better) (or 0 to 100%). The target is 90% (or 1 year of rationing out of 10 on average). This target is grounded in the performance of the RWS with the planned WSIP 2018 at a level of demand of 265 mgd over historical hydrology.
- **Frequency of 20% rationing** – Frequency of years during which a mandatory 20% rationing is applied. Such a rationing is applied when the years of remaining supply (YRS) gets below 3.14 on July 1st. The target is for the frequency to remain below 5% of years (on average 1 out of 20 years). This target is grounded in the performance of the RWS with the planned WSIP 2018 at a level of demand of 265 mgd over historical hydrology.
- **Duration of rationing** – Duration of rationing is the consecutive number of years that rationing is imposed. This metric is further divided based on the rationing type: mandatory rationing only (i.e., 20% rationing) or any rationing (i.e., 10% voluntary rationing or 20% mandatory rationing). In this report, the focus is on the maximum duration across the nine realizations. The target for maximum duration of rationing is 5 years as in the 1987-1992 drought and 6.5 years over the SFPUC Design Drought.
- **Water storage reliability –** The frequency of years during which the years of remaining supply is above 2.3 years of remaining supply. This target is grounded in the amount of storage in the RWS at the end of the 1987-1992 drought in the SFPUC Design Drought under WSIP 2018.
- **Maximum annual supply deficit** – Annual supply deficit is the annual sum of the deviations between pre-rationing demand and actual deliveries (Figure 4-9). The maximum value obtained across the 10 realizations is reported.
- **Maximum cumulative supply deficit through rationing events** – The cumulate supply deficit through a rationing event is total volume of deficit within a single rationing event, which can be cumulated across multiple years within a continuous sequence of rationing (Figure 4-9). The maximum value obtained across the 10 realizations is reported.

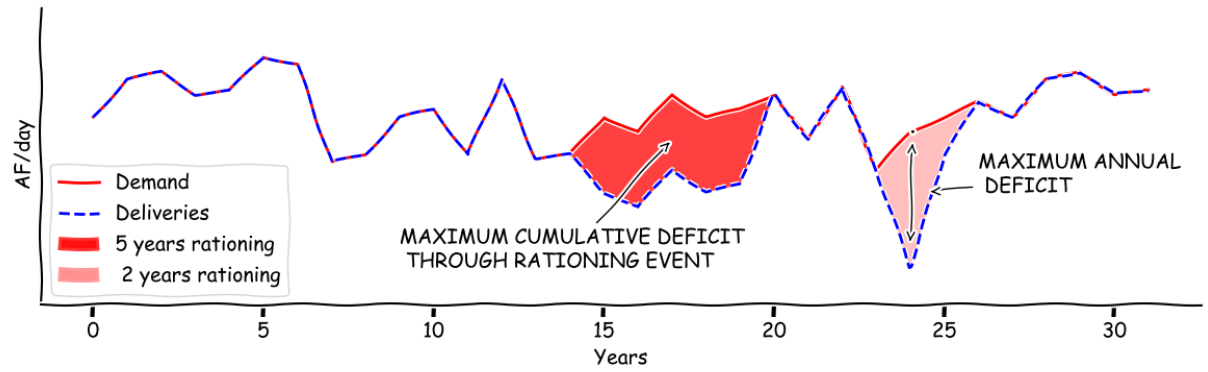


Figure 4-9. Sketch Illustrating the Maximum Annual Cumulative Deficit and the Maximum Cumulative Deficit through Rationing Event.

The example shows the metrics for only one realization.

Each of the above metrics are derived by considering the 9 stochastic realizations and the historical realization, and for each step change scenario and are presented herein in the form of climate response functions (see example of climate response functions in Figure 4-4 and Figure 4-5).

4.4.2 Upper Tuolumne River Ecosystem Program

The performance of the UTREP spill management program is assessed. This program is designed to support specific ecological functions in the Tuolumne River below Hetch Hetchy reservoir (SFPUC 2014). Generally, different spill volume ranges can be used to support specific species/functions as in Table 4-11. Note that "spill" here includes controlled releases via both the valves and the spillway drum gates.

For instance, 'good' years for the frog population below Hetch Hetchy reservoir are years when the total volume from April through August is lower than 9,800 AF. On the other hand, 'good' years for sediment transport are years when the total volume release to the river is larger than 178,000 AF.

Table 4-11. Primary Ecosystem Beneficiaries for Different Seasonal Spill Volumes from Hetch Hetchy Reservoir.

April-August spill volume threshold	Ecosystem beneficiary
< 9,800 AF	Frogs
> 9,800 AF	North Pond and Low Wetlands inundation
> 133,000 AF	High Wetlands inundation
> 178,000 AF	Sediment transport

The change in the frequency of years during which the spill volume gets above the thresholds specific to the different ecological functions (Table 4-11) under changes in climate and demand is evaluated in Section 5.4. The effect on the new UTREP base flows below O'Shaughnessy Dam (NAD1) on these frequencies is also discussed in this section.

4.4.3 Water Quality

In order to maintain water quality at acceptable levels, SFPUC needs to monitor water quality at different locations of the RWS and if contaminants are detected, trigger adapted operations, if needed. For instance, if the turbidity level of the Hetch Hetchy water exceeds the level of filtration avoidance at Tesla Portal, the water must be filtered at the Sunol Valley Water Treatment Plant (SVWTP) or

discharged. There is no direct violation with TOC but the combination of elevated TOC and chlorine disinfectant can lead to violation of DBPs threshold. Specific thresholds are set for turbidity and TOC:

- For turbidity at Tesla Portal, SFPUC's objective is to keep the level below 1 NTU. Serving water with a turbidity of 2 NTU is manageable in short-term but not acceptable in long term for some wholesale customers. The organizational goal is to provide the best possible product and 2 NTU water falls short of such a goal. The filtration avoidance threshold is 5 NTU and should not be exceed more than twice in 12 months.
- For TOC at Tesla Portal, a concentration of 2 mg/L is defined as a 'level of concern' while 3 mg/L could lead to violation of DBPs.

To assess the vulnerability of the RWS regarding these water quality indicators (Section 5.3), three metrics that match the SFPUC's objectives mentioned above were used. First, changes in daily mean of water quality indicators (turbidity and TOC) are obtained from the median prediction model.

Second, the change in reliability for not exceeding the above threshold is assessed. The reliability is the frequency of time the water quality threshold is not met at Tesla Portal and at OSHD. This criteria is further termed as reliability and is estimated following Equation 4-2.

- For turbidity, reliability represents the frequency of years for which maximum daily turbidity does not exceed 5 NTU at least twice in a 12-month rolling window. Following the discussion on the water quality prediction model (Section 3.5.4) three reliability metrics will be discussed. Each metric will be estimated using either the 50^{th}, 75^{th} or the 97.5^{th} predicted percentiles.
- For TOC, the reliability represents the frequency of years for which maximum daily TOC does not exceed a specific TOC threshold. Following the discussion on the water quality prediction model (Section 3.5.4), it is unlikely that predicted TOC gets values above or significantly above 3 mg/L. For this reason, the reliability regarding the TOC levels at the Tesla Portal will be assessed for two thresholds; 3 mg/L, which is the TOC level that might lead to violation of DBPs, and the 2 mg/L, which is level of concern for SFPUC.

4.4.4 Finance

Water rates are used to evaluate the impact of changing conditions in climate, demand, and expenditure on financial performance. The price setting logic of the finance model is based on cost recovery (see 3.6.2.2) and thus, changes in this value across different scenarios provides an indication of the cost of adaptation and of the financial impact of changing conditions on consumers.

4.4.4.1 Water Rate Increase

This metric is used to evaluate the rate at which water prices are required to increase from one year to the next in order to cover costs. Failure is defined here as an increase in price of more than 10% from one year to the next for either Wholesale or Retail customers. This metric is particularly useful when considering results of trend model runs and the impact of increasing capital and operational expenditure over time.

4.4.4.2 The Price of Water

The absolute price of water is a useful indicator when considering the results of the financial stress test under step change model runs in which capital and operational expenditure are constant over time.

CHAPTER 5

Vulnerability Assessment

Results for the most important performance metrics of interest, as well as important intermediary system variables for context, are presented here. All results discussed in this section have been obtained using the step change climate stress test across the nine stochastic realizations plus the historical realization, except when explicitly mentioned.

5.1 Effects of Climate Change and Reservoirs Inflows and Water Available to the City

The climate stress test was applied to all three watersheds of the Upcountry region (Section 5.1.1), East Bay region (Section 5.1.1.3), and the Peninsula region (Section 5.1.3) over the period of 2021 through 2070. Changes in temperature range from no change to +7°C. Change in precipitation change range from -40% to +40%. The internal climate variability is accounted for using 10 realizations. The effect of climate change on the frequency of hydrologic droughts on the Tuolumne River is also presented in Section 5.1.1.

5.1.1 Upcountry Region

5.1.1.1 Streamflow

The climate stress test was applied across all three watersheds in the Upcountry region (Hetch Hetchy, Cherry and Eleanor, and Don Pedro). Figure 5-1 shows the average monthly hydrograph for the baseline climate and +5°C warming. Figure 5-2 shows the average temporal shift of the center of mass of the hydrograph across the Upcountry region (i.e., broadly speaking, this metric shows the average temporal shift of the spring runoff due to warming conditions; negative values means that the spring runoff occurs earlier in the year).

The hydrology regimes of the Upcountry watersheds are significantly dominated by the snowpack accumulation and melt, with the exception of the Don Pedro accretion watershed for which the influence of snowpack dynamic is less. Therefore, an increase in temperature would result in less snowpack accumulation during winter season, as more precipitation will fall as rain rather than snow, leading to 1) a reduced spring runoff volume and magnitude and 2) an increase in winter flows. The spring runoff is also expected to occur earlier in the season (Figure 5-2). For Cherry Reservoir/Lake Eleanor watershed, a significant increase in flow during winter season highlights major change in the phase of the precipitation across this watershed (much more precipitation would fall as rain rather that snow). Such a significant increase in winter flows is not observed at the outlet of the Hetch Hetchy Reservoir watershed, likely because of its higher elevation range, which might still see most of the precipitation falling as snow. As mentioned above, the Don Pedro accretion watershed is a precipitation-driven watershed with minor to moderate snowpack influence. For this reason, the above described changes remain valid for this watershed, although their magnitude is lesser. The modifications of the monthly average cycle and temporality of the spring runoff at La Grange is a combination of the change discussed above for the three upstream watersheds (Figure 5-1 and Figure 5-2).

In summary, warming temperature has an effect on the timing of the annual hydrograph at the Upcountry reservoirs:

- A +2°C warming leads to a spring runoff arrival 10 days prior to the baseline temperature conditions.

Median of projections estimate warming around +2°C by 2040 with most projections and elicitations between +1°C and +4°C. At +4°C, the shift in timing would be closer to 20 days prior to baseline conditions. By 2070 RCP8.5, warming could reach around +4°C with most projections and elicitations between +3°C and +6°C.

- The shift in timing would be about the same for the inflow at Hetch Hetchy Reservoir.

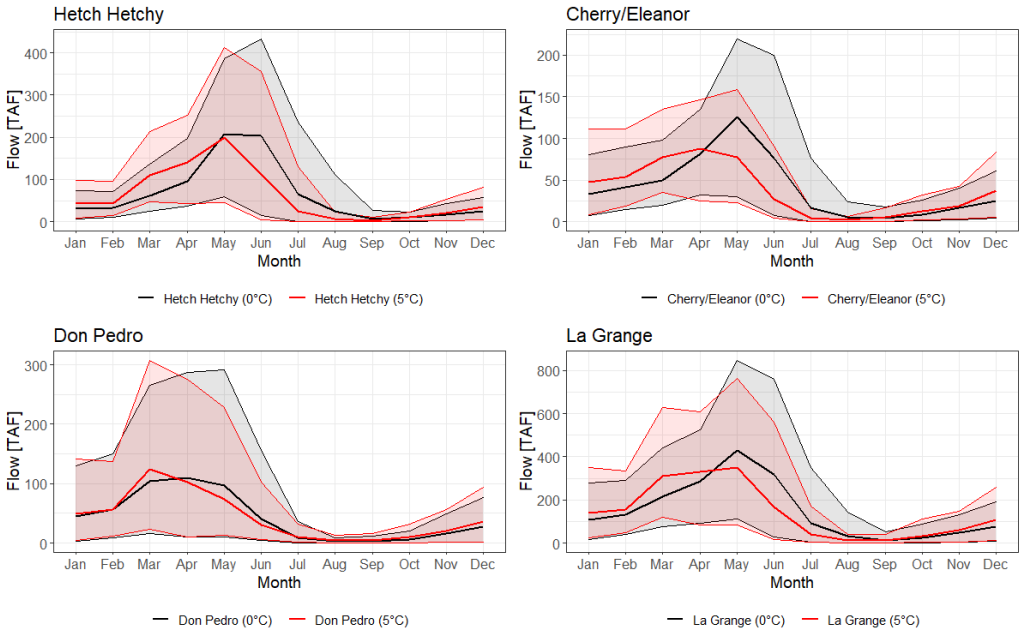


Figure 5-1. Effect of Warming Temperature on the Hydrograph across Upcountry Region.
Black and red curves show the results obtained for the baseline climate (no change in temperature and precipitation) and +5°C warming, respectively. The shaded areas show the deviation between the 5th and 95th percentiles, while the middle curves show the average.

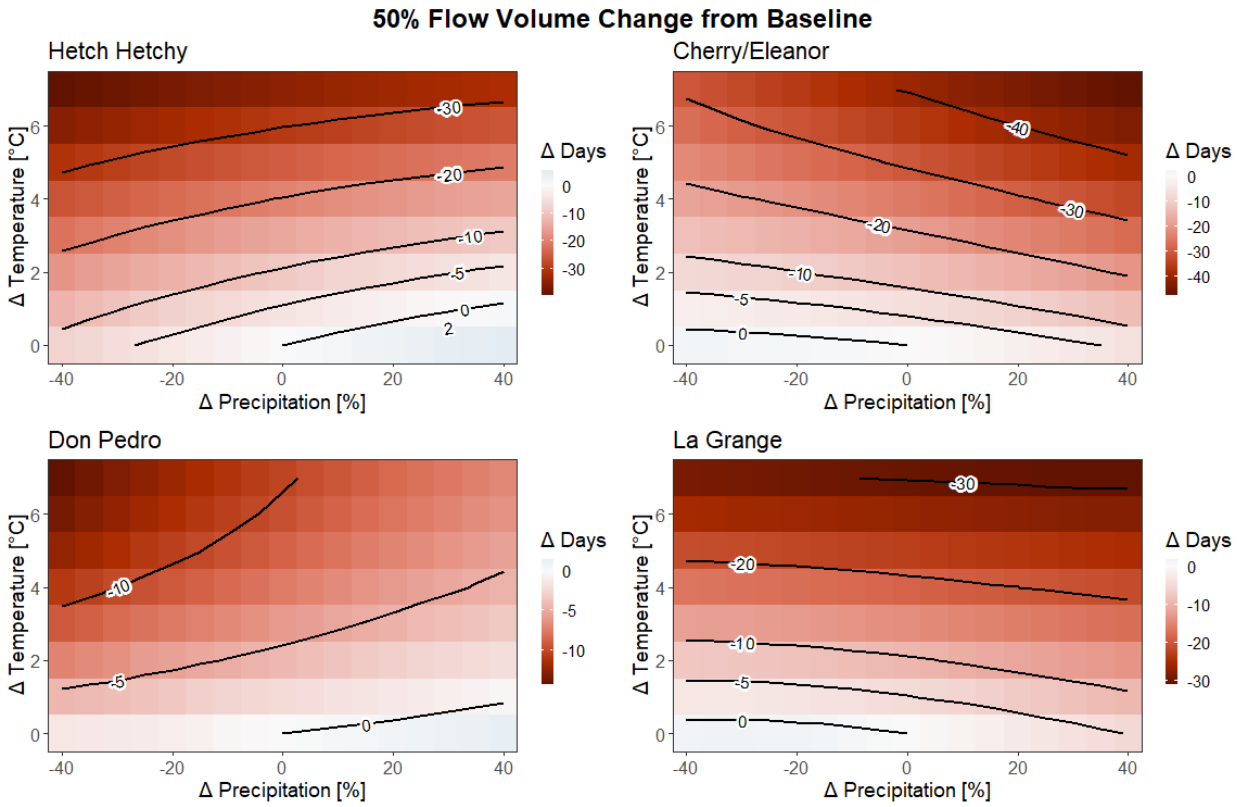


Figure 5-2. Change in Days of the Date at Which 50% of the Total Annual Volume Is Observed.
Negative values show that the center of mass of the hydrograph occurs earlier in the year.

Figure 5-3 illustrates the modification of the annual streamflow across the Upcountry region that would result from changing precipitation and warming temperature. The change in precipitation is by far the main driver of change in annual streamflow. More precipitation means larger streamflow across the region, and inversely. Warming temperature does reduce annual flow, although the absolute change is low, especially for Don Pedro accretion watershed. The difference between the response of Hetch Hetchy and Cherry/Eleanor watersheds, and the one of Don Pedro watershed, is that the formers seem energy constrained while the latter seems water constrained. For instance, it is understood that under baseline climate, the potential evapotranspiration (PET) over Don Pedro accretion watershed is not constraining the actual evapotranspiration (AET) what does is the quantity of water available for evapotranspiration. As such, even if an increase in temperature leads to an increased PET, AET cannot increase because of the lack of water. For the upstream watersheds (i.e., upstream Hetch Hetchy Reservoir and Cherry Reservoir/Lake Eleanor), the situation seems to be opposite. Under baseline climate conditions, PET is lower than the available water for the evapotranspiration. In this case, larger PET due to warmer temperature would lead to an increase in AET and eventually a decrease in streamflow.

Given the climate projections and elicitations, annual streamflow across the Upcountry region can either increase or decrease making predictions around effects on streamflow deeply uncertain. The uncertainty in change in streamflow is mainly driven by the uncertainty in the precipitation projections:

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation leads to small decrease of mean annual flow at Upcountry Reservoir. Most projections and elicitations fall between -5% and +5% change in mean annual precipitation by 2040, which would correspond to a change in inflow between 8% decrease and 8% increase at La Grange and 6%

decrease and 6% increase at Hetch Hetchy Reservoir.

- By 2070 RCP8.5, warming and precipitation change could reach around +4°C and 0% change, respectively, leading changes in annual inflow of a decrease of 2% at La Grange and a decrease of 3% at Hetch Hetchy Reservoir. Most projections and elicitations of warming are between +3°C and +6°C and for precipitation change between -15% and +15% resulting in potential change in inflow of a decrease of 24% at La Grange and a decrease of 21% at Hetch Hetchy Reservoir.
- By 2070, the uncertainty is such that the change of annual flow at La Grange roughly ranges from a 30% decrease to a 60% increase, and the changes at Hetch Hetchy Reservoir ranges from a 27% decrease to a 37% increase.

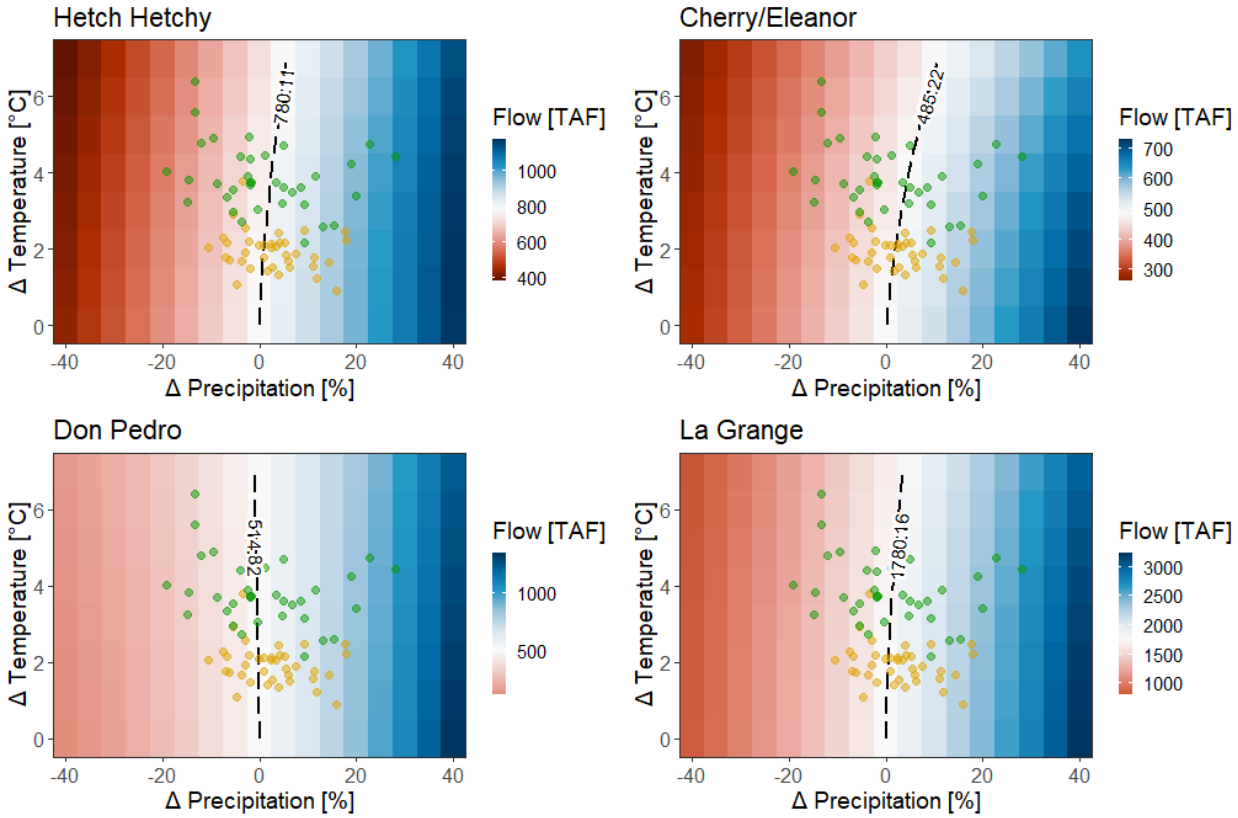


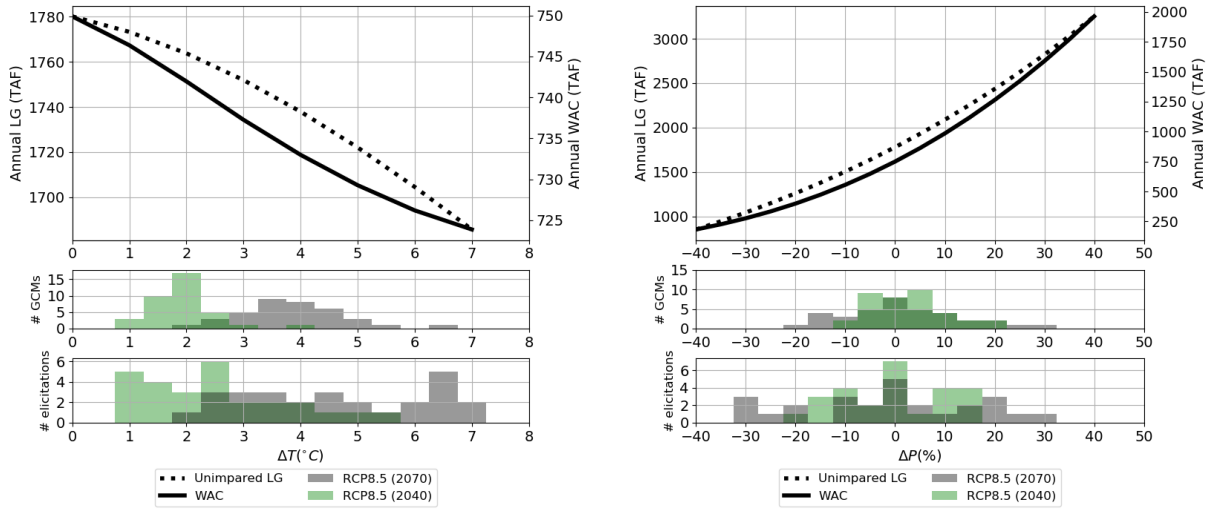
Figure 5-3. Climate Response Surface of the Annual Flow at Hetch Hetchy Reservoir, Cherry Reservoir/Lake Eleanor, Don Pedro Accretion, and Tuolumne River at La Grange.

For all response functions, white color shows no change, while blue and red colors represent an increase or a decrease in annual flow, respectively. The yellow and green dots over the response surfaces show the CMIP5 projections under the RCP8.5 for two 30-yr long periods: 2040 (2026-2055) and 2070 (2056-2095). Baseline is 1986-2005. See Section 2.2 for more details regarding the climate projections.

5.1.1.2 Water Available to the City (WAC)

The modifications of the Tuolumne River flow at La Grange under changing precipitation and warming temperature affects the annual values of WAC and its distribution through the years. Figure 5-4 illustrates the effects of warming temperature (left) and changing precipitation (right) on the Tuolumne River flow at La Grange and WAC. Following the response of the Tuolumne River flow at La Grange, precipitation change is the main driver of change of annual WAC. For example, a decrease in precipitation by 20% leads to a reduction of the annual average WAC from roughly 750 TAF to 400 TAF (i.e., a decrease by roughly 45%), while an increase by 20% precipitation leads to an increase of WAC from 750 TAF to 1260 TAF (i.e., +68%). At the annual scale, warming temperature slightly reduces the

WAC. For instance, a +5°C warming decreases the annual average WAC from roughly 750 TAF to 729 TAF (i.e., a decrease by -2.8%). Figure 5-5 shows the modification of the annual average WAC to change in both temperature and precipitation. It highlights no significant compound effect on the annual WAC.



a) Effect of temperature change b) Effect of precipitation change

Figure 5-4. Effect of Temperature and Precipitation Change on the Unimpaired Flow at La Grange and Water Available for the City (WAC).

Top panel shows unimpaired flow at La Grange (solid line) and WAC (dash line) with changes in temperature (ΔT) (a) and precipitation (ΔP) (b). Middle and bottom panels show the distribution of changes in temperature from CMIP5 projections (RCP8.5) and expert elicitations and for two 30-yr long periods centered in 2040 and 2070.

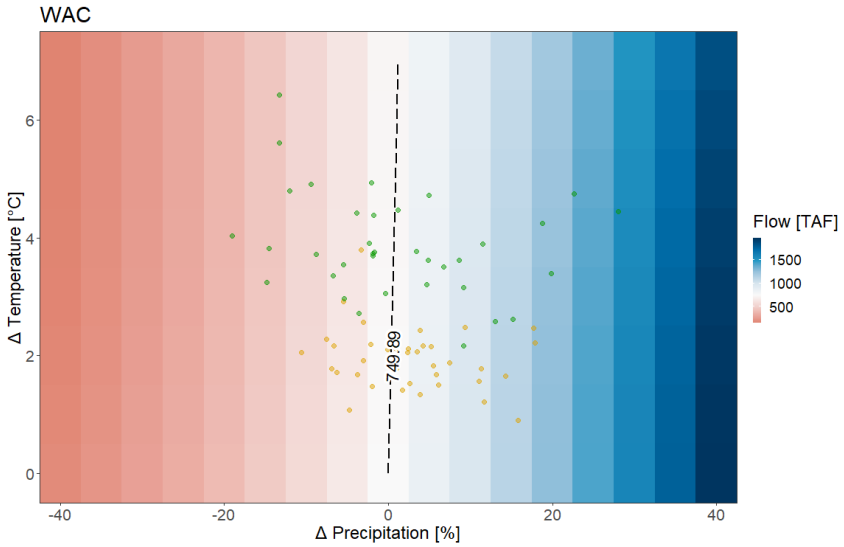


Figure 5-5. Climate Response Surface of the Annual WAC in Respect with Change in Precipitation (X-axis) and Temperature (Y-axis).

See Figure 5-3 caption for details.

It is important to note that behind the non-significant change in annual WAC that would follow from warming temperature (Figure 5-4, left), the WAC is actually expected to decrease for roughly 60% of the years, and increase for 40% of the years (Figure 5-6). As illustrated Figure 5-6, the distribution of change

in annual WAC is left skewed, which means that absolute values of reduction in WAC can get much larger values than for the increases in WAC.

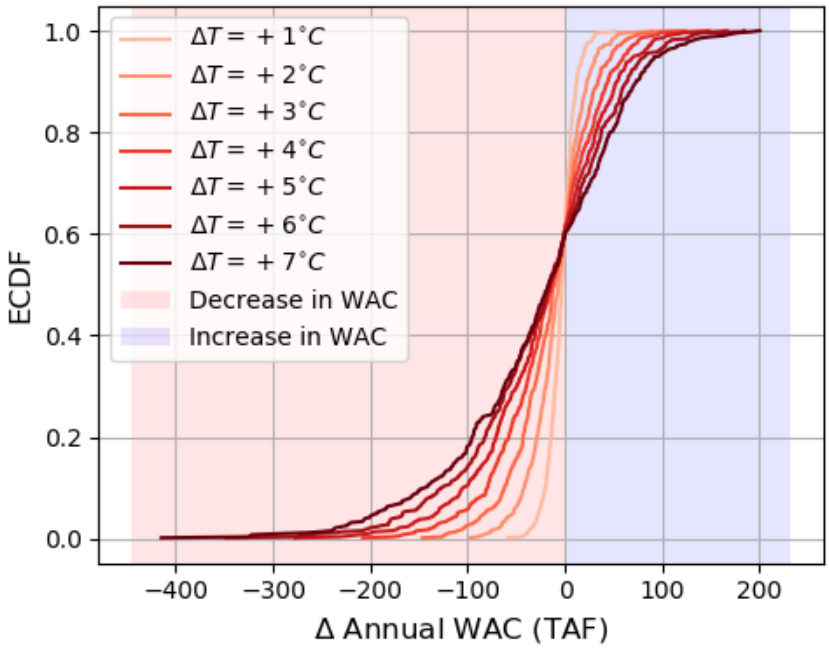


Figure 5-6. Empirical Cumulative Distribution Function (ECDF) of Change in Annual WAC Resulting from Warming Temperature.

Positive values (blue panel) shows years for which WAC increases and negative values (red panel) shows years for which WAC decreases. Results are obtained across the 9 stochastic realizations plus the historical realization.

Figure 5-7 illustrates the modification of the distribution of annual WAC under changing precipitation and temperature. Note that changes in all quantiles of the WAC distribution are not same. For instance, when compared the 20% precipitation reduction scenario (light red, Figure 5-7 left) with the baseline precipitation scenario (black, Figure 5-7 left), it is noted that the median value (i.e., 50% exceedance probability) decreases from 600 TAF to 300 TAF, which corresponds to a decrease by 50%. However, the quantile 10 of the distribution (i.e., 10% exceedance probability) decreases from 100 TAF to 40 TAF, which corresponds to a decrease by 60%. Figure 5-7 (left) also shows that under baseline climate conditions, roughly 25% of the years have a WAC values lower than 269 TAF while a reduction in precipitation by 20% increases this number to almost 45%, which also explains the significant increase in frequency of droughts.

The distribution of annual WAC obtained from the 510 realizations under warming conditions almost lines up with the one obtained under baseline conditions (Figure 5-7, right). This results is consistent with the lack of significant change in annual WAC obtained in Section 5.1.1.2 where only 10 realizations were used (Figure 5-4 and Figure 5-5). Despite the lack of significant change in annual WAC, the lower percentiles of the distribution (i.e., lower than 15th) slightly increases (Figure 5-7, right), which likely explains the reduction is drought frequency (i.e., increase in return period).

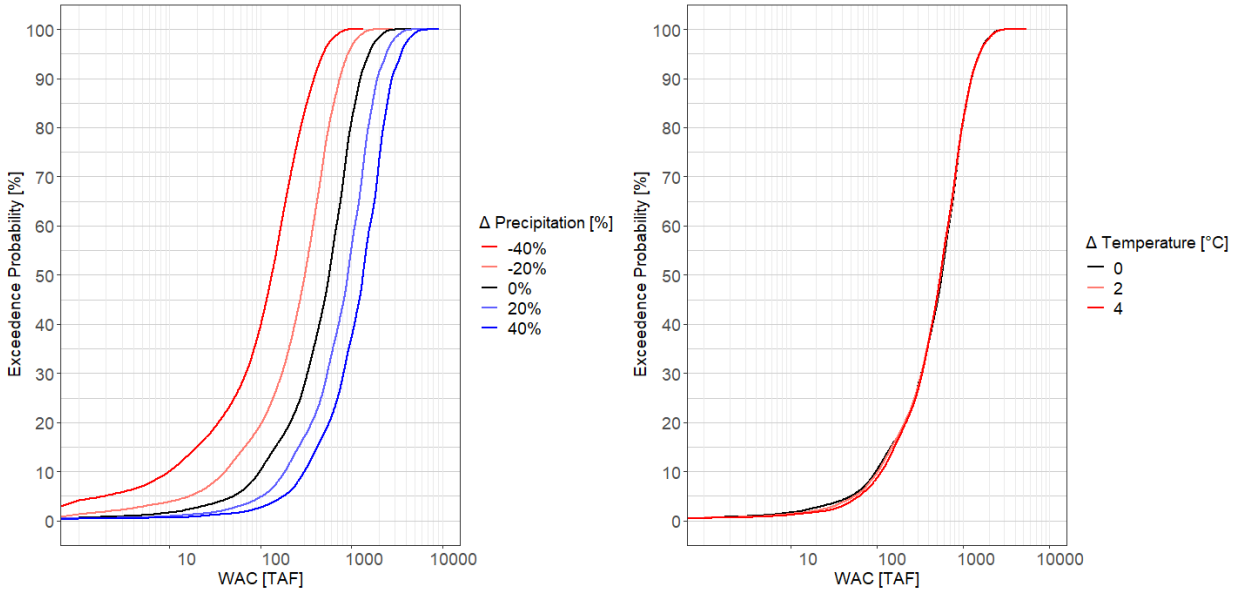


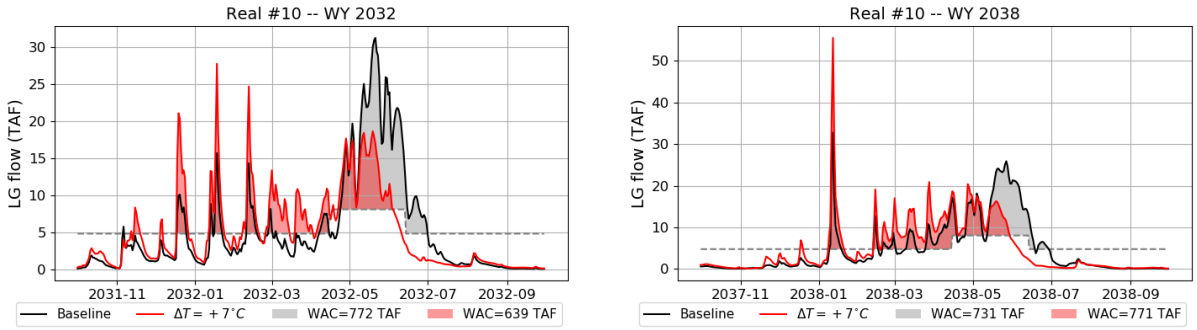
Figure 5-7. Effect of Precipitation (Left) and Temperature (Right) Changes on the Distribution of Annual WAC. Annual WAC values are based on water year. Values are shown across the 'stochastic realizations' dataset composed by 509 stochastic realization plus the historical realization.

Figure 5-8 shows the simulated hydrograph for six selected water years obtained from the realization 10 under baseline climate and +7°C warming. An extreme warming was considered in this figure for illustration only. It highlights that both increase and decrease in annual WAC can happened during normal, dry, and wet WAC years. It also shows that change in WAC roughly follows from three mechanisms:

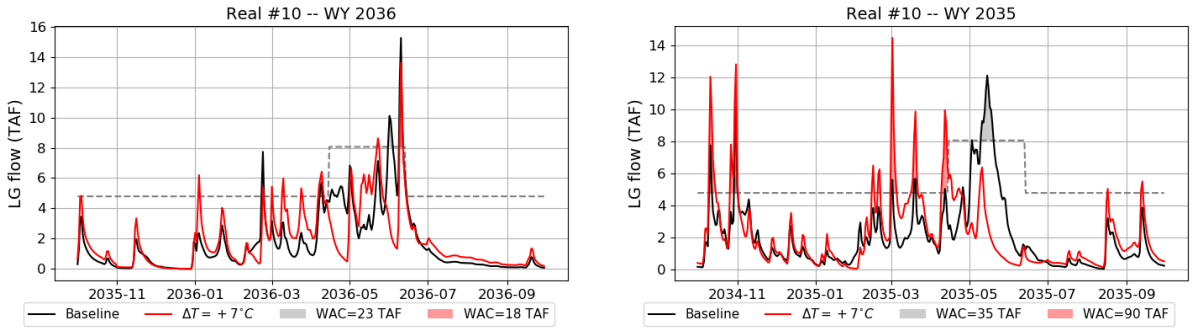
- The spring runoff volume tends to be smaller with warmer temperatures due to a reduced snowpack. The spring runoff volume tends also to spread over longer period. Both effect tends to reduce WAC during spring season.
- Spring runoff arrives earlier in the year because above freezing temperatures are seen earlier in the year. Given the higher maximum district entitlement from April 15th through June 13th, either a reduction or an increase in WAC is possible during spring / summer seasons.
- The winter floods tend to be more frequent and with larger magnitude because more precipitation fall as rain in winter, leading to an increase in WAC during this period.

All these mechanisms combine in a non-linear fashion and lead to either an increase or decrease in annual WAC. The change in WAC distribution within years illustrated with by example years in Figure 5-8 is summarized in Figure 5-9 and Figure 5-10. Both figures show scatterplots for each calendar month between monthly WAC values obtained for the current climate (i.e., no change in precipitation and temperature) and +3°C or +7°C warming, respectively. These figures show that winter months will likely get larger WAC values as temperature increase while summer months will see a significant reduction in WAC. As an example, under the extreme +7°C warming, August months are likely to have null WAC, while WAC during July months will be close to null value too.

“Normal” WAC Years



“Dry” WAC Years



“Wet” WAC Years

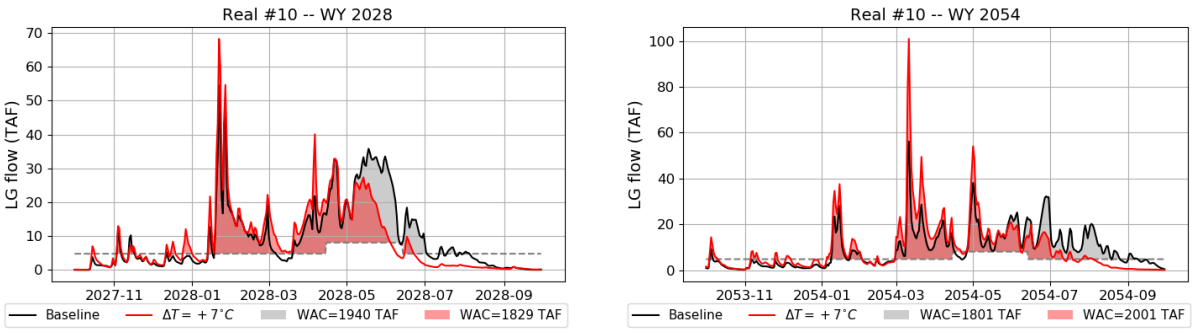


Figure 5-8. Illustration of the Effect of Temperature Change in the Temporality and Cumuli of WAC.
The example presented in this figure are all from the realization R10. Black color show the Tuolumne River flow at La Grange under historical climate while the red color shows the results obtained for an extreme warming of +7°C. Black and red shaded areas shows the WAC. Left and right columns show years for which WAC either decreases or decreases with a +7°C warming, respectively. The first, second and third row illustrate years for which the annual WAC is close to the average across the 10 realization (≈750 TAF), low and large, respectively.

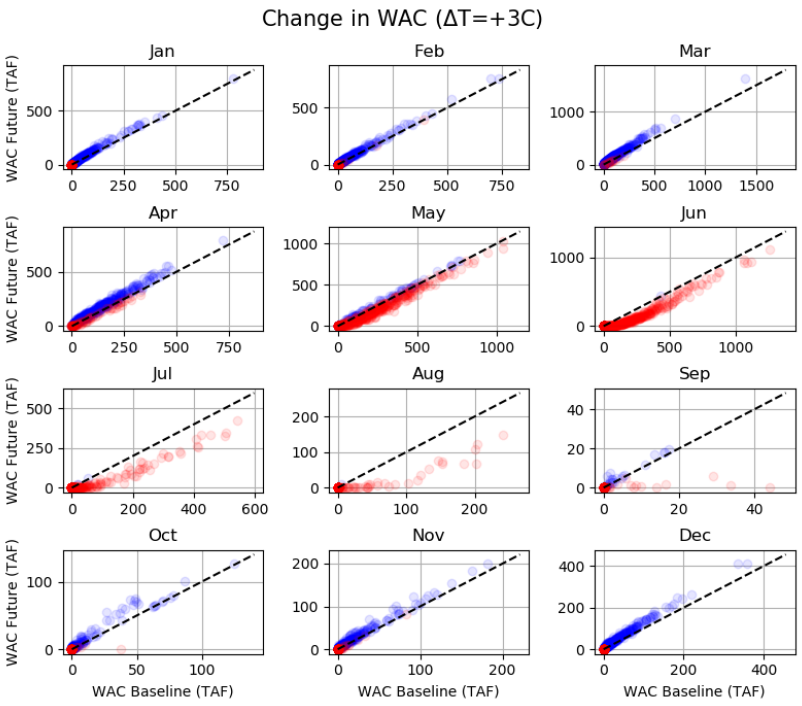


Figure 5-9. Change in Monthly Distribution of WAC Resulting from a Warming by +3°C.
The panels show for each calendar month the scatter plot between monthly WAC obtained under baseline climate (x-axis) and +3°C warming (y-axis). The black dash-line is the 'no change' line. Blue and red colors are used to highlight months for which the WAC either increases or decreases, respectively. Results shown are obtained across the 10 realizations, meaning that each scatterplot shows 500 data points (50 years x 10 realizations)

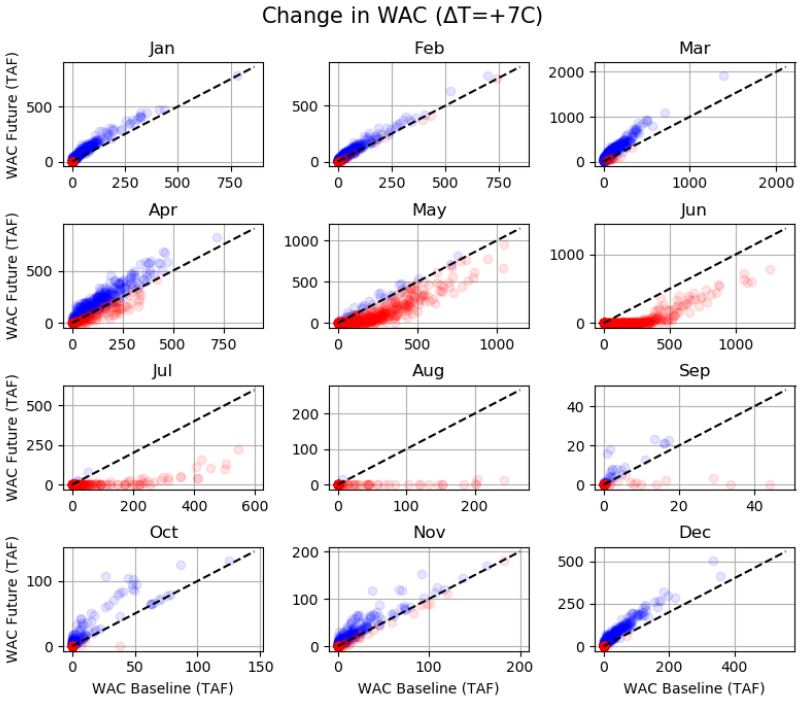


Figure 5-10. Same as Figure 5-9 but for a Temperature Scenario of +7°C.

In summary, precipitation change has an effect on the mean annual WAC volume.

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in no significant change in mean annual WAC volume, but WAC would decrease in 60% of the years and in 20% of them, the decrease would be more than 30 TAF.
- By 2040, most projections and elicitations of precipitation change fall between -5% and +5% which would correspond to a change in mean annual WAC between a decrease of 110 TAF and an increase of 120 TAF. Also, most projections and elicitations of warming are between +1°C and +4°C. At +4°C, WAC would decrease by more than 50 TAF in 20% of the years.
- WAC would decrease from May through August and would increase in the other months.
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual precipitation results in a 2.5% decrease in mean annual WAC volume. Most projections and elicitations of precipitation change are between -15% and +15% resulting in change in mean annual WAC volume between a decrease of 40% and an increase of 45%. . Most projections and elicitations of warming range between +3°C and +6°C. At +6°C, WAC would decrease by more than 80 TAF in 20% of the years.

Note that the PRMS hydrologic model used to simulate the streamflow on the Tuolumne watershed in response to precipitation and temperature overestimates streamflow during dry years. For example, the flow computed at the Tuolumne River at La Grange (foot note: the location where flow is allocated between the Districts and San Francisco) is overestimated, and therefore San Francisco's allocation is also overestimated by about 482,000 acre-feet during the drought sequence 1987-1992 (observed is 813,000 acre-feet versus simulated is 1,295,000 acre-feet). This volume of water is significant and is about equal to the volume of rationing that was required by customers during the six-year period, and so overstating the RWS water supply reliability.

5.1.1.3 Drought Severity and Duration

The modification of drought severity and duration following from change in precipitation and temperature are presented in this section.

Changes in Severity

Figure 5-11 illustrates the change in drought severity for a sample of scenarios of precipitation changes (i.e., -10% and -20%) and temperature changes (i.e., +2°C and +4°C), and for the two considered drought thresholds. Table 5-1 gives the estimated return periods for the historical droughts for the precipitation and temperature scenarios.

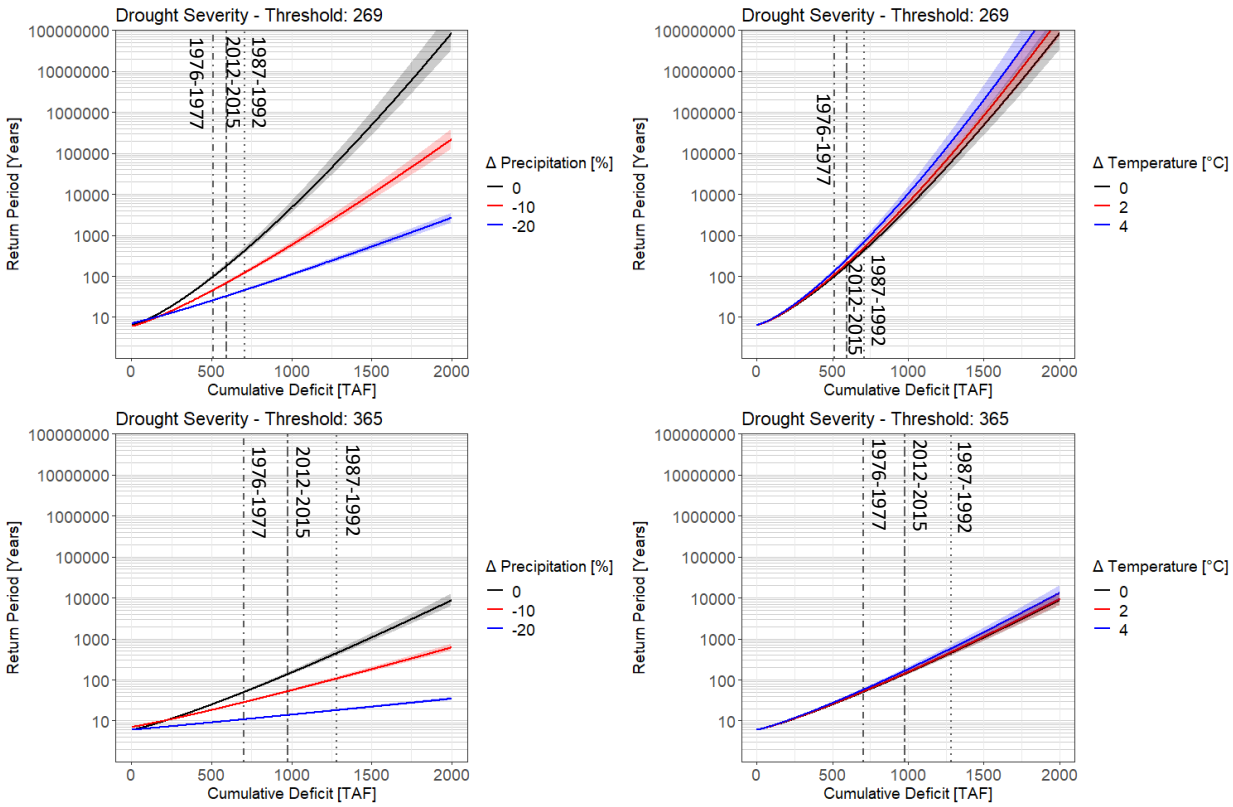


Figure 5-11. Effect of Precipitation (Left) and Temperature (Right) Change on the Drought Severity Frequency. Results are shown with 269 TAF (top) and 365 TAF (bottom) drought threshold. The x-axis shows that the absolute value of the cumulative deficit and the y-axis shows the estimated return period (years). The 95% confidence intervals are shown in shaded areas and are obtained using the bootstrap method. Vertical lines shows the severity of the historical droughts obtained using the historical dataset.

Table 5-1. Effect of Precipitation and Temperature Change on the Return Periods Associated with the Severity of the Historic Droughts.

Return periods are round off to the nearest 5 years.

Threshold [TAF]	Drought Event	Changes in Precipitation			Changes in Temperature [°C]		
		0%	-10%	-20%	0	+2	+4
269	1976-1977	100	45	25	100	105	130
	1987-1992	420	120	45	420	495	675
	2012-2015	180	70	35	180	200	260
365	1976-1977	50	30	10	50	50	60
	1987-1992	445	105	20	445	470	575
	2012-2015	135	55	15	140	145	165

Precipitation change is a significant driver of change for the severity of drought. Any reduction in precipitation leads to an increase in frequency of drought severity (i.e., a reduction of the return period of any severity level). For instance, considering the threshold of 269 TAF and a 10% precipitation reduction, the severity associated with the 1976-77 becomes roughly twice as more frequent (from 98

to 46 years return period). Similarly, the return period associated with the 2012-2015 drought drops from 178 to 69 years. The changes in return periods for the severities associated with the historical events under a 20% precipitation reduction scenario are more drastic. All three highlighted historical droughts see their return period fall below 50 years, including the severity associated with the 1987-1992 droughts whose estimated return period under baseline climate scenario is larger than 400 years.

The effect of warming temperature appears to be opposite with an increase in return period for all historical droughts. For instance, the return period of the severity associated with the 1987-92 drought increase from 420 to 494 years under 2°C warming.

Similar results are obtained for the 365 TAF threshold, although the change following from a reduction in precipitation are more important, and in the case of temperature warming, less important.

Changes in Duration

Figure 5-12 illustrates the change in drought duration for a sample of scenarios of precipitation changes (i.e., -10% and -20%) and temperature changes (i.e., +2°C and +4°C), and for the two considered drought thresholds. Table 5-2 gives the estimated return periods for the historical droughts for the scenarios shown in Figure 5-12.

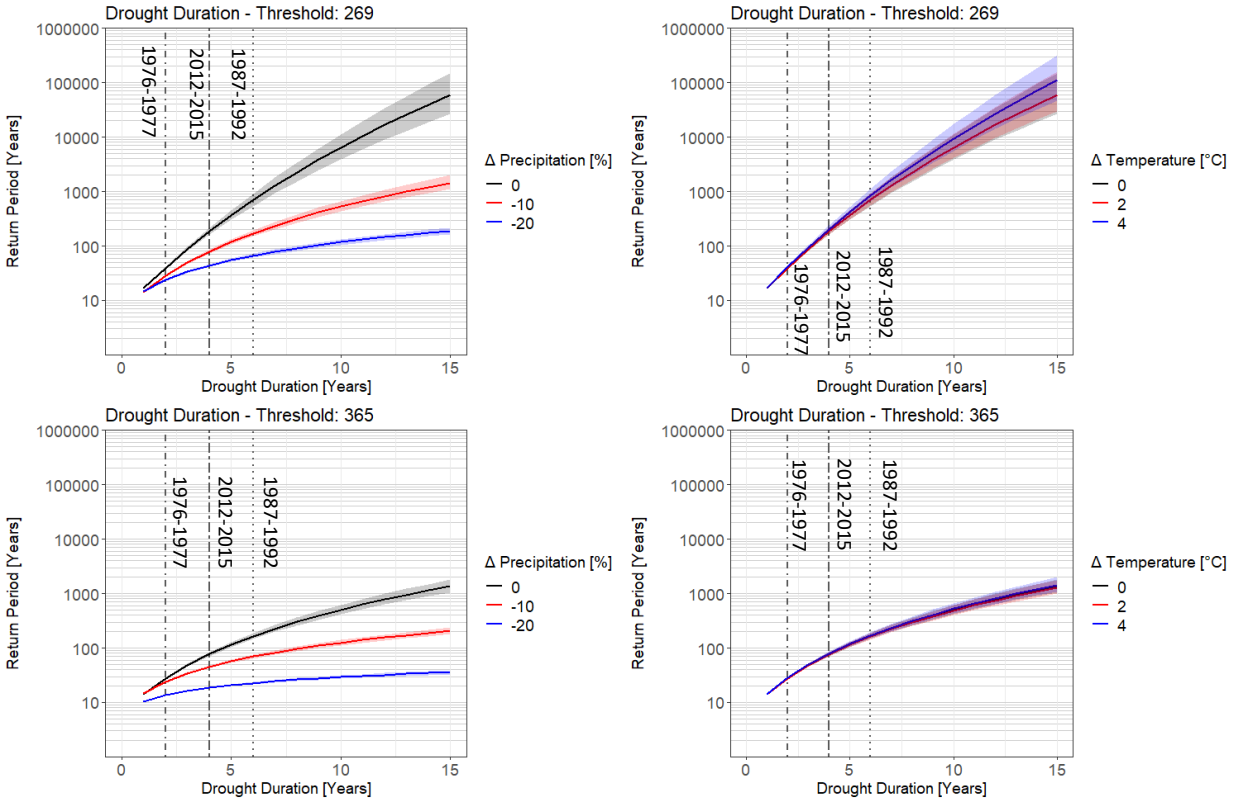


Figure 5-12. Effect of Precipitation (Left) and Temperature (Right) Change on the Drought Duration Frequency. Results are shown with 269 TAF (top) and 365 TAF (bottom) drought threshold. The x-axis shows that the drought duration (years) and the y-axis shows the estimated return period (years). The 95% confidence intervals are shown in shaded areas and are obtained using the bootstrap method. Vertical lines shows the duration of the historical droughts obtained using the historical dataset.

Table 5-2. Effect of Precipitation and Temperature Change on the Return Periods Associated with the Duration of the Historic Droughts.

Threshold [TAF]	Drought Event	Changes in Precipitation			Changes in Temperature [°C]		
		0%	-10%	-20%	0	+2	+4
269	1976-1977	40	30	25	40	40	40
	1987-1992	700	170	65	700	700	840
	2012-2015	185	80	45	185	185	205
365	1976-1977	30	25	15	30	30	30
	1987-1992	165	70	25	165	160	170
	2012-2015	75	45	20	75	75	80

Similar to drought severity, precipitation is the main driver of change in drought duration. Any decrease in precipitation makes droughts of any duration more frequent, especially the long ones. For instance, with 269 TAF threshold, the return period of 2-yr long drought (i.e., 1976-1977) drops from 40 years to 28 years or 24 years if precipitation decreases by either 10 or 20%, respectively. For long droughts, for instance of the 1987-1992 (i.e., 6 years), the return period drops from 700 years to 170 or 70 years with a reduction in precipitation of either 10 or 20%, respectively.

Warming temperature tends to slightly reduce the frequency of droughts with specific duration, although the signal is hardly significant. For instance, using 269 TAF threshold and a warming scenario of +4°C, the return period of the 1987-92 drought duration (i.e., 6 years) increases from 700 years to 850 years.

Similar results are obtained for the 365 TAF threshold: a major increase in frequency when precipitation decreases, and an insignificant signal under warming conditions.

Changes in Joint Return Period

The modification of the frequency of drought when accounting for the dependence between severity and duration was conducted using copula functions (Section 3.2.4). The survival Clayton copula was used with the 269 TAF threshold while the Gumbel copula was used for the 365 TAF threshold. Note that the return periods obtained under baseline climate (i.e., no change in precipitation and temperature) are slightly different from the ones presented in Table 3-19 because the 'combined' dataset was used in Chapter 3 rather than the 'stochastic realizations' dataset'.

Table 5-3 details for the considered climate scenarios and drought threshold the modification of the dependence (i.e., correlation) between drought severity and duration by mean of the Kendall's rank correlation coefficient (τ^{data}). It is noted that decreasing precipitation increases the dependence between drought severity and duration (i.e., τ^{data} increases) while increasing precipitation tends to decrease the dependence between drought severity and duration (i.e., τ^{data} decreases). Warming temperature tends to slightly decrease the correlation, although the changes are not as significant as for precipitation changes. Although the fitted copula functions for each climate scenarios and thresholds reproduce fairly well the Kendall's rank correlation, they consistently underestimates the correlation (τ^{cop} compared to τ^{data}).

Table 5-3. Effect of Precipitation and Temperature Change on the Dependence between Drought Severity and Duration.

The Kendall's rank correlation coefficients obtained from the fitted copula functions (τ^{cop}) and estimated from the identified drought events (τ^{data}) are used to assess the dependence between drought severity and duration. The fitted copula parameter is given for information only.

Threshold	Precipitation Scenario	Temperature Scenario	Copula	τ^{cop}	τ^{data}	Copula Parameter
269	ΔP=-20%	ΔT=0°C	Survival Clayton	0.674	0.719	2.272
	ΔP=-20%	ΔT=4°C		0.685	0.726	2.387
	ΔP=0%	ΔT=0°C		0.464	0.567	1.732
	ΔP=0%	ΔT=4°C		0.452	0.557	1.651
	ΔP=+20%	ΔT=0°C		0.366	0.451	1.738
	ΔP=+20%	ΔT=4°C		0.334	0.420	1.005
365	ΔP=-20%	ΔT=0°C	Gumbel	0.758	0.782	2.93
	ΔP=-20%	ΔT=4°C		0.754	0.772	2.64
	ΔP=0%	ΔT=0°C		0.558	0.632	2.26
	ΔP=0%	ΔT=4°C		0.552	0.625	2.23
	ΔP=+20%	ΔT=0°C		0.421	0.522	2.10
	ΔP=+20%	ΔT=4°C		0.399	0.498	1.81

The effect of precipitation and temperature change on the joint return period of drought severity and duration is presented in Table 5-4. The results are consistent with the results discussed for drought severity and duration taken separately. Without surprise, increasing precipitation significantly decreases the frequency of drought (i.e., increases the return period). With a 269 TAF threshold and under 20% increase precipitation, the reduction in frequency is such that the return periods associated with the considered historical droughts become all roughly larger than 20,000 years. However, decreasing precipitation leads to significant increase in drought frequency. Under 20% precipitation reduction scenario, the return period of all historical droughts gets below 100 years.

The effect on temperature change on the drought frequency appears to be not linear whether or not it is associated to either a decrease or an increase in precipitation. When combined with a decrease in precipitation, the effect of temperature is not significant.

The main results remain broadly consistent when using 365 TAF threshold. Note that the drought are more frequent using this threshold and the effect of temperature change remain low when compared with the increase in frequency following from a reduction in precipitation.

Table 5-4. Effect of Precipitation Change (ΔP, %) and Temperature Change (ΔT, °C) on the Joint Return Period (Years) Associated with the Severity and Duration of the Historic Droughts for Drought Thresholds of 269 TAF and 365 TAF.

Threshold	Event	Joint Return Period (years)					
		ΔP=-20%		ΔP=0%		ΔP=+20%	
		ΔT=0°C	ΔT=4°C	ΔT=0°C	ΔT=4°C	ΔT=0°C	ΔT=4°C
269	1976-1977	35	35	110	140	19,300	855
	1987-1992	75	70	855	1,160	728,510	1,532,685
	2012-2015	50	50	270	355	61,685	4,655
365	1976-1977	15	45	60	65	12,075	425
	1987-1992	25	75	505	635	702,410	62,225
	2012-2015	20	60	170	195	78,515	4,085

In summary, models indicate that hydrologic drought severity on the Tuolumne River for SFPUC will significantly increase with reduction in precipitation but not necessarily with temperature increase. Note that the models used (combination of weather generator and hydrologic model) tend to underestimate both the frequency of occurrence and severity of droughts, as compared to the observed record (not enough droughts occur in the simulated record compared to the historical record). Therefore, it is assumed that the frequency of occurrence and severity of hydrologic droughts in the simulated record, with climate change, is also underestimated, overstating the RWS water supply reliability.

5.1.2 East Bay Region

Figure 5-13 illustrates the change in annual streamflow across the three watersheds in the East Bay region (Calaveras, ACDD and San Antonio) that would result from change in precipitation and temperature. The precipitation is the main driver of change in annual flow across the region. The effect of warming temperature on annual flow is more important than across the Upcountry region, which results from a more important role of the evapotranspiration at lower elevation. For example, the reduction of annual flow that follow from a +4°C would cancel out an increase in precipitation by 10% across the region.

At the horizon 2040s (i.e., 2026-2055), the climate projections do not show a clear preference toward an increase or a decrease in annual flow (the yellow dots on the response functions are roughly distributed evenly on both side of the no change black line in Figure 5-13). However, at the horizon 2070s (2056-2085), the climate projections seem to show a higher likelihood toward a decrease than an increase (more green dots on the response functions lie on the left side of the no change line).

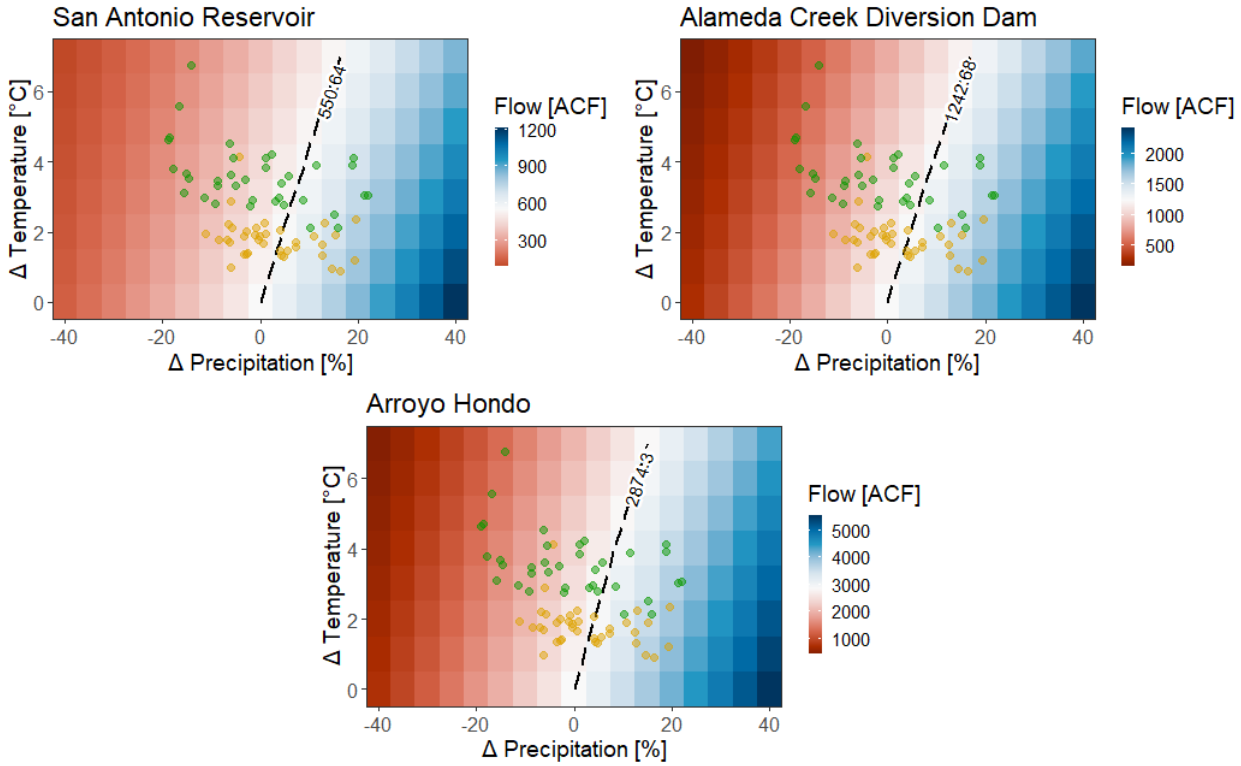


Figure 5-13. Climate Response Surface for Mean and Standard Deviation of Monthly Flow at San Antonio Reservoir Inflow, Alameda Creek Diversion Dam (ACDD) Inflow, and Arroyo Hondo above Calaveras Reservoir across All Ten Realizations.

Blue represents increase in streamflow and red represents decrease in streamflow. GCM dots are shown with various GCMs for RCP8.5 with green centered in 2040 and black centered in 2070. The black contour line shows baseline climate response.

In summary, mean annual reservoir inflows in the East Bay region are sensitive to both changes in precipitation and temperature. The sensitivity to temperature is due to soil water availability for increase in evapotranspiration which leads to a reduction in runoff. Results below are presented for inflow at Calaveras reservoir (Arroyo Hondo) the largest watershed in the East Bay.

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in 9% reduction in mean annual inflow.
- By 2040, most projections and elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5% which would correspond to a maximum change in mean annual inflow between a decrease of 27% and an increase of 7%.
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual precipitation results 17% decrease in mean annual inflow. Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15% resulting in change in mean annual inflow between a decrease of 50% and an increase of 33%.

5.1.3 Peninsula Hydrology

Figure 5-14 illustrates the change in annual streamflow across the Peninsula region that would follow from changes in precipitation and temperature. The response of the Peninsula watersheds is very similar to the response of the watersheds in East Bat region. The change in precipitation is the main driver of change in annual streamflow. In addition, similar to East Bay region, warming temperature tends to

decrease streamflow so that, an increase in temperature by 4°C would cancel out an increase in precipitation by 10%.

Similar to East Bay region, climate projections do not indicate any preference toward an increase or a decrease by the horizon 2040s (2026-2055) but a decrease by 2070s (2056-2085) is more likely.

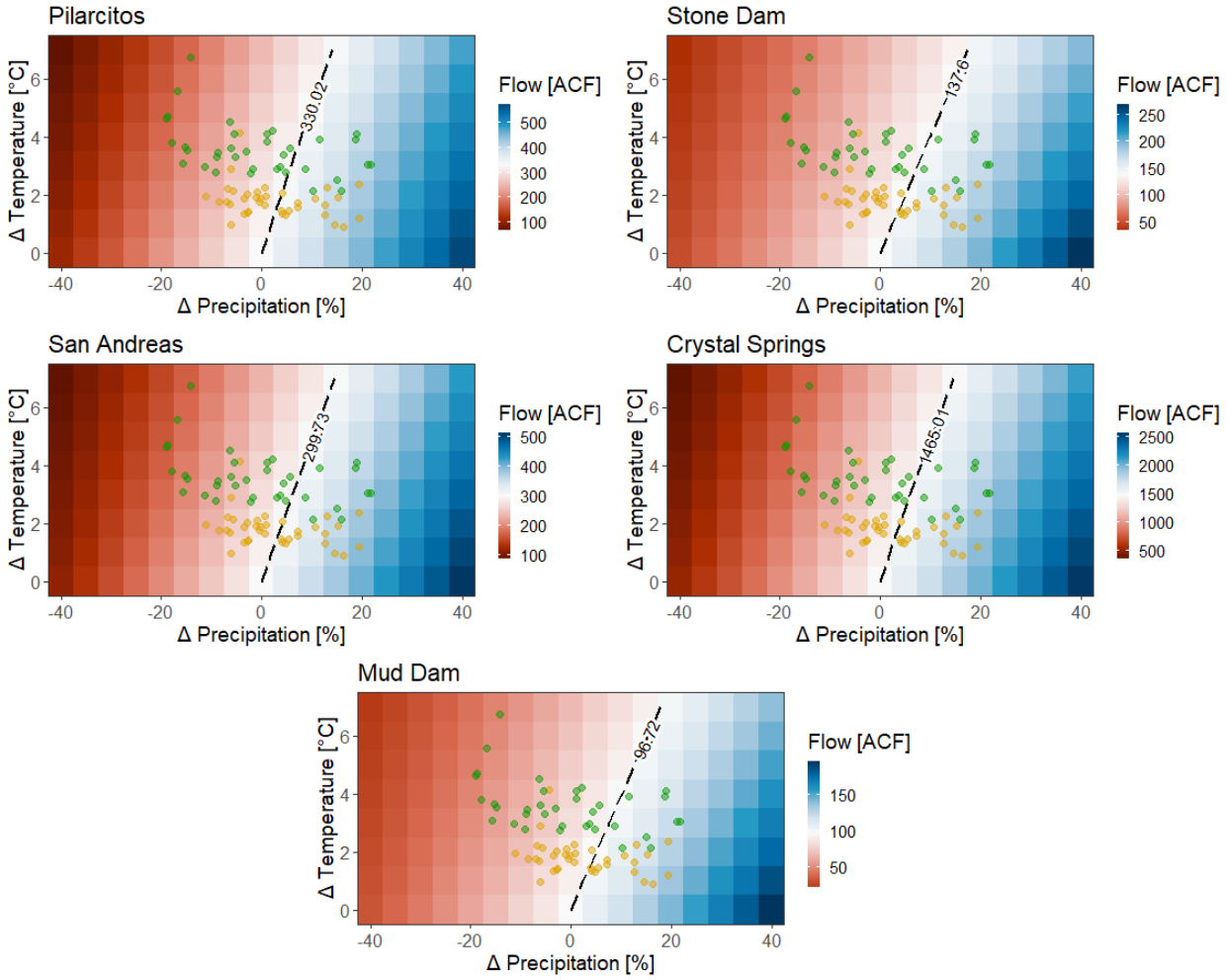


Figure 5-14. Total Annual Streamflow (mm) under Temperature and Precipitation Change Respectively for Crystal Spring Reservoir, San Andreas Reservoir, San Mateo Creek, Pilarcitos Reservoir, and Stone Dam Sub-watersheds.

Dots are GCM projections.

In summary, mean annual reservoir inflows in the Peninsula region are sensitive to both changes in precipitation and temperature. The sensitivity to temperature is due to soil water availability for increase in evapotranspiration which leads to a reduction in runoff. Results below are presented for inflow at Crystal Springs reservoir the largest watershed in the Peninsula.

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in 7% reduction in mean annual inflow.
- By 2040, most projections and elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5% which would correspond to a maximum change in mean annual inflow between a decrease of 23% and an increase of 10%.
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual

precipitation results 14% decrease in mean annual inflow. Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15% resulting in change in mean annual inflow between a decrease of 46% and an increase of 29%.

5.2 Compound Effects of Climate and Demand Change on Water Supply

5.2.1 Virtual Water

To better understand and contextualize the vulnerability assessment results for water supply, it is important to keep track of the requirement for virtual water. As noted in Section 3.4.2, virtual water is water that can enter the system to both prevent the system model from crashing and reveal where, when and how much extra water may be needed even after rationing.

Under baseline climate and demand conditions, virtual water is not needed by SFWSM (Figure 5-15). Beside the scenarios with large decrease in precipitation (i.e., 30% reduction or more), for which virtual water is required for up to 30% of the years depending on the actual precipitation reduction and temperature increase (Figure 5-15), virtual water is needed only during a specific drought event within the ninth realization (this event is described further in this section and in Figure 5-17). As demand increases, the need for virtual water becomes more frequent. For example, with +30% demand, all climate realizations result in virtual water needed at some point to prevent the system model from failing completely (including under baseline climate for which almost 10% of the years need virtual water). With +45% demand, all combinations of temperature and precipitation changes require virtual water for at least 25% of the years (Figure 5-15).

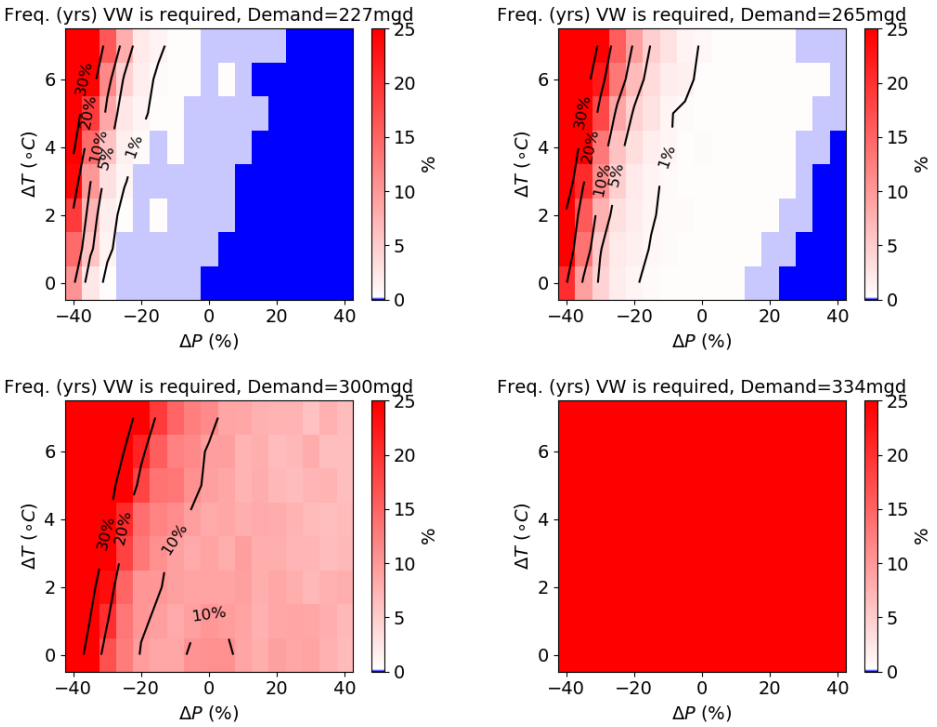


Figure 5-15. Frequency (Percentage of Years) of Virtual Water Requirement.
Results are shown for baseline demand (227 mgd), +15% demand (265 mgd), +30% demand (300 mgd) and +45% demand (334 mgd) scenarios under different temperature and precipitation scenarios.

The annual maximum virtual water requirement obtained across the ten realizations is as much as about 80 TAF under the baseline demand scenario (ΔP=-40% and ΔT=+7 °C) and about 180 TAF with +45% demand (ΔP=-40%) (Figure 5-16). This indicates a potential maximum total water supply deficit of about 80-180 TAF that needs to be somehow accounted for under extreme precipitation decrease. With 30% and 45% increase in demand, significant volume of additional water would be required to satisfy system demands even with much more precipitation and no temperature increase.

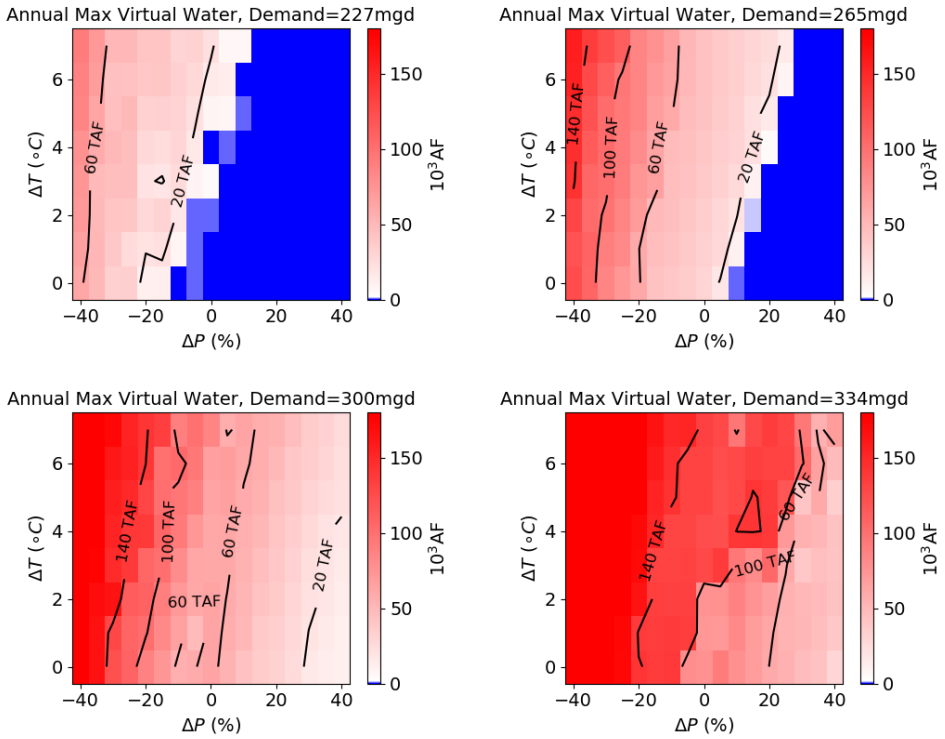


Figure 5-16. Annual Maximum Virtual Water Requirement.
Results are shown for baseline demand (227 mgd), +15% demand (265 mgd), +30% demand (300 mgd) and +45% demand (334 mgd) scenarios under different temperature and precipitation scenarios.

Figure 5-17 illustrates the two-year drought event within the ninth realization for which virtual water is needed by the system as soon as temperature increases or precipitation decreases. The specific drought event spans from the water year 2062 through 2063. During the water year 2062, not only the Water Available to the City (WAC) is null, the unimpaired flows at La Grange are also almost null too (Figure 5-17a), which means that the water stored at Water Bank cannot be used to fill Hetch Hetchy Reservoir. As a result, Hetch Hetchy Reservoir storage reduces significantly and eventually runs dry by the end of the water year 2062 (Figure 5-17b). The water year 2063 is also very dry with unimpaired flow at La Grange barely larger than the maximum district entitlement (dash grey line in Figure 5-17a). Such a low inflow across Upcountry region only allows Hetch Hetchy Reservoir to fill above (respectively below) 100 TAF under baseline climate (respectively under +3°C temperature scenario). As a result, Hetch Hetchy Reservoir runs dry again during the second half of the water year 2063 through February 2064.

Meanwhile, the Sunol Valley Water Treatment plant (SVWTP) is significantly utilized (not shown) and Calaveras and San Antonio storage reduces significantly (Figure 5-17d and Figure 5-17e). Under baseline climate, Calaveras Reservoir runs dry at the beginning of the water year 2064 and water release from San Antonio Reservoir maintains a continuous flow to SVWTP. However, under +3°C warming scenario, Calaveras Reservoir runs dry significantly earlier, which forces San Antonio to contribute alone to SVWTP

for a longer period, which leads the reservoir drawdown to dead pool too, interrupting the production at SVWTP and leading to deficit (Figure 5-17f).

It is important to note that this failure of the system occurs while water bank is nearly full, highlighting the vulnerability of the RWS to a single very severe drought during which Hetch Hetchy Reservoir gets to storage level that prevent the reservoir for sending water to the Bay area. In the case illustrated in Figure 5-17, a second dry year even leads to actual deficit (i.e., requirement for virtual water).

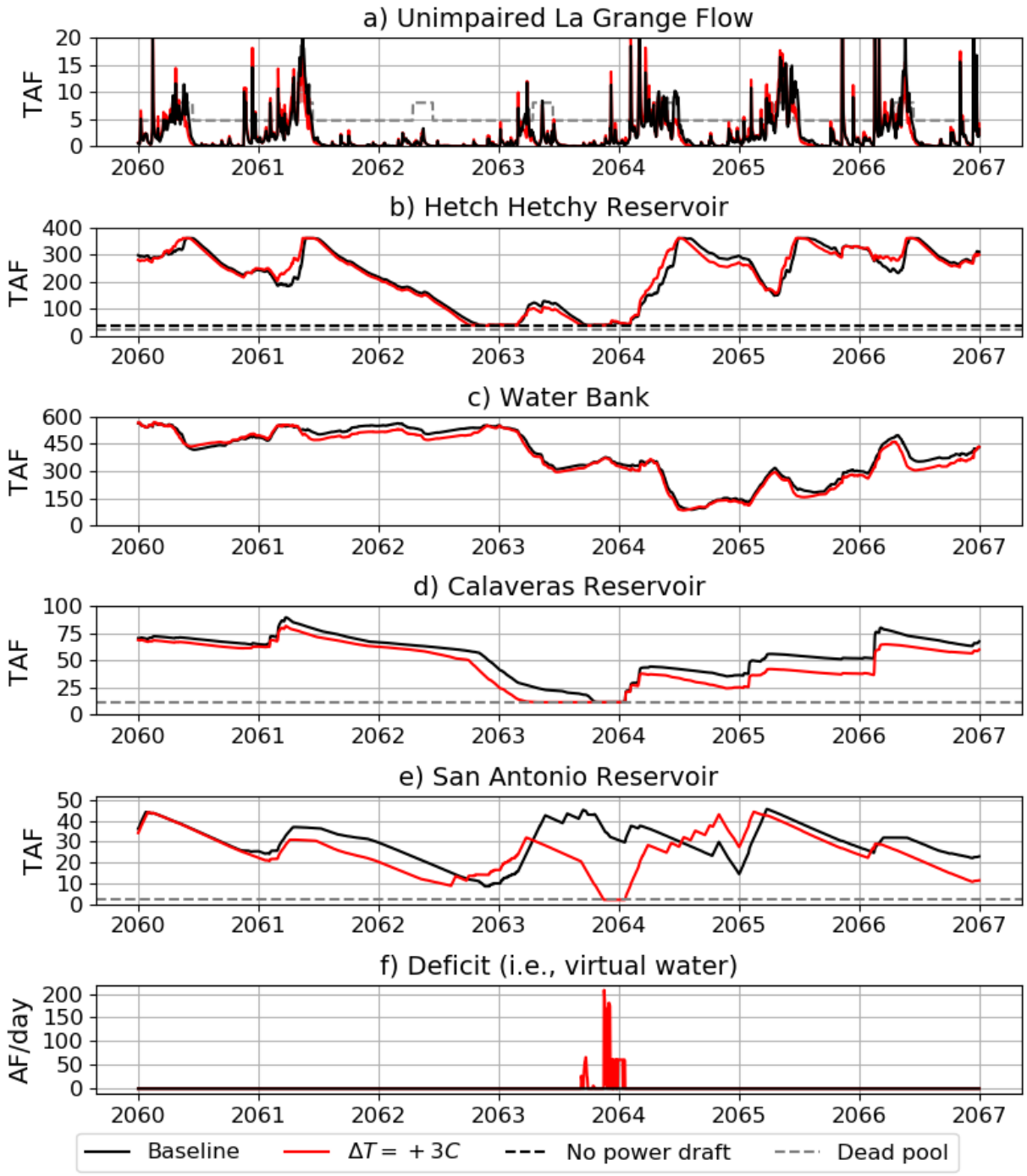


Figure 5-17. Illustration of the Drought Event within Realization Nine That Leads to Virtual Water Requirement under Baseline Demand Condition and Minor Change in Temperature and/or Precipitation.
The figure specifically highlights the +3°C warming scenario (red lines), but similar conclusion could be made for other warming scenarios and/or decrease in precipitation. Black lines show results obtained under baseline demand and climate. X-ticks are located on January 1st.

The inclusion of virtual water in the system model does not affect the remaining performance metrics such as reliability or duration of deficit. When virtual water has contributed to deliveries, performance metrics based on a measure of deficit discount virtual water first before the metrics are calculated. In other words, virtual water does not reduce the value of the deficit.

Several key performance metrics are presented in the sub-sections below. Each metric is accompanied by a similar set of four complementary figures, as follows:

- The univariate response of the metric to temperature (ΔT)—with precipitation at the baseline level—across four different demand scenarios (e.g., Figure 5-18)
- The univariate response of the metric to precipitation (ΔP)—with temperature at the baseline level—across the same four demand scenarios (e.g., Figure 5-19)
- Multivariate response surfaces spanning both temperature and precipitation for each of the four demand scenarios (e.g., Figure 5-20)
- Boxplots of the overall anticipated long-term value of the performance metric for each demand levels across all temperature/precipitation values in 2040 and 2070 (e.g., Figure 5-21)

In addition, the first three of these plot types also include histograms of the relevant change factor (i.e., ΔT, ΔP or both) across all GCMs and expert elicitations for to 30-year long periods centered in 2040 and 2070.

Some plots include reference to alternative HHLSM (WSIP 2018 265 version) (e.g., Figure 5-18). The SFPUC Water System Improvement Program 2018 (WSIP 2018) variant from the programmatic environmental impact report (PEIR) for the WSIP program. This was the program alternative that was selected during WSIP planning and is the underlying benchmark for comparison for SFPUC's future planning efforts. Modeled at 265 mgd demand, which corresponds to an increase by 15% of the baseline used in this study.

5.2.2 Water Delivery Reliability

Reliability of water delivered to SFPUC customers is an important metrics of the RWS system performance. In general, SFPUC aims to maintain a system-wide reliability of 90% (meaning a frequency of rationing of 10% or rationing in 1 out of 10 years on average) for any given demand level. This reliability target is grounded in the performance of the RWS with the planned WSIP 2018 at a level of demand of 265 mgd over historical hydrology. To calculate system reliability for RWS delivery, the system-wide aggregation of demand and delivery were used for X_{target} and X_t, respectively, in Equations 4-1 and 4-2. Results for water delivery reliability are shown in Figure 5-18 through Figure 5-21. From these figures, several observations can be made, as follows.

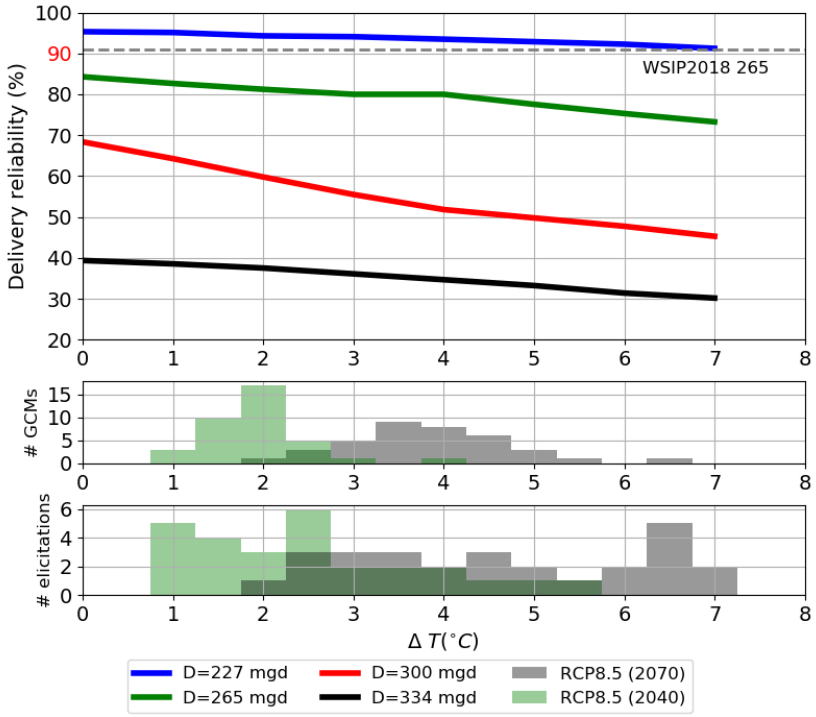


Figure 5-18. Effect of Temperature Change on Water Delivery Reliability (%).
Top panel shows water delivery reliability across all temperature (ΔT) scenarios and a subset of four demand scenarios (227 mgd, baseline SFPUC demand; 265 mgd [+15%]; 300 mgd [+30%]; and 334 mgd [+45%]). The dash-grey line shows reliability obtained via HHLSM WSIP 2018. The 90% target is indicated in red color in the y-axis. Middle and bottom panels show the distribution of changes in temperature obtained from CMIP5 projections (RCP8.5) and expert elicitations and for two 30-yr long periods centered in 2040 and 2070.

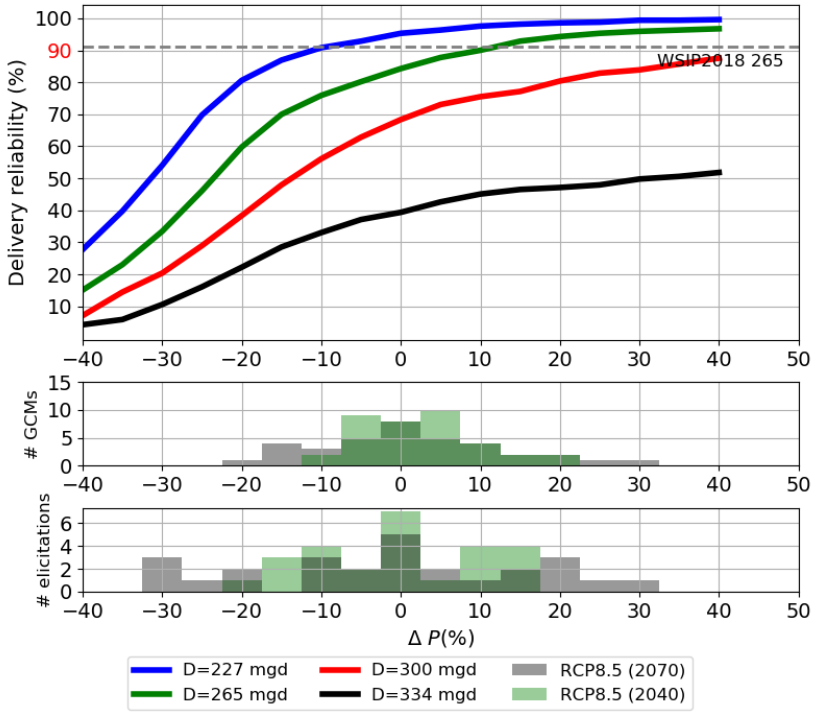


Figure 5-19. Effect of Precipitation Change on Water Delivery Reliability (%).
See Figure 5-18 caption for details explanation of the figure components.

Under baseline demand (227 mgd) and precipitation, the warming temperature alone decreases water delivery reliability (blue curve in Figure 5-18). The main reason for change in water delivery reliability under warming conditions is the reduction in inflow to RWS reservoirs due to larger evaporation. For instance, the largest temperature change (i.e., +7°C) decreases the water delivery reliability by nearly 5%. The effect of temperature on delivery reliability increases when demand increases. With +30% demand scenario (i.e., 300 mgd), a moderate increase in temperature by 2°C decreases water delivery reliability by almost 10% while barely no decrease in reliability is obtained for a demand of 227 mgd. With the same demand of 300 mgd, the water delivery reliability decreases by more than 20% for an increase in temperature by 7°C while it decreases by less than 5% under 227 mgd demand scenario. The increase in vulnerability with temperature most likely results from the decrease in streamflow due to larger evapotranspiration across the RWS watersheds and to the decrease in WAC during specific years (Section 5.1).

There is significant sensitivity of delivery reliability to change in precipitation (Figure 5-19). Specifically, a decrease in precipitation by 10% would bring the delivery reliability below the target value of 90%. Whether the climate projections indicate low probability for such a change in precipitation for the 2040 period (i.e., two GCM projections indicate a decrease in precipitation by 10% or more), results from the expert elicitation shows a different picture, with a larger number of elicitations showing in a decrease by 10% or more by 2040. For the 2070 period, 10% reduction in precipitation (and its associated decrease in water delivery reliability) becomes more likely from the perspective of both the climate projections and the expert elicitation.

There is a significant sensitivity of delivery reliability to change in demand (Figure 5-18 and Figure 5-19). More specifically, if demand increases by 15%, the 90% reliability target is not met unless precipitation increases by 10% (up to 15% if temperature increase by 5°C (Figure 5-20). If demand increases by 30% or more, the 90% target cannot be met, even if precipitation increases by 40%.

Combining the joint effect of precipitation, temperature and demand changes (Figure 5-20) with climate projections (Figure 5-21), demand appears to be the main driver of change in water delivery reliability and the 90% reliability target is very likely not to be met if demand increases by more 15% or more.

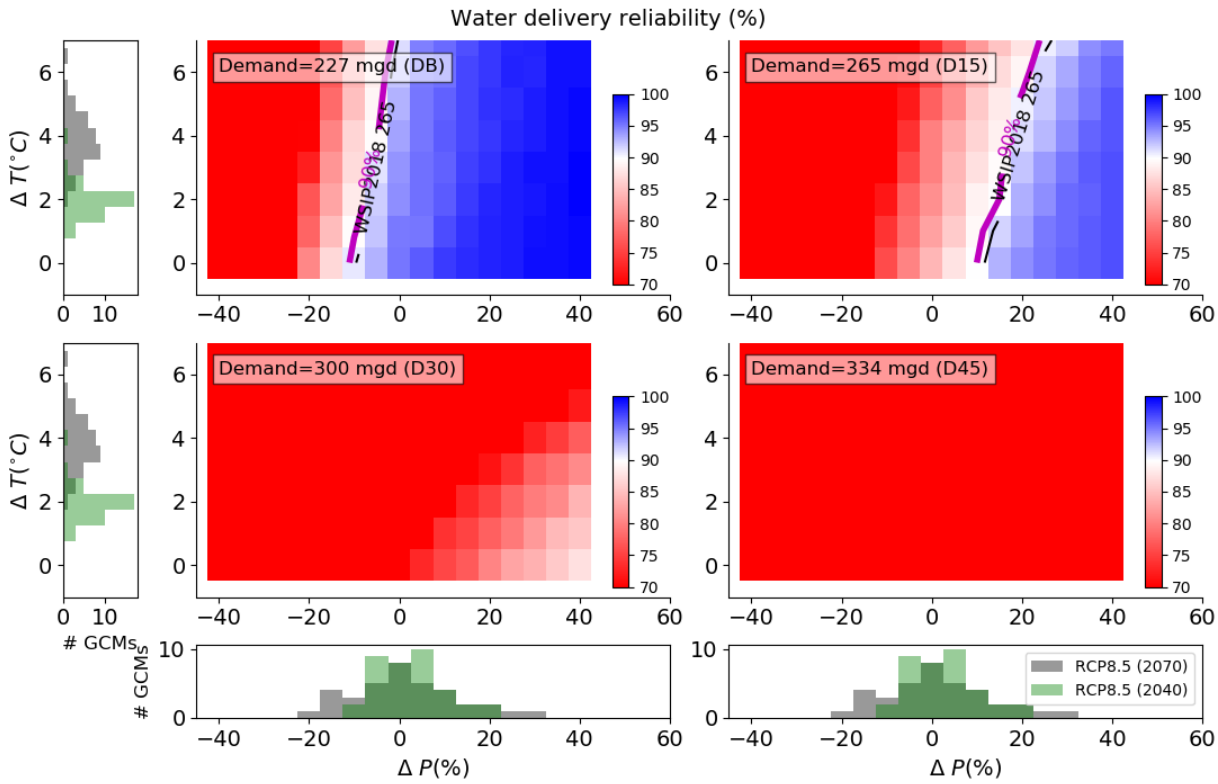


Figure 5-20. Combined Effect of Temperature (ΔT), Precipitation (ΔP) and Demand Changes on the Water Delivery Reliability (Color Map).

The black contour lines show the reliability obtained from HHLSM WSIP 2018 and the purple line shows the 90% reliability target. The distribution plots show expected changes in precipitation (horizontal) and temperature (vertical) from CMIP5 climate projections (RCP8.5). Two 30-yr long periods centered in 2040 and 2070 are considered.

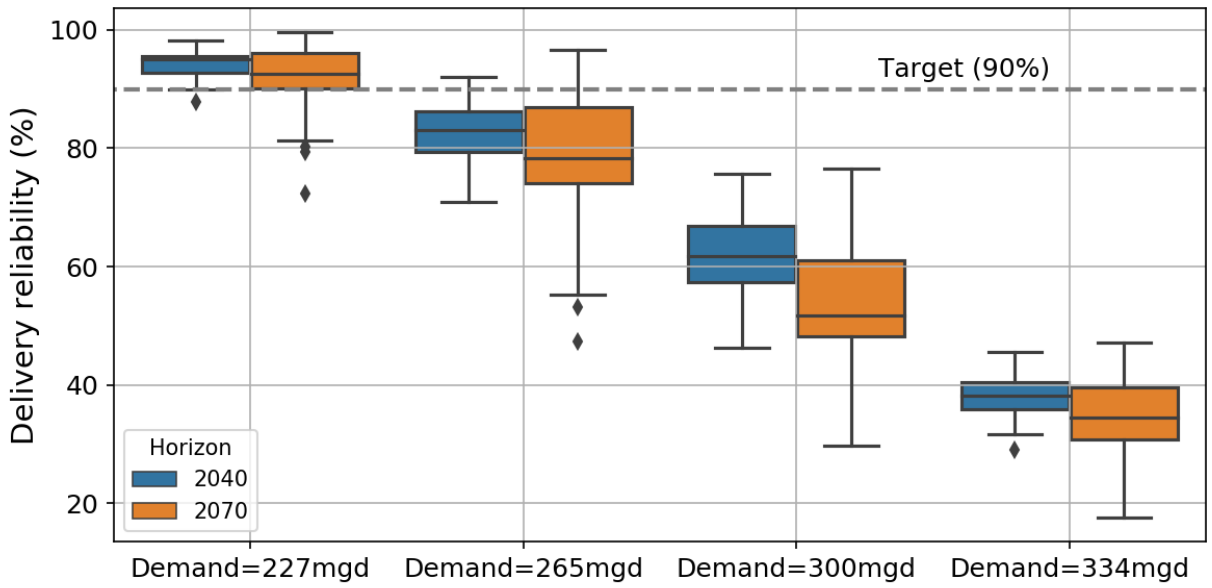


Figure 5-21. Expected Change in Water Delivery Reliability When Combining the Vulnerability of the System with the Climate Projections.

The 90% target is shown with the dashed grey line.

In summary, water delivery reliability is more sensitive to demand change than to change in temperature and precipitation.

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in meeting the reliability target with 227 mgd demand and results in a doubling of rationing frequency with a demand of 265 mgd (reliability around 80%, rationing in 1 out of 5 years on average).
- By 2040, most projections and elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5%. At +4°C and -5%, the reliability meets the target with 227 mgd of demand and fails the target with 265 mgd of demand (reliability around 75%, rationing in 1 out of 4 years on average).
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual precipitation results in meeting the reliability target for 227 mgd demand and fails the target for 265 mgd demand (reliability also around 80%) Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15%. At +6°C and -15%, the reliability fails the target with 227 mgd demand (reliability around 75%, rationing in 1 out of 4 years) and fails the target with 265 mgd demand (reliability around 50%, rationing in 1 out of 2 years on average).

5.2.3 Frequency of 20% Rationing

SFPUC level-of-service ascribes to no more than 20% demand rationing under drought conditions. In SFWSM, a 20% reduction in deliveries is applied on July 1st if the years of remaining supply is below 3.14 years. For the analysis of performance, the target is 20% rationing in less than 5% of the years. This target is grounded in the performance of the RWS with the planned WSIP 2018 at a level of demand of 265 mgd over historical hydrology. Results that relates to the frequency of 20% rationing are shown in Figure 5-22 through Figure 5-25. From these figures, several observations can be made, as follows.

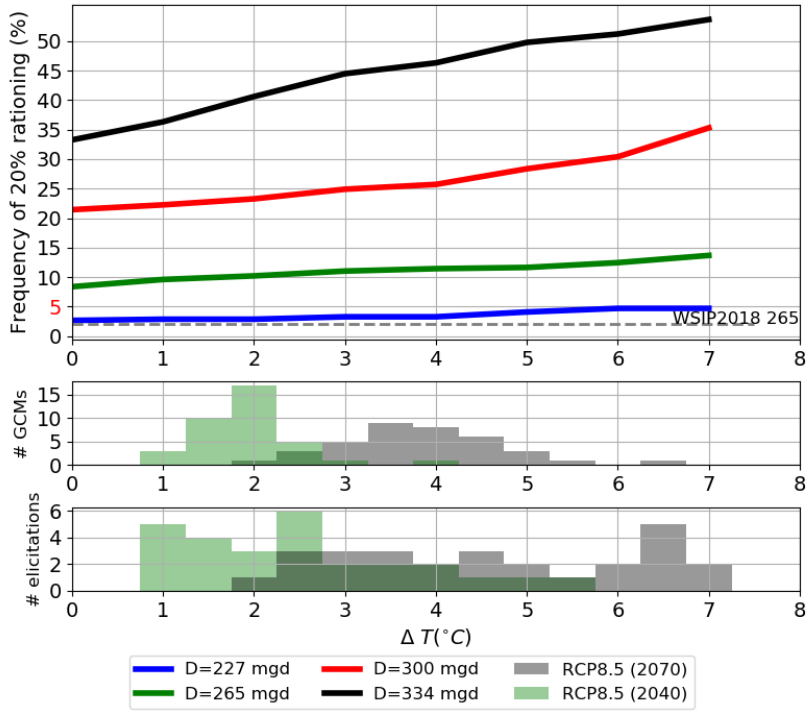


Figure 5-22. Effect of Temperature Change on the Frequency of 20% Rationing (%).
Top panel shows frequency of 20% rationing across all temperature (ΔT) scenarios and a subset of four demand scenarios (227 mgd, baseline SFPUC demand; 265 mgd [+15%]; 300 mgd [+30%]; and 334 mgd [+45%]). The dash-grey line shows frequency obtained via HHLSM WSIP 2018 265 mgd. The 5% target is indicated in red color in the y-axis. Middle and bottom panels show the distribution of changes in temperature obtained from CMIP5 projections (RCP8.5) and expert elicitations and for two 30-yr long periods centered in 2040 and 2070.

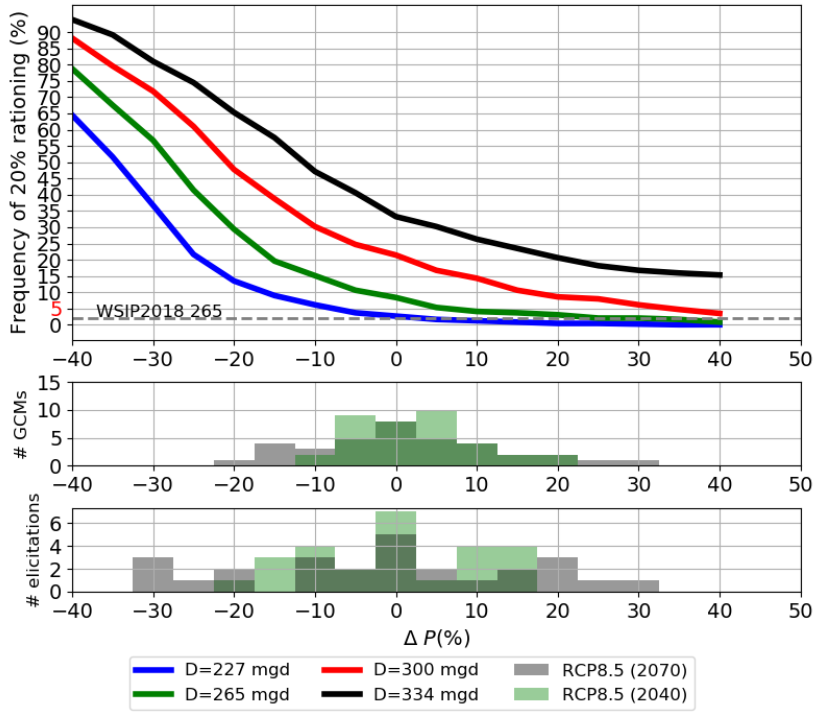


Figure 5-23. Effect of Precipitation Change on the Frequency of 20% Rationing (%).
See Figure 5-22 caption for details explanation of the figure components.

Under baseline demand scenario of 227 mgd, the frequency of 20% rationing is not significantly sensitive to temperature change as it increases from about 3% to 5% when temperature raises by 7°C. (Figure 5-22). However, the sensitivity to temperature change increases quickly when demand increases.

The sensitivity of the frequency of 20% rationing to change in precipitation increases with demand (Figure 5-23). Under baseline demand, the frequency of 20% rationing gets larger than the 5% target if precipitation decreases by more than 5%, which, as mentioned above, is possible (respectively likely) given the climate projections and elicitation of experts by 2040 (respectively 2070). For scenarios with larger demand, any decrease in precipitation significantly increases the frequency of rationing at 20%. For example, considering a demand of 265 mgd (i.e., +15% from baseline), a decrease in precipitation by 10% raises the frequency from 8% to 15%.

There is a significant sensitivity of delivery reliability to change in demand. For instance, if demand increases by 15%, the frequency of 20% rationing increases from 3% to 8%. An increase in precipitation by 10% is required to maintain the frequency below the target of 5%. If demand increase by 30%, frequency of 20% rationing becomes 21%. In this case, an unrealistic 35% increase in precipitation (i.e., given the climate projections and the elicitation of experts) is required to maintain the frequency of 20% rationing below 5%.

Combining the joint effect of precipitation, temperature and demand changes (Figure 5-24) with climate projections (Figure 5-25), demand appears to be the main driver of change in frequency of 20% rationing. The 5% target is likely not to be met if demand increase by 15% or more. Climate change tends to increase the uncertainty (see the spread of the boxplot in Figure 5-25).

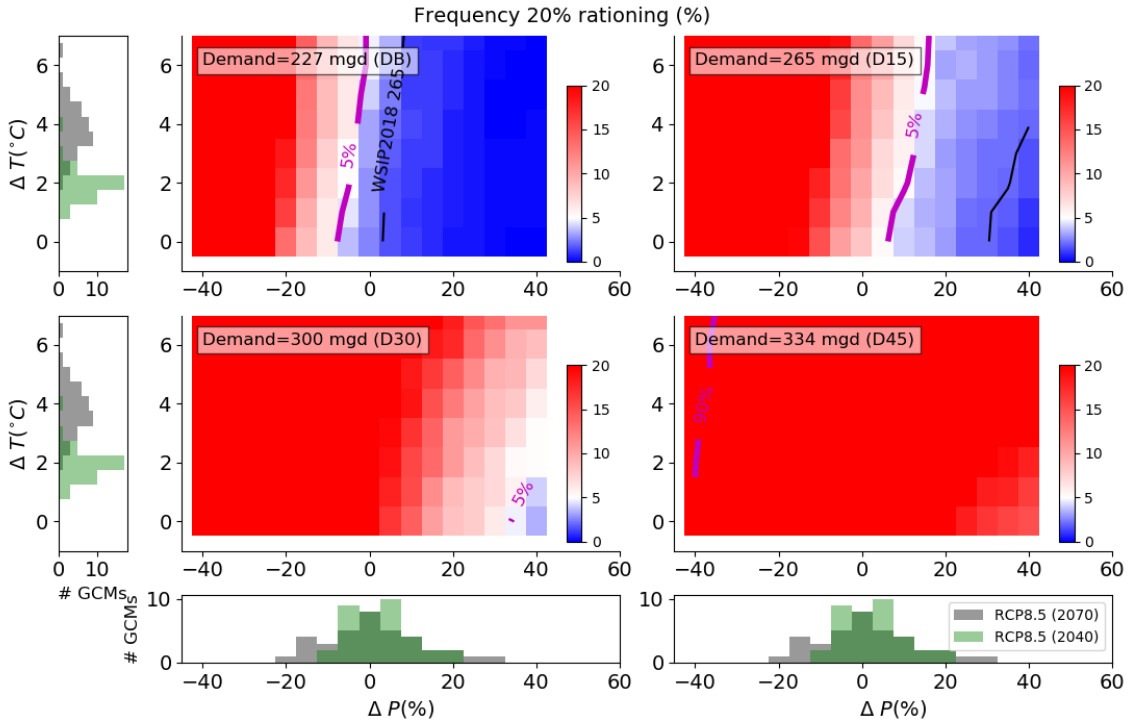


Figure 5-24. Combined Effect of Temperature (ΔT), Precipitation (ΔP), and Demand Changes on the Frequency of 20% Rationing (Color Map).

The black contour lines show the frequency obtained from HHLSM WSIP 2018 and the purple line shows the 5% frequency target. The distribution plots show expected changes in precipitation (horizontal) and temperature (vertical) from CMIP5 climate projections (RCP8.5). Two 30-yr long periods centered in 2040 and 2070 are considered.

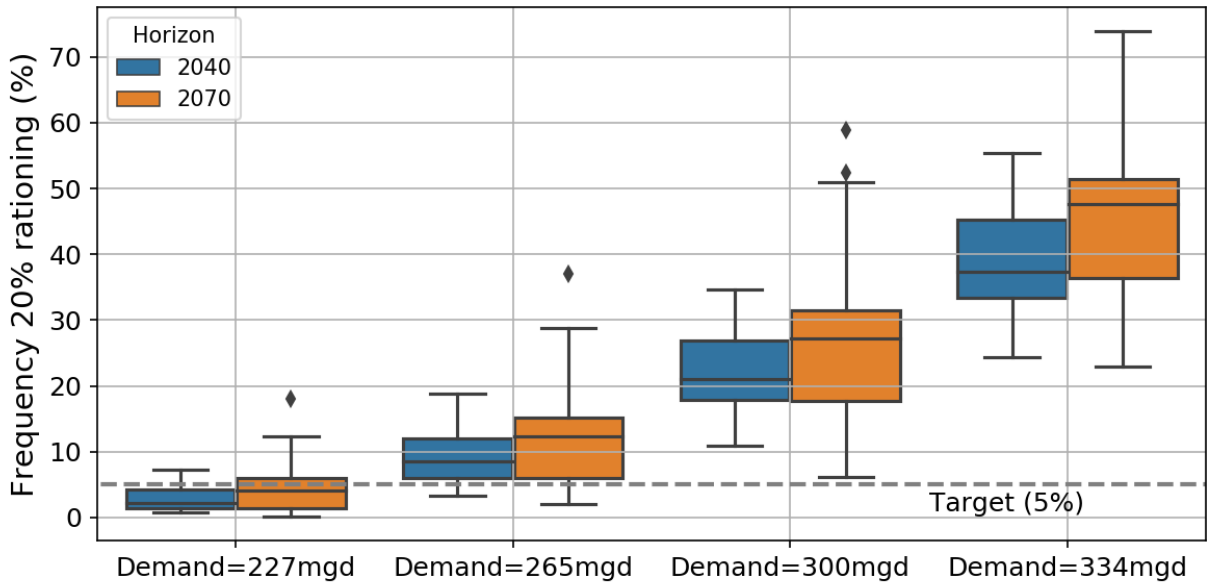


Figure 5-25. Expected Change in Frequency of 20% Rationing When Combining the Vulnerability of the System with the Climate Projections.

The 5% target is shown with the dash grey line.

In summary, frequency of 20% rationing is more sensitive to demand change than to change in temperature and precipitation.

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in meeting the 5% target with 227 mgd demand (i.e., frequency equal to 3%) and failing the 5% target with a demand of 265 (frequency around 10%, which means 20% rationing in 1 out of 10 years on average) which corresponds to a threefold increase in the frequency of rationing.
- By 2040, most projections and elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5%. At +4°C and -5%, the frequency does not meet the 5% target with 227 mgd of demand (i.e., 6.5%) and fails even more with 265 mgd of demand (frequency around 12%, 20% rationing in about 1 out of 8 years on average).
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual precipitation in meeting the 5% target for 227 mgd demand (i.e., frequency around 3%) and failing the 5% target for 265 mgd demand (frequency around 11%) Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15%. At +6°C and -15%, the frequency fails the target with 227 mgd demand (frequency around 15%, 20% rationing in about 1 out of 10 years) and fails the target with 265 mgd demand (frequency larger than 30%; 20% rationing in 1 out of 3 years on average).

5.2.4 Duration of Rationing

Duration of rationing includes the total number of consecutive years of rationing in any given climate/demand scenario across climate realizations. The maximum duration is compared with simulation results of HHLSM for the performance of the RWS with the planned WSIP 2018 at a level of

demand of 265 mgd, which are 5 consecutive years using the historical record and 6.5 years during Design Drought[7].

Figure 5-26 and Figure 5-27 illustrate the increase in rationing associated with a decrease in precipitation. Figure 5-26 is for current demand condition (i.e., 227 mgd) and Figure 5-27 for an increase in demand by 30% (i.e., 300 mgd). It is noted that multiple short rationing periods can merge into a single giant rationing period when precipitation decreases (Figure 5-27).

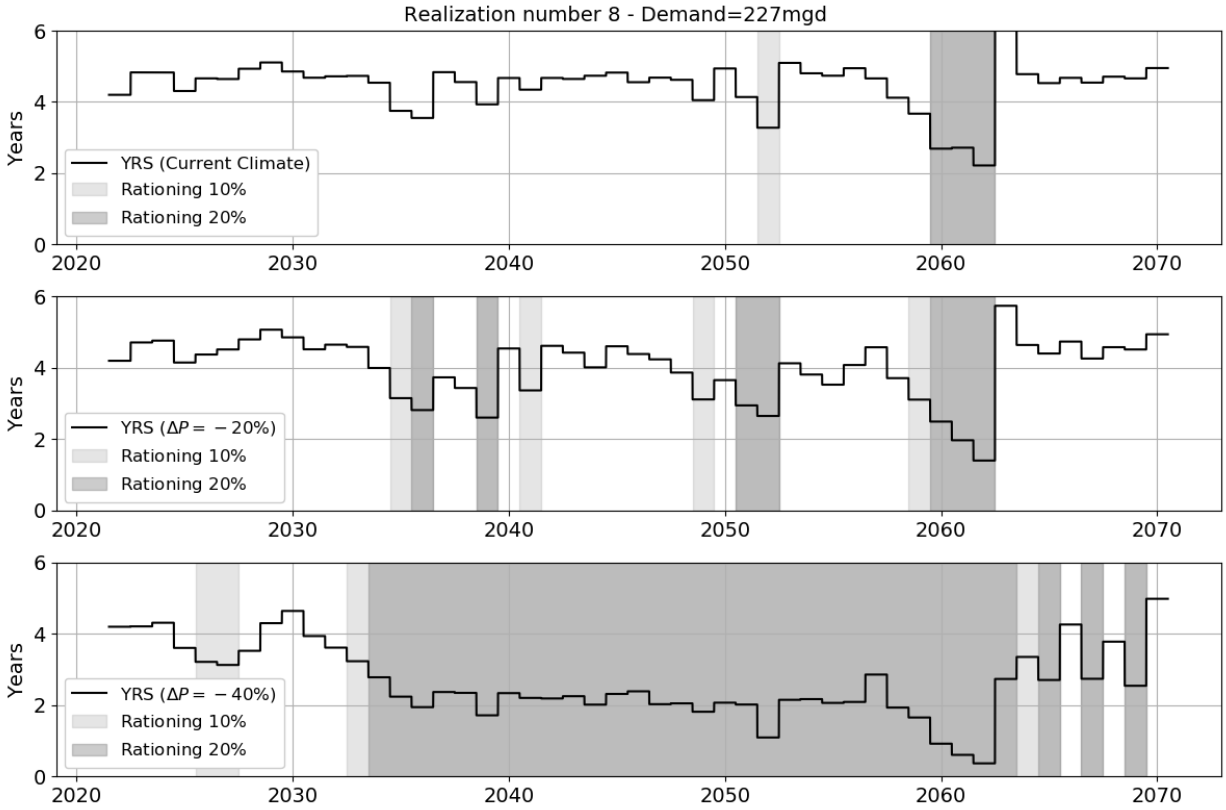


Figure 5-26. Illustration of the Impact of Precipitation Change on the Rationing.
From top to bottom, the subplots show the Years of Remaining Supply (YRS) for the current climate, a decrease in precipitation by 20% and a decrease in precipitation by 40%. The light and dark grey areas highlight the periods of rationing with 10% (i.e., YRS<3.41 years) and 20% (YRS<3.14 years) respectively. For this example, the Realization number 6 is considered under current demand conditions (i.e., 227 mgd).

[7] SFPUC uses a synthetic drought for long-term planning. This drought is the 1987-92 drought with 1976-77 appended, creating an 8.5 year drought. Note that in SFWSM, the duration of rationing is integer and cannot take value such as 1.5 years, or the duration of rationing in the design drought, which is 6.5 years.

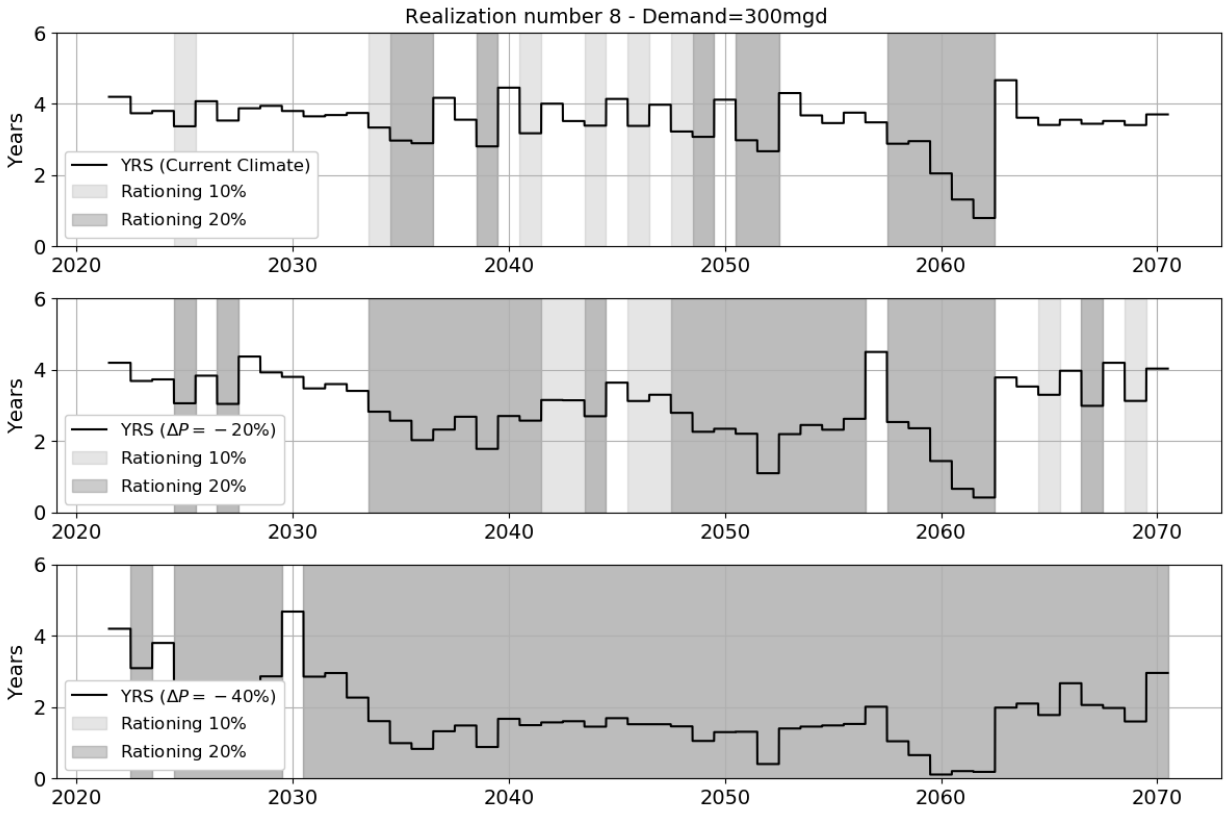


Figure 5-27. Illustration of the Impact of Precipitation Change on the Rationing.
From top to bottom, the subplots show the Years of Remaining Supply (YRS) for the current climate, a decrease in precipitation by 20% and a decrease in precipitation by 40%. The light and dark grey areas highlight the periods of rationing with 10% (i.e., YRS<3.41 years) and 20% (YRS<3.14 years) respectively. For this example, the Realization number 6 is considered under current demand conditions (i.e., 300 mgd).

Maximum duration of rationing is considered here, and presented in Figure 5-28 through Figure 5-31. From these figures, several observations can be made, as follows.

The effect of temperature alone on the maximum duration of rationing is such that under baseline demand scenario the duration increases from 3 to 4 years when temperature increases by 2°C and then to 5 years if temperature increases by 4°C or more (Figure 5-28).

There is a significant sensitivity of the maximum duration of rationing to change in precipitation (Figure 5-29). Under baseline demand (227 mgd), the maximum duration of rationing could get to 5 years if precipitation decreases by 10%. As discussed above, such a change in precipitation becomes more likely toward the second half of the century.

There is a significant sensitivity of the maximum duration of rationing to change in demand. More specifically, if demand increases by 15%, the maximum duration of rationing increases from 3 to 5 years. For the 300 mgd demand scenario (i.e., an increase by 30% from the baseline), the maximum duration gets to 8 years, which exceeds the duration of rationing in the design drought (i.e., 6.5 years).

Combining the joint effect of precipitation, temperature and demand changes (Figure 5-30) with climate projections (Figure 5-31), demand appears to be the main driver of change in maximum duration of rationing. If demand increases by 15% (265 mgd), the large majority of the climate projections indicates that the maximum duration of rationing could get above the target of 5 years by 2040.

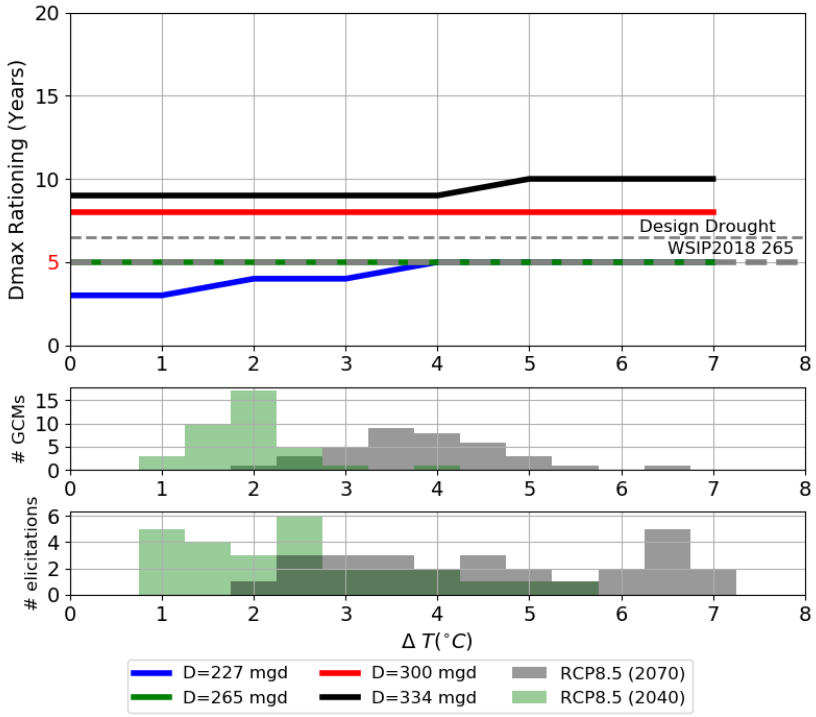


Figure 5-28. Effect of Temperature Change on the Maximum Duration of Rationing (Years).
Top panel shows the maximum duration of rationing across all temperature (ΔT) scenarios and a subset of four demand scenarios (227 mgd, baseline SFPUC demand; 265 mgd [+15%]; 300 mgd [+30%]; and 334 mgd [+45%]). The dash-grey lines show the rationing duration obtained via HHLSM WSIP 2018 for the design drought. Middle and bottom panels show the distribution of changes in temperature obtained from CMIP5 projections (RCP8.5) and expert elicitations and for two 30-yr long periods centered in 2040 and 2070.

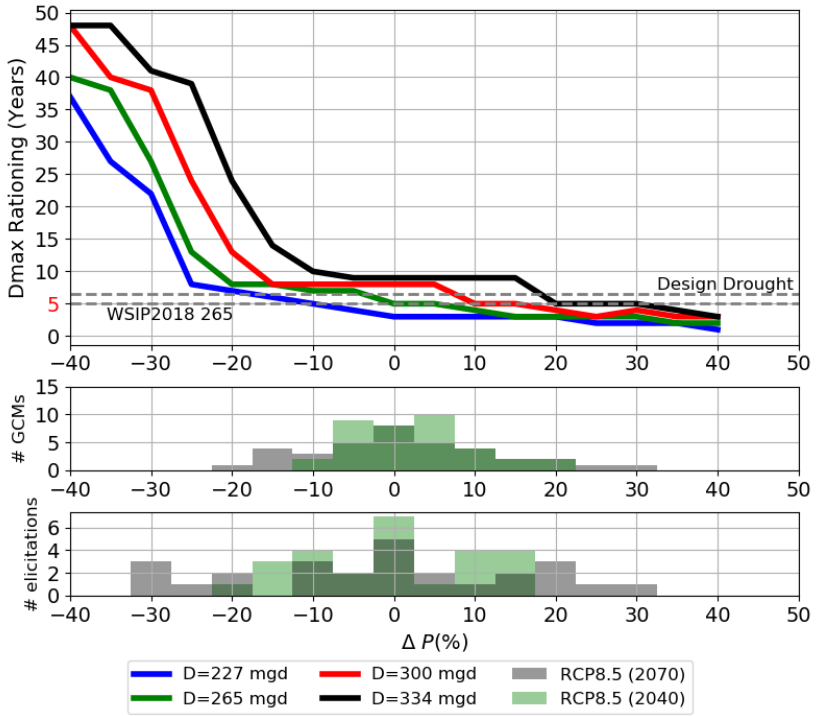


Figure 5-29. Effect of Precipitation Change on the Maximum Drought Duration (Years).
See Figure 5-28 caption for details explanation of the figure components.

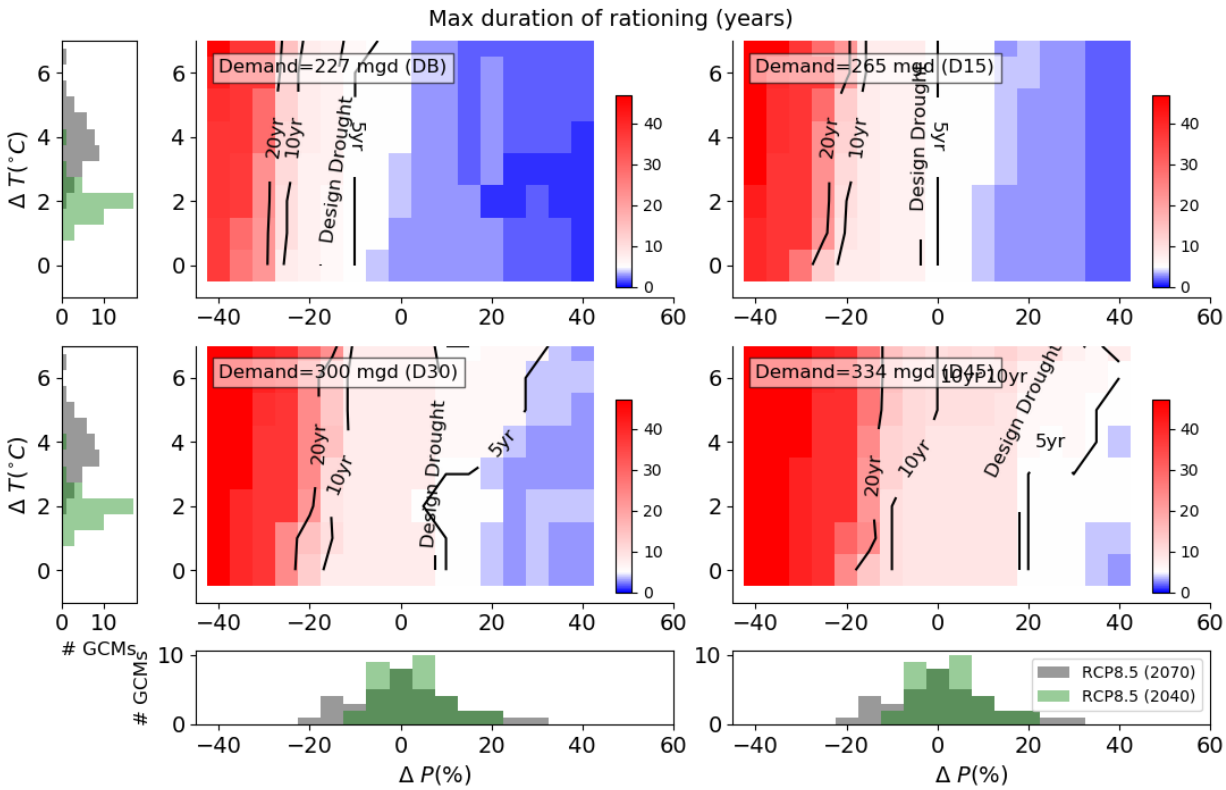


Figure 5-30. Combined Effect of Temperature (ΔT), Precipitation (ΔP), and Demand Changes on the Maximum Duration of Rationing (Color Map).

To provide context, the black contour lines show the several durations and the duration of rationing in design drought. The distribution plots show expected changes in precipitation (horizontal) and temperature (vertical) from CMIP5 climate projections (RCP8.5). Two 30-yr long periods centered in 2040 and 2070 are considered.

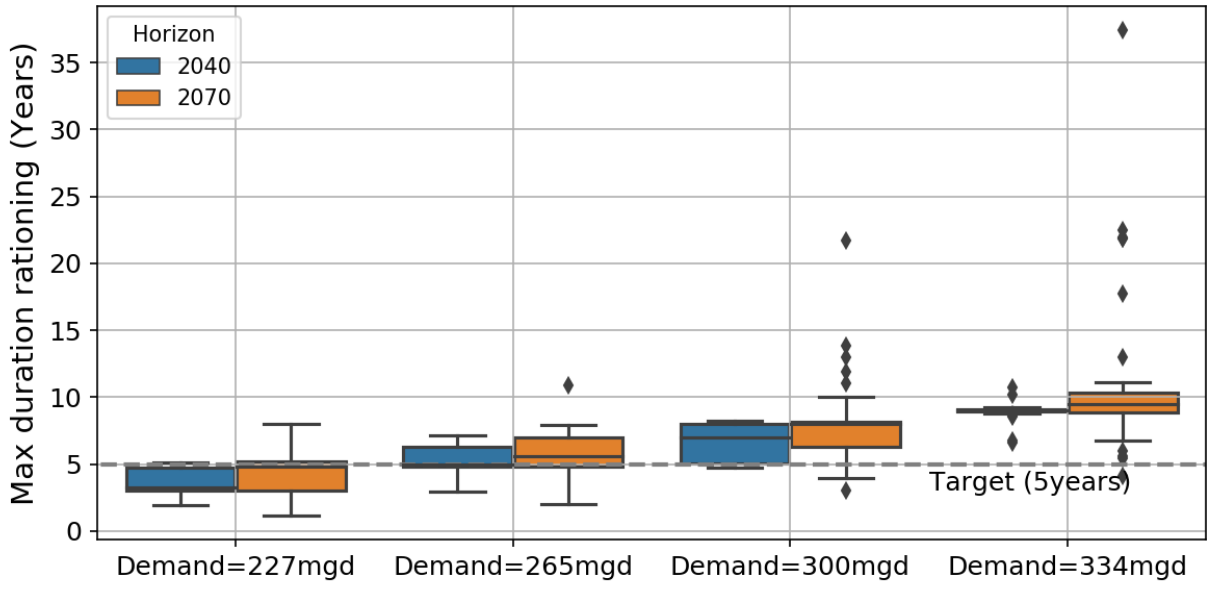


Figure 5-31. Expected Change in Maximum Duration of Rationing When Combining the Vulnerability of the System with the Climate Projections.

The 5 years target is indicated with dashed grey line.

In summary, at a demand of 227 mgd, maximum duration of rationing remains below 5 years for a most climate projections and elicitations by 2040. This is not the case for a demand of 265 mgd with a reduction in precipitation of 5%.

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in meeting the 5-year target for both 227 mgd and 265 mgd demand.
- By 2040, most projections and elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5%. At +4°C and -5%, the maximum duration meets the 5-year target with 227 mgd of demand and fails the target with 265 mgd of demand (maximum duration of 7 years).
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual precipitation results in meeting the 5-year target for both 227 mgd and 265 mgd demand. Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15%. At +6°C and -15%, the maximum duration would fail the 5-year target for both 227 mgd and 265 mgd demand.

5.2.5 Water Storage Reliability (above 2.3 Years of Remaining Supply)

Years of Remaining Supply (YRS) refers to the number of years water can be delivered at post-rationing demand and instream flow requirements from total stored water. This target is grounded in the amount of storage in the RWS at the end of the 1987-1992 drought in the SFPUC Design Drought under WSIP 2018. Here, the focus is on the frequency of years during which the system storage is above 2.3 years of remaining supply (noted storage reliability), with results shown in Figure 5-32 through Figure 5-35. Although there is no defined target for SFPUC, a 99% target is here assumed.

Under 227 mgd and 265 mgd demand (i.e., baseline and +15%), the storage reliability is not sensitive to temperature change alone (Figure 5-32). Storage reliability becomes slightly sensitive to temperature change when demand increases. For a demand of 227 mgd, the target is met for all plausible warming alone but for a demand of 265 mgd, the RWS fails the target for warming above +4°C.

There is some sensitivity to precipitation change (Figure 5-33). Under baseline demand, the storage reliability gets below the 99% target if precipitation decrease by 15% or more, which is possible but not really likely given the climate projections, although the expert elicitations deemed that a decrease likely by 2070. However, if demand increases by 15% (265 mgd), the storage reliability is below the 99% as soon as precipitation decreases by 5%, which is more likely as shown by both the climate projections and expert elicitations.

Storage reliability is sensitive to demand change. Under baseline climate, storage reliability is below target if demand increases by 30%. For this demand scenario, an increase in precipitation by at least 10% is required to keep the storage reliability above the 99% target.

Combining the joint effect of precipitation, temperature and demand changes (Figure 5-34) with climate projections (Figure 5-35), demand appears to be the main driver of storage reliability. If demand increases by 30% (300 mgd), the storage reliability cannot meet the 99% target.

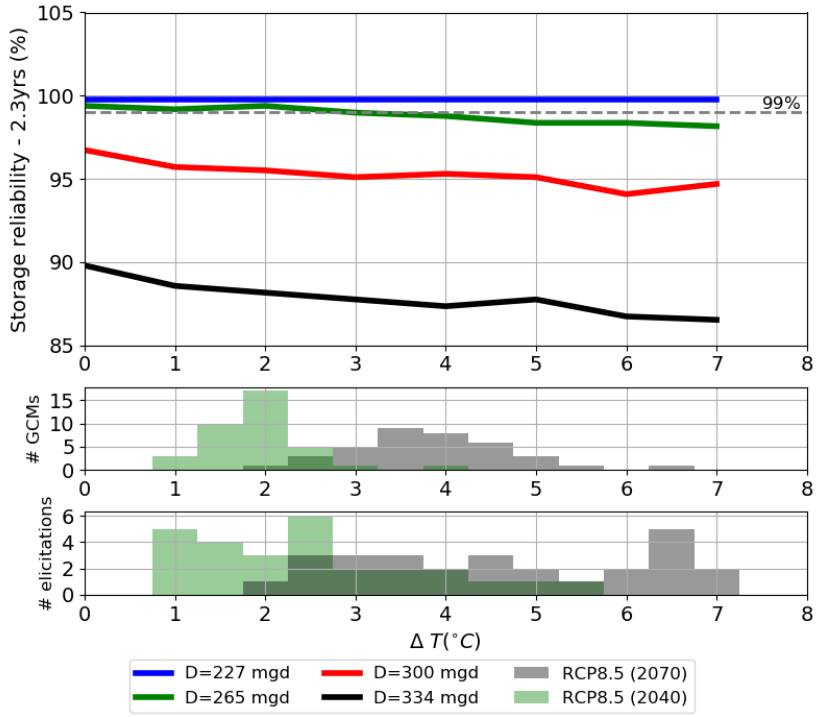


Figure 5-32. Effect of Temperature Change on the Storage Reliability (2.3 yrs).
Top panel shows the maximum reliabilities across all temperature (ΔT) scenarios and a subset of four demand scenarios (227 mgd, baseline SFPUC demand; 265 mgd [+15%]; 300 mgd [+30%]; and 334 mgd [+45%]). The dash-grey line shows the 99% target. Middle and bottom panels show the distribution of changes in temperature obtained from CMIP5 projections (RCP8.5) and expert elicitations and for two 30-yr long periods centered in 2040 and 2070.

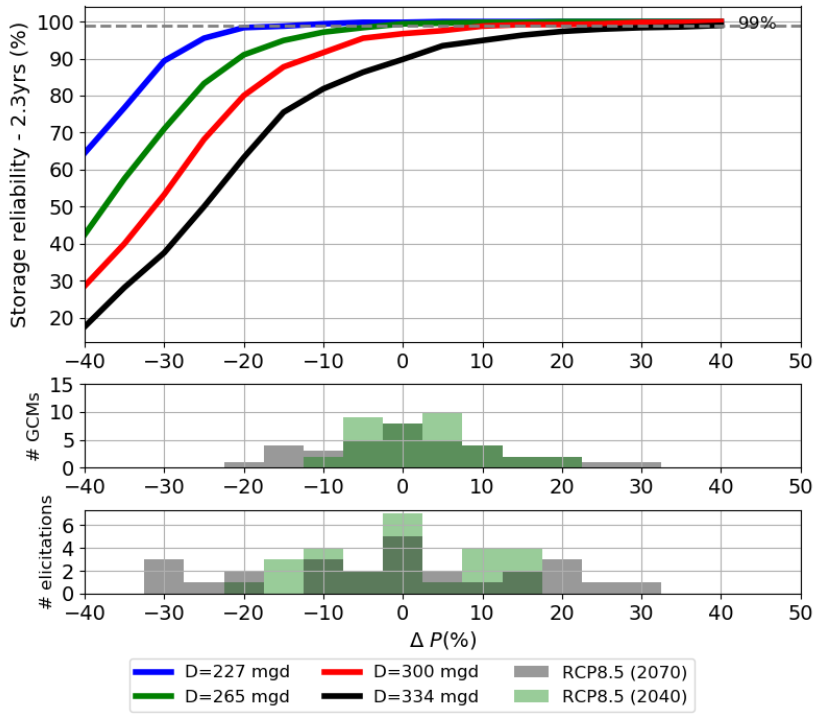


Figure 5-33. Effect of Precipitation Change on the Storage Reliability (2.3 yrs).
See Figure 5-32 caption for details explanation of the figure components.

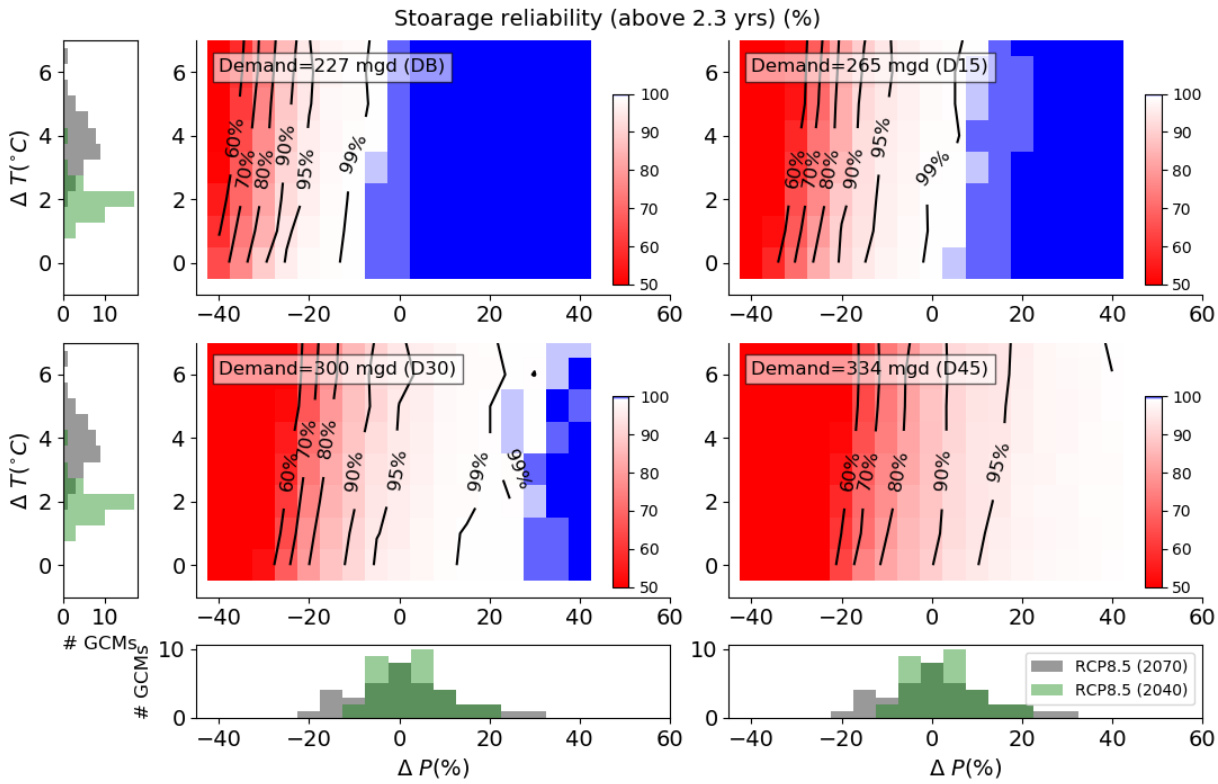


Figure 5-34. Combined Effect of Temperature (ΔT), Precipitation (ΔP), and Demand Changes on the Storage Reliability (above 2.3 yrs) (Color Map).

To ease the reading, the black contour lines show the several percentage. The distribution plots show expected changes in precipitation (horizontal) and temperature (vertical) from CMIP5 climate projections (RCP8.5). Two 30-yr long periods centered in 2040 and 2070 are considered.

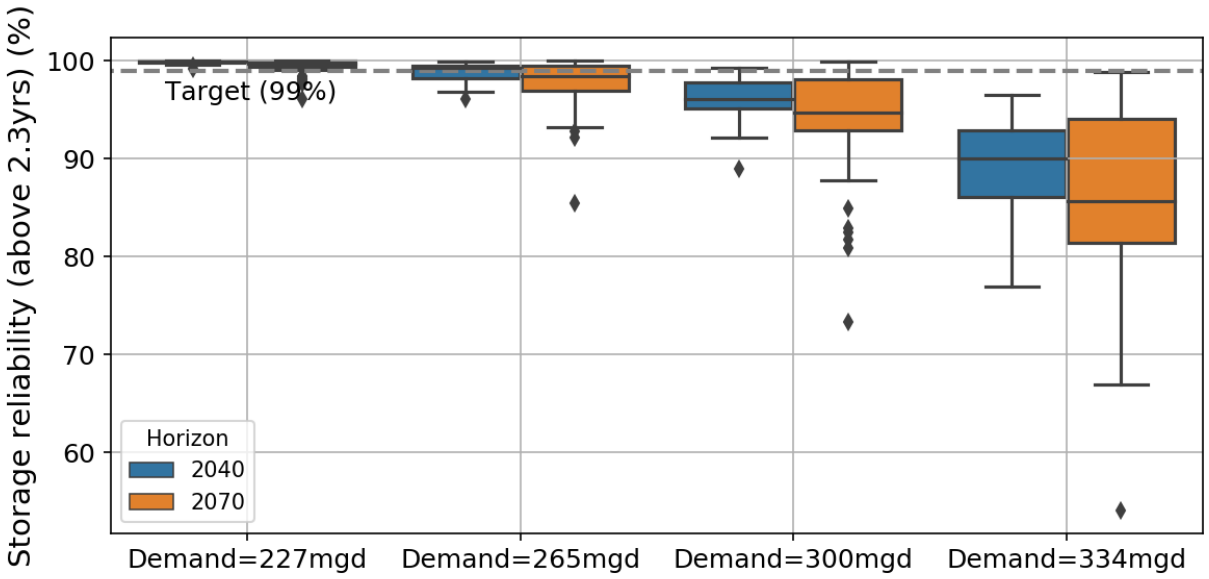


Figure 5-35. Expected Change in Storage Reliability When Combining the Vulnerability of the System with the Climate Projections.

The dash-grey line shows the 99% target.

In summary, storage reliability is more sensitive to demand change than change in temperature and precipitation.

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in meeting the storage reliability target for both 227 mgd and 265 mgd demand.
- By 2040, most projections and elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5%. At +4°C and -5%, the storage reliability meets the target with 227 mgd of demand but fails the 99% reliability target with 265 mgd of demand (i.e., 97% storage reliability).
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual precipitation results in meeting the 99% storage reliability target for both 227 mgd but not for 265 mgd demand (i.e., 98.7% reliability). Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15%. At +6°C and -15%, the storage reliability fails the target for both 227 mgd and 265 mgd demand.

5.2.6 Maximum Annual Supply Deficit

The annual supply deficit is the annual sum of the deviations between pre-rationing demand and actual deliveries (i.e., demand rationing plus virtual water volumes). The maximum value obtained across the 10 realizations is reported here, with results shown in Figure 5-36 through Figure 5-39. The supply deficit is compared to 59.3 TAF grounded in the performance of the RWS with the planned WSIP 2018 at a level of demand of 265 mgd.

Under baseline demand, the maximum annual supply deficit is not sensitive to temperature change alone (Figure 5-36). When demand increases the deficit increases with temperature, but the relationship appears to be neither significant nor consistent.

There is a significant sensitivity of the maximum annual supply deficit to change in precipitation (Figure 5-37). With a demand of 265 mgd (+15% increase from the baseline demand), a decrease in precipitation by 10% leads to an annual deficit (roughly 100 TAF) larger than one fourth of Hetch Hetchy reservoir storage capacity (about 90 TAF).

There is a significant sensitivity of the maximum annual supply deficit to change in demand. For instance, if demand increases by 30% (i.e., from 227 mgd to 300 mgd), the maximum annual supply deficit doubles. It triples if demand increases by 45% (from 227 mgd to 334 mgd).

Combining the joint effect of precipitation, temperature and demand changes (Figure 5-38) with climate projections (Figure 5-39), demand appears to be the main driver for the maximum annual cumulative deficit. For instance, if demand increases by 45% (334 mgd), the maximum annual supply deficit could be multiplied by three. In addition, climate change will increase the uncertainty of change in deficit.

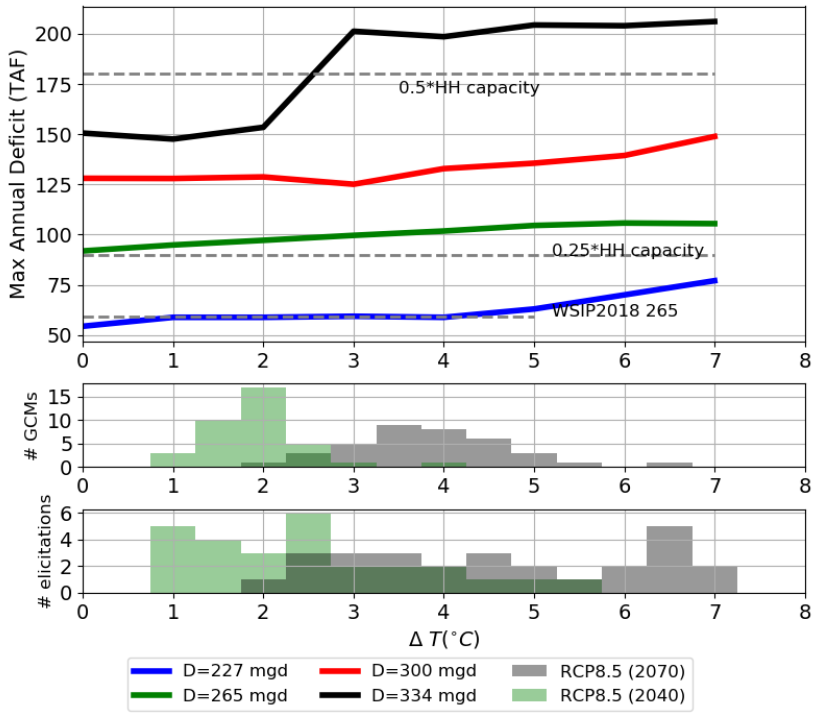


Figure 5-36. Effect of Temperature Change on the Maximum Annual Deficit (TAF).
Top panel shows the maximum annual deficit across all temperature (ΔT) scenarios and a subset of four demand scenarios (227 mgd, baseline SFPUC demand; 265 mgd [+15%]; 300 mgd [+30%]; and 334 mgd [+45%]). The dash-grey lines show the deficit obtained from HHSLM WSIP 2018 and for context two fractions of Hetch Hetchy storage. Middle and bottom panels show the distribution of changes in temperature obtained from CMIP5 projections (RCP8.5) and expert elicitations and for two 30-yr long periods centered in 2040 and 2070.

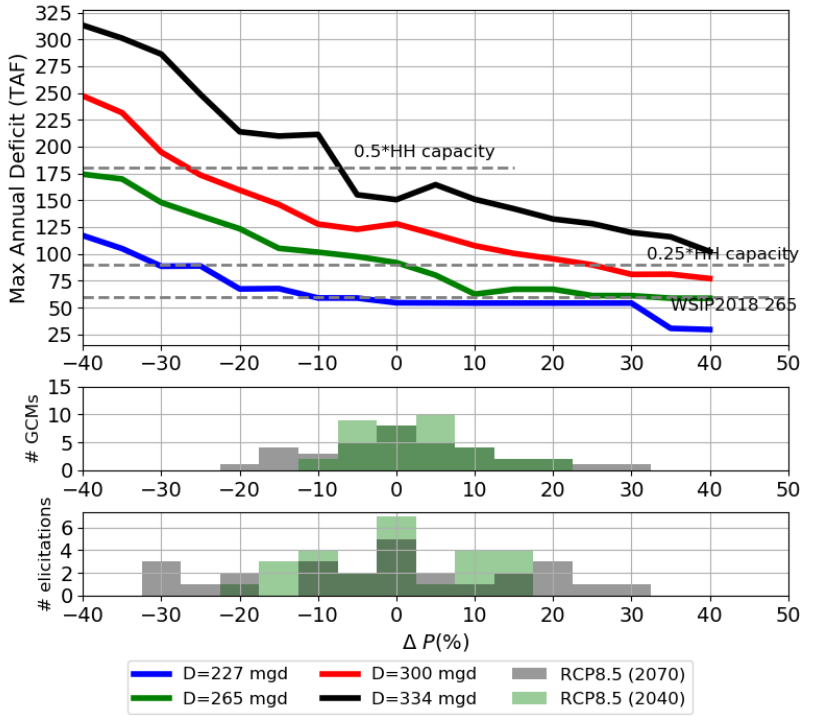


Figure 5-37. Effect of Precipitation Change on the Maximum Annual Deficit (TAF).
See Figure 5-36 caption for details explanation of the figure components.

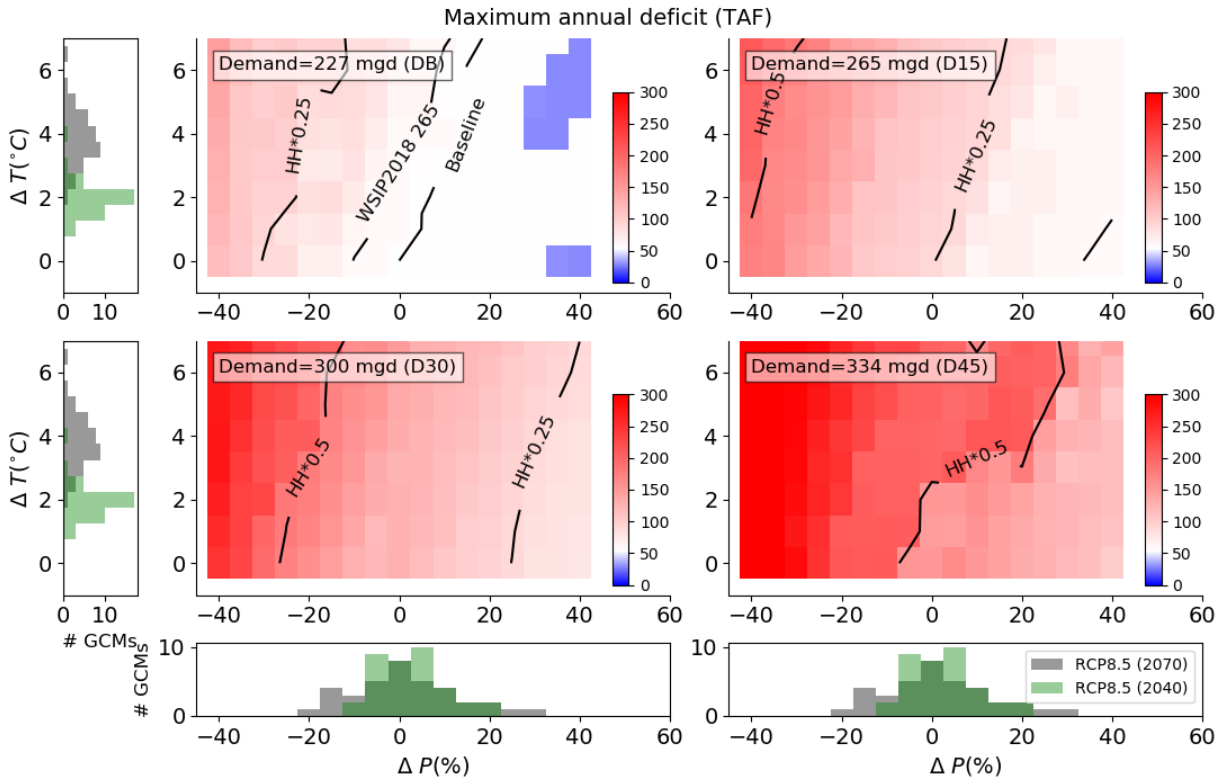


Figure 5-38. Combined Effect of Temperature (ΔT), Precipitation (ΔP), and Demand Changes on the Maximum Annual Deficit (Color Map).

To ease the reading, the black contour lines show the value obtained for the baseline, HHLSM WSIP 2018 and two fractions of Hetch Hetchy storage capacity. The distribution plots show expected changes in precipitation (horizontal) and temperature (vertical) from CMIP5 climate projections (RCP8.5). Two 30-yr long periods centered in 2040 and 2070 are considered.

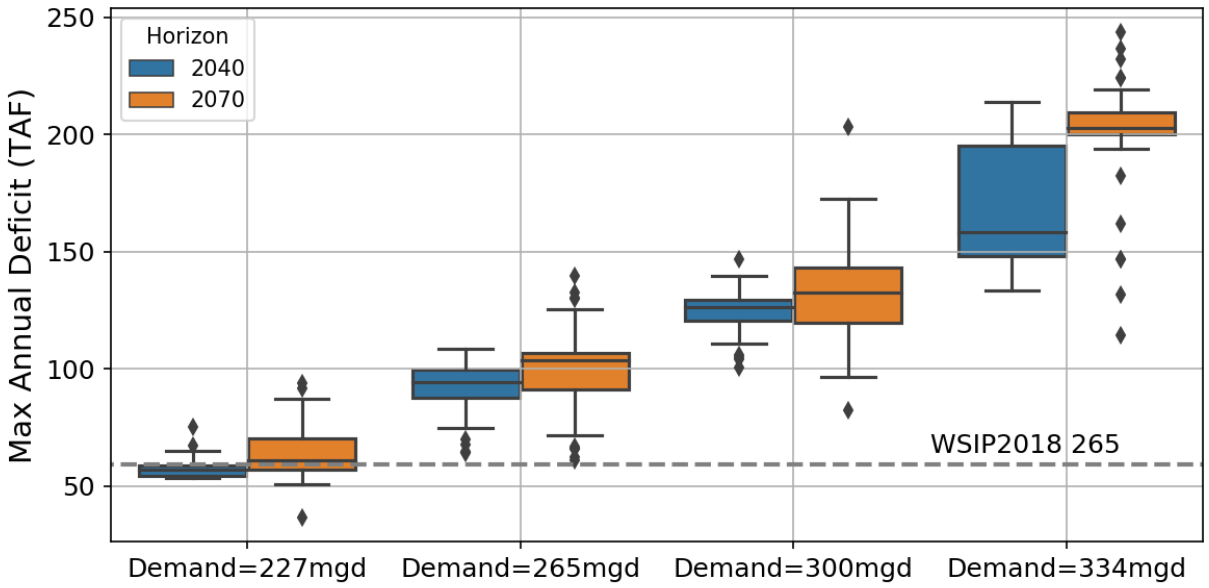


Figure 5-39. Expected Change in Maximum Annual Deficit When Combining the Vulnerability of the System with the Climate Projections.

The dash-grey line shows the simulated deficit with HHLSM WSIP 2018.

In summary, maximum annual supply deficit is more sensitive to demand change than change in temperature and precipitation.

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in maximum annual supply deficit of about 60 TAF for a 227 mgd demand (equivalent rationing of 24%) and about 97 TAF for a 265 mgd demand (equivalent rationing of 34%).
- By 2040, most projections and elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5%. At +4°C and -5%, the maximum annual supply deficits are around 70 TAF for a 227 mgd demand and about 107 TAF for a 265 mgd demand .
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual precipitation results maximum annual supply deficit of about 60 TAF and 100 TAF for 227 mgd and 265 mgd demand, respectively. Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15%. At +6°C and -15%, the maximum annual supply deficit is about 95 TAF for a 227 mgd demand (equivalent rationing of 37%) and 134 TAF (equivalent rationing of 45%) for a 265 mgd demand.

5.2.7 Maximum Cumulative Supply Deficit through Rationing Events

Cumulative supply deficit through drought events is the sum of daily deviations between pre-rationing demand and actual deliveries through a rationing event, which can last multiple years within a continuous sequence of rationing. The maximum values across the 10 realizations are reported here, with relevant results shown in Figure 5-40 through Figure 5-43.

There is no apparent influence of temperature change on maximum cumulative deficit, except for demand greater than 265 mgd (Figure 5-40).

There is a significant sensitivity to change in precipitation (Figure 5-41). For instance, under current demand condition (227 mgd), if precipitation decreases by 15%, the maximum cumulative deficit more than doubles and would nearly correspond to twice the full Hetch Hetchy storage capacity. With a decrease in precipitation by 20% and greater, the maximum cumulative deficit increases linearly with increase in demand.

Combining the joint effect of precipitation, temperature and demand changes (Figure 5-42) with climate projections (Figure 5-43), demand appears to be the main driver for the cumulative deficit through rationing events. For instance, if demand increases by 45% (334 mgd), the maximum cumulative deficit through rationing events could be multiplied by eight. In addition, climate change would further increase the cumulative deficit.

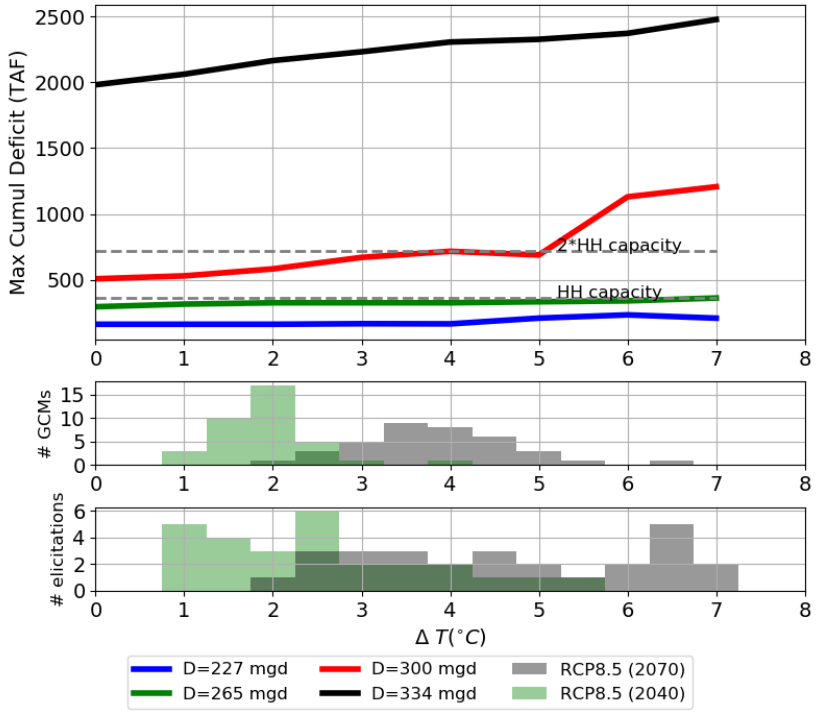


Figure 5-40. Effect of Temperature Change on the Maximum Cumulative Deficit through Rationing Event (TAF). Top panel shows the maximum cumulative deficit across all temperature (ΔT) scenarios and a subset of four demand scenarios (227 mgd, baseline SFPUC demand; 265 mgd [+15%]; 300 mgd [+30%]; and 334 mgd [+45%]). The dash-grey lines show the deficit obtained from HHSLM WSIP 2018 and for context two fractions of Hetch Hetchy storage. Middle and bottom panels show the distribution of changes in temperature obtained from CMIP5 projections (RCP8.5) and expert elicitations and for two 30-yr long periods centered in 2040 and 2070.

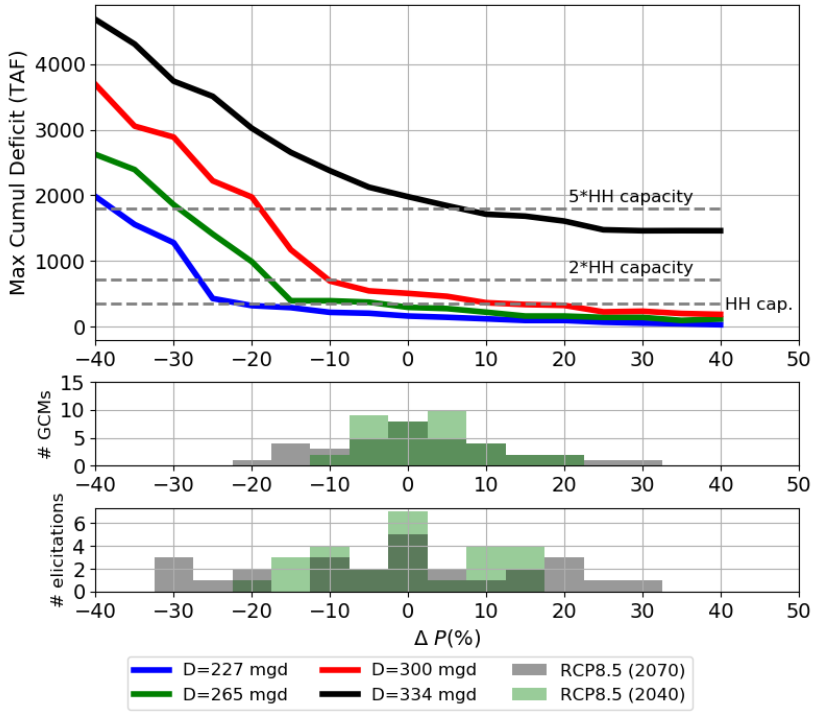


Figure 5-41. Effect of Precipitation Change on the Maximum Annual Deficit (TAF). See Figure 5-40 caption for details explanation of the figure components.

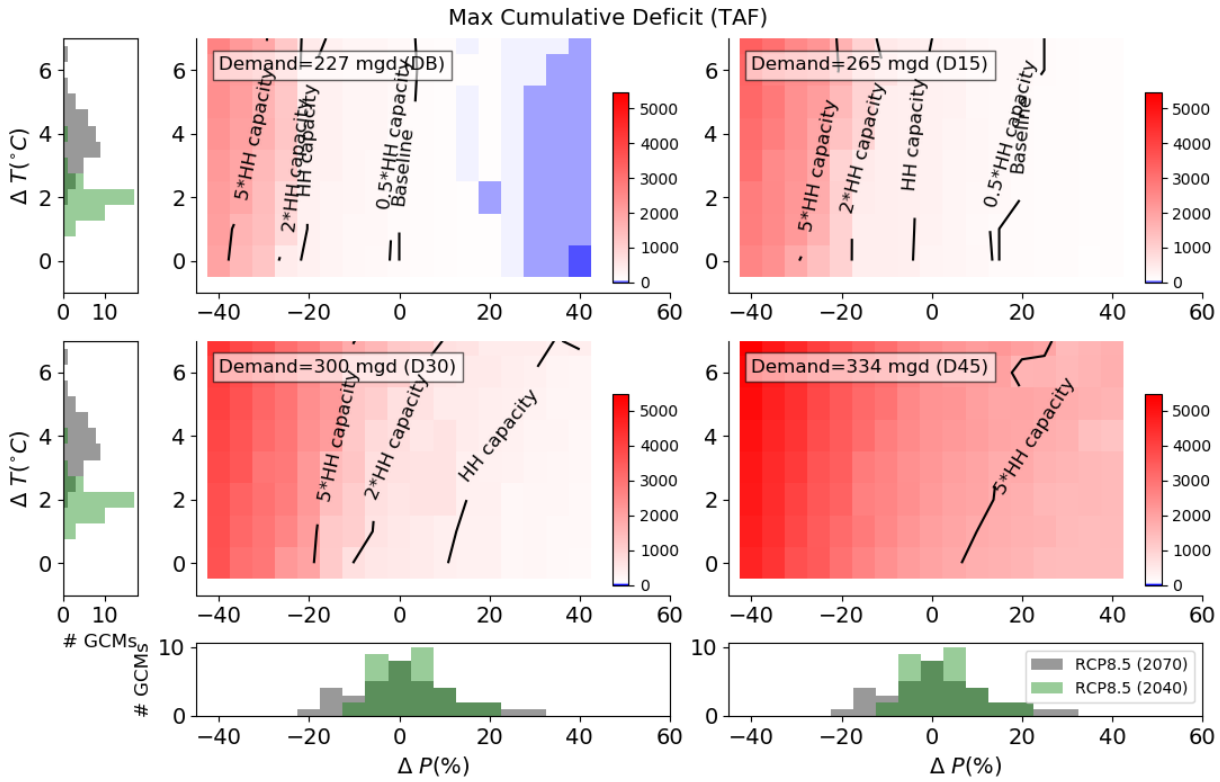


Figure 5-42. Combined Effect of Temperature (ΔT), Precipitation (ΔP), and Demand Changes on the Maximum Cumulative Deficit (Color Map).

To ease the reading, the black contour lines show the value obtained for the baseline, and three fractions of Hetch Hetchy storage capacity. The distribution plots show expected changes in precipitation (horizontal) and temperature (vertical) from CMIP5 climate projections (RCP8.5). Two 30-yr long periods centered in 2040 and 2070 are considered.

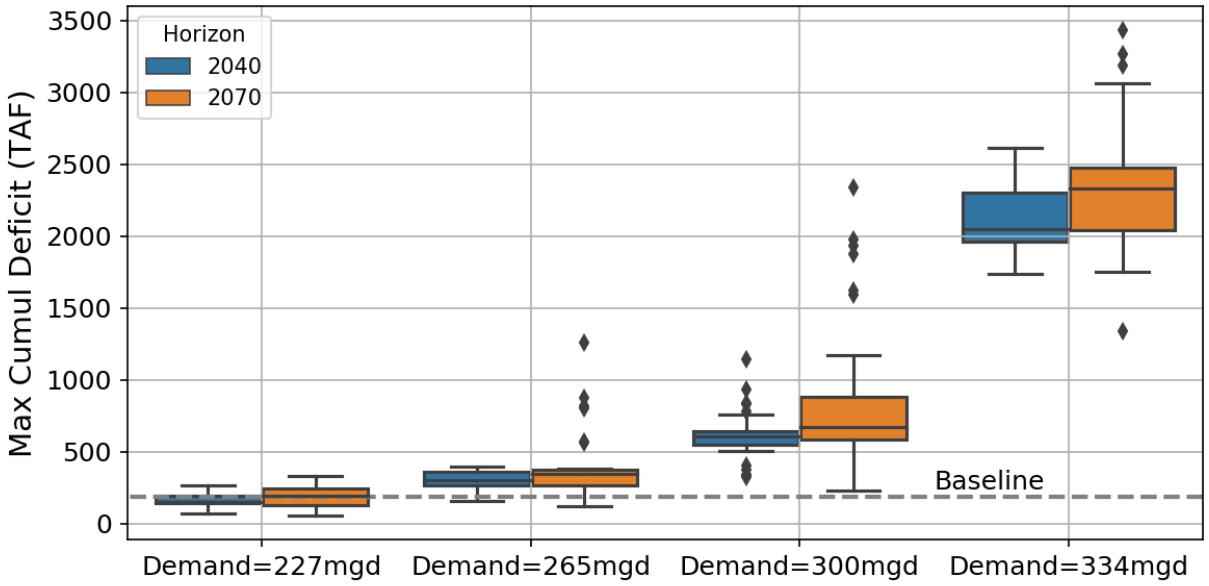


Figure 5-43. Expected Change in Maximum Cumulative Deficit When Combining the Vulnerability of the System with the Climate Projections.

The baseline (i.e., dash-grey line) is the shows the deficit obtained with SFWSM with no climate or demand change.

In summary, maximum annual supply deficit is more sensitive to demand change than change in temperature and precipitation.

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in maximum cumulative supply deficit of 162 TAF for a 227 mgd demand and of 325 TAF for a 265 mgd demand.
- By 2040, most projections and elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5%. At +4°C and -5%, the maximum cumulative supply deficits are around 257 TAF for a 227 mgd demand and about 373 TAF for a 265 mgd demand.
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual precipitation results maximum cumulative supply deficit of about 165 TAF and 325 TAF for 227 mgd and 265 mgd demand, respectively. Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15%. At +6°C and -15%, the maximum cumulative supply deficit is about 320 TAF and 991 TAF for 227 mgd and 265 mgd demand, respectively.

5.2.8 Demand Narrative-Based Scenarios

SFPUC is interested in understanding what the performance implications would be of capping supply to permanent customers at their ISG level (NAD1.1). Under this assumption, SFPUC is also interested in the performance implications of making Santa Clara and San Jose permanent customers (NAD1.2). See Section 4.3.3 for description of the demand narrative-based scenarios.

The water delivery reliability is presented for each of the demand narratives (Figure 5-44) across a range of potential system-wide demand by 2070 for three precipitation change scenarios of -30%, 0 and +30%. The demand in 2070 is presented as a distribution that accounts for the uncertainty in drivers of demand such as income, population, and others as well as in the elasticities of demand (HRG 2020a). The “Reference” distribution (black line) assumes uncertainty in both wholesale and retail demand. For NAD1.1 and NAD1.2 distributions, wholesale demands are maintained constant at 184 mgd and 199.5 mgd (Table 4-8) (Suburban retail demand is also maintained constant for simplification). Wholesale demands equal their ISG and no change in demand is simulated for suburban retail customers. The dispersion for NAD1.1 and NAD1.2 narratives (blue and red lines) comes from the uncertainty stemming from retail demand in San Francisco, which explains the reduced spread of the red and blue distributions compared to the uncertainty stemming from all customers (black distribution). The NAD1.2 distribution (purple) has the same shape as the NAD1.2 distribution (green) but shifted by 15.5 mgd, which is the ISG of Santa Clara and San Jose. There is no notable difference in system performance when permanent wholesale customer’s deliveries are capped to their ISG, even when SC and SJ are included. No significant differences are observed for the other metrics (not shown).

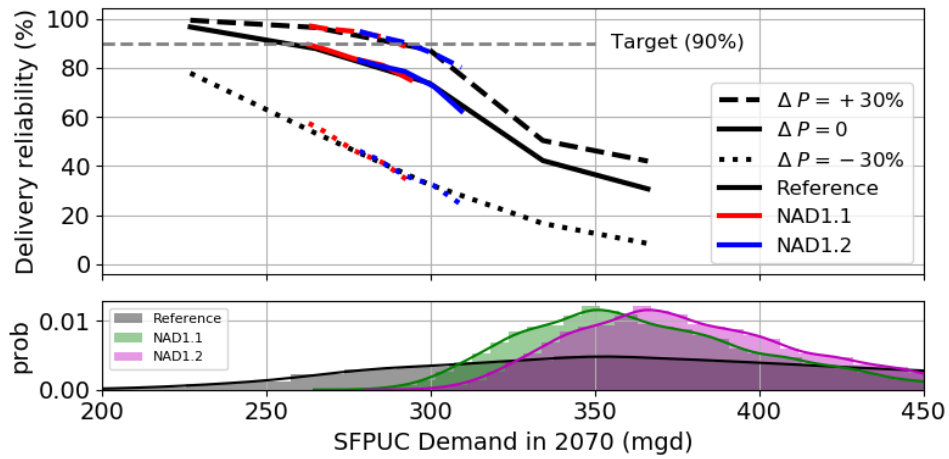


Figure 5-44. Water Delivery Reliability across a Range of Demand Scenarios and a Subset of Precipitation Scenarios (-30%, no Change, and +30% Precipitation).

The distribution in the bottom plots show the expected demand at the time horizon 2070. These distributions account for the uncertainty in the exogenous drivers of demand (e.g., income, population, etc.) and elasticities of demand; see the Technical Report 3 (HRG 2020a). The dash-grey line shows water delivery reliability target.

5.3 Compound Effects of Climate Change and Instream Flow Requirements on Water Supply

Instream flow requirements (IFRs) represent a demand for water on the RWS. In discussing the vulnerability of the system and alternative IFRs, the impact of climate change on total system IFR (i.e., an aggregation of IFRs across the entire system) for existing IFRs is shown below. The impact on water supply is evaluated with the single metric of water delivery reliability with a target of 90% or rationing in 1 out of 10 years on average (Section 4.4.1). A climate response surface showing the mean annual aggregated system IFRs is shown in Figure 5-45. As expected, an increase in precipitation results in an increase in IFR, as most IFRs—and certainly the dominant IFRs—depend directly on precipitation and/or runoff to determine the IFR value. An increase in temperature decreases IFRs. The decrease in IFR obtained when temperature increases results from location where the IFR depends on runoff and/or the SJVI index. As warming temperature reduces runoff due to larger evapotranspiration, the IFRs decrease. However, IFRs are far more sensitive to precipitation changes than temperature changes within the ranges considered. For instance, the reduction in IFR volume resulting from a warming by 7°C is similar to the one obtained with 10% reduction in precipitation under baseline temperature. Because of this and the fact that GCM projections do not have a clear indication of precipitation change, GCMs do not indicate a general trajectory for changes in IFR volume.

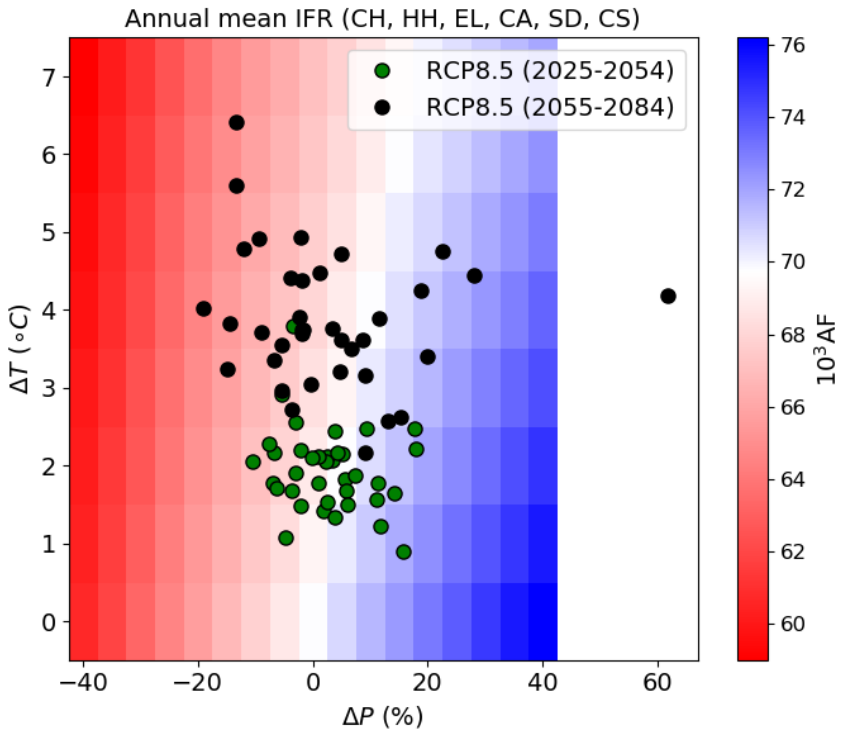


Figure 5-45. Climate Response Surface for Mean Annual Aggregated System Instream Flow Requirements (IFR). The color map shows the annual volume to be provided from Hetch Hetchy (HH), Cherry (CH), Eleanor (EL), Calaveras (CA), Stone Dam (SD) and Crystal Springs (CS) reservoirs. Dots in colors show the CMIP5 projections (RCP8.5) for two 30-yr long periods centered in 2040 and 2070.

5.3.1 UTREP Base Flows below O'Shaughnessy Dam (NAE1.1)

The annual IFR below O'Shaughnessy Dam increases significantly with the UTREP base flows, with the difference dependent on mean annual precipitation (Figure 5-46, left). UTREP base flows are more sensitive to changes in precipitation than the current IFR scheme (Figure 5-46), due to the higher range in base flows across water year types and most of the additional water allocated during wetter years (Figure 4-7).

The UTREP base flows reduce annual spill from Hetch Hetchy reservoir compared to the current IFR, with the absolute reduction increasing with more precipitation and decreasing with less precipitation (Figure 5-46, right). This is due to both the overall reduction in reservoir level (there is less water overall to spill) and the closer alignment of base flows with natural runoff patterns. As such, the losses through spill under baseline are replaced by more losses through IFR under NAE1.1 narrative, which has no impact on water delivery reliability (Figure 5-47 and Figure 5-48).

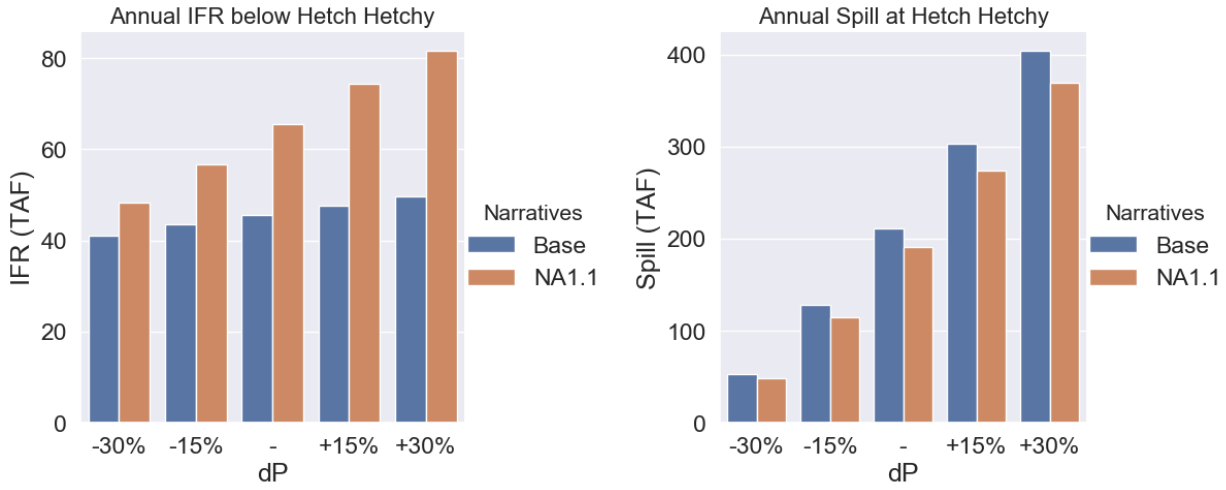


Figure 5-46. Annual IFR (Left) and Spill (Right) below O'Shaughnessy Dam for Each of Base IFR and Narrative (NA1.1) Scenarios with Changes in Precipitation (ΔP).

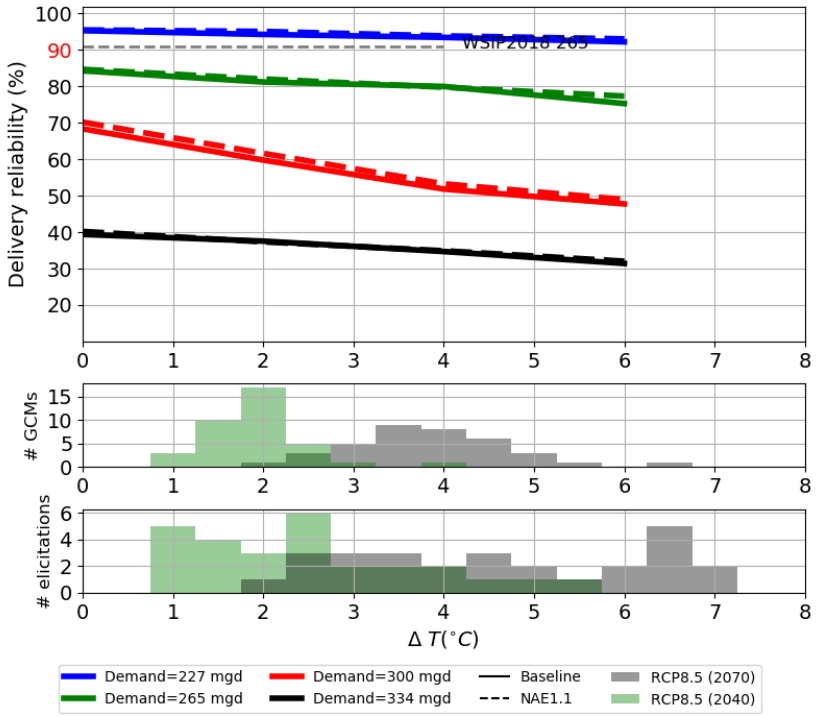


Figure 5-47. Water Delivery Reliability with the Reference and NAE1.1 Narratives across All Temperature (ΔT) Scenarios and a Subset of Four Demand Scenarios (227 mgd, Which Corresponds SFPUC Baseline Demand, 265 mgd [+15% Increase], 300 mgd [+30%] and 334 mgd [+45%]).

The dash-grey lines show reliabilities obtained HHLSM WSIP 2018. Middle and bottom panels show the distribution of changes in temperature obtained from CMIP5 projections (RCP8.5) and expert elicitations and for two 30-yr long periods centered in 2040 and 2070.

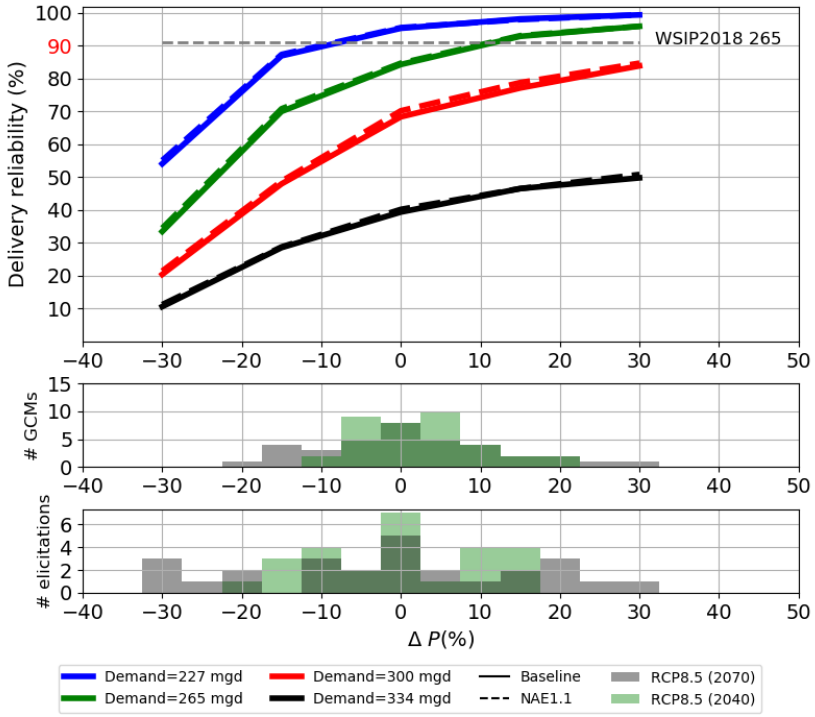


Figure 5-48. Water Delivery Reliability with the Reference and NAE1.1 Narratives across All Temperature (ΔT) Scenarios and a Subset of Four Demand Scenarios.

See Figure 5-47 caption for details explanation of the figure components.

In summary, UTREP baseflows have a negligible impact on water delivery reliability for all plausible changes in mean annual temperature and precipitation.

5.3.2 New IFR below Don Pedro Dam (NAE2.1 and NAE2.2)

The state-amended Water Quality Control Plan (aka state-amended WQCP or NAE2.1) represents a very large increase over the existing IFR below Don Pedro, while the Alternative proposal (NAE2.2), while still much larger than the existing baseline IFR, is substantially lower (Figure 5-49). Noting that the baseline (existing) release for SFPUC is about 50 cfs from June through September, or about 4 TAF/year, the state-amended WQCP represents more than a 50-fold increase above the current baseline with no increase or decrease in historical precipitation. Meanwhile, the Alternative proposal represents only about a 10-fold increase. Importantly, the state-amended WQCP is more sensitive to precipitation than the Alternative proposal. In contrast, the Alternative proposal is far less sensitive to precipitation, with relatively modest changes in total annual environmental releases with changes in overall precipitation. It is also noted that spill from the Water Bank contributes to the IFR.

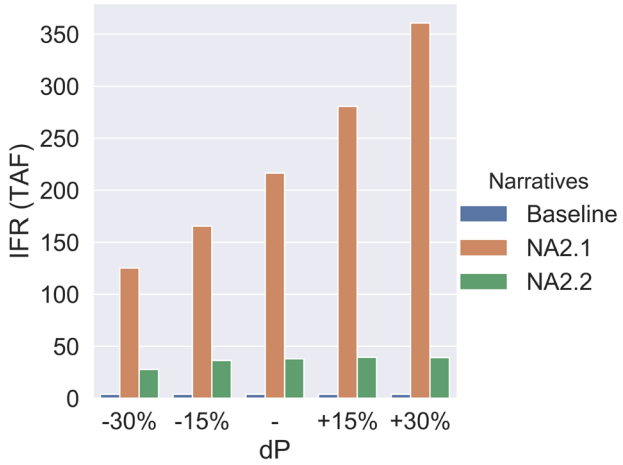


Figure 5-49. Annual IFR from Water Bank.
NAE2.1 Is the State Amended WQCP and NAE2.2 Is the Alternative Proposal.

To make the comparison between NAE2.1 and NAE2.2 easier, only two demand scenarios (227 and 300 mgd) are represented in figures Figure 5-50 and Figure 5-51. Both state-amended WQCP (NAE2.1) and Alternative proposal (NAE2.2) lead to a significant decrease in system performance for current climate and demand condition. The decrease in performance is significantly larger for NAE2.1. For instance, under current climate and demand, the water delivery reliability target of 90% is not met with NAE2.1, while it is still met for NAE2.2. Increasing temperatures and decreasing precipitation both lead to drop in performance. Decrease in performance is also larger for NAE2.1 than for NAE2.2.

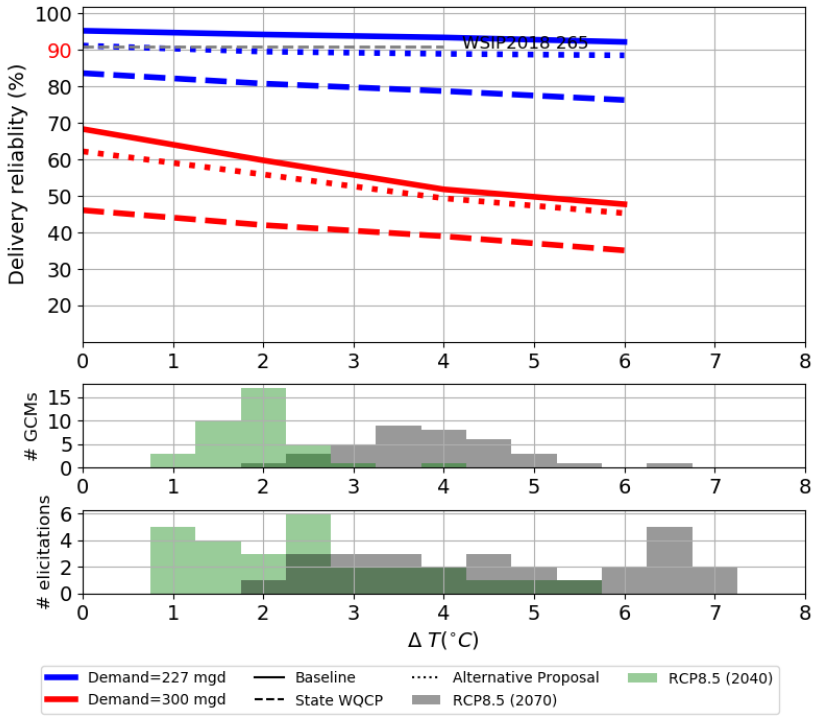


Figure 5-50. Water Delivery Reliability for the Reference, NAE2.1, and NAE2.2 Narratives across All Temperature (ΔT) Scenarios and a Subset of Two Demand Scenarios (227 mgd, Which Corresponds to Current SFPUC Demand, and 300 mgd [+30%]).
The dash-grey lines show reliabilities obtained from HHLSM WSIP 2018. The 90% reliability target is shown is red in the y-axis.

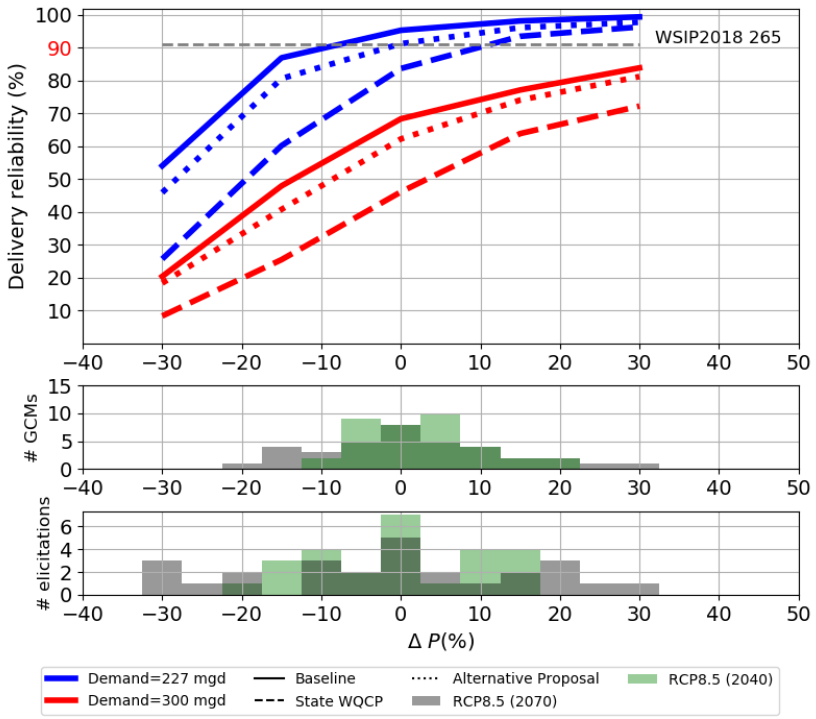


Figure 5-51. Water Delivery Reliability for the Reference, NAE2.1, and NAE2.2 Narratives across All Precipitation (ΔP) Scenarios and a Subset of Two Demand Scenarios (227 mgd, Which Corresponds to Current SFPUC Demand, and 300 mgd [+30%]).

The dash-grey lines show reliabilities obtained from HHLSM WSIP 2018. The 90% reliability target is shown is red in the y-axis.

In summary, state-amended WQCP and Alternative Proposal both have a significant effect on water delivery reliability.

- At a demand of 277 mgd, the water delivery reliability with the state-amended WQCP in current climate (reliability around 85%, rationing in 1 out of 6.5 years on average) is roughly the equivalent of a change in mean annual precipitation of about -15%. The water reliability with the Alternative Proposal in current climate meets the target.
- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in failing the reliability target with 227 mgd demand for both the state-amended WQCP and the Alternative Proposals (State WQCP: reliability around 80%, rationing in 1 out of 5 years on average; Alternative Proposal: reliability just below the target of 90%).
- By 2040, most projections and elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5%. At +4°C and -5% and 227 mgd demand, the reliability for the State WQCP and the Alternative Proposals fail the target (State WQCP: reliability around 75%, rationing in 1 out 4 years on average; Alternative Proposal: reliability around 85%, rationing in 1 out of 6.5 years on average).
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual precipitation results in failing the reliability target for 227 mgd demand for both the State WQCP and the Alternate Proposals (reliability with the Alternate Proposal is slightly below 90%). Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15%. At +6°C and -15% and 227 mgd, the reliability with the State WQCP drops slightly below 50% (rationing in 1 out of 2 years on average) while the reliability with the Alternative Proposal drops to about 70% (rationing in 1 out of 3 years).

5.3.3 New IFR below Turner Dam (II) (NAE4.2)

The outcome of this narrative is straightforward (Figure 5-52 through Figure 5-54): There is very little change in system performance as a result of a new IFR below Turner Dam. There is little effect on delivery reliability.

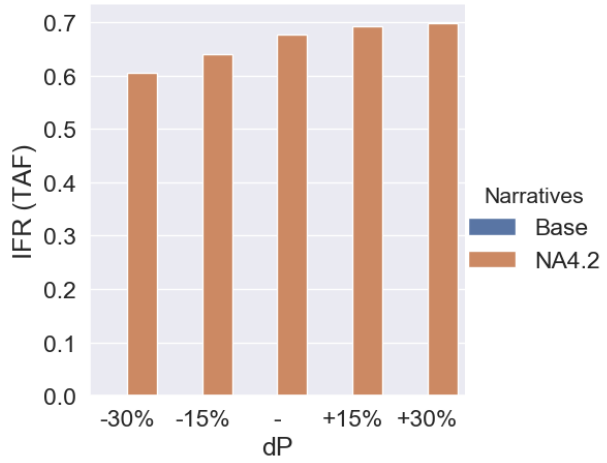


Figure 5-52. Annual IFR below Turner Dam (San Antonio Reservoir).

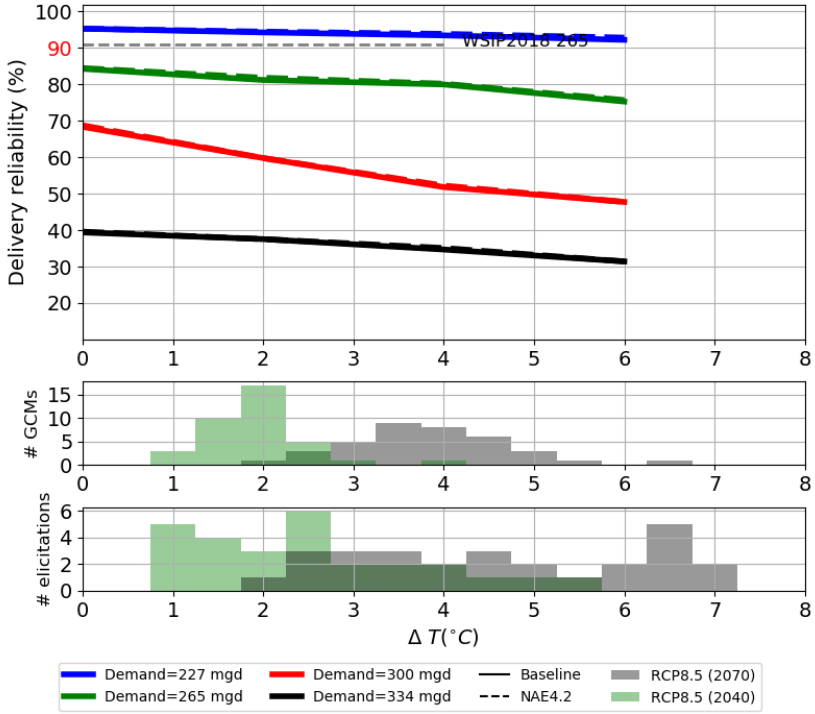


Figure 5-53. Water Delivery Reliability for the Reference and NAE4.2 Narratives across All Temperature (ΔT) Scenarios and a Subset of Four Demand Scenarios (227 mgd, Which Corresponds to Current SFPUC Demand, 265 mgd [+15% Increase], 300 mgd [+30%] and 334 mgd [+45%]).

The dash-grey lines show reliabilities obtained from HHLSM WSIP 2018.

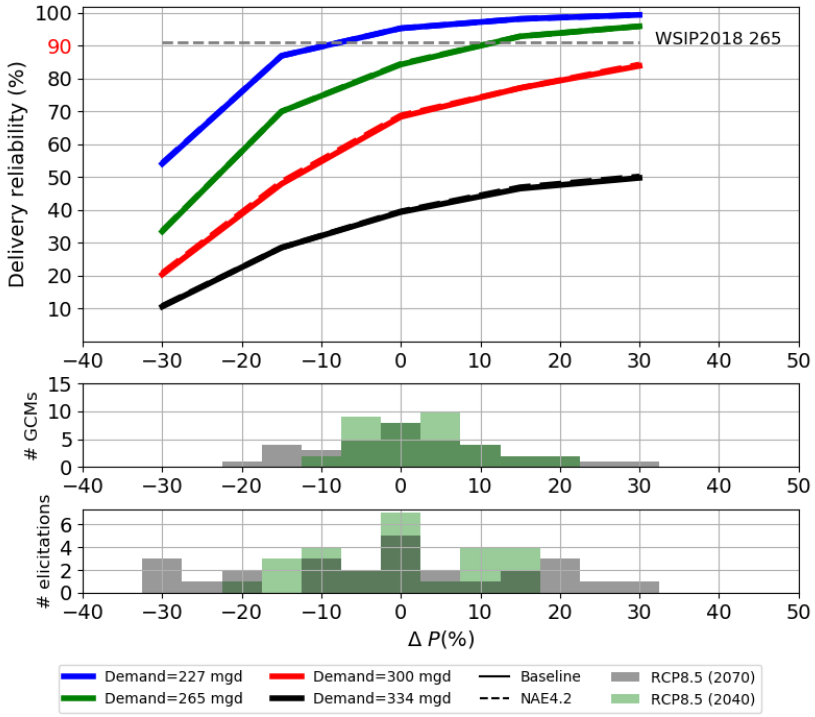


Figure 5-54. Water Delivery Reliability for the Reference and NAE4.2 Narratives across All Precipitation (ΔP) Scenarios and a Subset of Four Demand Scenarios (227 mgd, Which Corresponds to Current SFPUC Demand, 265 mgd [+15% Increase], 300 mgd (+30%) and 334 mgd [+45%]).
The dash-grey lines show reliabilities obtained from HHLSM WSIP 2018.

In summary, this new IFR has a negligible impact on water delivery reliability for all plausible changes in mean annual temperature and precipitation.

5.3.4 New IFR below Stone Dam (NAE5.1)

This IFR is relatively insensitive to changes in precipitation (Figure 5-55). Overall, it represents about a 30% increase over the existing voluntary release of 1.5 cfs. However, the IFR at Pilarcitos is the only IFR not satisfied, even for the current climate conditions.

Water delivery reliability is not sensitive to change in demand (all lines overlap in Figure 5-56 and Figure 5-57), which makes sense as Pilarcitos contribution to system deliveries is relatively little, even for current demand conditions.

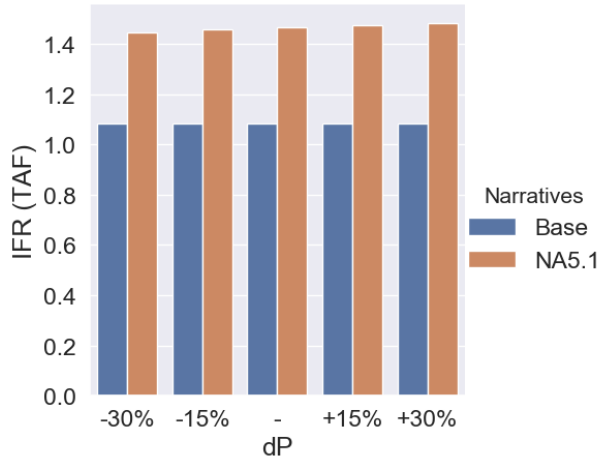


Figure 5-55. Annual IFR below Stone Dam.

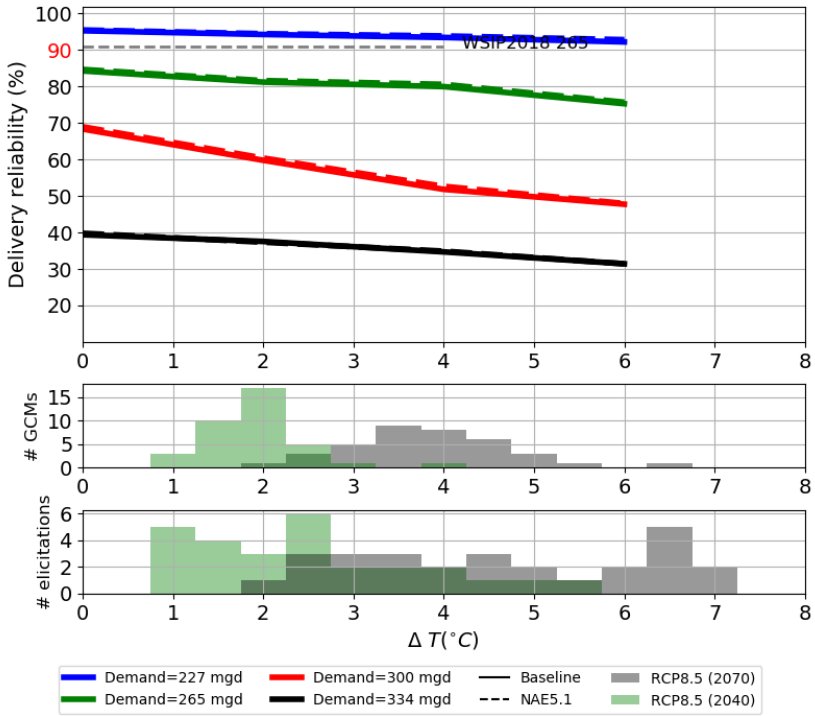


Figure 5-56. Water Delivery Reliability for the Reference and NAE5.1 Narratives across All Temperature (ΔT) Scenarios and a Subset of Four Demand Scenarios (227 mgd, Which Corresponds to Current SFPUC Demand, 265 mgd [+15% Increase], 300 mgd [+30%] and 334 mgd [+45%]).
The dash-grey lines show reliabilities obtained from HHLSM WSIP 2018.

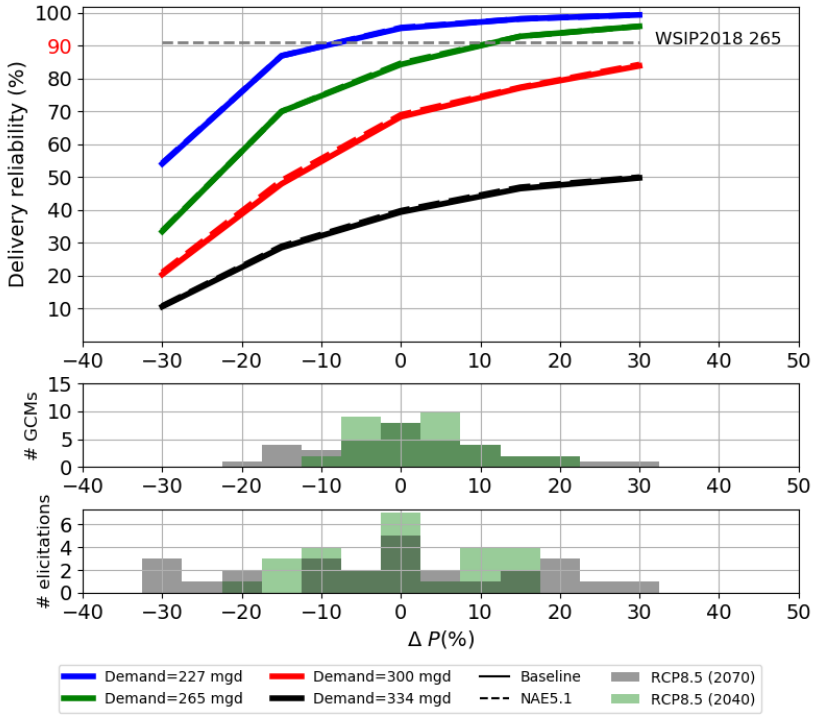


Figure 5-57. Water Delivery Reliability for the Reference and NAE5.1 Narratives across All Precipitation (ΔP) Scenarios and a Subset of Four Demand Scenarios (227 mgd, Which Corresponds to Current SFPUC Demand, 265 mgd [+15% Increase], 300 mgd [+30%] and 334 mgd [+45%]).
The dash-grey lines show reliabilities obtained from HHLSM WSIP 2018.

In summary, this new IFR has a negligible impact on water delivery reliability for all plausible changes in mean annual temperature and precipitation.

5.4 Effects of Climate Change on Upper Tuolumne River Ecosystem Program

The UTREP is designed to support specific ecological functions in the Tuolumne River below Hetch Hetchy Reservoir. This section is meant to illustrate the use of SFWSM outputs for assessing ecological performance metrics. Various ecological functions may benefit or suffer from different spill volumes (Table 5-5). The frequency of years for which simulated spill volumes from April through August reach the various thresholds under current climate conditions is shown in Table 5-5.

Table 5-5. Frequency of Seasonal Spill Volumes (from April through August) from Hetch Hetchy Reservoir under Current Climate Conditions for Identified Ecological Functions.
Spill includes releases from both the valves and drum gates.

April-August spill volume threshold	Ecosystem beneficiary	Frequency Spill above threshold
< 9,800 AF	Frogs	16.2%
> 9,800 AF	North Pond and Low Wetlands inundation	83.8%
> 133,000 AF	High Wetlands inundation	65.2%
> 178,000 AF	Sediment transport	55.2%

The effects of temperature and precipitation changes combined with demand change are presented in Figure 5-58 Figure 5-59, respectively. Figure 5-58 illustrates that, although temperature changes do affect the frequency of spill below Hetch Hetchy Reservoir, the expected change from current climate is small. Note that the frequency of years with large spill volume (i.e., larger than 133 TAF) is shown to slightly increase for the first few degrees of warming before decreasing if temperature keeps increasing. This behavior could be link to the reduction of the spring runoff with temperature warming across the Upcountry region (Figure 5-1). From Figure 5-58, the impact of demand change on the spill volume frequency is clear and significant. Larger water demand means that more water is transferred to the Bay area through the San Joaquin Pipelines, which results in lower reservoir levels at Hetch Hetchy (not shown) and reduces the spill. As a result, the frequency of low spill years that are beneficial for frogs is expected to increase if demand increases. On the other hand, the frequency of years will spill larger than 9.8 TAF is expected to decrease significantly.

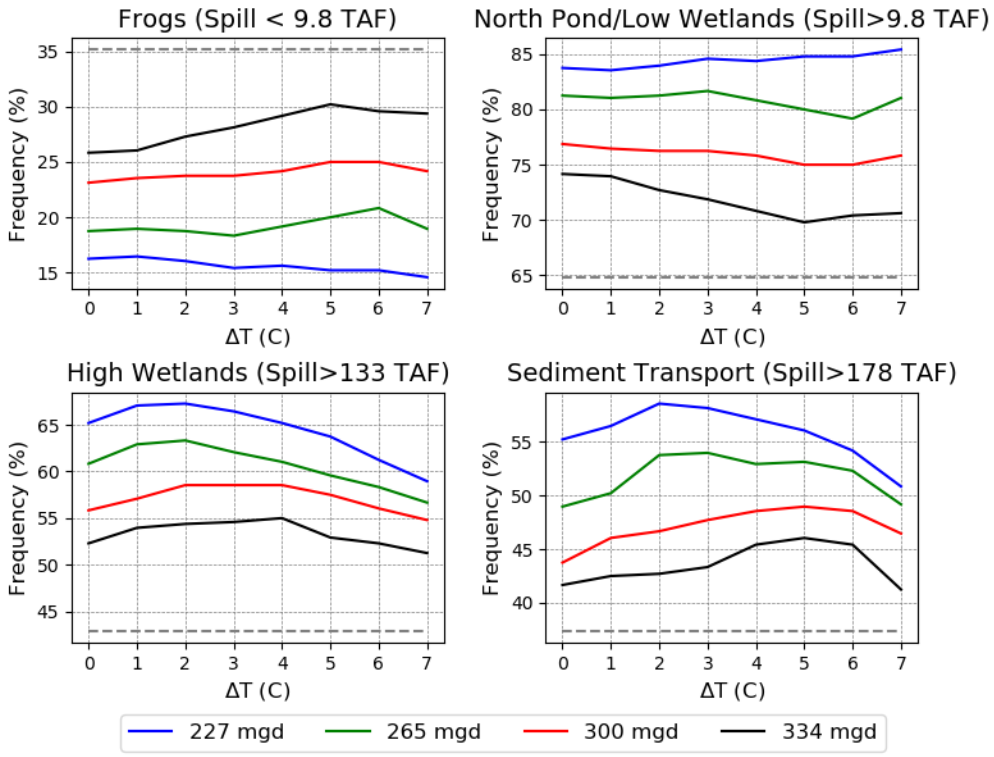


Figure 5-58. UTREP Spill Management Ecological Performance in Response to Temperature (X-axis) and Demand (Color Lines) Changes.

Spill volumes below O'Shaughnessy Dam includes releases via valves and drum gates from April through August. The grey dash lines show the frequency calculated from HHSLM WSIP 2018 at 265 mgd demand.

The frequency of years for which spill volume from April through August reaches various thresholds with changes in precipitation is illustrated in Figure 5-59. With respect to Figure 5-58, the results shown in Figure 5-59 highlight, without much surprise, that change in precipitation is a key driver of change in the frequency of spill for the various considered volume (larger precipitation reduces the frequency of low spill years; and inversely increases the frequency of high spill years). Change in demand does affect the frequency of spill as well (i.e., increases the frequency of low spill years and decreases the frequency of high spill years). Whether precipitation or demand changes will be the main driver of change in spill will depend on the pace of change of each driver.

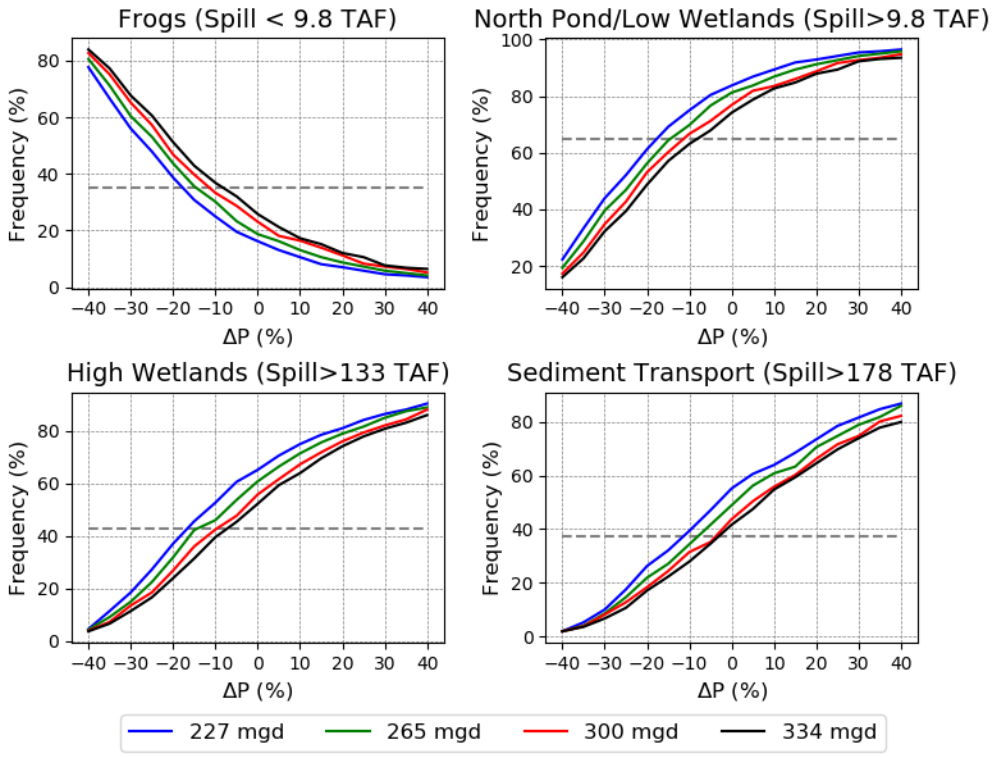


Figure 5-59. UTREP Spill Management Ecological Performance in Response to Precipitation (X-axis) and Demand (Color Lines) Changes.

Spill volumes below O'Shaughnessy Dam includes releases via valves and drum gates from April through August. The grey dash lines show the frequency calculated from HHSLM WSIP 2018.

Figure 5-60 and Figure 5-61 are respectively similar to Figure 5-58 and Figure 5-59 except that on the former the frequency of spill obtained under the narrative NAE1.1 (UTREP base flows) is shown with dashed lines. It is noted that, when comparing to the baseline IFR and under warming temperature conditions (Figure 5-60), the new IFR scheme from narrative NAE1.1 tends to decrease the frequency of low spill years (i.e., with spill volume lower than 9.8 TAF) and increase the frequency of moderated spill volume (i.e., in-between 9.8 and 133 TAF). The change in frequency of years with large spill volume (i.e., larger than 133 TAF) is less straightforward (ranging from a significant decrease under baseline temperature to a slight increase for significant warming). Figure 5-61 illustrates that the effect of the implementation of the UTREP base flows on the frequency of spill is minor when compared to the effect stemming from precipitation and demand change.

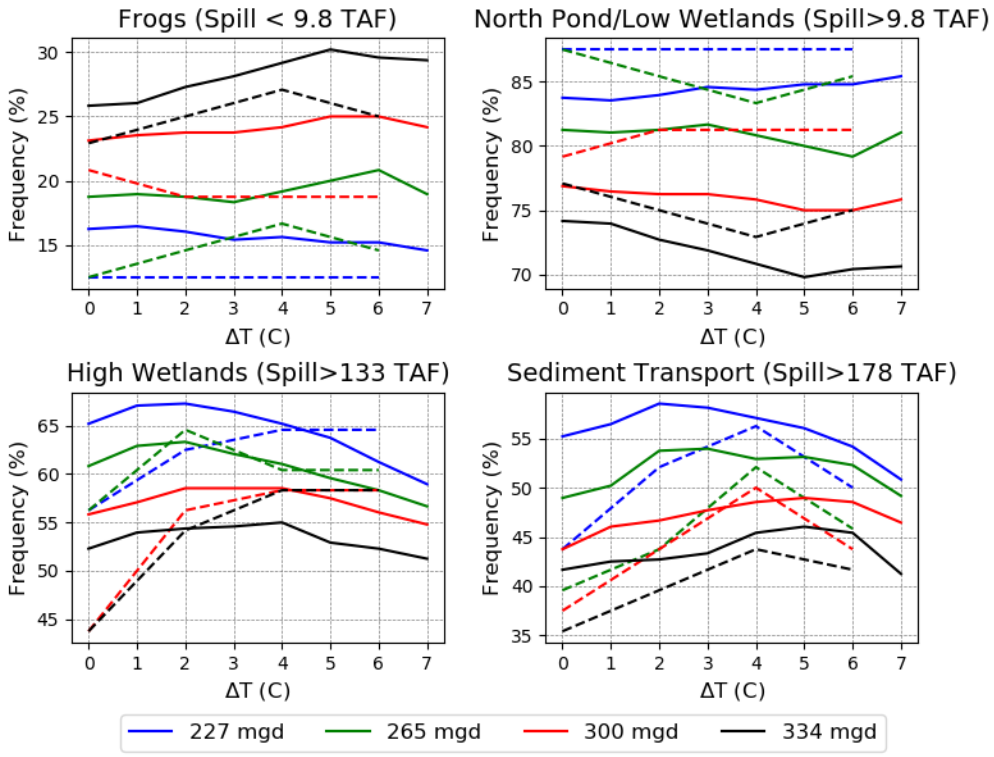


Figure 5-60. Same as Figure 5-58 Except That the Results Obtained with the Implementation of the UTREP Base Flows (Narrative NAE1.1; Section 5.3.1) Are Shown with Dash Lines.

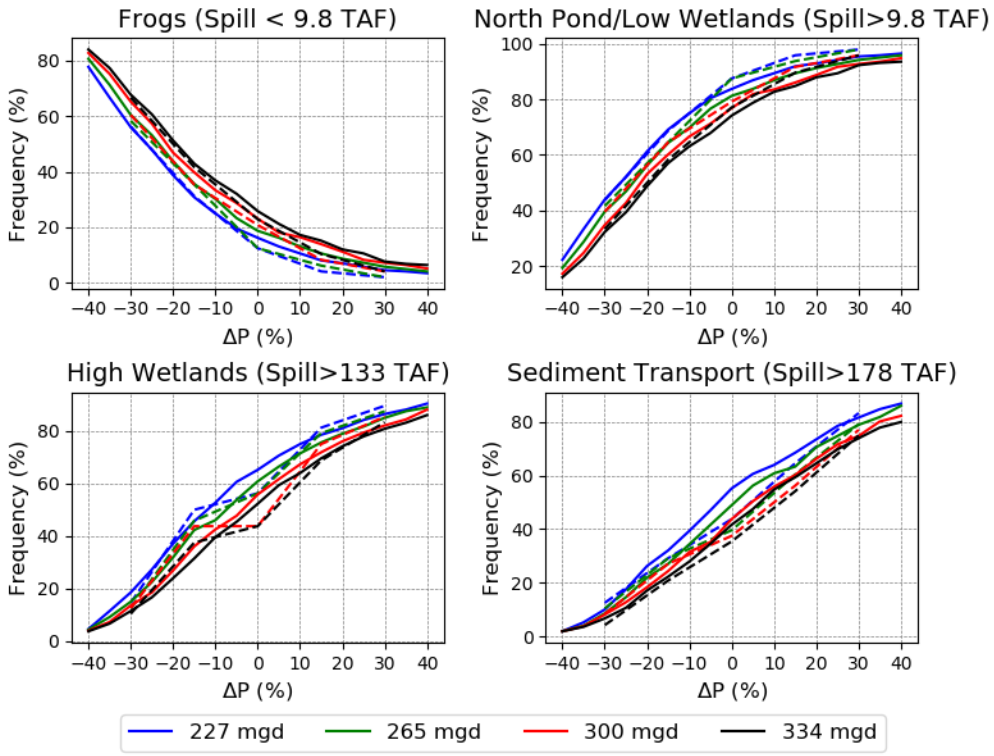


Figure 5-61. Same as Figure 5-59 Except That the Results Obtained with the Implementation of the UTREP Base Flows (Narrative NAE1.1; Section 5.3.1) Are Shown with Dash Lines.

5.5 Effects of Climate Change on Turbidity and TOC

The potential change in water quality indicators that would follow from climate and demand change are discussed in this section. The water quality model is data-driven and represents a non-linear, machine learning approach that simulates water quality parameters on a daily time step, thus adding capability in addition to the existing regression models currently used by SFPUC. Due to limited variability in the available data, the model has several limitations (Section 3.5.4). As such, it is worth restating some the limitations that may affect the results of the climate stress test:

- The quantity and variability of the available data for training was rather small, which may result in reduced model robustness when applied to out-of-sample dataset.
- The number of extreme events is very small in the training and validation datasets, which increases the uncertainty regarding the predicted turbidity and TOC values during extreme events.

Potential change in turbidity at the O'Shaughnessy Dam and Tesla Portal are presented in Section 5.5.1. Change in TOC are discussed in Section 5.5.2. Because high turbidity and TOC values could be potentially disruptive for the RWS, results are shown for the median prediction (i.e., the 50th percentile), the upper inter-quartile (75th) and one extreme percentile (97.5th). The reliability metrics obtained using the 50th, 75th and 97.5th percentiles are discussed below. Section 5.5.3 shows the results of the water quality narrative NAI1.1 (Table 4-6) that summarizes the assessment of the RWS vulnerability to water quality events.

5.5.1 Turbidity at the O'Shaughnessy Dam and the Tesla Portal

Turbidity is a measure of the light scattering in water from suspended solids and, as such, is an indicator of the concentration of suspended sediments. Higher turbidity reduces disinfection potential from ultraviolet and sodium hypochlorite. The filtration avoidance threshold is 5 NTU and shall not be exceeded more than twice in a 12-month period.

5.5.1.1 Turbidity at the O'Shaughnessy Dam

Figure 5-62 illustrates the effects of precipitation (left) and temperature (right) change on the average predicted 50th, 75th and 97.5th percentiles of the turbidity values at the O'Shaughnessy Dam. Precipitation change is the main driver of change for turbidity at the O'Shaughnessy Dam. The left panel of Figure 5-62 shows that higher precipitation leads to significant increases in turbidity at the O'Shaughnessy Dam. Such an increase in turbidity could be explained by an increase in streamflow leading to an increased erosion of the waterways and an increased volume of materials in suspension (Rossi et al. 2016). A decrease in precipitation would have minor effect on the average turbidity values. These results are consistent with the literature (Mukundan et al. 2013; Samal et al. 2013; Rossi et al. 2016) and with the existing SFPUC water quality model which uses inflow as the primary predictor. The right panel of Figure 5-62 demonstrates that warming temperature reduces turbidity, which could be linked to the reduction of the spring runoff volume and magnitude that should reduce the transport of materials from the watershed to the reservoir. Increasing urban water demand tends to lead to a decrease in turbidity at the dam, although not significantly. A possible explanation could be a lower average levels of the Hetch Hetchy Reservoir at the beginning of spring runoff. Lower average levels could result from higher flow in the San Joaquin Pipelines to deliver to a larger demand in the Bay area.

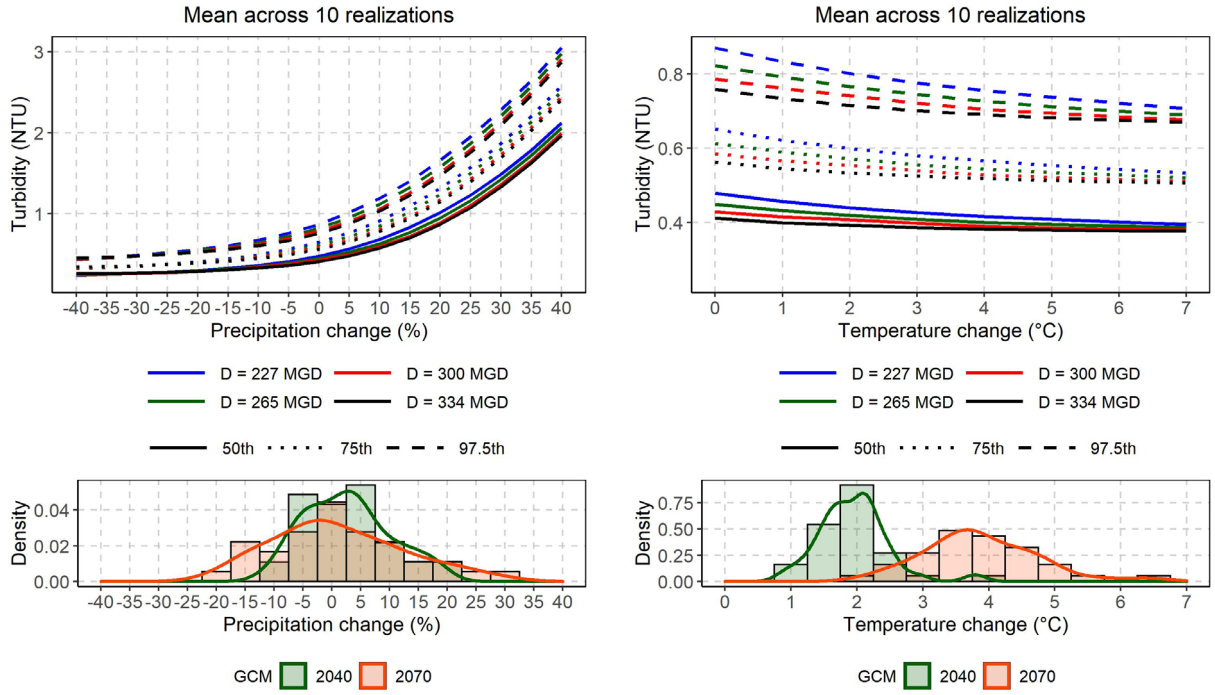


Figure 5-62. Effects of Precipitation (Left), Temperature (Right), and Demand (Colors) Changes on Daily Average Values of the Predicted 50th (Curve), 75th (Dotted Curve), and 97.5th (Dashed Curve) Percentiles of Predicted Turbidity Values at O'Shaughnessy Dam.

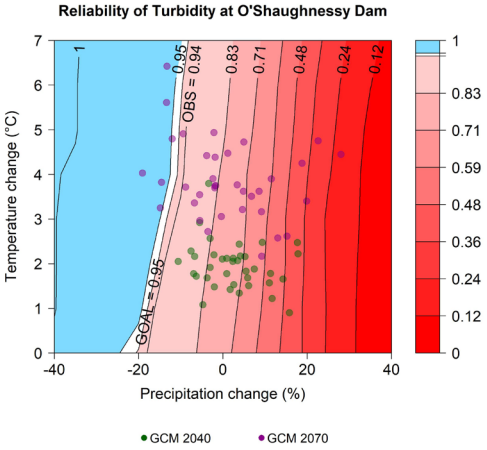
The bottom panel shows the expected change in precipitation and temperature obtained from the considered GCMs.

Figure 5-63 shows the change in reliability metric (i.e., the fraction of years that do not have daily turbidity over 5 NTU at least twice over a 12 month-window) obtained for a system demand of 227 mgd from the 50th (Figure 5-63a), 75th (Figure 5-63c) and 97.5th (Figure 5-63e) percentiles; and respectively for Figure 5-63b, Figure 5-63d, and Figure 5-63f when a demand of 334 mgd is considered. It is noted that the reliability obtained from each of the predicted percentiles does not align with the reliability estimated from the historical observations. The reasons for such discrepancies are several. The most important ones include i) the lack of robustness of the reliability value assessed from the historical observations because of limited observations, ii) the hydro-meteorological time series used as input to the water quality model shows a larger variety of hydrological conditions than those experienced in the historical period, iii) the baseline demand used for the climate stress test does not align with the historical demand during the period during which turbidity data is available at the O'Shaughnessy Dam. Despite these discrepancies, the future evolution of the reliability obtained from the three abovementioned percentiles of the predicted turbidity provides insights on the potential change in turbidity levels at the Dam. Thus, the results should be interpreted in terms of the relative change and direction of change.

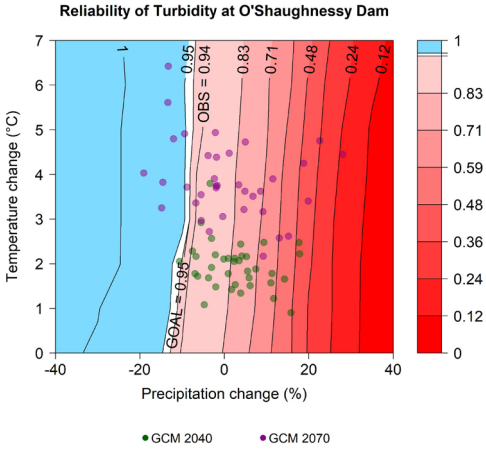
Under baseline climate and demand scenarios, the reliability obtained with the 50th percentile of the predicted turbidity (Figure 5-63a) is close to 75%, which means that for at least 25% of the 12 month-long rolling windows turbidity values reach 5 NTU at least twice. Precipitation change is the main driver of change in reliability. An increase by 20% precipitation would decrease the above reliability by roughly 40%. On the other hand, a decrease in precipitation by 20% would result in an increase in reliability by roughly 15%. Warming temperature increases the reliability. For example, an increase in temperature by +6°C could increase the reliability by 10% under baseline demand conditions. The effect of demand

change of reliability is less significant. For instance, an increase in demand by 45% (i.e., from 227 mgd to 334 mgd) under bassline climate conditions would only increase reliability by about 7% (Figure 5-63b compared to Figure 5-63a). These results are consistent with changes in turbidity discussed in Figure 5-62. The reliability obtained with higher percentiles (i.e., 75^{th} and 97.5^{th}) show lower reliability values.

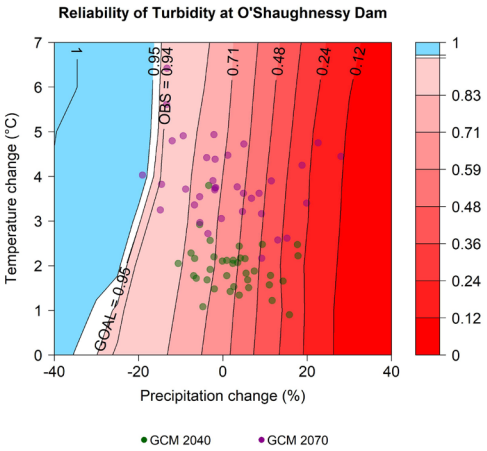
a) 50th precentile prediction (Demand = 227 mgd)



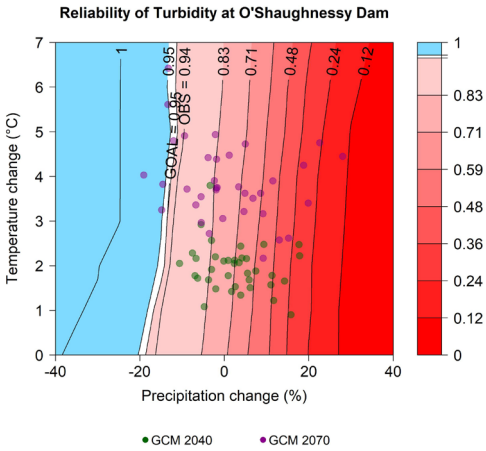
b) 50th precentile prediction (Demand = 334 mgd)



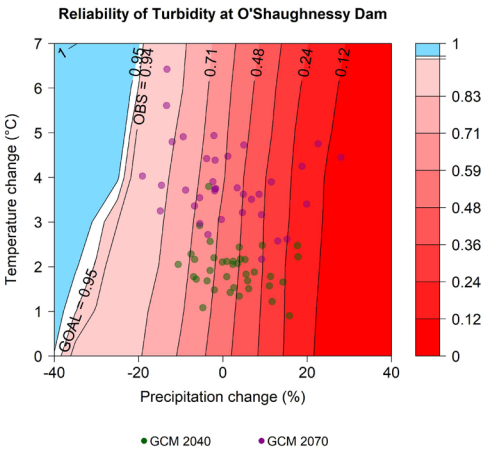
c) 75th precentile prediction (Demand = 227 mgd)



d) 75th precentile prediction (Demand = 334 mgd)



e) 97.5th precentile prediction (Demand = 227 mgd)



f) 97.5th precentile prediction (Demand = 334 mgd)

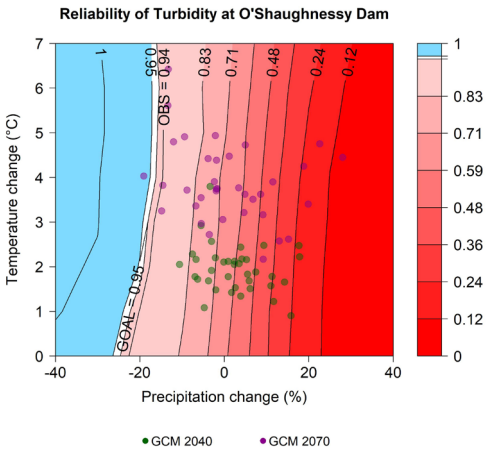


Figure 5-63. Climate Response Function (CRF) Showing the Reliability Metrics Obtained Using the 75th Percentile (Left) and the 97.5th Percentile (Right) of the Predicted Turbidity at the O'Shaughnessy Dam.

Reliability is here defined as the fraction of years for which daily turbidity maximum at O'Shaughnessy Dam does not exceed 5 NTU at least twice in a 12-month rolling window. For each CRF, x- and y-axes are showing the relative change in precipitation and absolute change in temperature, respectively. Blue, white, and red colors show a decrease, no change, and an increase in turbidity levels, respectively. Green and purple dots on the CRF shows the CMIP5 projections for the 2040 and 2070 future periods.

5.5.1.2 Turbidity at the Tesla Portal

Figure 5-64 illustrates the effects of precipitation (left), temperature (right) and demand (colors) change on the average predicted 50th, 75th and 97.5th percentiles of the turbidity levels at the Tesla Portal. The simulated response at Tesla Portal in terms of change in turbidity values to change in precipitation is similar to that obtained at the O'Shaughnessy Dam; increasing precipitation leads to higher average turbidity values at the Tesla Portal. However, the magnitude of the response is much less at Tesla Portal than it is at the O'Shaughnessy Dam, highlighting the buffer effect of the Hetch Hetchy Reservoir. Results also show that the effect of temperature change on the turbidity levels at the Tesla Portal is none. This is relatively different than the effect of temperature on the turbidity levels at the O'Shaughnessy Dam, where a minor to moderate decrease in turbidity is simulated under warming conditions. A possible explanation is the buffer effect from the reservoir, especially when change in turbidity at the Dam are small. Interestingly, the turbidity levels at Tesla Portal tend to increase when demand to the Bay area increases, which is an opposite response to what has been obtained at the Dam (Figure 3-43). This opposite response could be explained by the fact that with larger demand, the residence time in the reservoir decreases due to larger demand in the Bay, which could reduce the efficiency of the sedimentation process taking place in the reservoir.

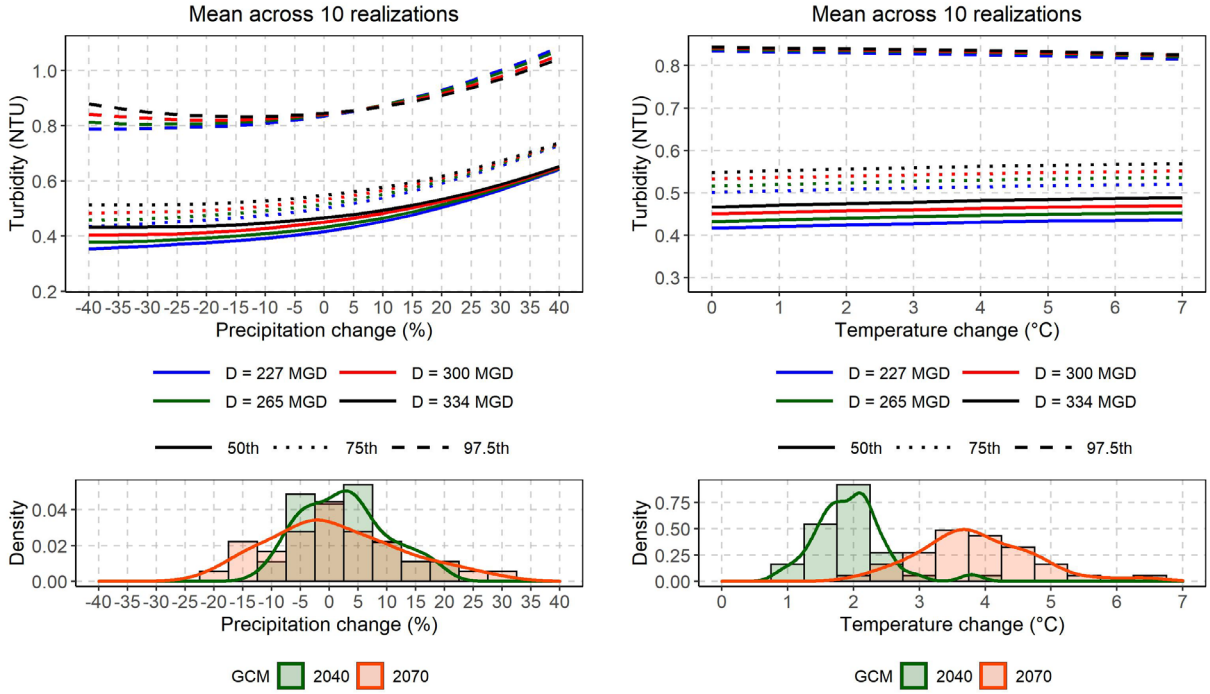
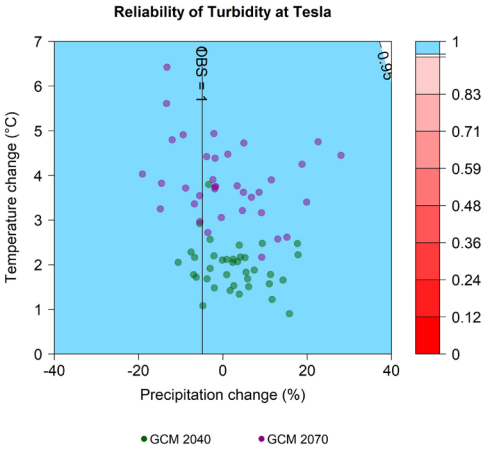


Figure 5-64. Effects of Precipitation (Left), Temperature (Right), and Demand (Colors) Changes on Daily Average (Top) and Daily Maximum (Middle) Values of the Median Prediction of Turbidity at Tesla Portal.
The bottom panel shows the expected change in precipitation and temperature obtained from the considered GCMs.

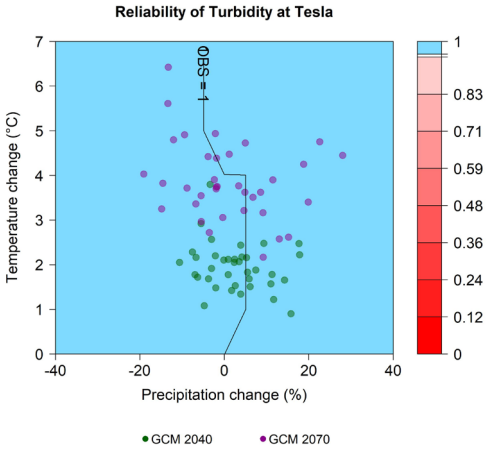
Figure 5-65 shows the reliability metrics estimated from 50th, 75th and 97.5th percentiles of the predicted turbidity. For all considered percentiles, the reliability is 100% under baseline conditions. The reliability obtained from the extreme 97.5th percentiles of the predicted turbidity is expected to decrease below the 95% target only for very large increase in precipitation. Given the range of the CMIP5 projections, only 3 GCM simulations lead to such a decrease in reliability by 2070 under baseline demand conditions, and only 2 under a 334 mgd demand scenarios (i.e., +45% demand from baseline). As such, the Hetch

Hetchy RWS does not appear vulnerable to turbidity events above 5 NTU occurring more than twice over a 12-month rolling-window.

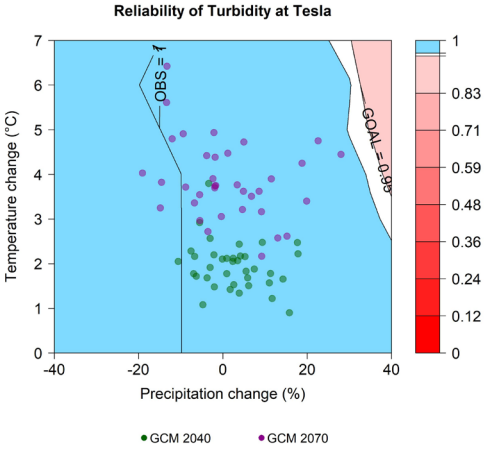
a) 50th precentile prediction (Demand = 227 mgd)



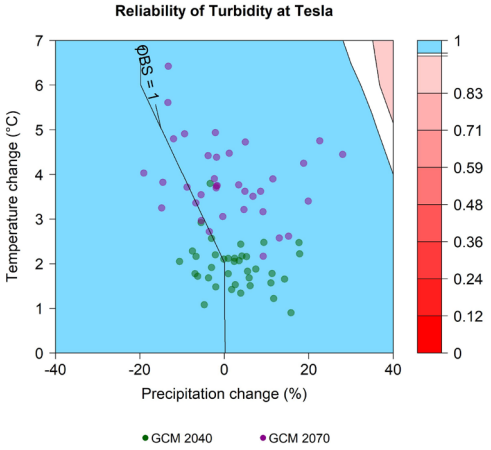
b) 50th precentile prediction (Demand = 334 mgd)



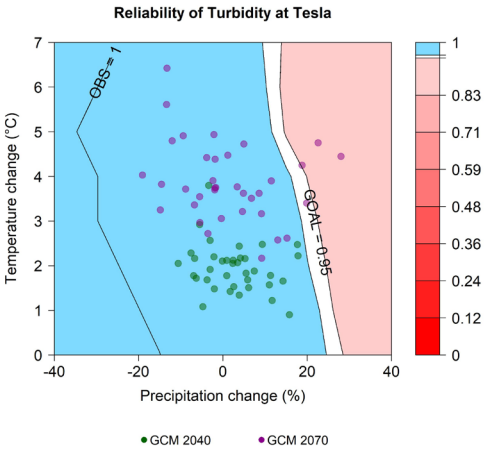
c) 75th precentile prediction (Demand = 227 mgd)



d) 75th precentile prediction (Demand = 334 mgd)



e) 97.5th precentile prediction (Demand = 227 mgd)



f) 97.5th precentile prediction (Demand = 334 mgd)

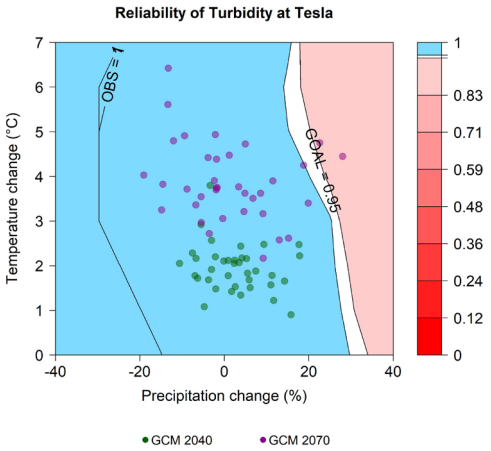


Figure 5-65. Effects of Precipitation (Left), Temperature (Right), and Demand (Colors) Changes in Reliability in Regards with Turbidity at Tesla Portal.

Reliability is here defined as the fraction of years for which daily turbidity maximum at Tesla Portal does not exceed 5 NTU at least twice in a 12-month rolling window.

5.5.2 TOC at the Tesla Portal

TOC is the amount of carbon in organic compounds that is present in the water. It is a surrogate for organic contaminants (natural organic substances) or precursors of disinfection byproducts. For TOC at the Tesla Portal, a concentration of 2 mg/L is defined as a 'level of concern' while 3 mg/L requires treating the Hetch Hetchy water at SVWTP.

Figure 5-66 illustrates the effects of precipitation (left), temperature (right) and demand (colors) changes on the average predicted 50th, 75th and 97.5th percentiles of the predicted TOC at the Tesla Portal. Precipitation change is the main driver of change for TOC at the Tesla Portal. Not like turbidity, the response of TOC to changes in precipitation is not monotonic as TOC tends to decrease when both precipitation increases or decreases. For the former, increased dilution resulting from more water could decrease TOC concentration; while for the latter, reduction in the intensity of rainfall could reduce the transport of organic material from the basin to the lake. The response in terms of maximum TOC value is different, with a significant decrease when precipitation decreases, and inversely, an increase when precipitation increases, although the maximum simulated TOC levels at Tesla Portal stay below 3 mg/L whatever the climate and demand conditions (not shown). Warming temperatures do not alter TOC at Tesla Portal. Increasing urban water demand tends to slightly reduced TOC at Tesla Portal.

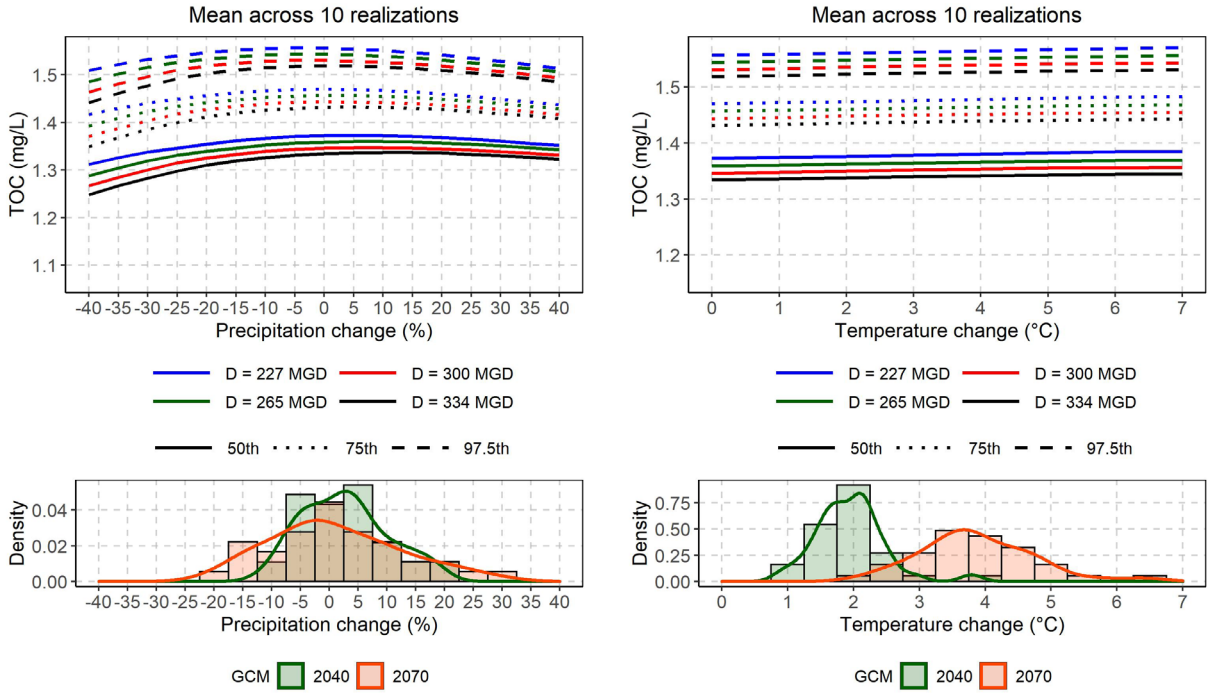
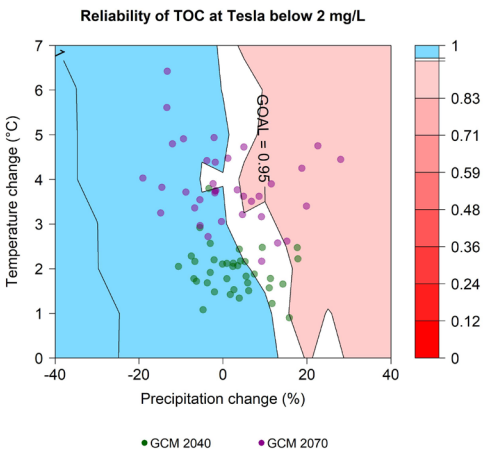


Figure 5-66. Effects of Precipitation (Left), Temperature (Right), and Demand (Colors) Changes on Daily Average (Top) and Daily Maximum (Middle) Values of the Median Prediction of TOC at Tesla Portal.
The bottom panel shows the expected change in precipitation and temperature obtained from the considered GCMs.

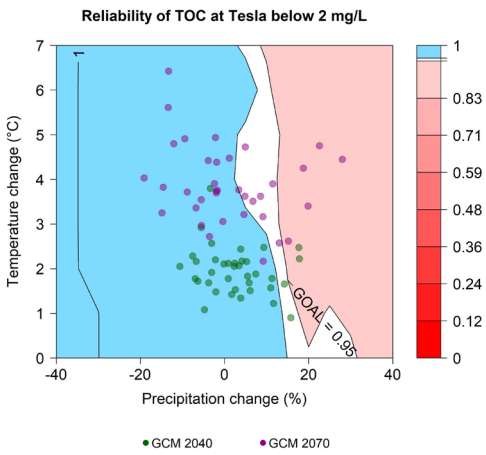
The reliability metric for 3 mg/L level is not shown because it is always above 95% whatever the climate and demand conditions. While this could reflect the correct response of the RWS in terms of TOC levels at Tesla Portal, it is important to bear in mind that because no TOC values larger than 3 mg/L were included in the model training dataset, it can simply be really difficult for a data-driven model to predict TOC values above this threshold.

Figure 5-67 shows the annual reliability of TOC above the 2 mg/L threshold as an indication of the potential changes in the frequency of events leading to high TOC values. Under current climate conditions, the reliability estimated from the 50th percentile of the predicted TOC at Tesla Portal is 1; or 100% (i.e., no simulated TOC values are above 2 mg/L under baseline climate conditions). The reliability is found to fall below the 95% target when there is a significant increase in precipitation (i.e., larger than 15% increase). Under the baseline climate, reliability metrics obtained for the 75th and 97.5th percentiles are below the 95% target, highlighting that TOC is likely to get above 2 mg/L during high and extreme events. It is noted that the reliability metric obtained for the 97.5th percentile of the TOC values at Tesla Portal is close to the observed reliability (i.e., 75%). Increasing precipitation leads to a very modest decrease in reliability. Overall, there is an indication of falling reliability relative to TOC targets when mean precipitation and temperatures increase, with the larger effect due to precipitation.

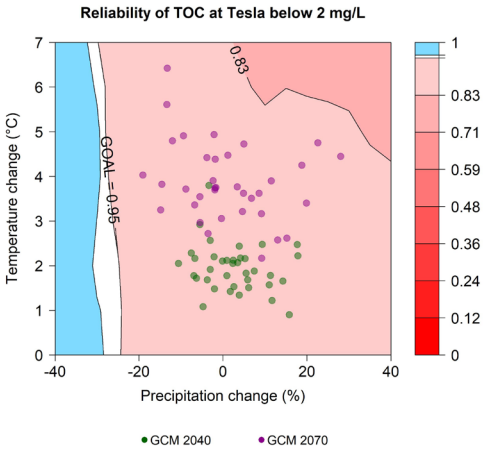
a) 50th precentile prediction (Demand = 227 mgd)



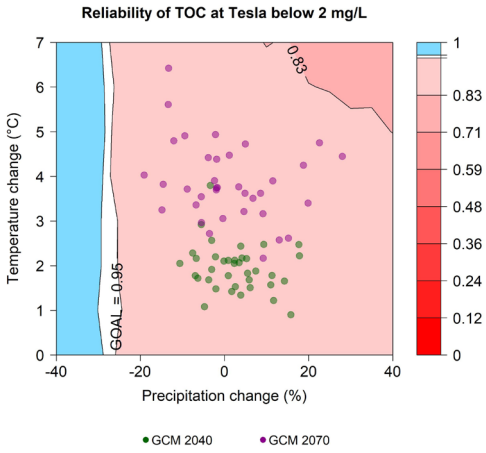
b) 50th precentile prediction (Demand = 334 mgd)



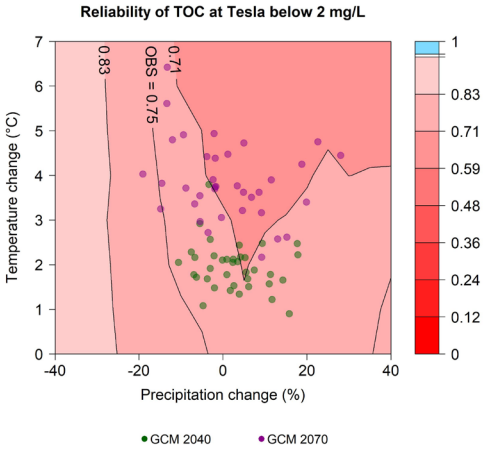
c) 75th precentile prediction (Demand = 227 mgd)



d) 75th precentile prediction (Demand = 334 mgd)



e) 97.5th precentile prediction (Demand = 227 mgd)



f) 97.5th precentile prediction (Demand = 334 mgd)

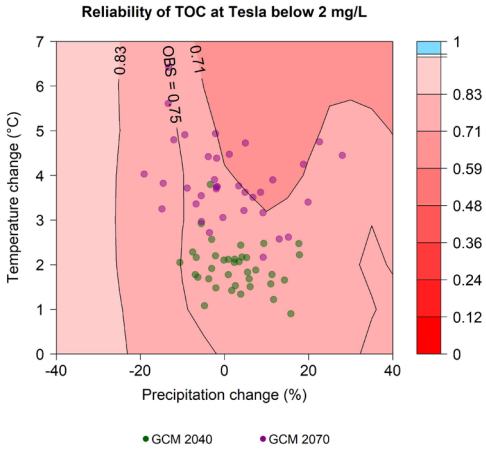


Figure 5-67. Effects of Precipitation (Left), Temperature (Right), and Demand (Colors) Changes in Reliability in Regards with TOC Concentration at Tesla Portal.

Reliability represents the fraction of years for which TOC does not exceed 3 mg/L. Reliability is calculated using the median prediction of TOC at Tesla Portal.

5.5.3 Vulnerability of Water Delivery to High TOC/Turbidity Levels of Hetch Hetchy Water

When either turbidity or TOC levels cause an exceedance of the filtration avoidance criteria limits or DBP targets, respectively, water from Hetch Hetchy may have to be diverted to the SVWTP at a rate of 90 MGD (capacity of the San Antonio pump station). This section illustrates the potential vulnerability of the water delivery during such an event through simulations of SFWSM using the water quality narrative NAI1.1 described in Section 4.3.2.1. The system response to both TOC and turbidity extremes are investigated. High turbidity events are more likely to lead to an acute health risk (if water is not diverted) and could lead to a filtration avoidance violation and potentially the loss of the filtration avoidance waiver, which could have a significant financial impact on SFPUC.

Two water quality event have been selected using the output the water quality prediction model for TOC at Tesla Portal. The two largest simulated TOC peaks using current climate conditions have been selected (Table 5-6). The date at which the system starts treating Hetch Hetchy water has been defined as the time when TOC value exceeds 2 mg/L and eventually reaches the peak value. Two water quality events have been selected using the predicted turbidity value at Tesla Portal. The first event presented in Table 5-7 is from the realization 10 (i.e., the historical realization) and corresponds to the 1997 turbidity event. For both turbidity and TOC events, one winter event and one late springs/summer event were selected, highlighting that high turbidity and TOC events can happen in both season. It is noted that for events occurring during late spring/summer season, the maximum yield of the system provided by SVWTP and HTWP (i.e., 921 AF/day) is not large enough to cover the daily peak demand during the period of the outage, which means that without demand rationing and/or importation of water to the system, deficit is unavoidable.

Table 5-6. Characteristics of the Start Dates of the Water Quality Event Narratives Associated with High TOC Values at Tesla.

The maximum system yield is the aggregated capacity treatment capacity from the Sunol Valley (160 mgd) and Harry Tracey (140 mgd) water treatment plants. Demand values are given for the baseline demand and climate conditions. The results of the 50th predicted quantile was used to select the date.

Start date of the failure	TOC values at the start date (mg/L)	TOC value at the peak of the event (mg/L)	Date of the peak TOC	Climate Realization	Maximum system yield during outage (AF/day)	Max daily demand during the outage period (AF/day)	Total demand during outage period (TAF)
2/2/2043	2.89	2.89	2/2/2043	6	921	721	31
6/1/2065	2.05	2.09	6/2/2065	8		1076	44

Table 5-7. Characteristics of the Start Dates of the Water Quality Event Narratives Associated with High Turbidity Values at Tesla.

The maximum system yield is the aggregated capacity treatment capacity from the Sunol Valley (160 mgd) and Harry Tracey (140 mgd) water treatment plants. Demand values are given for the baseline demand and climate conditions. The results of the 97.5th predicted quantile was used to select the event. The event starting on 1/4/2056 is from the historical realization (R10) and corresponds to the 1997 turbidity event discussed in Sections 3.5. Note that the turbidity peak at Tesla Portal occurred on February 3rd, 1997; while it does occur on February 4th in the narrative. This marginal difference follows from the sequence of leap year in the climate realization that differs from the sequence of leap year in the historical record.

Start date of the failure	Turbidity values at the start date (mg/L)	Turbidity value at the peak of the event (mg/L)	Date of the turbidity peak	Climate Realization	Maximum system yield during outage (AF/day)	Max daily demand during the outage period (AF/day)	Total demand during outage period (TAF)
1/4/2056	10.8	10.8	1/4/2056	10	921	680	33
6/5/2058	5.1	6.5	6/9/2058	5		1000	45

Figure 5-68 illustrates the response of the RWS during a water quality outage following a peak in TOC under baseline climate and demand conditions (second row Table 5-6). During the outage, the flow through the SJPL is reduced to 90 mgd (i.e., 276 AF/day). SVWTP outflow is at maximum capacity during the duration of the outage; so is HTWP, although not shown in Figure 5-68. San Antonio Reservoir cannot send water to SVWTP during the duration of the event because the San Antonio pump station is used to transfer water from Hetch Hetchy (via the Alameda Siphons) to SVWTP. As a result, San Antonio level is higher than it would have been without the outage. Conversely, Calaveras contributes more to SVWTP than it would have without the outage, which leads to a more important drawdown of the reservoir through the event. The deficit during the event follows from the mismatch between the maximum system yield and the peak demand during the outage period, as indicated in Table 5-6.

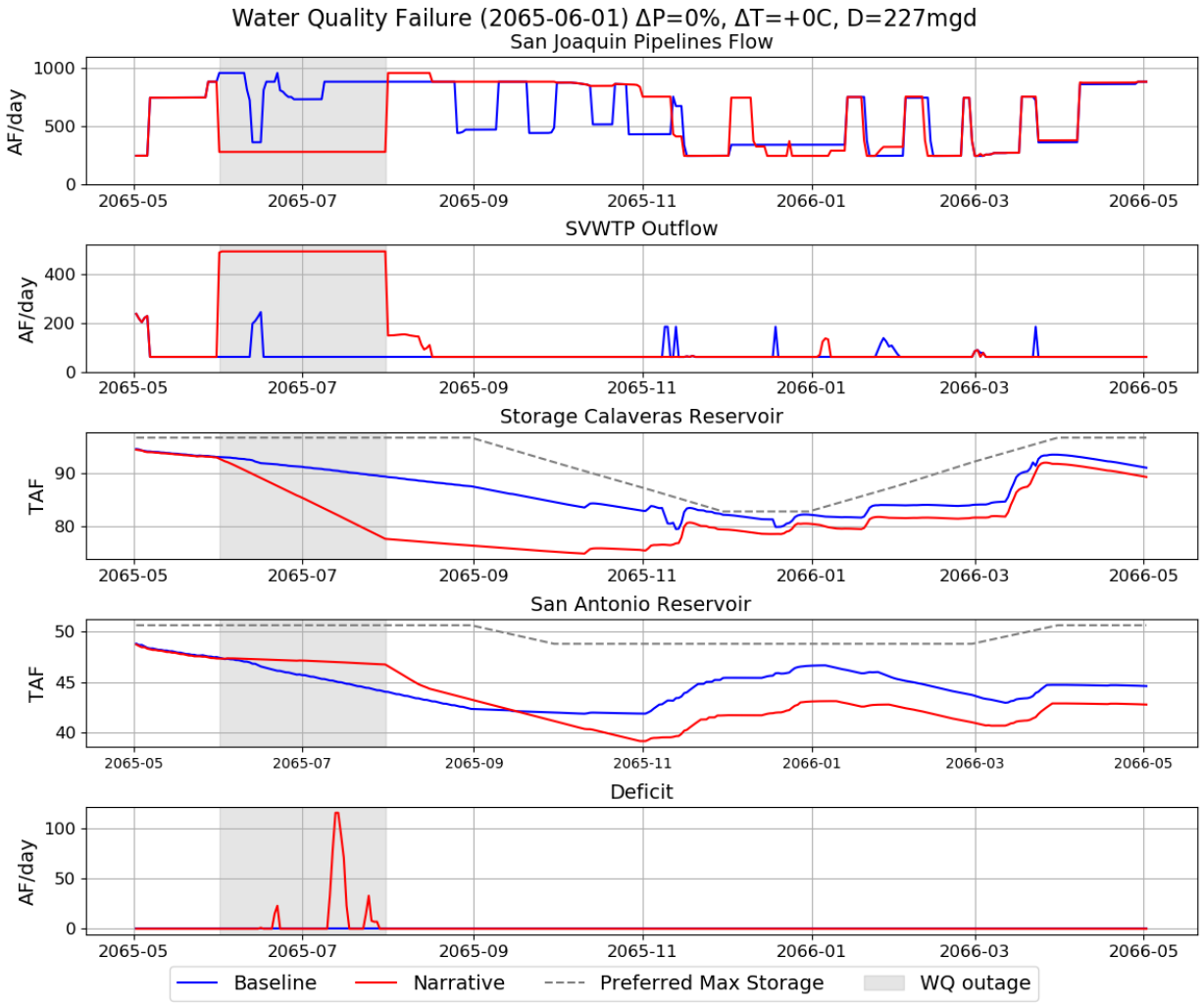


Figure 5-68. System Response during a Water Quality Outage.
From top to bottom, figure panels focus on the operations of San Joaquin Pipelines, the outflow from the Sunol Valley Water Treatment Plant (SVWTP) and Calaveras and San Antonio Reservoirs. The bottom panel illustrates the deficit. For each panel, the colors separate the baseline simulation (i.e., no outage; blue) from the simulation for which the outage was implemented (red). The grey dash lines show the preferred maximum storage for Calaveras and San Antonio Reservoirs. The date of the outage, the precipitation (ΔP), temperature (ΔT) and demand (D) scenarios are given in the figure title. The grey shaded area shows the duration of the outage.

Figure 5-69 illustrates the response of the RWS during a water quality outage following a peak in turbidity under baseline climate and demand conditions. Turbidity events are 'wet' events which means that they occur during high flow events. In the example illustrated in Figure 5-69, Calaveras Reservoir is nearly full at the beginning of the event, while San Antonio Reservoir is at its preferred maximum storage. In this situation, diverting water from the Alameda Siphons from Hetch Hetchy to SVWTP would prevent release water from San Antonio Reservoir to SVWTP, because both operations require the use of the San Antonio pump station. In other words, diverting Hetch Hetchy water from the Alameda Siphons to SVWTP would likely cause San Antonio to spill, which is something SFPUC would avoid, as it is preferred to spill at Hetch Hetchy if needed. The reason is that spill from Hetch Hetchy can be store in Water Bank. Regarding the example Figure 5-69, the flows through the SJPLs are almost null for more than half of the outage period, so that San Antonio can release water to SVWTP and avoid spills. This

event occurs during the winter season; peak demand is significantly lower than the maximum system yield during the duration of the outage. As a result, no deficit is recorded throughout the outage.

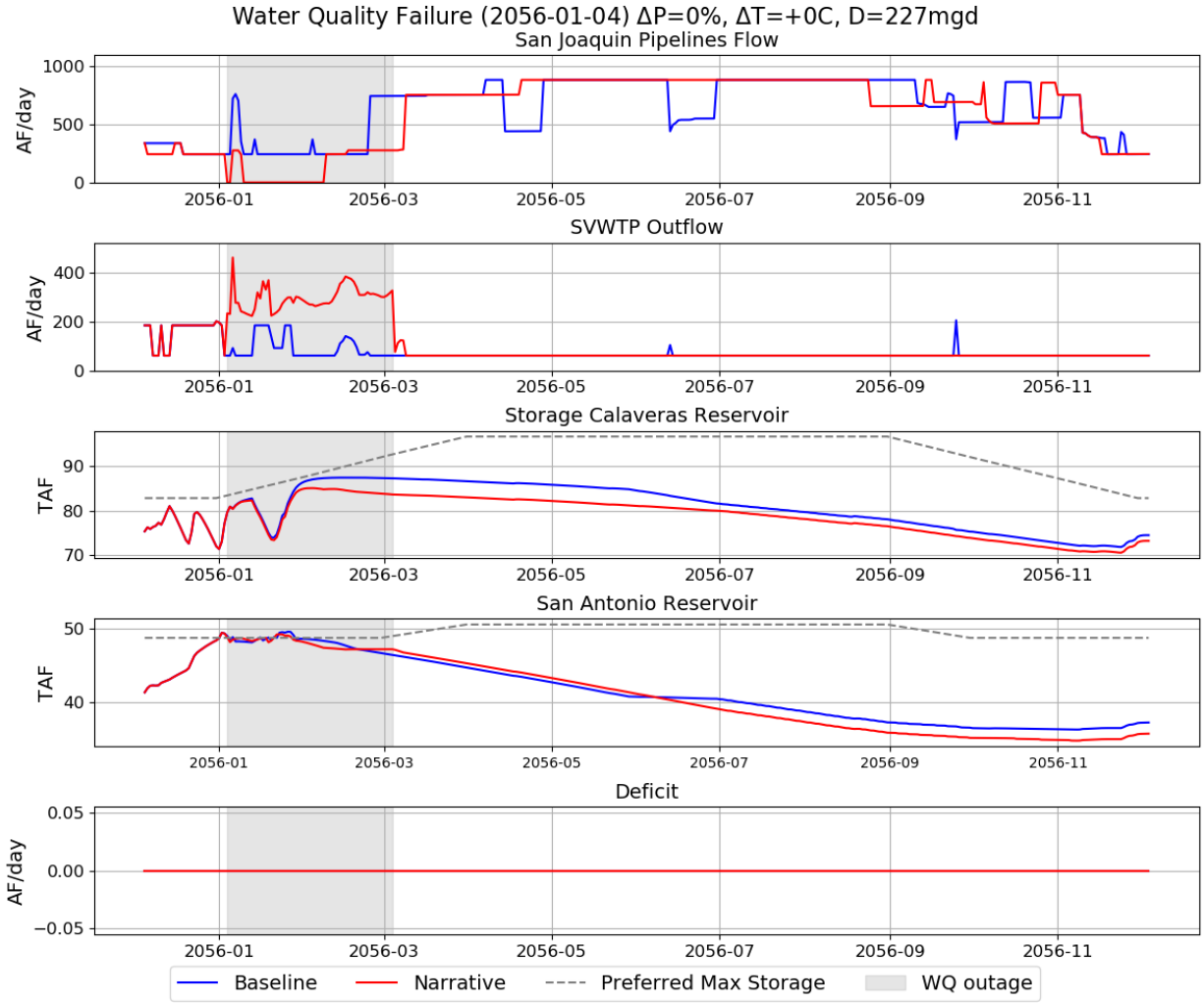


Figure 5-69. System Response during a Water Quality Outage.
From top to bottom, figure panels focus on the operations of San Joaquin Pipelines, the outflow from the Sunol Valley Water Treatment Plant (SVWTP) and Calaveras and San Antonio Reservoirs. The bottom panel illustrates the deficit. For each panel, the colors separate the baseline simulation (i.e., no outage; blue) from the simulation for which the outage was implemented (red). The grey dash lines show the preferred maximum storage for Calaveras and San Antonio Reservoirs. The date of the outage, the precipitation (ΔP), temperature (ΔT) and demand (D) scenarios are given in the figure title. The grey shaded area shows the duration of the outage.

Figure 5-70 illustrates for the four considered water quality outages the influence of precipitation and demand changes on the simulated deficit throughout the event. The left column shows the simulated deficit for the baseline (i.e., no outage), while the simulated deficits during the period of the outage are shown in the right column. Note that the deficit shown in the left column results only from either drought rationing or additional virtual water that SFWSM requires to either satisfy the deliveries and/or the LP constraints. The simulated deficits during the period of the outage are shown in the right column.

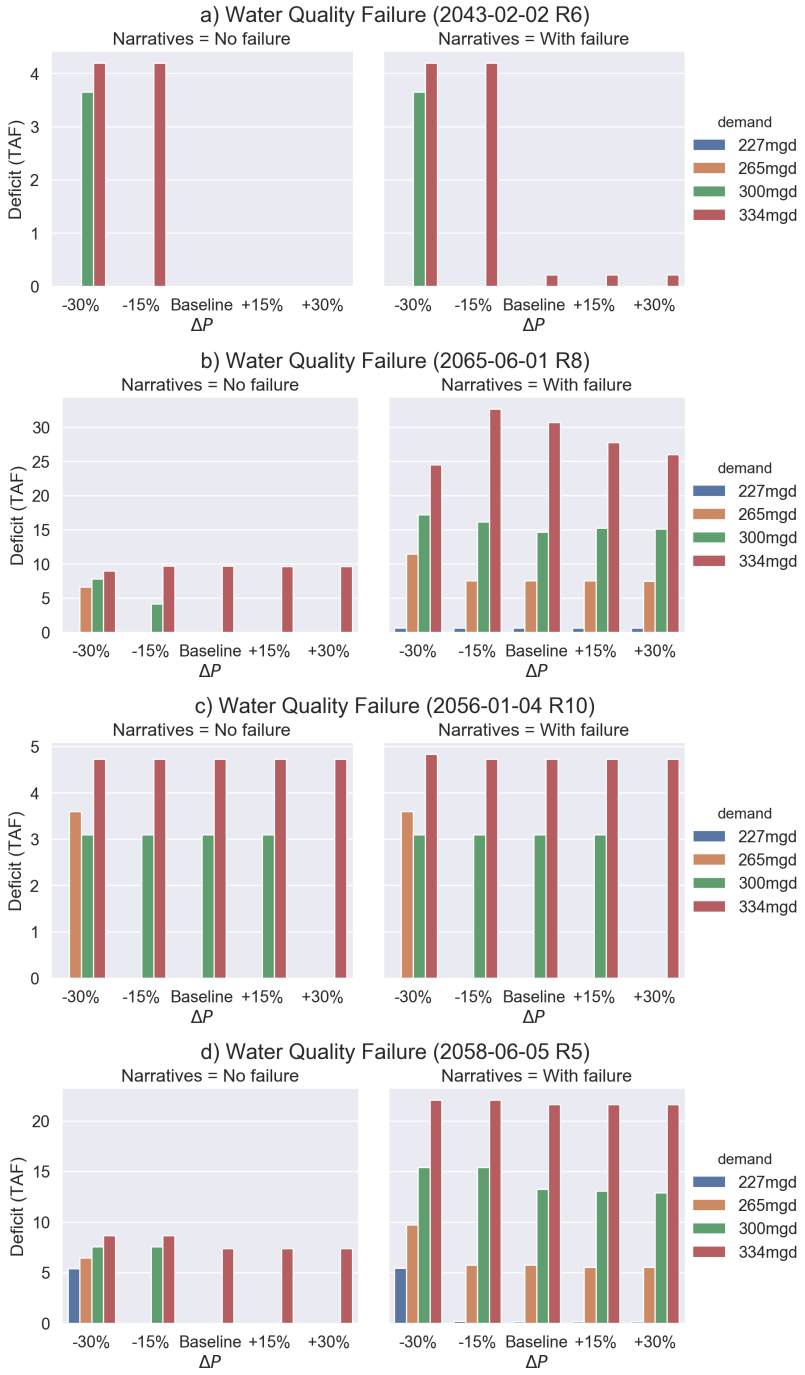


Figure 5-70. Deficit Resulting from a Water Quality Outage in Combination with Precipitation and Demand Changes.

Left panels show the simulated deficit without outage. Right panels show the simulated deficit with the outage. Daily deficits are summed across the duration of the outage (60 days from the beginning of the outage).

From Figure 5-70, several observations can be made:

- **Influence of the onset date of the events**: the timing of the water quality outage is an important driver of the vulnerability of the RWS to a water quality outage under baseline climate and demand. When occurring during the late spring/summer season, the available system yield is lower than peak daily demand. Given the results of the water quality model, late spring/summer outage could follow

from either TOC or turbidity events. The total deficit during the events are usually low.

- **Influence if precipitation change**: Precipitation change does not seem to affect significantly the capacity of the RWS to deliver demand during a water quality outage. The reason could be that water quality outage occur during or after big storms. As such, water shortage is not an issue.
- **Influence of demand change**: An increase in demand is likely to increase almost linearly the recorded deficit during a water quality outage. A significant increase in demand (i.e., 334 mgd) could even lead the system to not deliver full demand during the winter outages (Figure 5-70a).

In light of the results discussed above, the RWS is vulnerable to a water quality outage if it occurs during the high demand season. The considered water quality outages all occurred during wet events, which helps the system to keep delivering full demand even for scenario with decrease precipitation. Increasing demand is likely to lead to larger deficit, which would mostly be explained by an increased discrepancy between the peak demand and the maximum system yield that SVWTP and HTWP provide.

In summary, turbidity and TOC reliabilities at the Tesla Portal are more sensitive to the precipitation change than alterations in temperature and demand. Likely due to the dampening effect of the operation in the reservoir and pipeline operations, turbidity levels at the Tesla Portal are less sensitive to climate change than the ones at the O'Shaughnessy Dam.

5.5.3.1 Turbidity at Tesla Portal

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in meeting the turbidity reliability target at Tesla Portal.
- By 2040, most projections and elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5%. Within this uncertainty range, the turbidity reliability meets the 95% target.
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual precipitation results in meeting the 95% turbidity reliability target. Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15%. Under this range of uncertainty, the turbidity reliability meets the 95% target.

5.5.3.2 TOC at Tesla Portal

- By 2040, the median projections of +2°C warming combined with 0% change in mean annual precipitation results in meeting the TOC reliability target at Tesla Portal.
- By 2040, most projections and elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5%. Within this uncertainty range, the TOC reliability estimated from the 50th percentile of the predicted TOC meets the 95% target. TOC reliability estimated from the 75th and 97.5th percentiles do not meet the 95% target and highlight that during high and extreme events, TOC can get above 2 mg/L twice within a twelve month window.
- By 2070 RCP8.5, the median projections of about +4°C combined with 0% change in mean annual precipitation results in meeting the 95% TOC reliability target. Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15%. At +4°C and +15%, the TOC reliability obtained from the 50th percentile falls slightly below the 95%. TOC reliability estimated from the 75th and 97.5th percentiles are below the 95% target, highlighting the impact of high and extreme events on TOC values.

5.6 Compound Effects of Climate Change with Infrastructure Outages

The infrastructure narrative-based scenarios below represent the outage of critical facilities. Although it is reasonable to think that an emergency rationing could take place at the onset of some outage events,

no rationing other than the ones resulting for the drought rationing is applied (no modification in the drought rationing was made). Rationing the deliveries is one possible response to minimize water supply deficit that occurs throughout the duration of an outage, but others exist such as importing water from neighbor systems. Finding the optimal operational response to these outages requires a more comprehensive analysis that is beyond the scope of this vulnerability assessment.

For the Moccasin Switchyard outage (NAI3), the Earthquake along the Calaveras Fault (NAI4) and the wildfire across the Crystal Springs Reservoir watershed (NAI5), the deficit is presented throughout the duration of the event. The deficit is calculated as the difference between the demand prior drought rationing and the deliveries. Note that any virtual water is subtracted first from the deliveries prior the calculation of the deficit. For each of these narratives, a first figure is used to illustrate the RWS response throughout the outage period and bar plots for each event are used to show the change in deficit during the event with precipitation and demand changes. Note that compared to the effect of precipitation and demand changes, the incremental change in deficit stemming from warmer temperatures during outages (i.e., NAI3, NAI4 and NAI5) is not significant. As such, the change in deficit under warming conditions is not discussed further for the above mentioned outages.

5.6.1 New Dam Safety Regulations at Hetch Hetchy (NAI2)

Under baseline climate and 227 mgd demand, a reduction of Hetch Hetchy storage by 20% has little impact on water delivery reliability and storage reliability (see Section 4.4.1 for definition) (Figure 5-71 through Figure 5-74). Something worth noting, however, is that the delivery reliability falls below the 90% target if temperature increases by more than 2°C (Figure 5-71) or if precipitation decreases by 5% or more (Figure 5-72). Under baseline demand and climate, the storage reliability does not change when Hetch Hetchy reservoir storage is reduced by 20%. Warming temperature does not impact the storage reliability (Figure 5-73). On the other hand, a reduction of the precipitation by 5% gets the delivery reliability below target, while a decrease of the precipitation by 10% results in the storage reliability below target with the baseline Hetch Hetchy storage capacity (Figure 5-74). Similarly, delivery and storage reliability metrics decrease more when demand increases, highlighting an increased vulnerability of the RWS under a 20% storage capacity reduction at Hetch Hetchy Reservoir.

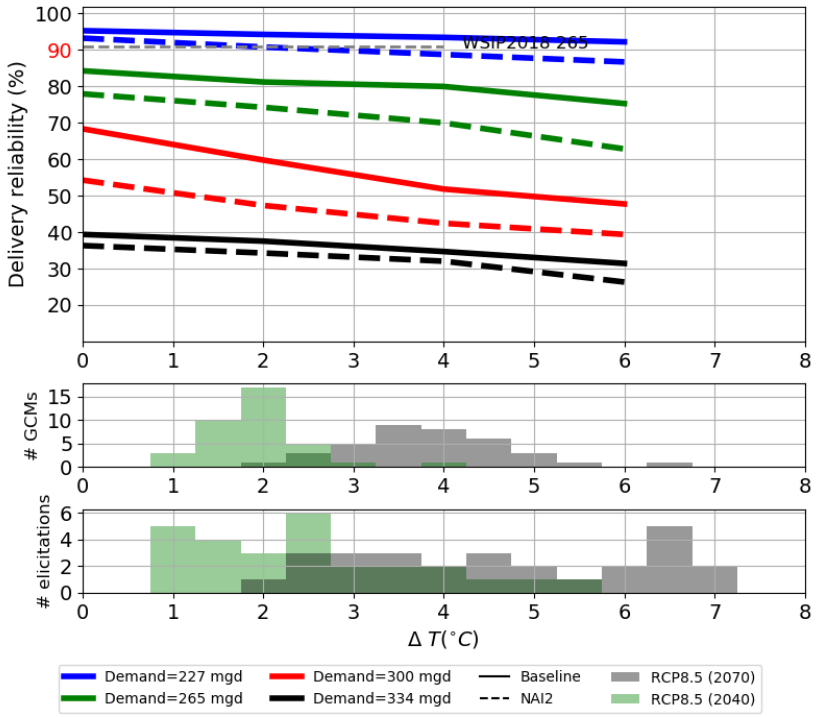


Figure 5-71. Water Delivery Reliability across All Temperature (ΔT) Scenarios and a Subset of Four Demand Scenarios (227 mgd, Which Corresponds to Current SFPUC Demand, 265 mgd [+15% Increase], 300 mgd [+30%] and 334 mgd [+45%]).

The dash color lines show the results obtained when new dam safety regulations are implemented (20% reduction of Hetch Hetchy storage). The dash-grey lines show reliabilities obtained from HHLSM WSIP 2018. The 90% reliability target in indicated in red color in the y-axis.

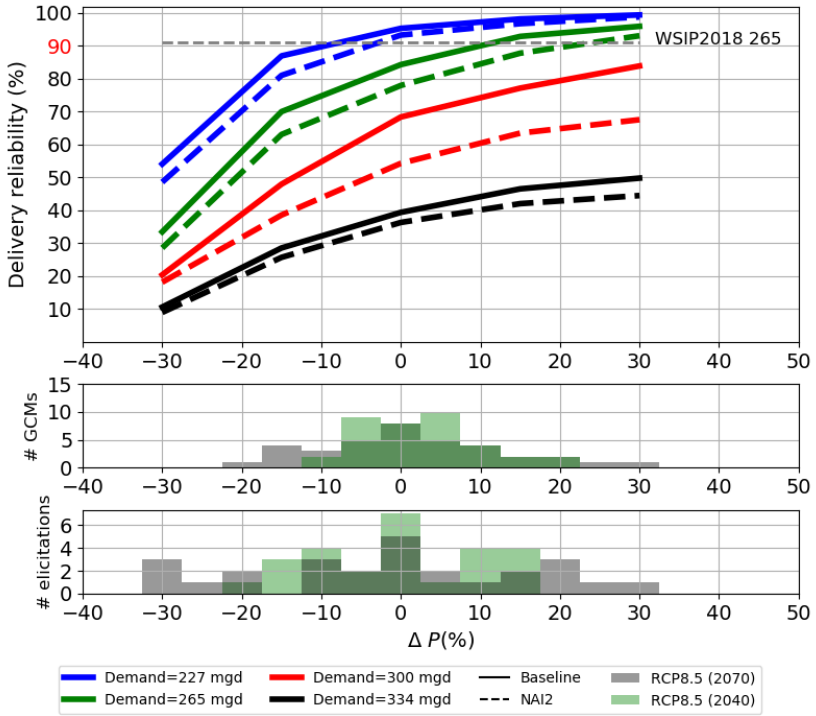


Figure 5-72. Same as Figure 5-71 for Change in Precipitation.

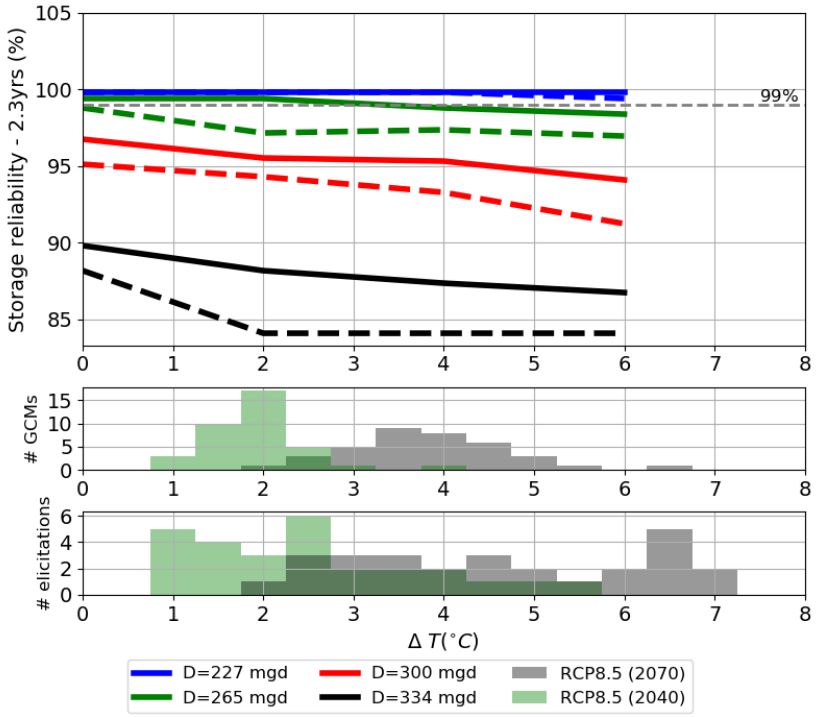


Figure 5-73. Same as Figure 5-70 but for Storage Reliability above 2.3 Years of Supply Remaining.
The 99% reliability target is indicated with the dash-grey line.

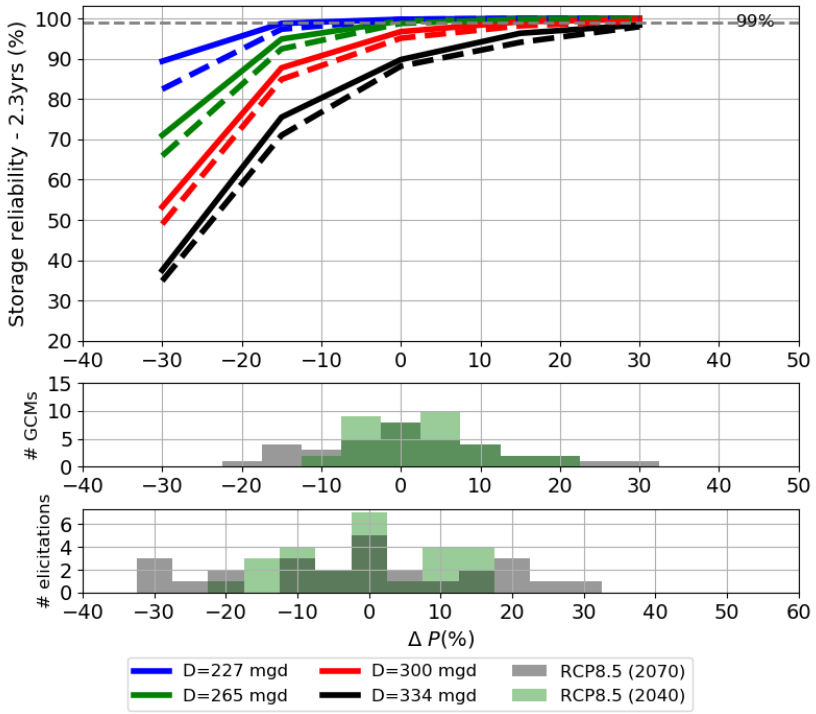


Figure 5-74. Same as Figure 5-72 but for Storage Reliability above 2.3 Years of Supply Remaining.

5.6.2 Major Failure at the Moccasin Switchyard (NAI3)

Under a Moccasin switchyard failure, Foothill Tunnel flow (and, therefore, SJPL flow) is completely curtailed for a duration of 365 days. For this narrative, the criteria for selecting the start dates of this

narrative focuses on the total volume of water demand during the event (i.e., either maximum or minimum total volume) and the level of storage in the local reservoirs (i.e., either maximum or minimum volume). Table 5-8 shows the selected dates and the realizations from which they are from. Note that during the Moccasin switchyard outage, only SVWTP and HTWTP provide water with a maximum yield of 921 AF/day (i.e., 300 mgd). This capacity being lower than the peak demand during the summer season, which frequently gets above 1,000 AF/day, deficits are unavoidable during this outage. As discussed below, climate and demand conditions are factors that increase the vulnerability of the RWS to the Moccasin switchyard outage.

Table 5-8. Characteristics of the Four Investigated Moccasin Switchyard Outage Narratives.
Local reservoirs are Calaveras, San Antonio, Crystal Springs, San Andreas and Pilarcitos. The maximum system yield is the aggregated capacity treatment capacity from the SVWTP (160 mgd) and HTWTP (140 mgd). Demand and initial storage values are given for the baseline 227 mgd demand and climate conditions.

Date of the outage	Selection Method	Realization	Maximum system yield during outage (AF/day)	Max daily demand during the outage period (AF/day)	Total demand during outage period (TAF)	Initial storage in local reservoirs (TAF)
6/30/2036	Minimum demand during the event (365 days)	5	921	1032	219	193
12/31/2049	Minimum storage local reservoirs	8		989	256	101
2/23/2057	Maximum storage local reservoirs	9		1106	265	236
8/20/2058	Maximum demand during the event (365 days)	9		1064	279	194

Figure 5-75 illustrates the RWS response during the Moccasin switchyard outage occurring on June 30th, 2036, from realization #5, which corresponds to the date where simulated demand throughout the outage is minimum across all ten realizations with baseline climate and demand conditions.

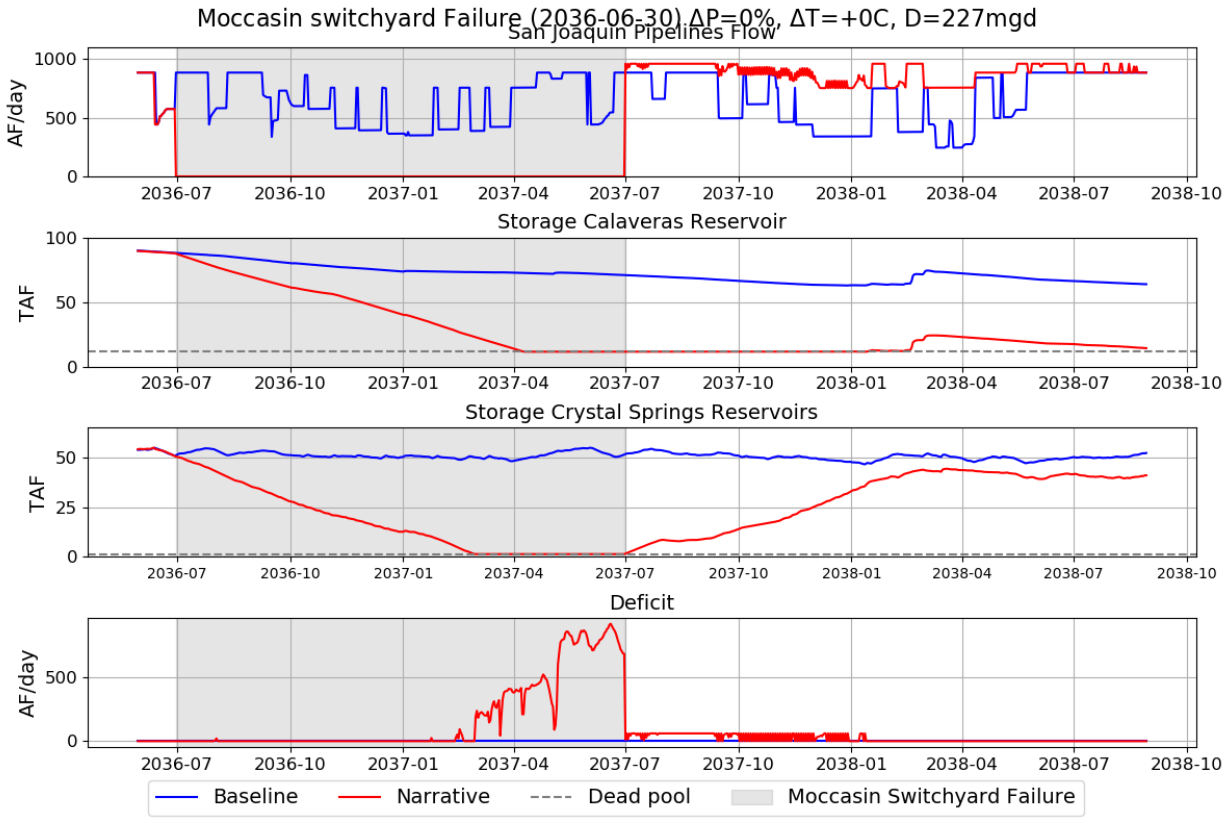


Figure 5-75. System Response after a Major Failure at the Moccasin Switchyard.
From top to bottom, figure panels focus on the operations of Upcountry, East Bay and Peninsula subsystems. The bottom panel illustrates the deficit. For each panel, the colors separate the baseline simulation (i.e., no outage; blue) from the simulation for which the outage was implemented (red). The grey dash lines show dead pool storage for Calaveras and Crystal Springs reservoirs. The date of the outage, the precipitation (ΔP), temperature (ΔT) and demand (D) scenarios are given in the figure title. The grey shaded area shows the duration of the outage.

During the Moccasin switchyard outage, SJPL flow remains at zero, and SVWTP and HTWTP increase their production to cope with the Moccasin switchyard failure. As a result, Calaveras and Crystal Springs Reservoirs are drawn down to supply the water treatment plants (Figure 5-75). They both reach their dead pool prior the end of the outage, which results into a sharp and large increase in deficit. At the end of the outage, SJPLs return online. Flows through the SJPLs are significantly larger than the ones obtained with the simulation without the outage. This is because at the end of the outage, storage in East Bay and Peninsula sub-systems are significantly below their emergency storage (i.e., 120 days of local demand, cf. Section 3.4.2). Crystal Springs starts refilling at the end of the outage period through transfer of water from Hetch Hetchy Reservoir. However, Calaveras Reservoir remains at dead pool level for several months after the end of the outage. The reason is that water from Hetch Hetchy Reservoir cannot be transfer to Calaveras Reservoir. Instead, the water is transferred to San Antonio Reservoir via the San Antonio pump station. During the refill, water from San Antonio Reservoir cannot be sent to SVWTP. Calaveras Reservoir is not able to satisfy the constraint of 61 AF/day to supply SVWTP, hence virtual water is required for few additional months by the LP solver to prevent SFWSM from crashing.

Figure 5-76 illustrates for the four outage onset dates the influence of precipitation and demand changes on the simulated deficit throughout the event (i.e., grey areas in Figure 5-75). The left column in Figure 5-76 shows the simulated deficit for the baseline (i.e., no outage) during the period of the outage. As such, the deficit shown in the left column results only from either drought rationing or additional

virtual water that SFWSM requires to either satisfy the deliveries and/or the LP constraints. The simulated deficits during the period of the outage are shown in the right column.

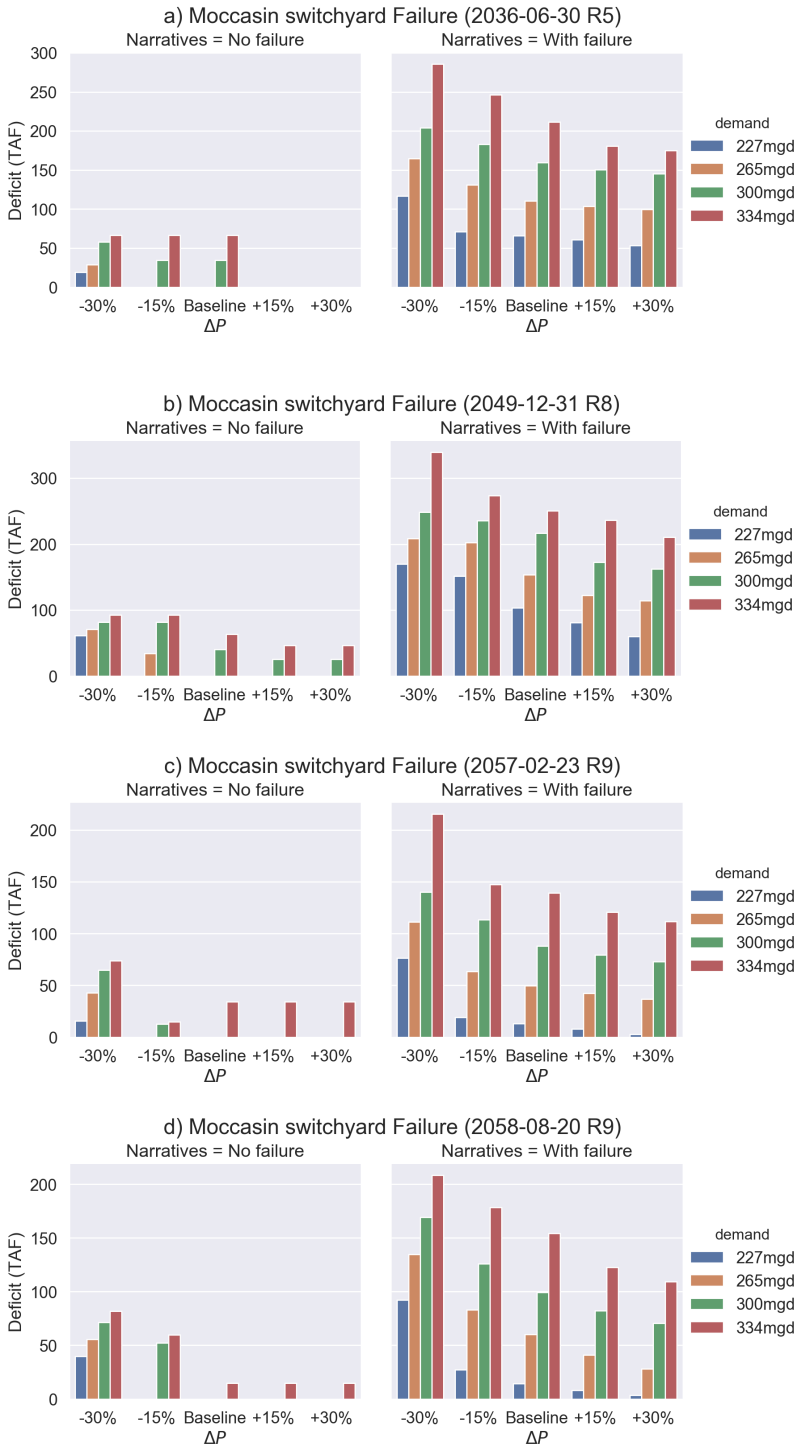


Figure 5-76. Deficit Resulting from the Outage of the Moccasin Switchyard in Combination with Precipitation and Demand Changes.

Left panels show the simulated deficit without outage. Right panels show the simulated deficit with the outage. Daily deficits are summed across the duration of the outage (365 days from the beginning of the outage).

From Figure 5-76, several observations can be made:

- **Influence of the onset date of the events**: Although only four onset dates were investigated for the Moccasin switchyard outage, significant differences among these outages are observed. For instance, under baseline climate and demand conditions, the simulated deficit during the outage period equal 14, 65, 13 and 103 TAF (from top to bottom in Figure 5-76). At a supply deficit of a 103 TAF this represents almost half of baseline demand volume over the outage. Differences among these deficits result from the initial state of the system at the onset of the outage (e.g., storage levels in the reservoirs, current drought rationing), the volume of demand the system must satisfy (Table 5-8) and the volume of inflow entering the local reservoirs throughout the outage period. The outage starting on December 31, 2049, occurs at the end of a 2-month long maintenance of the SJPL, expanding the total duration of SJPL being offline to 14 months, which likely explains the significant simulated deficit for this outage (i.e., 103 TAF).
- **Influence of precipitation change**: In general, any decrease in precipitation increases the simulated deficit throughout the outage. The reason is that with reduced precipitation across all system watersheds, inflow to reservoirs are reduced during the outage period. In addition, the local reservoirs storage is likely to be reduced at the onset of the outage, leading the local reservoirs running dry earlier during the event. Although not critical for the investigated Moccasin switchyard outages, some nonlinearities exist between precipitation change and simulated deficit. This is for instance illustrated for the outage starting on February 23rd, 2057 (Figure 5-76c) and for the +45% demand scenario (i.e., demand equal to 334 mgd). The simulated deficit obtained under -15% precipitation scenario is lower than the one obtained for the baseline precipitation and above. This difference is explained by different rationing sequences among the different precipitation change scenarios prior and during the outage.
- **Influence of demand change**: Across the four considered Moccasin switchyard outages, increasing demand increases almost linearly the simulated deficit throughout the duration of the event.

In light of the sensitivity discussed above, the RWS is vulnerable to an outage of the Moccasin switchyard. All four considered outages show water supply deficit, even under baseline demand and climate. Any reduction in precipitation or increase in demand will worsen the recorded deficit. Since the available system yield during the outage is lower than the demand during hot days in the summer season, even a major increase in precipitation will not prevent the RWS from water shortage.

5.6.3 Earthquake along the Calaveras Fault (NAI4)

Under this narrative-based scenario, the earthquake along the Calaveras fault brings SVWTP offline for 60 days and limits Alameda siphons flow to 368 AF/day (120 mgd) during the first 30 days, and to 491 AF/day (160 mgd) for the last 30 days. Note that in SFWSM, HTWTP can deliver water to some customers in South East Bay region (i.e., Burlingame, Hillsborough, Foster City and Estero Municipal Improvement District, Cal Water Services Mid-Peninsula, and Mid-Peninsula Peninsula Water District). Delivery to these customers with water from the Alameda Siphons is prioritized by setting a higher cost to their connection with HTWTP.

The criteria for selecting the start dates of the narratives focus on the total volume of water demand during the event (i.e., either maximum or minimum total volume) and the level of storage in the Peninsula reservoirs (i.e., either maximum or minimum volume). Table 5-9 shows the selected dates and the climate realizations from which they are from. Table 5-9 also reveals that when the earthquake occurs during the summer season, the daily peak demand is likely to be larger than the maximum system yield, which would result in deficit.

Table 5-9. Characteristics of the Start Dates of the Earthquake along Calaveras Fault Narratives.
Peninsula reservoirs are Crystal Springs, San Andreas and Pilarcitos. The maximum system yield is the aggregated capacity from the reduced Alameda Siphons capacity (120 mgd during the first 30 days and 160 mgd for the last 30 days) and the treatment capacity at Harry Tracy (140 mgd). Demand and initial storage values are given for the baseline demand and climate conditions.

Date of the outage	Selection Method	Realization	Maximum system yield during outage (AF/day)	Max daily demand during the outage period (AF/day)	Total demand during outage period (TAF)	Initial storage in PEN reservoirs (TAF)
12/16/2036	Minimum demand during the event (60 days)	5	798 (first 30 days) then 921 (last 30 days)	613	29	73
12/31/2049	Minimum storage in Peninsula reservoirs	8		685	33	52
2/8/2057	Maximum storage in Peninsula reservoirs	9		694	31	90
7/7/2057	Maximum demand during the event (60 days)	4		1005	60	72

Figure 5-77 illustrates the RWS response to an earthquake on July 7, 2057, from realization 4, which corresponds to the date where simulated demand for the next 60 days is maximum across all ten realizations and baseline climate conditions. At the onset of the event, the loss of SVWTP and reduction in Alameda Siphons capacity results in HTWTP increasing production to compensate this loss and satisfy demand. During this event, HTWTP delivered 4,236 AF to the South East Bay customers mentioned above, mostly during the first 30 days (not shown). During the first 30 days, the RWS is not able to deliver full demand, which correspond to the period with the highest demand (July month) and the lowest yield capacity (798AF/day), while several days have a larger daily demand (Table 5-9). During the last 30 days, the increase in Alameda Siphons capacity combined with a slightly lower demand throughout August – early September results in the RWS delivering full demand. The high flow through the SJPLs at the end of the event helps to refill Crystal Springs.

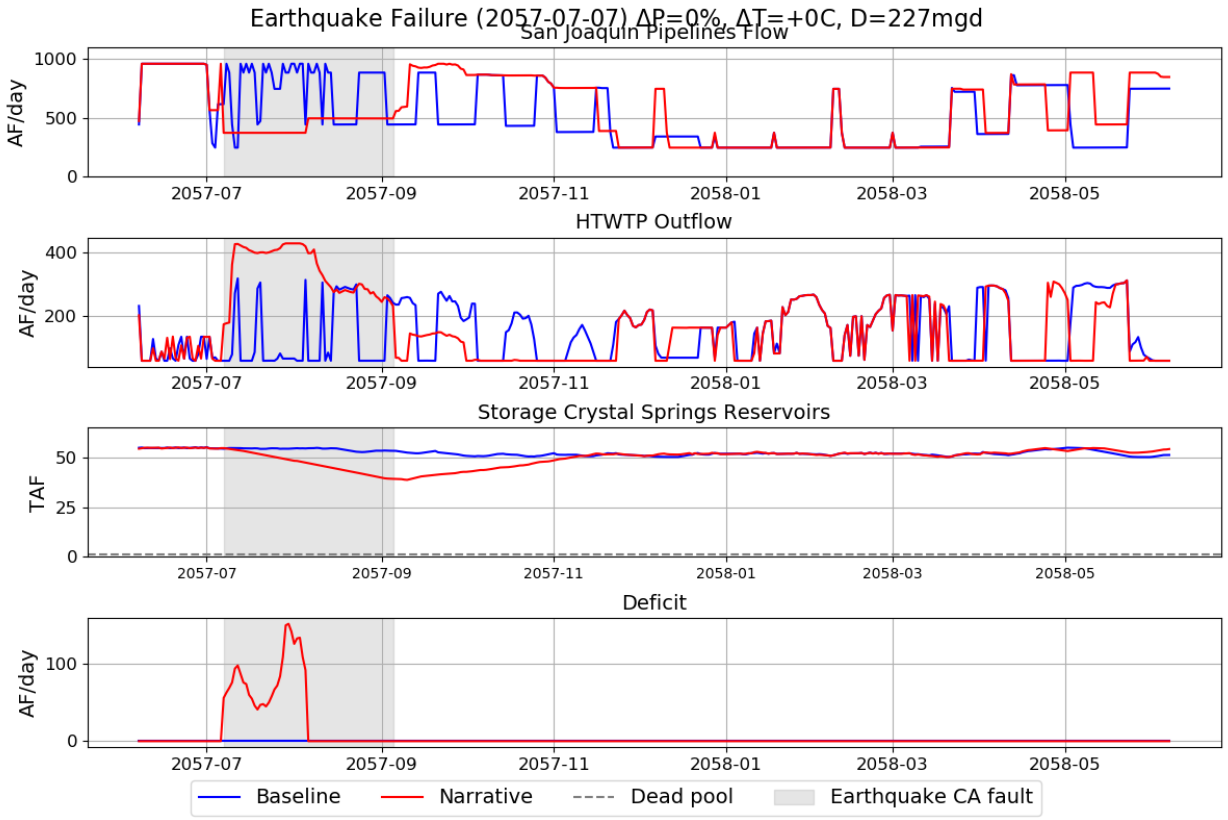


Figure 5-77. System Response after an Earthquake along the Calaveras Fault.
From top to bottom, figure panels focus on the operations of the SJPL, HTWTP and Crystal Springs Reservoir. The bottom panel illustrates the deficit. For each panel, the colors separate the baseline simulation (i.e., without the earthquake; blue) from the simulation for which the aftermath of the earthquake was implemented (red). The dash lines show dead pool storage for Crystal Springs Reservoirs. The date of the event, the precipitation (ΔP), temperature (ΔT) and demand (D) scenarios are given in the figure title. The grey shaded area shows the duration of the failure.

Figure 5-78 illustrates for the four considered earthquakes the influence of precipitation and demand changes on the simulated deficit throughout the event. Similar to Figure 5-76, the left column shows the simulated deficit for the baseline (i.e., no outage) while the simulated deficits during the period of the outage are shown in the right column.

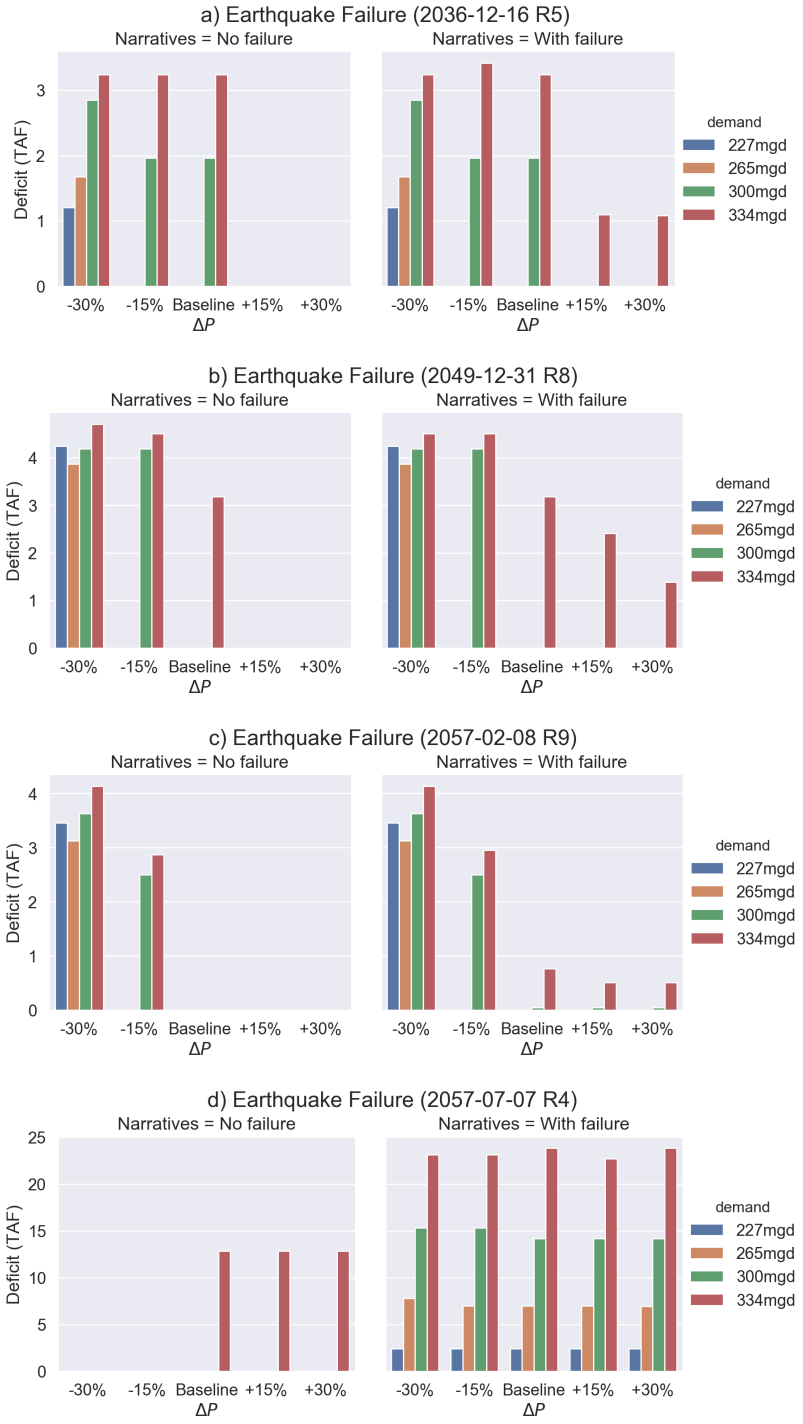


Figure 5-78. Deficit Resulting from an Earthquake along the Calaveras Fault in Combination with Precipitation and Demand Changes.

Left column shows the simulated deficit without outage. Right column shows the simulated deficit with the outage. Daily deficits are summed across the duration of the outage (60 days from the beginning of the outage). Note: In Figure 5-78d, the simulated deficit for the baseline simulation (left column) for the 334 mgd scenarios corresponds to a drought rationing of 10%. The rationing is not effective for the -30% and -15% precipitation scenarios, for which a rationing was effective two years prior the event.

From Figure 5-78, several observations can be made:

- **Influence of the onset date of the events**: the timing of the earthquake is an important driver of the vulnerability of the RWS to an earthquake along the Calaveras fault. Among the considered earthquake, only the one occurring during the summer season leads to deficit under baseline demand and climate conditions. This is due to a lower system yield when compared to the demand during very hot days. For the other three earthquake, deficits are only simulated for scenarios with reduced precipitation and/or increased in demand.
- **Influence of precipitation change**: Any decrease in precipitation tends to increase deficit, while an increase tends to reduce the deficit.
- **Influence of demand change**: Any increase in demand increases the deficit. Note that the drought rationing can differ from one precipitation/demand scenario to another, which sometimes leads to some non-linarites in the deficit, as it is the case for the earthquake occurring February 8th, 2057 (Figure 5-78d) for the -30% precipitation scenario.

Generally, the system is vulnerable to an earthquake along the Calaveras due to insufficient transfer capacity (i.e., via the Alameda Siphons) during the event if the latter occur during the summer season. This vulnerability increases when precipitation decreases and/or demand increases, which also makes the RWS vulnerable during seasons with lower seasonal demand (i.e., winter).

5.6.4 Major Fire across Crystal Springs Reservoirs Watershed (NAI5)

Under this narrative-based scenario, the Harry Tracy water treatment plant (HTWTP) is off for 365 days. The criteria for selecting the start dates of the narrative focus on the total volume of water demand during the event (i.e., either maximum or minimum total volume) and the level of storage in the Upcountry and East Bay reservoirs (i.e., either maximum or minimum volume). Candidate dates are selected only from the 'fire' season (i.e., from July 1st to October 30th). Table 5-10 shows the selected dates and the realizations from which they are from. Note that the system maximum yield is significantly larger than the daily peak demand during the summer season.

Table 5-10. Characteristics of the Start Dates of the Moccasin Switchyard Failure Narratives.
Upcountry reservoirs are Hetch Hetchy, Cherry and Eleanor and Water Bank. East Bay reservoirs are Calaveras and San Antonio. The maximum system yield is the aggregated capacity from the SJPL and SVWTP. Demand and initial storage values are given for the baseline demand and climate conditions. The maximum system yield is the sum of the SVWTP yield (160 mgd) and the SJPL capacity (312 mgd)

Date of the outage	Selection Method	Realization	Maximum system yield during outage (AF/day)	Max daily demand during the outage period (AF/day)	Total demand during outage period (TAF)	Initial storage in UC + EB reservoirs (TAF)
7/1/2036	Minimum demand during the event (365 days)	5	1451	1032	219	1342
10/31/2036	Minimum storage in Upcountry and East Bay reservoirs	8		1032	233	748
7/1/2057	Maximum storage in Upcountry and East Bay reservoirs	2		1106	251	1369
8/20/2058	Maximum demand during the event (365 days)	9		1065	279	1254

Figure 5-79 illustrates the RWS response when a wildfire across Crystal Springs watershed forces HTWTP to be offline on July 1st, 2036 (realization 5), which corresponds to the date and realization where simulated demand is the highest for the next 365 days under baseline conditions. During the HTWTP outage, the yield of SVWTP increases significantly, which results to larger drawdown of Calaveras reservoir. However, the outage of HTWTP for 365 days does not lead to any deficit. At the end of the event, Calaveras reservoir is more than half-full, which highlights that, even for the period of 365 days for which the largest demand was simulated, the RWS was pretty far to not delivering full demand.

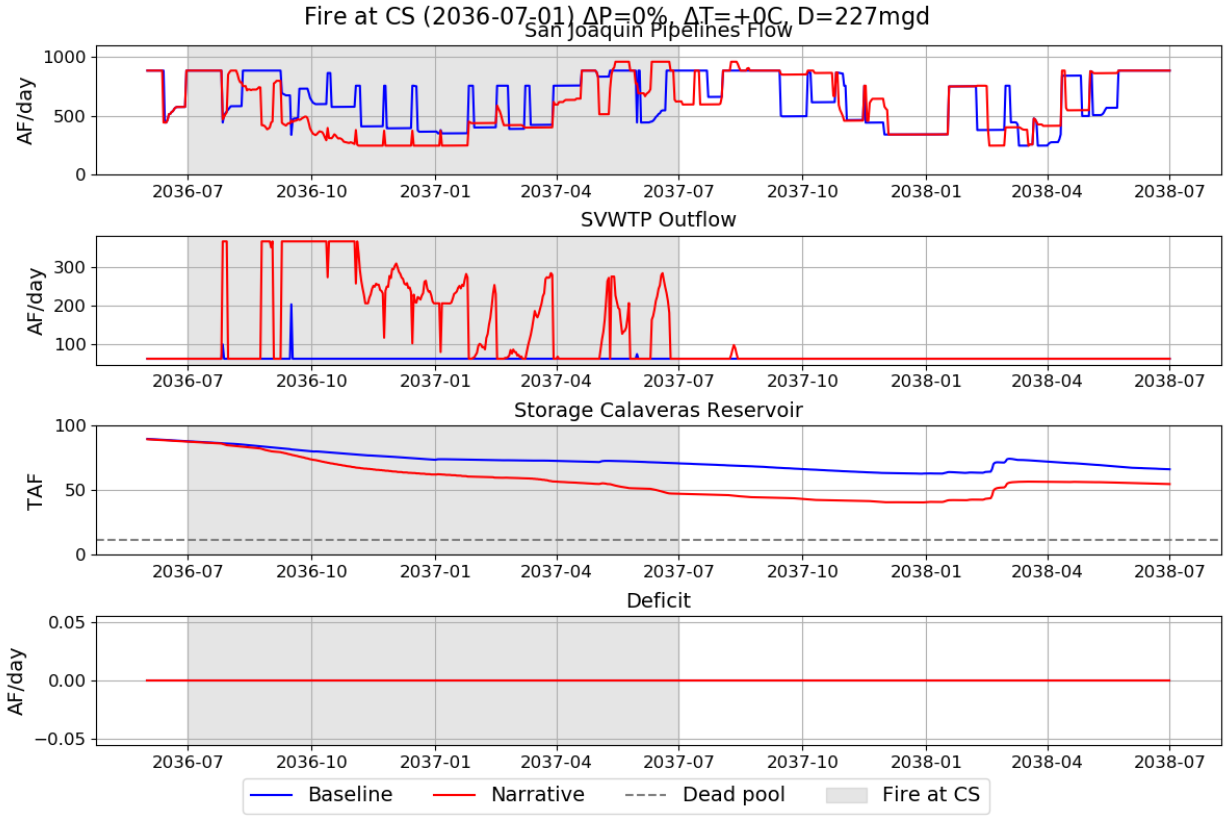


Figure 5-79. System Response after a Wildfire across Crystal Springs Reservoirs Watershed. From top to bottom, figure panels focus on the operations of the SJPL, SVWTP and Calaveras Reservoir. The bottom panel illustrates the deficit. For each panel, the colors separate the baseline simulation (i.e., without the wildfire; blue) from the simulation for which the aftermath of the wildfire was implemented (blue). The dash lines show dead pool storage for Calaveras Reservoir. The date of the event, the precipitation (ΔP), temperature (ΔT) and demand (D) scenarios are given in the figure title. The grey shaded area shows the duration of the failure.

Figure 5-80 illustrates for the four considered wildfires across the Crystal Springs Reservoirs watershed the influence of precipitation and demand changes on the simulated deficit throughout the event. Similar to Figure 5-76, the left column shows the simulated deficit for the baseline (i.e., no outage) while the simulated deficits during the period of the outage are shown in the right column.

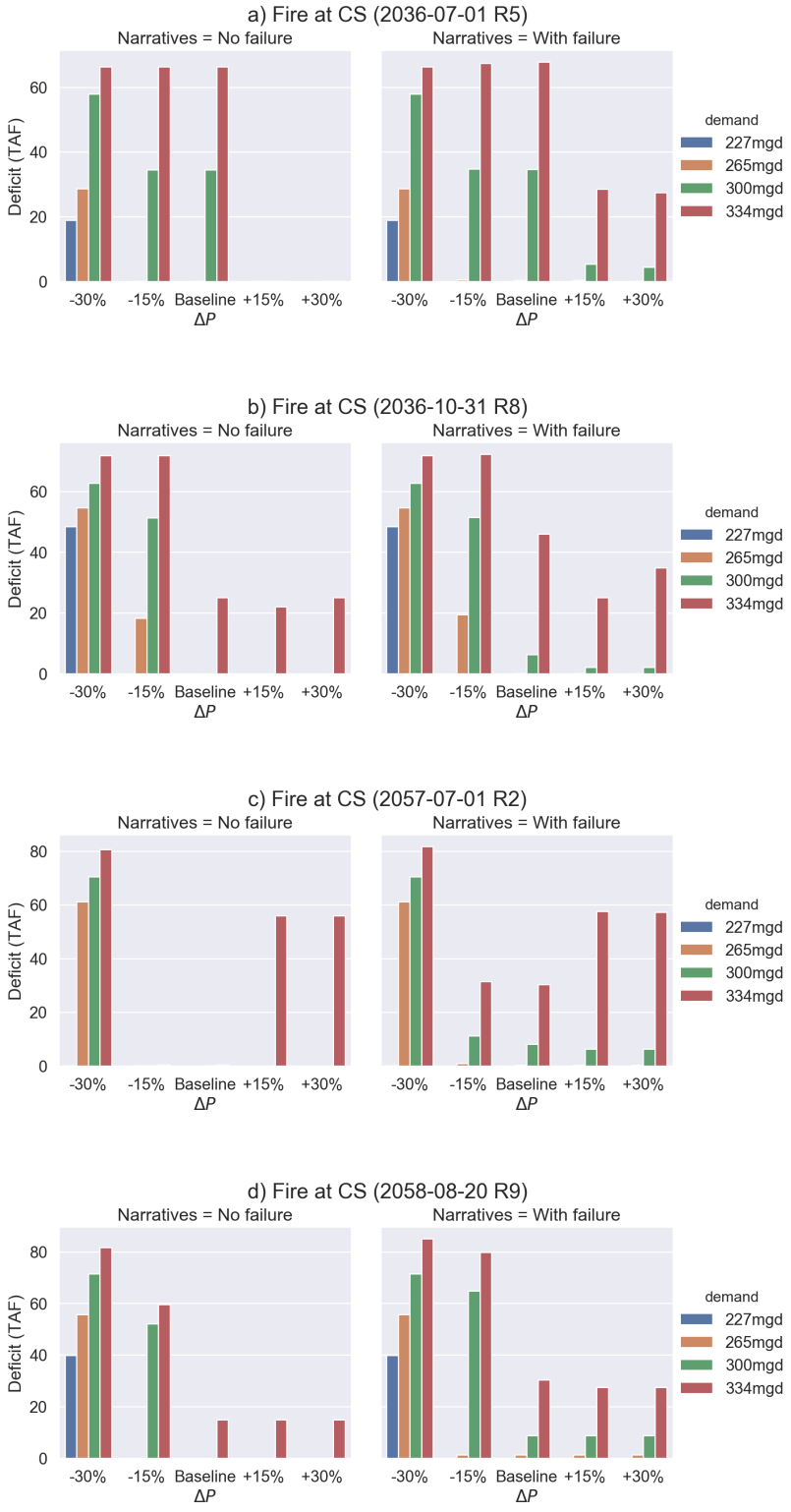


Figure 5-80. Deficit Resulting from a Wildfire across the Crystal Springs Watershed in Combination with Precipitation and Demand Changes.

Left panels show the simulated deficit without outage. Right panels show the simulated deficit with the outage. Daily deficits are summed across the duration of the outage (365 days from the beginning of the outage).

From Figure 5-80, several observations can be made:

- **Influence of the onset date of the events**: None of the four simulated HWTP outages leads to deficit under baseline demand and climate conditions.
- **Influence of precipitation change**: Decreasing precipitation tends to increase the deficit. Under baseline demand, deficits are simulated only if precipitation decreases by 30%. Similar to what has been discussed for the other outage narrative-based scenarios, some nonlinearities resulting from different sequences of drought rationing across the different precipitation scenarios may sometimes lead to an increased deficit when precipitation increases, although this only occur for high demand scenarios (Figure 5-80b and Figure 5-80c).
- **Influence of demand change**: Increasing demand leads to larger deficit. Under baseline climate, no deficit in simulated unless the demand increases by 30% or more (i.e., 300 mgd and above) for three of the simulated outages. For the fourth outages, a total deficit of 1259 AF is simulated under the +15% demand scenarios (i.e., 265 mgd).

No major vulnerability to water supply from a wildfire across Crystal Springs' watershed. Water shortages are only observed if demand increases by 30% or more and/or precipitation decrease significantly.

5.7 Effects of Climate and Demand Change on Finance

The results of the finance stress test against the metrics of the price of water and the rate of water price increase are discussed under both trend and step model runs in the sections below.

5.7.1 Step Model Runs

The results for the price rate increase metric are presented for a single weather realization, no change in temperature, 227 mgd demand, and under two conditions of precipitation (no change and -30%) in Figure 5-81. The effect of decreased precipitation, and thus in the frequency of drought rationing, is clear to see under stationary conditions provided by step model runs. Although the impact on the price of water is similar in absolute terms (middle graph in Figure 5-81), the relatively lower price of water for wholesale customers means that this change in price is much larger than for retail customers in percentage terms. The rate at which the price rate increase metric is violated increases under a decreased precipitation scenario. For example, the price rate increase metric for retail customers is not violated under the no change scenario (solid olive line in bottom graph of Figure 5-81) which increases to three times when precipitation decreases by 30% (dashed olive line in bottom graph of Figure 5-81).

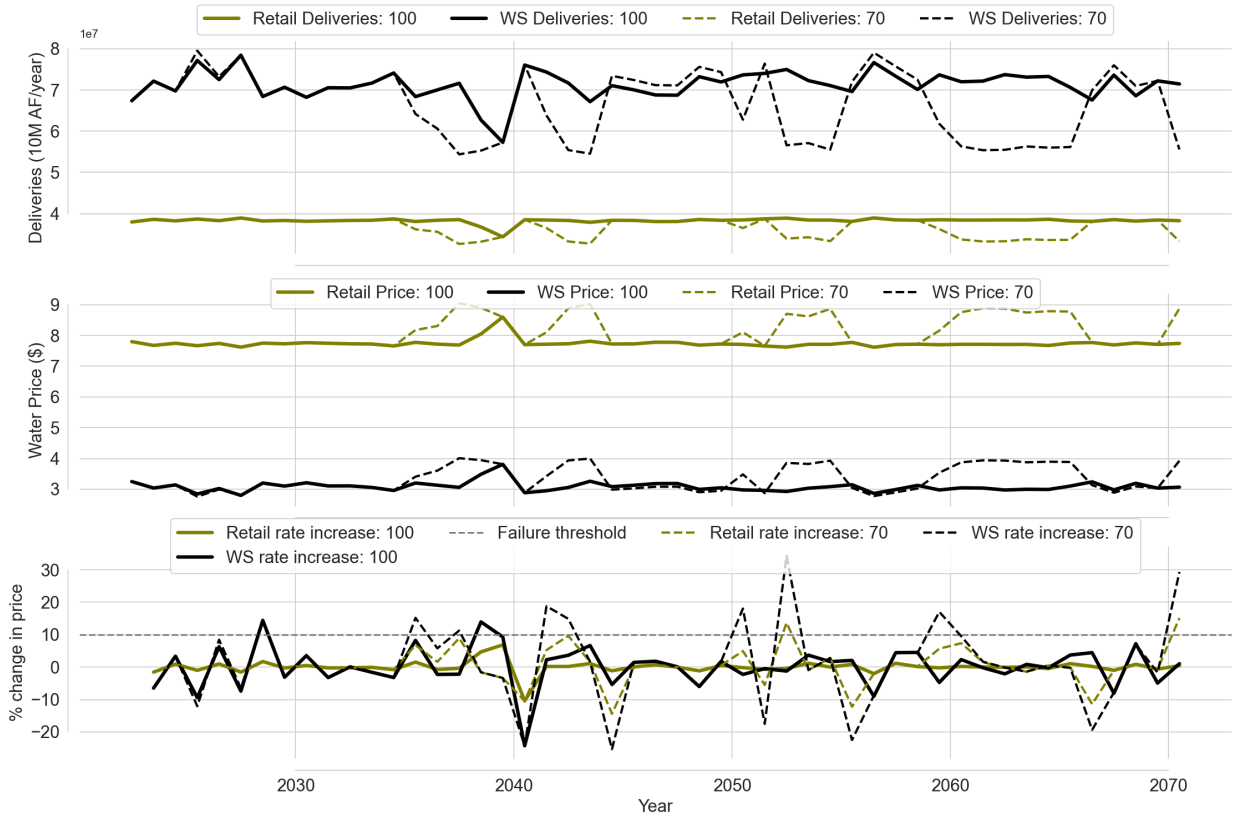


Figure 5-81. Results of Finance Model Stress Test under Step Model Runs for One Weather Realization (Realization 5) under Two Different Conditions of Precipitation: No Change (Solid Lines) and -30% (Dashed Lines), for Wholesale (Black) and Retail (Olive) Customers.

Failure thresholds are indicated in dashed grey lines. From top to bottom; annual deliveries in AF; price of water in $/ccf; and % increase in water rate from previous year. Note that after a drought, deliveries to wholesale customers are often larger than what they would have been without drought. This is due to an increased demand in the Peninsula region after a drought to refill the Westside Basin groundwater.

The same results are presented but under no change in precipitation, no change in temperature and two scenarios of demand (no change and +30% - 300 mgd) in Figure 5-82. Once again, one can observe the lower price of water in the 300 mgd demand scenario relative to the 227 mgd demand scenario. In this case, although the frequency of rationing is higher for retail customers in the 300 mgd scenario (dashed olive line in top graph of Figure 5-82), the magnitude of rationing is small, coupled with the relatively low price (dashed olive line in middle graph of Figure 5-82), results in a relatively stable rate of change in terms of price.

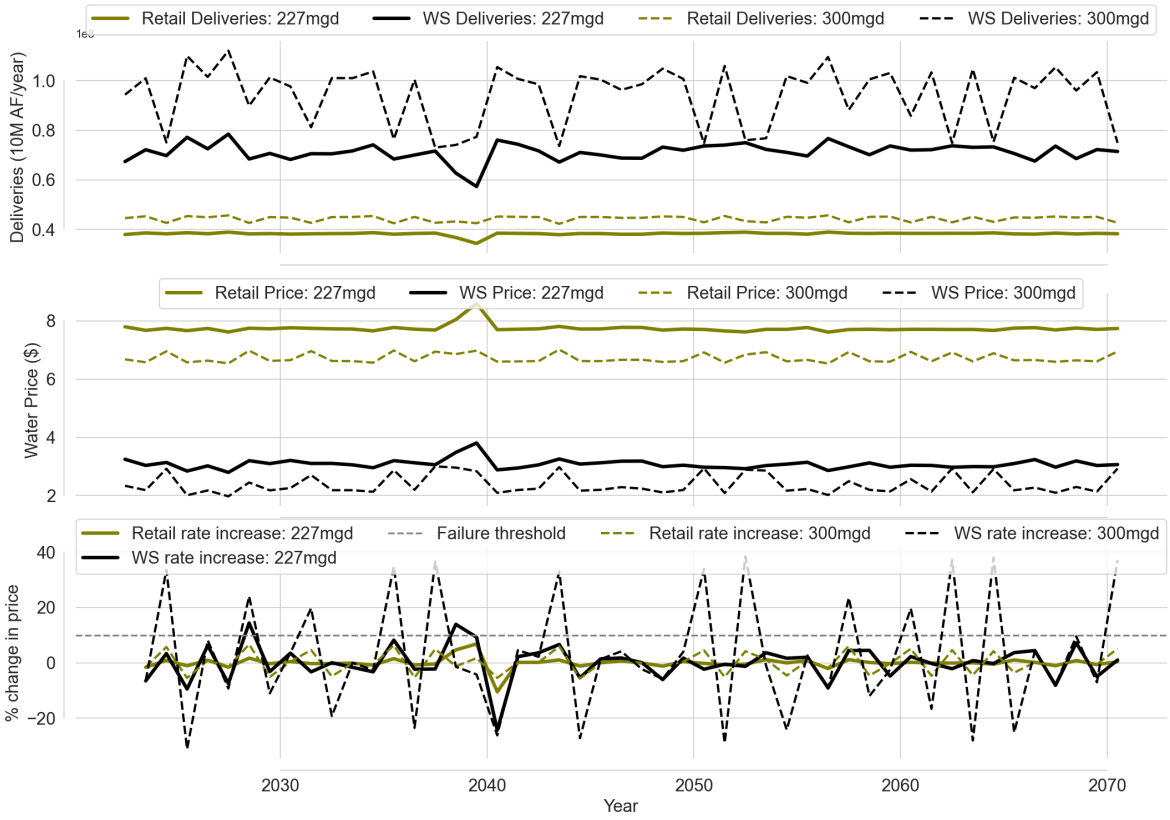


Figure 5-82. Results of Finance Model Stress Test under Step Model Runs for One Weather Realization (Realization 5) under Two Different Conditions of Demand: No Change (Solid Lines) and +30% (Dashed Lines), for Wholesale (Black) and Retail (Olive) Customers.

Failure thresholds are indicated in dashed grey lines. From top to bottom; annual deliveries in AF; price of water in $/ccf; and % increase in water rate from previous year.

Figure 5-83 shows the significant effect of different levels of annual CAPEX spend on the price of water for retail customers when all other variables are held constant. Annual CAPEX spend of $1.05B (dashed olive line in middle graph of Figure 5-83) results in a price of water more than twice that under 2020 CAPEX spend of $0.35B[8] (solid olive line in middle graph of Figure 5-83). The effect on the price of water for wholesale customers is less pronounced (black lines in middle graph of Figure 5-83) due to the way CAPEX spend is apportioned amongst retail and wholesale customers (see Figure 3-59), 75% of total annual CAPEX spend is included in the calculation of the price of water for retail customers. The rate of water price increase does not change between the two CAPEX scenarios given the same volume of water is delivered under the same conditions of demand, precipitation, and temperature.

[8] 2020 CAPEX spend refers to total CAPEX spend in 2020 in the SFPUC 10 yr financial plan

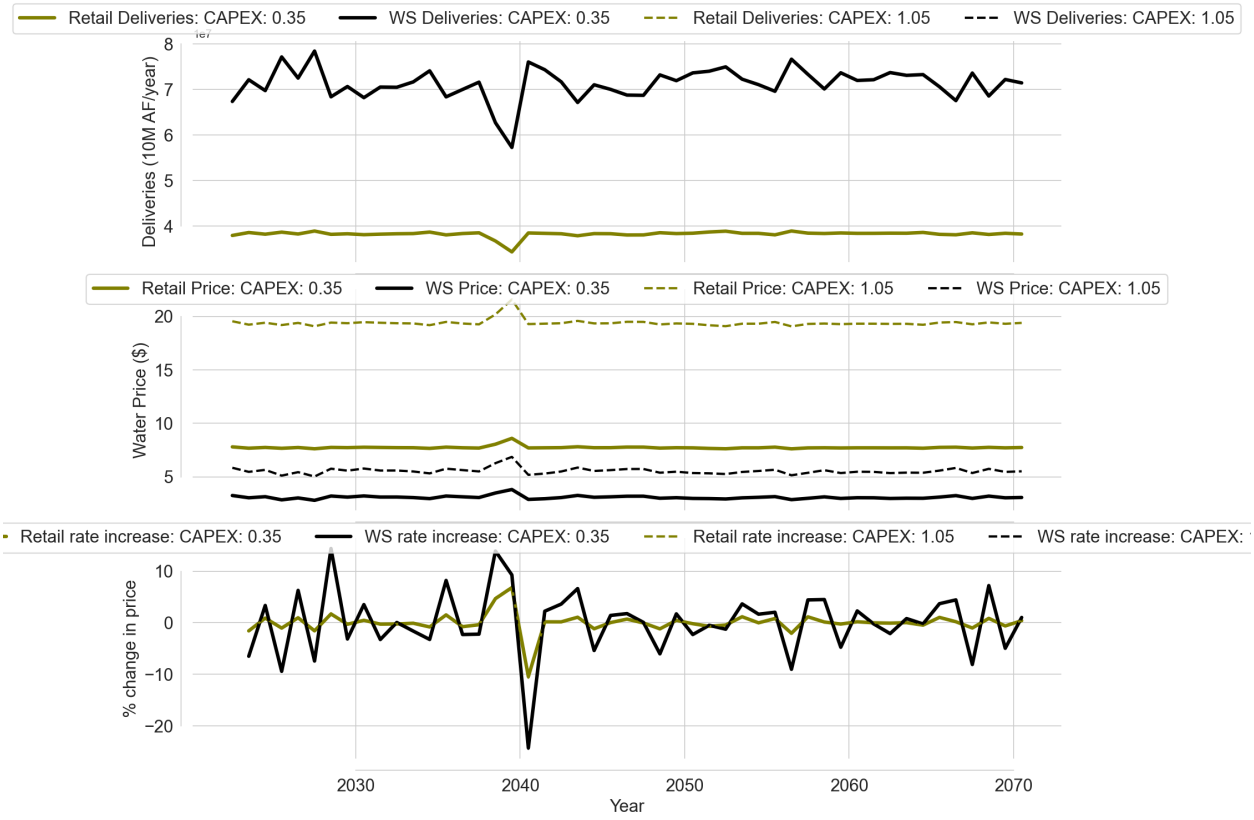


Figure 5-83. Results of Finance Model Stress Test under Step Model Runs for One Weather Realization (Realization 5) under Two Different Conditions of CAPEX: 2020 CAPEX Spend $0.35B (Solid Lines) and $1.05B (Dashed Lines), for Wholesale (Black) and Retail (Olive) Customers.

Failure thresholds are indicated in dashed grey lines**.** From top to bottom; annual deliveries in AF; price of water in $/ccf; and % increase in water rate from previous year.

The price of water is sensitive to both CAPEX and Demand with neither factor significantly outweighing the other. Figure 5-84 shows the response to changing Demand and CAPEX. As would seem logical significant growth in demand is required in order to mitigate increased CAPEX spend. Annual OPEX spend of $0.5B (representing expected OPEX spend in 2040 according the SFPUC 10-yr Financial Plan) relative to $0.29B (present day OPEX spend) compounds the effect of both decreasing demand and increasing annual CAPEX spend.

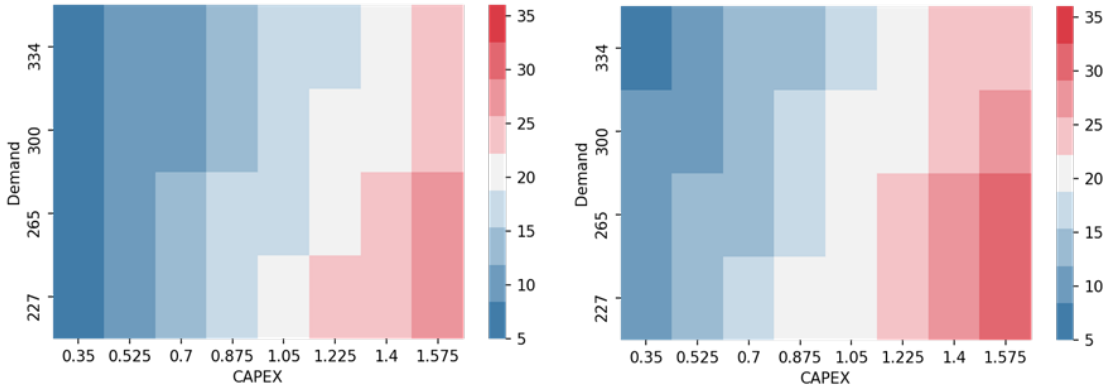


Figure 5-84. Response Surfaces Showing the Combined Effect of Demand and CAPEX Expenditure under $0.29B Annual OPEX Spend (Left), and $0.5B Annual OPEX Spend (Right) on the Price of Water ($/ccf).

The following key observation can be made from Figure 5-84:

- **Demand and CAPEX drive price response**: demand and the level of CAPEX spend has a significant impact on the price of water for customers. Results presented in previous sections here have shown how system performance deteriorates when precipitation decreases by more than 10%. Further, the potential introduction of new instream flow requirements would reduce the amount of available supply, compounding these effects. In a situation in which major capital investment is required to add additional supply to the system as a result of new instream flow requirements and/ or decreases in precipitation as a result of climate change, demand would need to increase significantly to mitigate substantial increases in the price of water for customers. For example, if annual CAPEX spend were to increase from the 2020 baseline of $350M to $525M under the $500M OPEX spend scenario (right hand side of Figure 5-84), demand would have to increase by 30% in order to maintain existing prices or else rise by ~50% from $10/ccf to $15/ccf. However, results presented above also show that system performance is very sensitive to even small increases in demand. Thus, in considering new capital investments, a trade off must be made between reliability and price.

5.7.2 Trend Model Runs

The results for the price rate increase metric are presented for a single weather realization, no change in temperature, no change in demand, and under two conditions of precipitation (no change and -30%) in Figure 5-85. One can clearly observe the price setting logic of cost recovery in these results – drought events cause short term reductions in deliveries and consequently, a spike in the price of water. Following the drought event, deliveries and prices return to their original trajectory. It is during these drought events that the price rate increase performance metric (i.e., no more than a 10% increase in price year on year) is violated. Thus, this metric is vulnerable to decreases in precipitation and under weather generator realizations that contain deep droughts. A decrease in precipitation that leads to more frequent drought rationing will result in a higher frequency of violations of this metric across the time horizon. One can observe this by comparing the price of water and change in price for the two precipitation scenarios in Figure 5-85. The scenario in which mean precipitation has decreased by 30% results in four drought events between 2040 and 2060 and resulting price rate increase violations in the early 2050s and early 2060s that are not observed in the no change scenario.

Wholesale prices are much more volatile than retail prices due to the relatively volatile nature of deliveries to these customers due to the imposition of drought restrictions - in accordance with the 2009 Water Service Agreement, wholesale customers are required to reduce their consumption during drought to a greater degree than retail customers. The result is that the rate increase metric is violated under almost all scenarios for wholesale customers– robustness of 99.6%. Retail price rate performs better but is still violated in just under half of all scenarios considered – robustness of 67.7%.

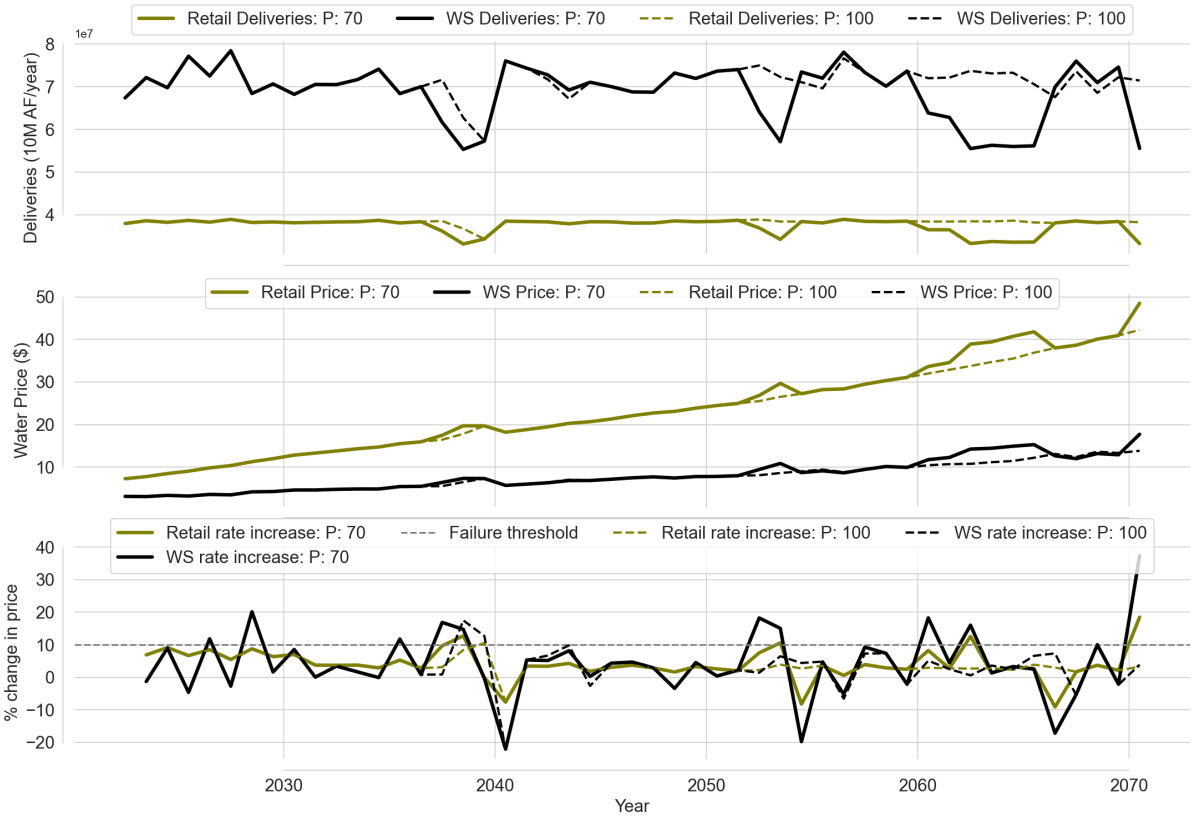


Figure 5-85. Results of Finance Model Stress Test under Trend Model Runs for One Weather Realization (Realization 5) under Two Different Conditions of Precipitation: No Change (Solid Lines) and -30% (Dashed lines), for Wholesale (Black) and Retail (Olive) Customers.

Failure thresholds are indicated in dashed grey lines. From top to bottom; annual deliveries in AF; price of water in $/ccf; and % increase in water rate from previous year.

The same results are presented but under no change in precipitation, no change in temperature and two scenarios of demand (no change and +30% - 300 mgd) in Figure 5-86. In this case, the price of water is lower in the 300 mgd demand scenario relative to the 227 mgd demand scenario given the same CAPEX and OPEX costs are spread across a larger volume of water delivered. The impact of demand on the rate of price increase is twofold: 1) the timing of the violation (higher demand leads to earlier imposition of drought rationing and thus a spike in the price of water earlier; 2) the lower demand scenario tends to experience fewer and less severe drought rationing events and thus violate the price rate increase metric less frequently.

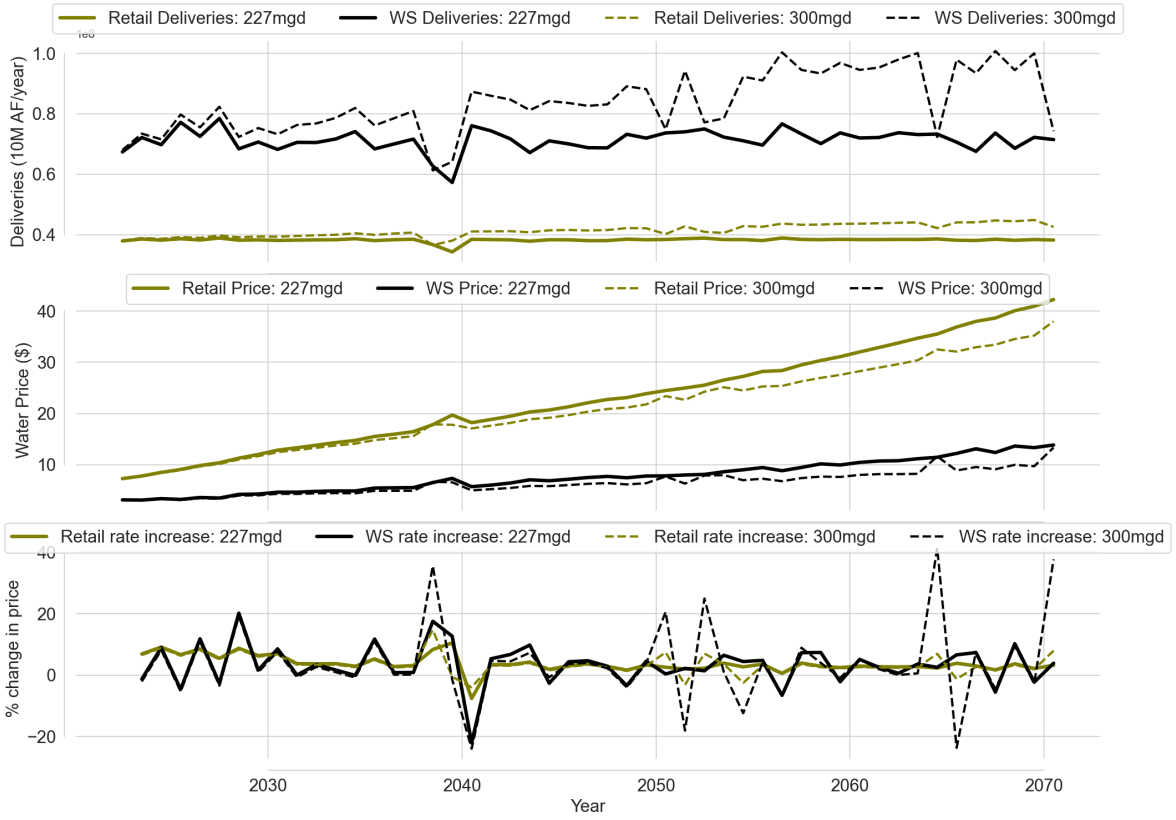


Figure 5-86. Results of Finance Model Stress Test under Trend Model Runs for One Weather Realization (Realization 5) under Two Different Conditions of Demand: No Change (Solid Lines) and +30% (Dashed Lines), for Wholesale (Black) and Retail (Olive) Customers.

Failure thresholds are indicated in dashed grey lines. From top to bottom; annual deliveries in AF; price of water in $/ccf; and % increase in water rate from previous year.

The results of the stress test for stress test for the trend model runs are presented in Figure 5-87 showing both the combined effect of demand and precipitation and of temperature and precipitation on the average rate at which price increases by year on year (in % terms). The presence of some non-linearities in these response surfaces is because of the trend model runs from which they are generated – i.e., price rate increase values are calculate under non-stationary conditions. In the case of temperature and precipitation (right side of Figure 5-87) a similar effect to results discussed in previous sections of this report. A clear signal is shown with regards to changes in precipitation - decreases in precipitation result in a higher rate of price increase on average and vice versa. However, the differences in the percentage points between the highest and lowest rate is small: 0.1%. The effect of temperature is limited with little difference between +5°C and current conditions when precipitation is held constant. The effect of demand is much more pronounced (left side of Figure 5-87). Demand shows a much clearer signal and largely outweighs any effect of changes in precipitation, i.e., the rate of price increase fall within the same bin across all precipitation scenarios under a single demand scenario. The range of price rate increase is also larger across demand scenarios – 0.7%.

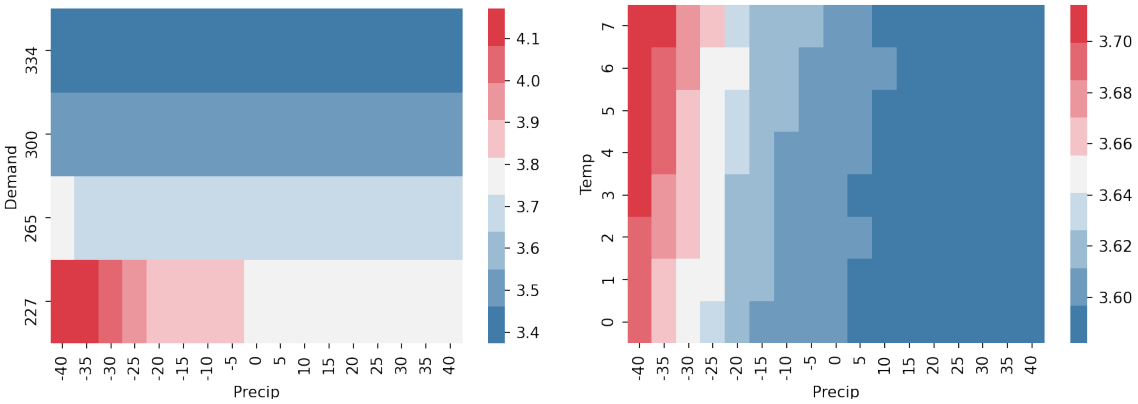


Figure 5-87. Response Surfaces Showing the Combined Effect of Demand and Precipitation (Left), and Temperature and Precipitation on the Average Price Rate Increase (% Change in Price from One Year to the Next).

The following key observation can be made from Figure 5-87:

- Influence of climate on rate of price increase: climate alone does not have a significant effect on the rate of price increase. The impact of climate on price will manifest through potential capital spend required to augment supply in the case of reduced streamflow.
- Demand will dominate price response: as discussed above, demand dominates the price response (left hand side of Figure 5-87). Counter to other performance metrics discussed in this report. Higher demand results in lower rates of price rate increase for customers.

CHAPTER 6

Prototype of Adaptation Pathways for Planning of Future Alternative Supply Option

This chapter summarizes the contribution to the LTVA from Deltares. The approach and results presented throughout this chapter are described in more details in the Technical Report by Stuparu et al. (2019).

6.1 Background and Objectives

The Dynamic Adaptive Policy Pathways (DAPP) (Haasnoot et al. 2013) approach, presented in Figure 6-1 recognizes that decisions made over time interact dynamically with the system of concern, and thus cannot be considered independently of each other. Therefore, DAPP explicitly considers the sequencing of decisions over time leading to proactive and dynamic planning in response to how the future actually unfolds. Within DAPP, alternative sequences of decisions or actions (adaptation pathways) are explored under multiple futures. These actions have uncertain design lives and may eventually fail to achieve their objectives as the conditions change (i.e., they reach an adaptation tipping point) or may not be feasibly implemented until certain conditions exist (i.e., they reach an opportunity tipping point). For example, an adaptation tipping point may occur as result of a decrease in precipitation resulting in insufficient water storage. An opportunity tipping point may occur when technological developments introduce a new technology (e.g., a new device to enable more efficient use of water) or a new regulation is put in place (e.g., direct potable reuse). DAPP supports planners in designing a dynamic adaptive plan that covers short-term actions, long-term options, and signposts to monitor to detect signals for implementation or adjustment of the plan. In the context of this study adaptation pathways were piloted for SFPUC on a simplified version of the system in order to provide examples of pathways and give recommendations for adaptation planning as a next phase of work. The scenarios developed for this pilot are not directly related to the vulnerability assessment presented in previous chapters.

Dynamic Adaptive Policy Pathways (DAPP) (Haasnoot et al. 2013) were applied on three levels for a stylized case of the SFPUC system. A stylized case is a simplified version of the real case but representative enough to provide insights in the complexities and challenges of the system.

An iterative and phased approach towards pathways analysis and implementation is proposed following three levels: 1) narrative pathways, 2) expert-based pathways and 3) model-based pathways (Figure 6-2). These three levels help to first prototype results and to specify elements of the analysis that need further attention in subsequent phases of work.

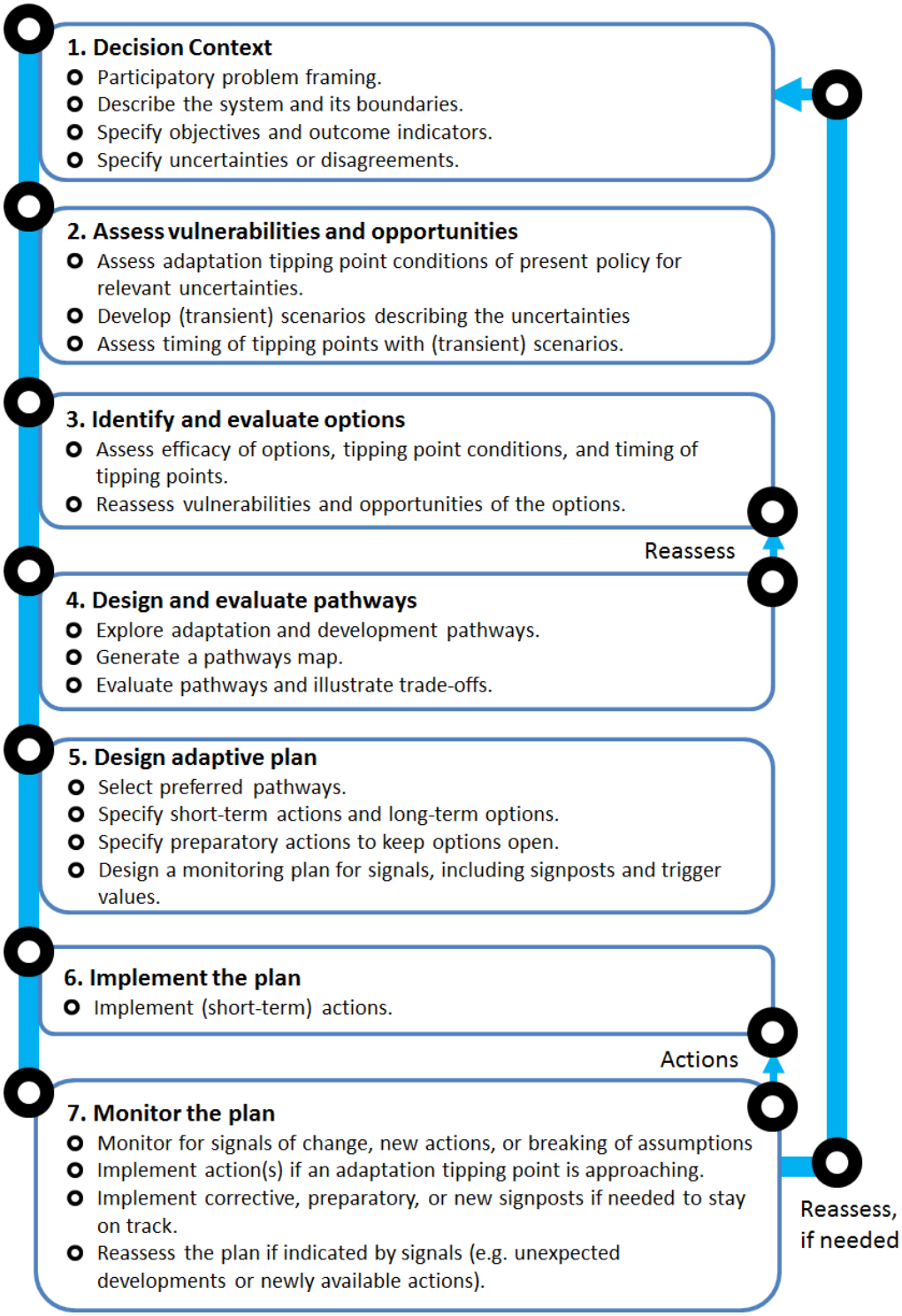


Figure 6-1. Overview of the Dynamic Adaptive Policy Pathways (DAPP) Approach.
Source: Marchau et al. 2019.

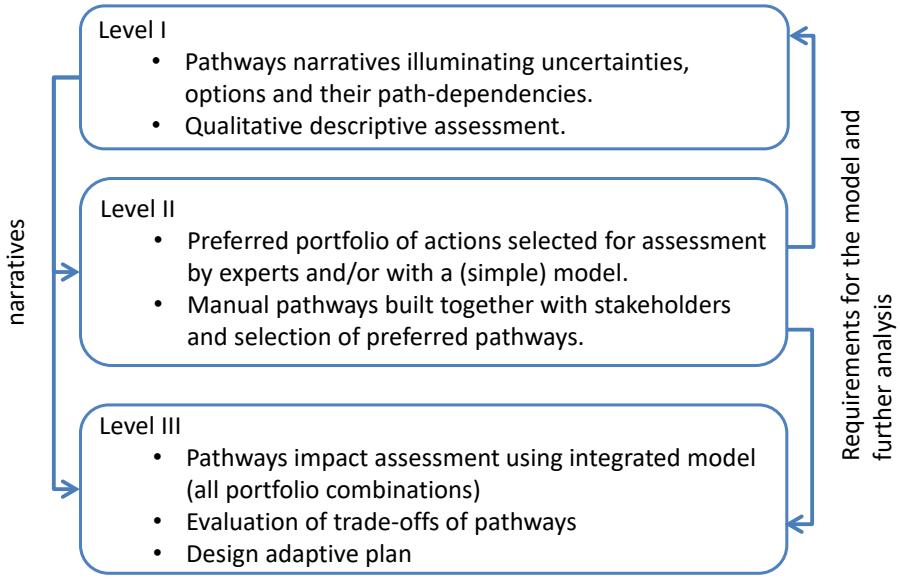


Figure 6-2. Levels of Analysis and Their Output and Purpose.

6.2 Level I: Narrative Pathways

Narratives were developed to explore scenarios and future strategies to deal with the uncertainties in decision making. Two relatively opposite scenarios, both representing an 'extreme' of possible future that cannot be influenced by SFPUC, were identified to explore the main policy relevant vulnerabilities of the system. Three uncertainties affecting system performance identified by SFPUC were considered to construct scenarios: climate change, water demand and instream flow regulations. The following presents scenarios using a narrative pathway to address the planning uncertainties.

- Narrative for scenario A - a future with significant climate change and moderate population growth and increased instream flow regulations:
  - Strategies:
    - Evaluate increased instream flows, decreased demand, water reuse, water transfers, and bulk water purchase in the short term. Implement most feasible in the long term.
    - Expansion of (above and below-ground) storage opportunities in the short term and the development of alternative water supplies (such as potable reuse and desalination) and basin transfers in the long term.
- Narrative for scenario B - a future with less significant climate change, large population growth, and only moderate focus on environmental concerns:
  - Strategies:
    - Water transfers are explored in the short term. Large scale centralized and decentralized reuse in the long term following changes to regulation to allow potable reuse.

6.3 Level II: Expert Pathways

Level II of the DAPP approach further enhances the narratives of Level I by including expert knowledge and existing information. The main source of information was the knowledge available within the project team and the water supply planning activities being implemented at the SFPUC. This information was used to develop a set of interventions and measures and combine them into strategies.

Based on the water supply work of the SFPUC, the following list of strategies and measures was developed:

- Water transfer (purchase of water from irrigation districts):
  - River diversion
  - In-basin transfer
  - Out-of basin transfer
  - Reduce water demand
- Water reuse (Assuming development of additional regulations to allow wastewater reuse to treated water augmentation standards)
  - Decentralized onsite water reuse, non-potable (by water users)
  - Centralized reuse, non-potable (by SFPUC)
  - Centralized reuse, potable (purified water) (by SFPUC)
- Desalinization
  - Regional desalinization
  - Local desalinization (large scale)

The performance of the measures was assessed in terms of their capability to reduce the water deficit. Examples using the above strategies and measures are included. Next alternative pathways were developed. First following the three strategies (Figure 6-3) and then through sequencing alternative measures (Figure 6-4).

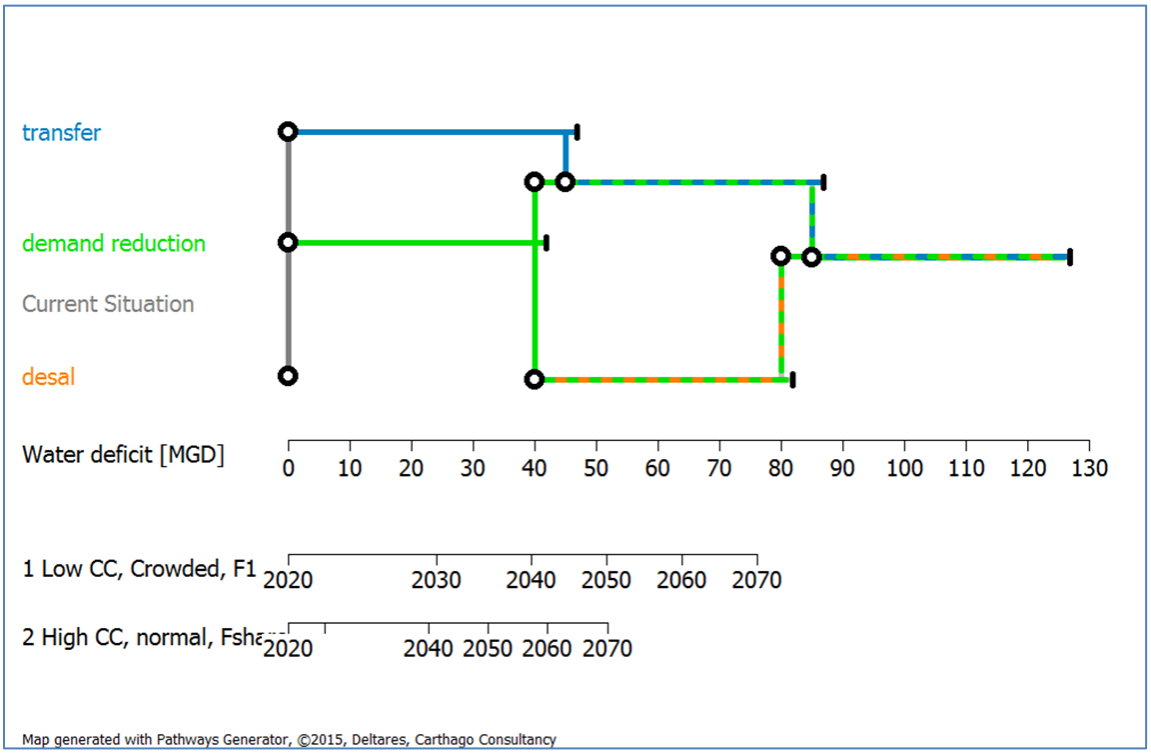


Figure 6-3. Pathways Focusing on Strategies.

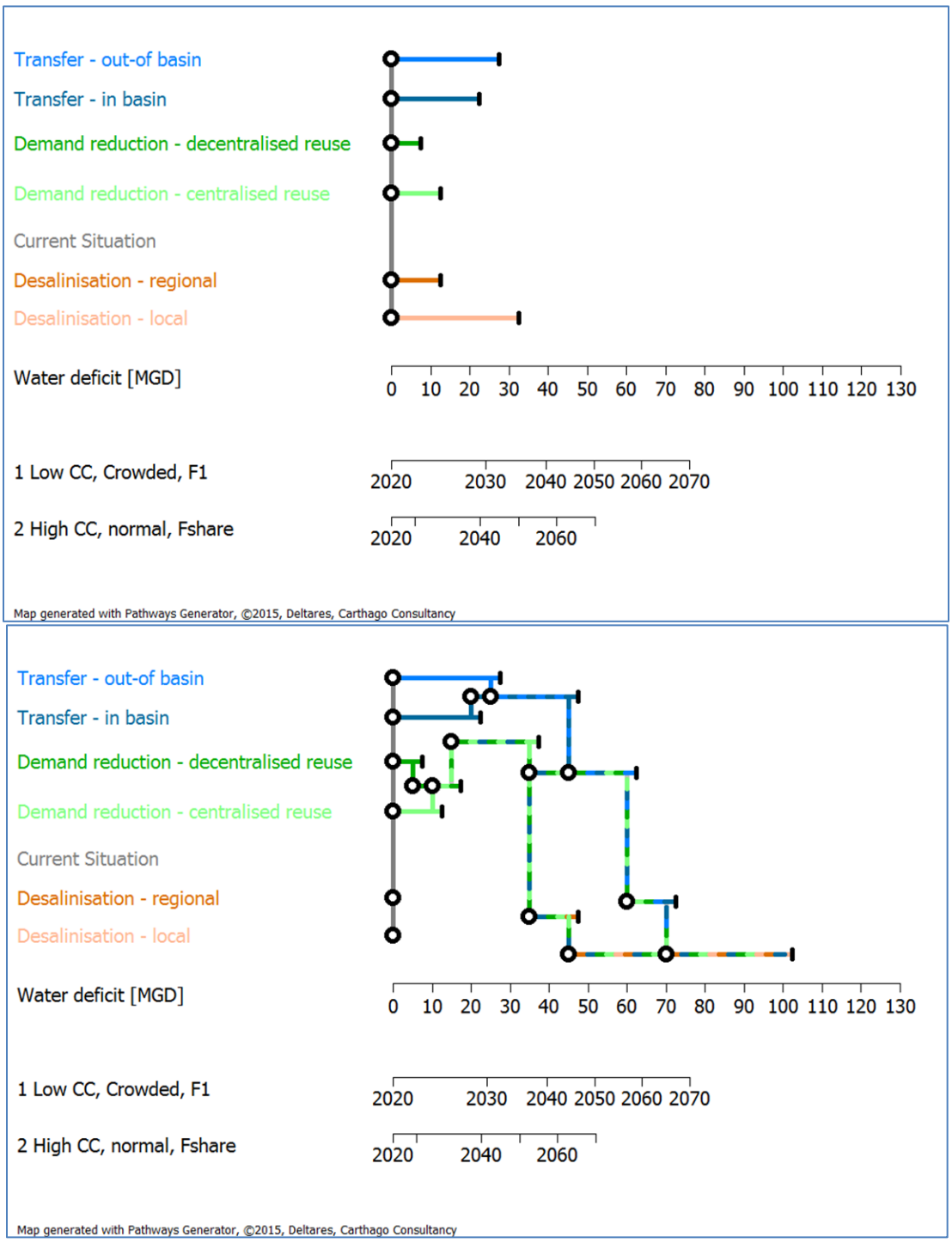


Figure 6-4. Efficiency of Individual Measures (Top) and Pathways Developed from the Water Supply Strategies Provided by the SFPUC (Bottom).

Key takeaways include:

- Scenario A: all actions are required post 2070
- Scenario B: a combination of demand reduction and water transfers is sufficient through to 2070
- One intervention alone is not enough to deal with the deficit.
- If significant climate change or high population growth is observed then preparations for the

development of alternative water supplies should be made.
- Demand reduction, water transfers, and decentralized onsite water reuse are low regret options.

6.4 Level III: Model-Based Pathways

Level III analysis involved the quantitative assessment of system vulnerability using a simplified model of the RWS. Furthermore, the model provides the means to explore uncertainty over time and identify not only which conditions may lead to vulnerability but also tipping points.

6.4.1 Modelling Framework

The core of the model focuses on the description of the water balance and includes the following processes: climate generator, hydrological models, reservoir system model, water rights, environmental water demand and urban water demand.

6.4.2 System Performance Indicators

System performance Indicators are defined to identify and capture the response of the system to changing conditions. Multiple metrics may be defined to obtain multiple perspectives in the system response to changing conditions. For this prototype study, two metrics have been used to assess the response of the SFPUC stylized case and are not related to the metrics in the vulnerability assessment presented in previous chapters:

- Total system reliability: The system reliability is the extent to which SFPUC is able to timely deliver sufficient water to its users. 100% reliability means that the system was able to deliver all requested water without the need for any demand rationing. Values lower than 100% means that demand rationing was needed. System reliability is calculated by dividing the number of times the demand rationing was not needed (therefore equal to zero) by the total number of time steps. System reliability should be higher than 90%, if this value is lower, then the level of service is not met (rationing in 1 out of 10 years on average).
- 10 years system reliability: Same definition as total system reliability but calculated over a 10-year rolling period instead of the entire period.

6.4.3 Vulnerability Assessment

6.4.3.1 Narrative for Scenario A

In a future with low climate change (+1.3°C, no change in precipitation), large population growth (300 mgd demand) and no increase in instream flow requirements

Simulations show the system reliability is equal to 59%, which is much lower than the desired 90% target. The 10-year reliability is smaller than 90% starting from the 3rd year in the future (tipping point).

Additional simulations evaluate the effect of a demand reduction of 5% to reduce the system vulnerability. This effect is significant: the total system reliability increases to 89%, and the timing of the tipping point as previously defined becomes equal to 26 years in the future. In the same way, the effect of other actions or combinations of actions can be quantified, which can then be used to update the pathways explored in level II or to explore additional pathways

6.4.3.2 Narrative for Scenario B

In a future with significant climate change (+2.7°C, 5% decrease in precipitation) and moderate population growth (270 mgd demand) and increased instream flow requirements (20,000 AF on the Tuolumne River).

Simulations show that the system reliability is equal to 91%, a little above the desired 90% target. The 10-year reliability is below the 90% target starting from the 46th year in the future (tipping point). Further simulations evaluate the effect of a transfer of 10MG to the Water Bank to reduce the system vulnerability. This effect is small compared to a reduction in demand: the total system reliability increases to 94%, and the timing of the tipping point stays equal to 46 years in the future.

6.5 Conclusions and Recommendations

The stylized case indicates that the RWS system is well equipped to deal with normal natural variation under present-day climate conditions. The system seems to be robust to many forms of climate change. However, when climatological changes are combined with changes in demand, the system is less robust.

This report describes a pilot application of adaptation pathways for long-term water system planning for the SFPUC. The objective of the study is to explore the use of the adaptation pathways approach (DAPP) as a method for formulating future management strategies for the SFPUC water supply system.

For piloting the adaptation pathways approach, a stylized example of the RWS has been developed and analyzed. The focus of the study is on the application of the DAPP method itself in order to understand the system' dynamics and form the building blocks for a follow-up more detailed analysis, and less on the validity of the resulting pathways for long term planning.

The development of the dynamic adaptation pathways followed the well-established DAPP approach, which was applied at three levels of pathway development:

- The first level is a narrative formulation of storylines, which act as a framework for further pathways exploration. It is a qualitative level, in which the emphasis lies on exploring the problem and solution space.
- A second level of pathways development is based on expert knowledge and existing information as sources for qualitative pathways development. This level results in a portfolio of possible pathways and promising actions.
- The findings of the second level are subsequently used as starting point for the third level, which consists of model-based quantitative pathways development.

These three levels reinforce one another, and a typical DAPP process can be mapped as an iterative process through these levels. All three levels were used in the SFPUC pilot study: the team developed narrative storylines for the stylized case, used these storylines for the qualitative assessment of scenario and pathways development, and further developed a simple simulation model for the water balance processes of the SFPUC water system. The simulation model was used to obtain the first results in the detailed quantitative pathways development. The models developed for the vulnerability analysis could be used to further develop pathways.

CHAPTER 7

Discussion and Outlook for Adaptation Planning

The original goal of the LTVA was to help quantitatively and qualitatively assess to what extent climate change will be a threat to the RWS in comparison to or in combination with other external drivers of change over the next 50 years (2020-2070). This goal was detailed in two more specific questions, which motivated the present study:

- Under what conditions and when will the RWS no longer meet system performance criteria?
- Is climate change the most important driver of vulnerability for the RWS and if not, what is?

This chapter presents the answers to these questions based on the results of the vulnerability assessment results presented in the previous chapter. The approach used to address these questions was to explore a range of future conditions through stress testing and specified scenarios to identify vulnerabilities and then assess the risk or level of concern associated with these vulnerabilities over the planning horizon. The external drivers of change selected for this vulnerability assessment are climate change (changes in mean annual temperature and precipitation), demand, instream flow requirements and unplanned infrastructure failures. Changes in climate and demand were explored through stress testing whereas instream flow requirements and infrastructure failures were explored through specified scenarios which were called narratives.

Overall, the analysis revealed that:

- climate change exacerbates impacts from other external drivers of change and is not the single most important driver of vulnerability for the RWS.
- the RWS at a baseline demand of 227 mgd is resilient to changes in climate and other external drivers, and
- the RWS is more vulnerable to changes in demand and instream flow requirement than changes in mean annual temperature and precipitation
- The RWS is vulnerable to changes to mean climate when demand or instream flow requirements increase

The following sections summarize the findings of this study.

7.1 Findings

7.1.1 Climate Projections for the RWS Regions

- **Warming is observed in the watersheds of the RWS**. Positive annual trends in minimum daily temperature over the Upcountry, East Bay, and Peninsula regions are observed and estimated at a warming of +1°C since the 1960s. The warming trend in East Bay region is the strongest. **No significant trends are observed for mean annual precipitation**.
- Since 85% of RWS water supply comes from the Tuolumne River, climate projections are presented in the report for the Upcountry region. Future climate is deeply uncertain, while climate projections provide an indication of signals from the latest available climate simulations. By 2040, the full range of projected changes in mean annual temperature varies between +1°C and +5.5°C and precipitation varies between -20% and +35% compared to 1986-2005 baseline. Ranges of projections for the East Bay and Peninsula regions are similar with lower extreme projections for warming but more extreme projections for drier climate.

- By 2040, in the Upcountry region, the median projections are +2°C warming and 0% change in mean annual precipitation. Most projections and expert elicitations of warming are between +1°C and +4°C and for precipitation change between -5% and +5%.
- By 2070 RCP8.5, in the Upcountry region, the median projections are about +4°C combined with 0% change in mean annual precipitation. Most projections and elicitations of warming range between +3°C and +6°C and of precipitation change between -15% and +15%.

7.1.2 Hydrology and Water Supply

- **In the Upcountry region, warming has a small effect on annual inflow volumes but affects timing of spring runoff.** This is not the case for the East Bay and Peninsula regions where warming does reduce annual inflow volumes due to evapotranspiration. Any change in mean annual precipitation would have a significant effect on annual inflows in all regions.
- To summarize the effect of warming temperature on the streamflow dynamic of the Tuolumne River at La Grange, it is noted that **+2°C warming leads to a spring runoff arrival 10 days prior** to the baseline temperature conditions, while a warming by **+4°C would lead to the spring runoff arrival up to 20 days prior** to baseline temperature conditions. The shift in timing is about the same for the inflow at Hetch Hetchy Reservoir.
- **By 2040, in the Upcountry region, the median projections do not result in significant changes to mean annual WAC volume**. Most projections and elicitations would correspond to changes in inflow between -8% and +8% at La Grange and -6% and +6% at Hetch Hetchy Reservoir. At +4°C, WAC would decrease by more than 50 TAF in 20% of the years.
- Severity and duration of WAC deficits (i.e., hydrologic droughts on the Tuolumne River) are sensitive to changes in mean annual precipitation. **A 10% decrease in precipitation causes more than doubling of the frequency of a WAC deficit with a severity similar to that experienced during the 1987-1992 drought.** The same is true for the duration of a 6-year drought. The expert elicitations concur and also revealed a tendency toward more frequent and more severe meteorological droughts.
- **By 2040, in the East Bay and Peninsula regions, the median projections lead to a decrease of mean annual flow of 9% (Arroyo Hondo) and 7% (Crystal Springs)**, respectively. Most projections and elicitations would correspond to changes in inflow between -27% and +7% at Arroyo Hondo and between -23% and +10% at Crystal Springs.
- Changes in hydrology due to climate change can affect the ability of the RWS to meet performance targets. **At a baseline demand of 227 mgd, the RWS can sustain warming of +4°C and -5% change in precipitation before failing to meet performance targets on delivery reliability, frequency of 20% rationing, storage reliability and duration of rationing.** At this level of demand, the RWS is resilient to all median climate projections for the planning horizon 2020-2070.
- Precipitation change is an important driver of change for RWS performance. **A decrease by 10% or more will significantly affect the RWS performance,** causing targets to be missed. The CMIP5 projections and expert elicitations show that such a change in precipitation is possible by 2040, although unlikely. However, the likelihood of such a change becomes more important toward 2070.
- **The RWS shows minor sensitivity to temperature change for the metrics evaluated in this study**. Most criteria stay above performance targets under warming conditions. However, warming conditions often magnify the decrease in system performance if precipitation or demand change.
- **The RWS is vulnerable to short and very severe drought, similar to the 1976-77 historical drought**. A similar, although severer, drought is included in the realization #9. Simulation results show that even under baseline climate and demand, the RWS might not be able to deliver full demand during the drought episode, even if a significant volume of water is still present in Water Bank and Cherry Reservoir. Such a drought episode could motivate changing drought response by triggering rationing

earlier than on July 1st.

7.1.3 Demand and Water Supply

Demand change appears to be a major driver of future RWS performance. The sensitivity to demand change is such that an increase by 15% (265 mgd) will lead to failing to meet the rationing frequency targets in current climate[9]. At 265 mgd demand, the rationing frequency targets would only be met if there is an increase in precipitation of 10%. If demand increases by 30%, the rationing target cannot be met even when precipitation increases by 40%, which is believed plausible but unlikely over the planning horizon.

7.1.4 Instream Flow Requirements and Water Supply

- **The RWS is particularly vulnerable to the State-amended WQCP new IFR below Don Pedro Dam** of 40% of full natural flow during February through July, which represents a significant reduction in water available for the RWS.
- **The state-amended WQCP causes a significant decrease in delivery reliability**. At a demand of 227 mgd, the water delivery reliability with the State WQCP in current climate (reliability around 85%, rationing in 1 out of 6.5 years on average) is the **equivalent of a decrease of 15% in mean annual precipitation**. The water reliability with the Alternative Proposal meets the water delivery reliability target under current climate conditions.
- Taken individually, additional IFRs considered in the East Bay and Peninsula regions will have small effects on the RWS performance in comparison to new IFRs in the Upcountry region.

7.1.5 Raw Water Quality In Hetch Hetchy Supply

- Raw Water Quality assessment tools were developed in this study but were limited by the lack of variability in the observations of historical water quality.
- General relationships between climate and water quality indicated that **turbidity is more responsive to changes in mean precipitation than to temperature, with increases causing increases in mean turbidity**.
- Similarly, TOC was generally less responsive to mean climate changes overall, and slightly more sensitive to precipitation changes than temperature changes.
- Water quality values at Tesla Portal are less sensitive to changes in mean climate than measurements at O'Shaughnessy Dam, presumably due to a dampening effect of the pipeline operations.
- In general, **raw water quality deterioration in the Hetch Hetchy water supply for turbidity and TOC as a result of mean climate changes does not appear to be a major concern**. However, changes in daily intensity of precipitation events was not directly evaluated.

7.1.6 Infrastructure Failures

- In addition to precipitation, demand, and IFR, it is also clear that the RWS is vulnerable to prolonged—albeit temporary—losses of or decreased capacity of specific infrastructure components in response to different one-time events. However, the degree to which these drivers are important cannot be generalized and depend on the specific nature of each change.
- Some infrastructure component failures represent a significant potential vulnerability to the system. In general, this assessment shows that failures related to importing water from the Upcountry are most important (e.g., a water quality event in Hetch Hetchy, an earthquake reducing SJPL capacity, or a major failure of the Moccasin Switchyard), especially when compounded with a reduction in

[9] The RWS configuration considered in this study is not based on the completed WSIP 2018.

ability to treat local water (e.g., unplanned outage of SVWTP), and/or a low local emergency storage reserve preceding the event.
- Decreases in precipitation and increases in demand exacerbate the vulnerability of the RWS to the infrastructure failures explored.
- The one Peninsula infrastructure failure narrative-based scenario examined—a loss of HTWTP capacity—indicated less vulnerability to loss of local water.
- This assessment is based only on the infrastructure component failures that the narratives explored. The system could be vulnerable in several ways to other infrastructure failures or combination of failures that were not explored in this assessment. It is also important to highlight that some outages reduce the system yield to a value lower than the daily peak demand. As such, depending on the timing of the event (i.e., high vs. low demand period), the capacity of the system to deliver full demand is jeopardize no matter what the current Bay Area reservoir storage.

7.1.7 Finance

- If major capital investment is required to add additional supply to the system as a result of climate change or IFRs, **demand would need to increase significantly to mitigate substantial increases in the price** of water.
- For example, if annual capital expenditures increased from the 2020 baseline $350M to $525M, demand would have to increase by 30% in order to maintain existing prices or else prices would rise by about 50%.
- Since demand increases also affect supply reliability, there is a **trade-off between reliability and the price** of water for customers.

7.1.8 Is Climate Change the Most Important Driver of Vulnerability for the RWS and If Not, What Is?

Climate change is not the single most important driver of vulnerability for the RWS. Under current RWS infrastructure conditions, either state-amended WQCP for additional IFR on the Tuolumne River or an increase in demand by 15% have significant impacts on the RWS performance that are equivalent to a decrease in mean annual precipitation of around 15%. Such a reduction in precipitation is outside of the median projections for both 2040 and 2070 RCP8.5 but within the range of plausible projections of both CMIP5 and expert elicitations. Temperature increases had less effect on overall performance over the range of CMIP5 projections and elicitations.

The state-amended WQCP is considered a short-term change. In terms of long-term change, the RWS is clearly most sensitive to—and vulnerable to—precipitation and demand rather than temperature. However, there is more uncertainty associated with these drivers. GCMs are inconclusive about precipitation trends (temperature trends, in contrast, are consistent). Likewise, there are multiple development and behavior scenarios that could lead to fundamentally different future demand. The RWS is also clearly vulnerable to infrastructure failure that affects the ability to deliver water from the Upcountry or East Bay.

This lack of certainty about precipitation and demand was a key driver of the methodological approach—namely, decision scaling—employed in the LTVA. This phase of the LTVA analysis quantified the potential vulnerability using systematic sampling of future climate changes, rather than being limited to the changes an ensemble of projections happened to show. Rather, GCM projections were included for bounding scenarios and state-of-the-science context.

7.2 Implications for Adaptation Planning

The questions that are relevant to a water utility regarding adaptation are: is adaptation required? Which actions should be taken? And when should action be taken? This study addressed the first question. Results indicate that action is very likely required over the planning horizon. The results of the vulnerability assessment revealed several concerns associated with climate change, as well as other drivers of change, and additional findings that likely require further investigation. In particular, the study has identified the vulnerabilities caused by decreases in precipitation, increases in demand, and changes in regulations for instream flow requirements. Together or independently, these drivers of change could cause the RWS to fail performance targets. For this reason, adaptation planning is indicated.

The next question relates to the actions that should be taken in response to these vulnerabilities. The modeling tools and framework that were used for this study provide the basis for answering this question. First, a list of possible adaptation measures should be generated from experts at SFPUC. These may include operational changes, demand management, new supply sources, new infrastructure, etc. Next, these options can be implemented in the models to enable the simulation of their effects on the system through the lenses of performance metrics. Then, the stress testing approach is used to quantify the reduction in vulnerability that each adaptation measure produces. With these results and other attributes of the measures (e.g., life-cycle cost, ease of implementation, political factors, ancillary benefits), a multi-objective trade-off analysis can be conducted to select promising measures. This provides the answer to the question of which actions should be taken.

The final question relates to the timing of adaptation measures. The choice of when to implement a measure depends on the benefit it provides and the regret associated with both implementing it when not necessary and choosing not to act when it was necessary. Thus, a consequence analysis provides the basic input to the decision, which itself builds from the benefits and costs of action estimated in the previous step of analysis. This also incorporates the reversibility of decisions and the potential from leveraging measures of early steps as in the use of real options analysis (e.g., Steinschneider and Brown 2012). Using decision scaling, the particular conditions under which a measure should be implemented can be identified as well as the associated monitoring variables (“indicators”) and “triggers” for timing of its implementation. However, the variability in typical variables that are relevant to water supply systems complicates this determination. Unlike a climate phenomenon such as sea level rise, which exhibits a strong monotonic trend, precipitation time series in particular are highly variable. This increases the required duration of a trend for it to become statistically significant. The use of consequence analysis improves this situation, because trend detection can be based on consequence significance instead of statistical significance. The use of climate projections in combination with observed trends is a promising approach to answering the question of when action should be initiated. This is an emerging research area, and the practical application of these approaches would be a landmark for any industry.

7.3 Limitations and Next Steps

This study is based on a number of assumptions, data and modeling choices that potentially lead to uncertainty in the results presented through this report. The sources of uncertainty should be known and well understood by the decision-makers when planning for future adaptation measures. Limitations are discussed at the end of all relevant sections. Those considered the most important are summarized below.

“All models are wrong, but some are useful”: this common aphorism in statistics also applies for all models developed and used for the LTVA. All developed models are useful and helped the assessment of

vulnerabilities of the RWS. However, models have limitations, and decision-makers should be aware of their consequences for the LTVA.

One important limitation is the low skill of PRMS model to reproduce the inter-annual variability of streamflow across the Upcountry region: annual streamflow variability as simulated by PRMS significantly underestimates the variability observed in the historical record, although the inter-annual precipitation variability is well reproduced by the weather generator. The consequence is that simulated droughts are likely to be less severe than they might be, and high flow years are likely to be less wet than they might be. Such a bias in variability ultimately transfers to all models using PRMS outputs. More specifically, it is possible that estimates of the return periods for the historical droughts may be overestimated (i.e., droughts may be more frequent than simulated). Therefore, it is assumed that the frequency, intensity, and sequences of demand rationing periods with climate change is underestimated, overstating the RWS water supply reliability.

Another modeling limitation is the fact that SFWSM operates reservoirs using streamflow forecasts with perfect foresight, which is a common modeling choice in water resource system analysis (e.g., François et al. 2015). However, this could be critical for the Upcountry region where real-time reservoir operations mostly rely on seasonal streamflow forecasts informed by snow-pack observations across the upstream watersheds. Reduced snowpack due to warming temperatures should reduce the skill of SFPUC operators to predict streamflow at the entrance of their reservoirs in the Upcountry region. As such, reservoir operators will have to rely more on seasonal precipitation forecasts that are commonly conceded to be less skillful than streamflow forecasts based on snowpack. This could potentially lead to situations where operations of the main Upcountry reservoirs are flawed, which could potentially impact the performance of the SFPUC system. Such a potential, although uncertain, decrease in forecast skill due to reduced snowpack is not accounted for in SFWSM, which could potentially lead to an overestimate of the system performance which in turn would underestimate the identified vulnerabilities.

Regarding water demand, SFPUC requests its customers to reduce their consumption during severe droughts. Mandatory rationing usually leads customers to adapt to such conditions. Adaptations range from change in behavior to home-improvements leading to reduced consumption (e.g., installation of flow restrictors). Analysis of water bills for years following historical droughts show that customers' demand usually does not bounce back immediately once the rationing has been cancelled by SFPUC. For instance, the current water demand for the SFPUC RWS has not bounced back yet to its level prior the 2012-2015 drought. Such a reduction in demand following rationing events is not implemented in SFWSM. Although the effect on the simulated system-wide performance is not expected to be significantly impacted, this could lead to a slight overestimate of the system vulnerability as demand would be higher than it should, which is a factor of vulnerability. In addition, the necessary increases in the price of water during and after drought (given CAPEX and OPEX costs must be covered by reduced sales of water) may cause demand for water to reduce still further in response – the so called 'death spiral.' In order to explore these relationships further, the finance, demand and system models need to be fully coupled to allow dynamic feedback between them. Accounting for this feedback effect between water price and demand requires coupling SFWSM with the finance model, which has not been done.

Regarding water quality, the reduced number of high turbidity and TOC events in the historical records makes it difficult for data-driven approaches to predict extreme values. The UMass team overcame this difficulty by coupling a neural network with a quantile regression approach that allows prediction of high quantiles of the distribution that are usually not observed. Although this modeling is a success, the

interpretation of the results is more difficult, especially from stakeholders with limited knowledge in statistics.

The assessment of the vulnerability regarding short term shocks deserves more attention. More specifically, a small number of infrastructure failure narrative-based scenarios (i.e., 5) have been investigated through the LTVA. While the selected narratives have been carefully selected by SFPUC personnel, these five narratives represent only a small sample of the many infrastructure failures that could occur across the system. Further assessment of the vulnerability needs to be linked with other ongoing works of SFPUC on this topic. In addition, the LTVA highlighted that the time at which these failures occur has a significant influence on the capacity of the system to cope with such short-term shocks, which also deserves more attention. Finally, the vulnerability to compound infrastructure failures have not been investigated.

The effect of change in instream flow requirements (IFRs) have been investigated below each major reservoir of the SFPUC RWS. Besides the obvious vulnerability of the system to the IFR proposals downstream of Don Pedro reservoir, the vulnerability to change in IFRs is not significant. However, combination of changes in IFRs have not been investigated.

The results presented in this document indicate that climate change will significantly impact RWS water supply performance, however, due to the shortcomings of the models, these results are likely understating the impact of climate change on the RWS. Next steps will include improvements on these models to address the bias in the current versions.

References

Albano, C.M., M.I. McCarthy, M.D. Dettinger, and S.A. McAfee. 2021. “Techniques For Constructing Climate Scenarios for Stress Test Applications.” *Clim. Change*, 164. https://doi.org/10.1007/s10584-021-02985-6.

Allen, D., S. Arthur, H. Haynes, and V. Olive. 2017. “Multiple Rainfall Event Pollution Transport by Sustainable Drainage Systems: The Fate of Fine Sediment Pollution.” *Int. J. Environ. Sci. Technol*., 14: 639–652. https://doi.org/10.1007/s13762-016-1177-y.

Andrews, E.D., R.C. Antweiler, P.J. Neiman, and F.M. Ralph. 2004. "Influence of ENSO on Flood Frequency along the California Coast.” *J. Clim*. 17: 337–348.

Bamber, J.L., M. Oppenheimer, R.E. Kopp, W.P. Aspinall, and R.M. Cooke. 2019. “Ice Sheet Contributions to Future Sea-Level Rise from Structured Expert Judgment.” *Proc. Natl. Acad. Sci*., 116: 11195–11200. https://doi.org/10.1073/pnas.1817205116.

Barnett, T.P., D.W. Pierce, H.G. Hidalgo, C. Bonfils, B.D. Santer, T. Das, G. Bala, A.W. Wood, T. Nozawa, and A.A. Mirin. 2008. “Human-Induced Changes in the Hydrology of the Western United States.” *Science,* 319 (5866): 1080–1083.

BAWSCA (Bay Area Water Supply and Conservation Agency). 2021. “Annual Survey.” Accessed October 26, 2021. https://bawsca.org/water/supply/survey.

Behnke, R. 2011. The Role of Elevation on Temperature Trends in the Western United States: A Comparison of Two Statistical Methods. Master Thesis, University of Wisconsin-Madison. http://www.aos.wisc.edu/uwaosjournal/Volume16/Behnke_MS_Thesis.pdf.

Bonaccorso, B., A. Cancelliere, and G. Rossi. 2003. “An Analytical Formulation of Return Period of Drought Severity.” *Stochastic Environmental Research and Risk Assessment*, 17: 157–174. https://doi.org/10.1007/s00477-003-0127-7.

Bonfils, C., P.B. Duffy, B.D. Santer, T.M.L. Wigley, D.B. Lobell, T.J. Phillips, and C. Doutriaux. 2008. “Identification of External Influences on Temperatures in California.” *Climatic Change*, 87 (S1): 43–55. https://doi.org/10.1007/s10584-007-9374-9.

Brattle Group. 2018. *Socioeconomic Impacts of Water Shortages within the Hetch Hetchy Regional Water System Service Area*. The Brattle Group.

Brown, C., and A. Dufour. 2017. *Long-Term Vulnerability Assessment and Adaptation Planning for the San Francisco Public Utilities Commission Water Enterprise – Technical Memorandum #1 – Identification of Long Term Sources of Vulnerability*. n.p.

Brown, C., Y. Ghile, M. Laverty, and K. Li. 2012. “Decision Scaling: Linking Bottom-Up Vulnerability Analysis with Climate Projections in the Water Sector.” *Water Resour. Res*., 48 (W09537). https://doi.org/10.1029/2011WR011212.

Brown, C., and R.L. Wilby. 2012. “An Alternate Approach to Assessing Climate Risks.” *Eos, Transactions American Geophysical Union*, 93 (41): 401–2. https://doi.org/10.1029/2012EO410001.

Burnash, J.C.R., R.L. Ferrai, and R.A. McGuire. 1973. *A Generalized Streamflow Simulation System*. NWS and California Dept. of Water Res., Sacramento, California.

CA DWR (California Department of Water Resources). 2015. "California Climate Science and Data - For Water Resources Management." Sacramento, CA. https://cawaterlibrary.net/document/california-climate-science-and-data-for-water-resources-management/.

CA DWR (California Department of Water Resources). 2019. *Decision Scaling Evaluation of Climate Change Driven Hydrologic Risk to the State Water Project, Final Report*. California Department of Water Resources, Sacramento, CA https://water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/All-Programs/Climate-Change-Program/Climate-Action-Plan/Files/CAP-III-Decision-Scaling-Vulnerability-Assessment.pdf.

Cannon, A.J. 2018. "Non-crossing Nonlinear Regression Quantiles by Monotone Composite Quantile Regression Neural Network, with Application to Rainfall Extremes." *Stoch. Environ. Res. Risk Assess.*, 32 : 3207–3225. https://doi.org/10.1007/s00477-018-1573-6.

Cardil, A., M. Rodrigues, J. Ramirez, S. de-Miguel, C.A. Silva, M. Mariani, and D. Ascoli. 2020. "Coupled Effects of Climate Teleconnections on Drought, Santa Ana Winds and Wildfires in Southern California." *Sci. Total Environ*., 142788. https://doi.org/10.1016/j.scitotenv.2020.142788.

Cha, Y., M.-H. Park, S.-H. Lee, J.H. Kim, and K.H. Cho. 2016. "Modeling Spatiotemporal Bacterial Variability with Meteorological and Watershed Land-Use Characteristics." *Water Res*., 100: 306–315. https://doi.org/10.1016/j.watres.2016.05.024.

Clayton, D.G. 1978. "A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence." *Biometrika*, 65 (1): 141–151. https://doi.org/10.1093/biomet/65.1.141.

Cole, T.M., and S.A. Wells. 2002. *CE-QUAL-W2: A Two Dimensional, Laterally Averaged, Hydrodynamic and Water Quality Model, Version 3.1*. Instruction Report EL-2002-1. U.S. Army Engineering and Research Development Center, Vicksburg, Miss.

Cunningham, P., J. Carney, and S. Jacob. 2000. "Stability Problems with Artificial Neural Networks and the Ensemble Solution." *Artificial Intelligence in Medicine,* 20 (3): 217–25. https://doi.org/10.1016/S0933-3657(00)00065-8.

Deser, C., R. Knutti, S. Solomon, and A.S. Phillips. 2012. "Communication of the Role of Natural Variability in Future North American Climate." *Nature Climate Change*, 2 (11): 775–79. https://doi.org/10.1038/nclimate1562.

Dettinger, M.D., D.R. Cayan, H.F. Diaz, and D.M. Meko. 1998. "North–South Precipitation Patterns in Western North America on Interannual-to-Decadal Timescales." *Journal of Climate*, 11 (12): 3095–3111.

Diffenbaugh, N.S., D.L. Swain, and D. Touma. 2015. "Anthropogenic Warming Has Increased Drought Risk in California." *Proc. Natl. Acad. Sci*., 112: 3931–3936. https://doi.org/10.1073/pnas.1422385112.

Efbrazil. 2020. "All Forcing Agents' Atmospheric CO2-Equivalent Concentrations (in Parts-per-Million-by-Volume (ppmv)) According to the Four RCPs Used by the Fifth IPCC Assessment Report to Make Predictions." Wikimedia Commons. https://commons.wikimedia.org/w/index.php?curid=87801257.

Flato, G., J. Marotzke, B. Abiodun, P. Braconnot, S.C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason and M. Rummukainen. 2013. "Evaluation of Climate Models." In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY.

François, B., B. Hingray, J.D. Creutin, and F. Hendrickx. 2015. "Estimating Water System Performance Under Climate Change: Influence of the Management Strategy Modeling." *Water Resources Management,* 29 (13): 4903–18. https://doi.org/10.1007/s11269-015-1097-5.

Genest, C., and A.-C. Favre. 2007. "Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask." *J. Hydrol. Eng*., 12: 347–368. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347).

Gräler, B., M.J. van den Berg, S. Vandenberghe, A. Petroselli, S. Grimaldi, B. De Baets, and N.E.C. Verhoest. 2013. "Multivariate Return Periods in Hydrology: A Critical and Practical Review Focusing on Synthetic Design Hydrograph Estimation." *Hydrol. Earth Syst. Sci*., 17: 1281–1296. https://doi.org/10.5194/hess-17-1281-2013.

Gumbel, É J. 1960. "Distributions des Valeurs Extrêmes en Plusieurs Dimensions." *Publ. Inst. Stat. Univ. Paris*, 9: 171–173.

Gupta, H.V., H. Kling, K.K. Yilmaz, and G.F. Martinez. 2009. "Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications for Improving Hydrological Modelling." *Journal of Hydrology*, 377 (1–2): 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Haasnoot, M., J.H. Kwakkel, W.E. Walker, and J. ter Maat. 2013. "Dynamic Adaptive Policy Pathways: A Method for Crafting Robust Decisions for a Deeply Uncertain World." *Global Environmental Change,* 23 (2): 485–98. https://doi.org/10.1016/j.gloenvcha.2012.12.006.

Hashimoto, T., J.R. Stedinger, and D.P. Loucks. 1982. "Reliability, Resiliency, and Vulnerability Criteria for Water Resource System Performance Evaluation." *Water Resources Research* 18 (1): 14. https://doi.org/10.1029/WR018i001p00014.

HRG (Hydrosystems Research Group). 2018. *Technical Report 1: Weather Generator Module*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.

HRG (Hydrosystems Research Group). 2021a. *Technical Report 2: Hydrologic Modeling Module*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.

HRG (Hydrosystems Research Group). 2020a. *Technical Report 3: Urban Water Demand*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.

HRG (Hydrosystems Research Group). 2021b. *Technical Report 4: San Francisco Water System Model*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.

HRG (Hydrosystems Research Group). 2021c. *Technical Report 5: Raw Water Quality Model*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.

HRG (Hydrosystems Research Group). 2020b. *Technical Report 6: Finance Module*. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.

Hydrocomp, San Francisco Public Utilities Commission, and Turlock Irrigation District. 2012. *Sensitivity of Upper Tuolumne River Flow to Climate Change Scenarios*. SFPUC.

IPCC (Intergovernmental Panel on Climate Change). n.d. "AR5 GCM Data." Accessed October 26, 2021. http://www.ipcc-data.org/sim/gcm_monthly/AR5/.

Kiefer, J., B. Dziegielewski, and J. Henderson. 2013. *Changes in Water Use Under Regional Climate Change Scenarios*. Water Research Foundation.

Killam, D., A. Bui, S. LaDochy, P. Ramirez, J. Willis, and W. Patzert. 2014. "California Getting Wetter to the North, Drier to the South: Natural Variability or Climate Change?" *Climate*, 2 (3): 168–80. https://doi.org/10.3390/cli2030168.

Kjeldsen, T., and D. Rosbjerg,. 2004. "Choice of Reliability, Resilience and Vulnerability Estimators for Risk Assessments of Water Resources Systems." *Hydrological Sciences Journal*, 5 (49): 755–67.

Kwon, H.-H., U. Lall, and A.F. Khalil. 2007. "Stochastic Simulation Model for Nonstationary Time Series Using an Autoregressive Wavelet Decomposition: Applications to Rainfall and Temperature: Regional Calibration of Catchment Models." *Water Resources Research,* 43 (5). https://doi.org/10.1029/2006WR005258.

Lempert, R., S. Berry, and S. Tanverakul. 2019. *Climate Information Workshop Goals, Process and Results*. The RAND Corporation.

Livneh, B., T.J. Bohn, D.W. Pierce, F. Munoz-Arriola, B. Nijssen, R. Vose, D.R. Cayan, and L. Brekke. 2015. "A Spatially Comprehensive, Hydrometeorological Data Set for Mexico, the U.S., and Southern Canada 1950–2013." *Scientific Data*, 2: 150042. https://doi.org/10.1038/sdata.2015.42.

Maidment, D.R., S.-P. Miaou, And M.M. Crawford. 1985. "Transfer Function Models of Daily Urban Water Use." *Water Resources Research*, 21: 425–432. https://doi.org/10.1029/WR021i004p00425.

Maidment, D.R., and S.P. Miaou. 1986. "Daily Water Use in Nine Cities." *Water Resources Research*, 22 (6): 845–51.

Marchau, V.A.W.J., W.E. Walker, P.J.T.M. Bloemen, and S.W. Popper (Eds.). 2019. *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing.

Markstrom, S.L., R.S. Regan, L.E., Hay, R.J. Viger, R.M.T. Webb, R.A. Payn, and J.H. LaFontaine. 2015. *PRMS-IV, the Precipitation-Runoff Modeling System, Version 4*. U.S. Geological Survey Techniques and Methods, book 6.

Meko, D.M., C.A. Woodhouse, and R. Touchan. 2014. *Klamath/San Joaquin/Sacramento Hydroclimatic Reconstructions from Tree Rings*. Draft Report to the California Department of Water Resources; University of Arizona: Tucson, AZ.

Moss, R.H., J.A. Edmonds, K.A. Hibbard, M.R. Manning, S.K. Rose, D.P. van Vuuren, T.R. Carter, S. Emori, M. Kainuma, T. Kram, G.A. Meehl, J.F.B. Mitchell, N. Nakicenovic, K. Riahi, S.J. Smith, R.J. Stouffer, A.M. Thomson, J.P. Weyant, and T.J. Wilbanks. 2010. "The Next Generation of Scenarios for Climate Change Research and Assessment." *Nature*, 463: 747–756. https://doi.org/10.1038/nature08823.

Mukundan, R., D.C. Pierson, L. Wang, A.H. Matonse, N.R. Samal, M.S. Zion, and E.M. Schneiderman. 2013. "Effect of Projected Changes in Winter Streamflow on Stream Turbidity, Esopus Creek Watershed in New York, USA." *Hydrological Processes* 27 (21): 3014–23. https://doi.org/10.1002/hyp.9824.

Nash, J.E., and J.V. Sutcliffe. 1970. "River Flow Forecasting through Conceptual Models Part I — A Discussion of Principles." Journal of Hydrology 10 (3): 282–90. https://doi.org/10.1016/0022-1694(70)90255-6.

NCAR (National Center of Atmospheric Research). 2018. *Technical Report: Climate Change Storyline.* National Center of Atmospheric Research, Boulder, CO.

ONESF. n.d. "Water System Improvement Program." Accessed October 26, 2021. https://onesanfrancisco.org/resiliency/water-system-improvement-program.

Parsons (Parsons Corporation). 2006. *WSIP System Assessment for Levels of Service Objectives*. SFPUC.

Poff, N.L., C.M. Brown, T.E. Grantham, J.H. Matthews, M.A. Palmer, C.M. Spence, R.L. Wilby, M. Haasnoot, G.F. Mendoza, K.C. Dominique, and A. Baeza. 2015. "Sustainable Water Management under Future Uncertainty with Eco-engineering Decision Scaling." *Nat. Clim. Change*, 6: 25–34. https://doi.org/10.1038/nclimate2765.

Rossi, N., L. DeCristofaro, S. Steinschneider, C. Brown, and R. Palmer. 2016. "Potential Impacts of Changes in Climate on Turbidity in New York City's Ashokan Reservoir." *Journal of Water Resources Planning and Management*, 142 (3): 04015066. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000614.

Samal, N.R., A.H. Matonse, R. Mukundan, M.S. Zion, D.C. Pierson, R.K. Gelda, and E.M. Schneiderman. 2013. "Modelling Potential Effects of Climate Change on Winter Turbidity Loading in the Ashokan Reservoir, NY." *Hydrological Processes*, 27 (21): 3061–74. https://doi.org/10.1002/hyp.9910.

SFPUC (San Francisco Public Utilities Commission). 2016. *2015 Urban Water Management Plan for the City and County of San Francisco*. San Francisco Public Utilities Commission.

SFPUC (San Francisco Public Utilities Commission). 2009. *Water Supply Agreement between the City of San Francisco and Wholesale Customers in Alameda County, San Mateo County and Santa Clara County*. SFPUC.

SFPUC (San Francisco Public Utilities Commission). 2014. *The Upper Tuolumne River Ecosystem Program. O'Shaughnessy Dam Instream Flow Management Plan*. San Francisco Public Utilities Commission, San Francisco, CA https://uppertuolumneriver.wordpress.com/library/draft-oshaughnessy-dam-instream-flow-management-plan/.

Shukla, S., A. Steinemann, S.F. Iacobellis, and D.R. Cayan. 2015. "Annual Drought in California: Association with Monthly Precipitation and Climate Phases." *J. Appl. Meteorol. Climatol*., 54: 2273–2281.

Silva A.T., and M.M. Portela. 2012. "Disaggregation Modelling of Monthly Streamflows Using a New Approach of the Method of Fragments." *Hydrol Sci J*, 57: 942–955. doi: 10.1080/02626667.2012.686695.

Sklar, A. 1959. “Fonctions de Répartition à n Dimensions et Leurs Marges.” *Publ. Inst. Stat. Univ. Paris*, 8 : 229–231.

Stahl, K., J.-P. Vidal, J. Hannaford, E. Tijdeman, G. Laaha, T. Gauster, and L.M. Tallaksen. 2020. "The Challenges of Hydrological Drought Definition, Quantification and Communication: An Interdisciplinary Perspective.” *Proc. Int. Assoc. Hydrol. Sci*., 383: 291–295. https://doi.org/10.5194/piahs-383-291-2020.

Steinschneider, S., and C. Brown. 2012. “Dynamic Reservoir Management with Real-Option Risk Hedging as a Robust Adaptation to Nonstationary Climate.” *Water Resour. Res*., 48. https://doi.org/10.1029/2011WR011540.

Steinschneider, S., and C. Brown. 2013. “A Semiparametric Multivariate, Multisite Weather Generator with Low-Frequency Variability for Use in Climate Risk Assessments: Weather Generator for Climate Risk.” *Water Resources Research*, 49 (11): 7205–20. https://doi.org/10.1002/wrcr.20528.

Stocker, T.F., D. Qin, G.-K. Plattner, L.V. Alexander, S.K. Allen, N.L. Bindoff, F.-M. Bréon, J.A. Church, U. Cubasch, S. Emori, P. Forster, P. Friedlingstein, N. Gillett, J.M. Gregory, D.L. Hartmann, E. Jansen, B. Kirtman, R. Knutti, K. Krishna Kumar, P. Lemke, J. Marotzke, V. Masson-Delmotte, G.A. Meehl, I.I. Mokhov, S. Piao, V. Ramaswamy, D. Randall, M. Rhein, M. Rojas, C. Sabine, D. Shindell, L.D. Talley, D.G. Vaughan and S.-P. Xie. 2013. “Technical Summary.” In *Climate change 2013: The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Swain, D.L., B. Langenbrunner, J.D. Neelin, and A. Hall. 2018. “Increasing Precipitation Volatility in Twenty-First-Century California.” *Nat. Clim. Change*, 8: 427–433. https://doi.org/10.1038/s41558-018-0140-y.

Stuparu D., A. Dufour, W. van Deursen, and M. Haasnoot. 2019. *Piloting Adaptation Pathways for San Francisco Water*. Deltares

Taylor, K.E., R.J. Stouffer, and G.A. Meehl. 2012. “An Overview of CMIP5 and the Experiment Design.” *Bulletin of the American Meteorological Society*, 93 (4): 485–98. https://doi.org/10.1175/BAMS-D-11-00094.1.

Tebaldi, C., J.M. Arblaster, and R. Knutti. 2011. “Mapping Model Agreement on Future Climate Projections.” *Geophysical Research Letters*, 38 (23).

Tomlinson, J.E., J.H. Arnott, and J.J. Harou. 2020. “A Water Resource Simulator in Python.” *Environmental Modelling & Software*. https://doi.org/10.1016/j.envsoft.2020.104635.

Torrence, C., and G.P. Compo. 1998. “A Practical Guide to Wavelet Analysis.” *Bull. Am. Meteorol. Soc*., 79: 61–78.

van Vuuren, D.P., J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G.C. Hurtt, T. Kram, V. Krey, J.-F. Lamarque, T. Masui, M. Meinshausen, N. Nakicenovic, S.J. Smith, and S.K. Rose. 2011. “The Representative Concentration Pathways: An Overview.” *Clim. Change*, 109: 5–31. https://doi.org/10.1007/s10584-011-0148-z.

Vicuna, S., E.P. Maurer, B. Joyce, J.A. Dracup, and D. Purkey. 2007. “The Sensitivity of California Water Resources to Climate Change Scenarios.” *J. Am. Water Resour. Assoc.*, 43: 482–498. https://doi.org/10.1111/j.1752-1688.2007.00038.x.

Whateley, S., and C. Brown. 2016. “Assessing the Relative Effects of Emissions, Climate Means, and Variability on Large Water Supply Systems: Climate Uncertainty and Water Supply.” *Geophysical Research Letters*, 43 (21): 11,329-11,338. https://doi.org/10.1002/2016GL070241.

Whateley, S., S. Steinschneider, and C. Brown. 2016. “Selecting Stochastic Climate Realizations to Efficiently Explore a Wide Range of Climate Risk to Water Resource Systems.” *J. Water Resour. Plan. Manag.*, 142: 6016002. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000631.

Wise, R.M., I. Fazey, M.S. Smith, S.E. Park, H.C. Eakin, E.R.M.A. Van Garderen, and B. Campbell. 2014. “Reconceptualising Adaptation to Climate Change as Part of Pathways of Change and Response.” *Global Environmental Change*, 28 (September): 325–36. https://doi.org/10.1016/j.gloenvcha.2013.12.002.

Yevjevich, V. 1967. *Objective Approach to Definitions and Investigations of Continental Hydrologic Droughts; Hydrology Paper 23*. Colorado State University: Fort Collins, CO. https://mountainscholar.org/bitstream/handle/10217/61303/HydrologyPapers_n23.pdf?sequence=1.

THE
Water
Research
FOUNDATION

advancing the science of water®

1199 North Fairfax Street, Suite 900
Alexandria, VA 22314-1445

6666 West Quincy Avenue
Denver, CO 80235-3098

www.waterrf.org | info@waterrf.org