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

Technical Report 4: San Francisco Water System Model [FINAL]

Prepared for:

San Francisco Public Utilities Commission

Prepared by:

University of Massachusetts Amherst, Hydrosystems Research Group

May, 2021

Summary

This Technical Report describes the methods used to develop a simulation model of San Francisco's water supply system built for the Long-Term Vulnerability Assessment and Adaptation Planning for the San Francisco Public Utilities Commission (SFPUC) Water Enterprise (Vulnerability Assessment). The model, called the San Francisco Water System Model (SFWSM), consists of scripts and data designed to simulate the operations of the SFPUC water system under a range of environmental, policy, and other conditions. SFWSM simulates operations at the daily timestep using linear programming-based simulation. It includes all major infrastructure between and including reservoirs and demand centers. Demand centers include all wholesale, retail customers in San Francisco and large suburban retail customers of SFPUC, including each of the 27 Bay Area Water Service Contract Association (BAWSCA) customers and smaller suburban retail customers. Agricultural demand for Turlock and Modesto Irrigation Districts (TID and MID) through the canals is included in the model. Modelled operations include reservoirs, water deliveries, instream flow requirements, hydropower, water treatment quantity, groundwater, and numerous supporting programs and projects (the Water Bank at Don Pedro (Water Bank), the Upper Tuolumne River Ecological Program, the Alameda Creek Recapture Project, etc.). Water quality is not currently represented. This report describes modeling goals, modeling scope, generalized modeling framework, included system infrastructure, operations simulation methods, and model validation. It also identifies limitations as needed.

Table of Contents

1 Introduction and overview

1.1 Background and motivation

The San Francisco Public Utility Commission's (SFPUC) operates the Hetch Hetchy Regional Water System (RWS, Figure 1) to deliver fresh water to over 2.7 million people within the counties of San Francisco, San Mateo, Santa Clara, Alameda, San Joaquin and Tuolumne. Climate change and other changing conditions may jeopardize the future ability of the RWS to meet SFPUC's desired water supply performance targets. To help address this concern, SFPUC partnered with the University of Massachusetts Amherst to help identify key vulnerabilities of the RWS to long term change in climate and other conditions, under a project called the "Long-Term Vulnerability Assessment and Adaptation Planning for the SFPUC Water Enterprise" (LTVA or Vulnerability Assessment). The objective of the LVTA is, in part, to design and execute an exhaustive vulnerability assessment that provides a comprehensive understanding of the expected water system performance relative to goals and expectations of the system. The general approach to achieve this objective, as outlined in the LTVA Detailed Analytical Plan, is to develop a suite of interconnected computer models and supporting analytical modules representing important processes involved in the long-term planning of RWS and to then use them along with a scenario discovery approach to quantitatively assess system vulnerability. One of these models is a simulation model of the RWS that can mimic the response of the physical storage, distribution and treatment system to operational drivers and constraints, such as inflow hydrology, water demand, operational policy, and so on. Such a simulation model can then be used to help assess the long-term vulnerability of the SFPUC Water Enterprise to long term changes in climate conditions, water demand, and policy drivers with respect to water system operations.

SFPUC currently uses a simulation model of the RWS called the Hetch Hetchy / Local Simulation Model (HH/LSM or HHLSM), developed originally with Fortran and later converted to Excel/VBA (SFPUC, 2009a). HHLSM uses a monthly time step with five delivery regions. The monthly time step of HHLSM model limits the ability to 1) represent the current operations of the different reservoirs that comprise the SFPUC system (e.g. timing and volume of uncontrolled spill) and 2) evaluate the effects of climate change (e.g., shifts in the timing of runoff anticipated with a warmer temperature) on the system performance. The coarse spatial scope also prevents evaluating the effect of changes on specific wholesale customers. The motivation of the new model described here is to resolve these constraints, in addition to generally updating the modeling approach for use in the LTVA.

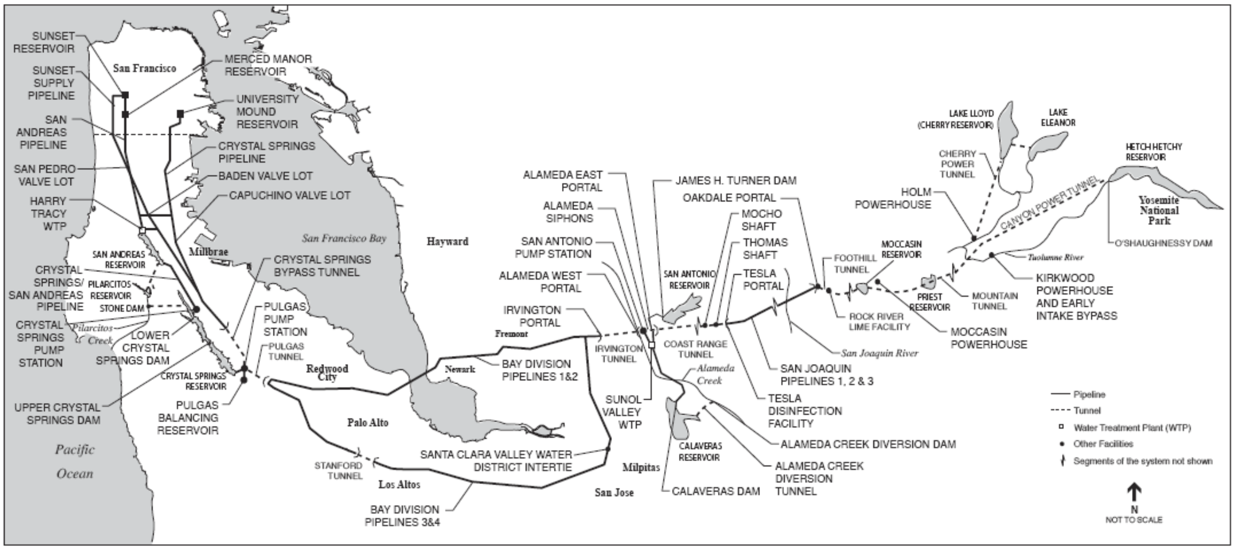

Figure 1. Schematic of the SFPUC Hetch Hetchy Regional Water System.

1.2 Modeling goals and objectives

Generally, the broader goals of the water system model are to:

- Support the goals of the Vulnerability Assessment
- Simulate (represent) system operations under different climate, demand and operational assumptions
- Have a transparent, readable code base to facilitate transfer to SFPUC personnel for further improvements/modifications
- Be reasonably fast to account for a large number of scenarios of climate, demand, regulations and infrastructure outage events.

To inform the Vulnerability Assessment, the main purpose of the water system model is to evaluate the impact on water delivery reliability and other metrics with variations in hydrologic conditions, level of water demand, system configuration (shutdowns, outages, new projects, etc.) and various management policies with greater detail and accuracy than is possible in the existing HHLSM model.

The water system model scope and approach are described generally here, with a more detailed description below.

1.3 Water system model scope

The water system model described here is further denoted as the San Francisco Water System Model (SFWSM). It includes a more detailed representation of system facilities than HHLSM and operates at the daily time step. Table 1 lists key differences between HHLSM and SFWSM. Each of SFPUC's 26 wholesale customers and the City of San Francisco are included, for a total of 27 primary delivery points. Large suburban retail customers are also included. Most important system facilities are included (e.g., reservoirs, aqueducts, treatment plants, pump stations, etc.), with some aggregation/simplification as needed and able to reduce both computation time and model development time. For example, this aggregation applies to several parallel pipelines or tunnels such as for San Joaquin Pipelines, the Irvington Tunnels or the Bay Division Pipelines. The daily time step was selected based on both the need for greater detail, and the relatively fast speed of the modeling framework, described below.

Aside from these above components, the following specific operational aspects were identified as important to include in the system model's representation of infrastructure and operations:

- Urban water demand (retail and wholesale customers)
- Agricultural canal diversions to Modesto and Turlock Irrigation Districts.
- Hydrology
- Reservoir operations
- Instream flow requirements and Upper Tuolumne River Ecological Program snowmelt spill management
- Drought policy
- Specific supporting programs and projects, including:
    - Water Bank
    - Alameda Creek Recapture Project
    - Regional Groundwater Storage and Recovery Project
    - San Francisco Groundwater Project
- Streamflow forecast to operate hydropower generation and minimize spills

Infrastructure includes Don Pedro Reservoir—not strictly part of the RWS—to model canal deliveries to Modesto and Turlock Irrigation Districts, who have seniority water rights compared to SFPUC for water from the Tuolumne River, and for compliance with flow requirements downstream of Don Pedro.

Important operational policies related to water quantity are included in SFWSM (e.g., "water first, power second"; emergency storage for South East Bay (SEB) and Peninsula (PEN) regions; filling of San Antonio reservoir prior San Joaquin pipelines shut down for maintenance), as appropriate for the spatial and temporal scope to realistically represent system operations. However, HHLSM includes some optional operations that may not be of importance for the LTVA (i.e., Lower Cherry Aqueduct and treatment of Hetch Hetchy water at SVWTP[1]); the inclusion of these operations could be added to SFWSM if needed in the future.

Inputs to and outputs from SFWSM are summarized in Figure 2. Inputs to the system model include a range of parameters related to initial conditions (reservoir storage, groundwater storage,

[1] Treatment of Hetch Hetchy water at SVWTP occurs when water quality at Alameda East Portal does not meet filtration avoidance specifications. This operation is not included in validation or base case models but is added for studying infrastructure outage events.

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 2 are time series. Aggregated outputs, such as performance metrics and policy violation checks, can be calculated from these model outputs.

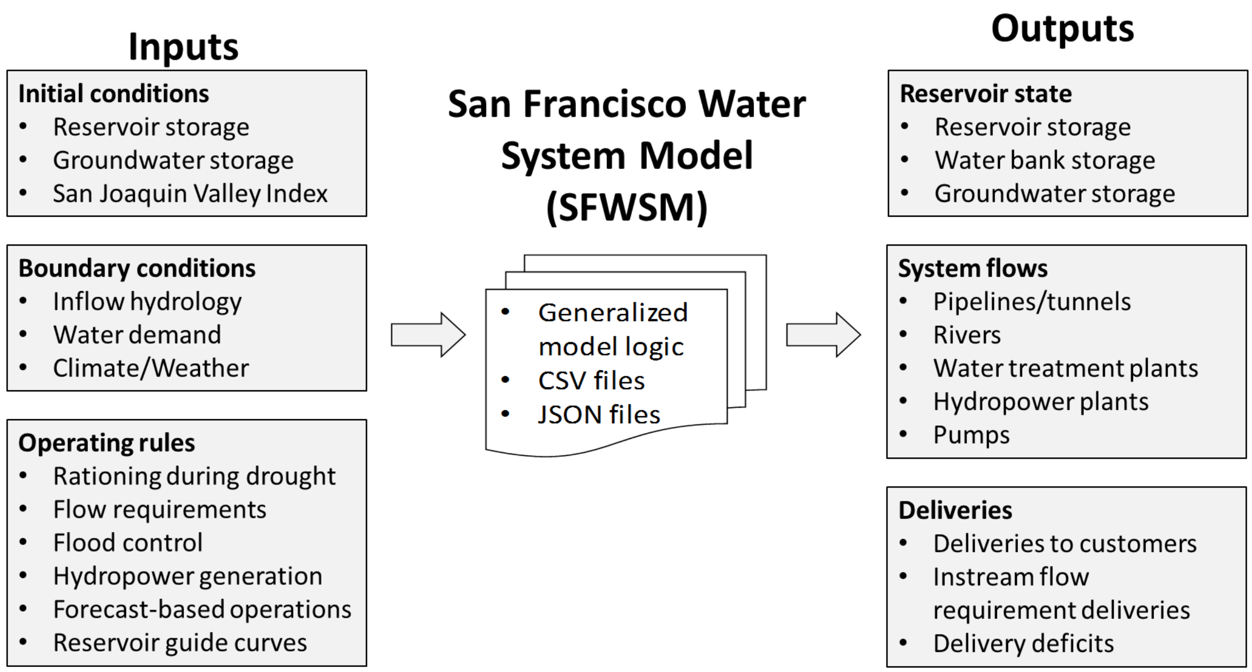

Figure 2. Inputs to and outputs from the San Francisco Water System Model.

1.4 Modeling approach

The core modeling approach of SFWSM is to use a linear programming (LP) optimization model run anew each time step. 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, subject to 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. This report describes these methods and values. We note here that "costs", which are described in detail below, are relative terms for simulation purposes, and do not have any monetary significance. The LP model is developed using *Pywr* (Tomlinson et al., 2020), a generalized network resource allocation model written in Python (https://github.com/pywr/pywr).

1.5 Model versions

As developed for the LTVA, SFWSM includes three configurations: a planning version and two validation versions. The planning version, intended for long term planning, represents current and anticipated future infrastructure and operations, and includes input from simulated future time series. In implementation, the planning version can be considered as the main version, as the validation version generally operates by disabling certain newer/future infrastructure and operations. The planning version will be able to be modified to include adaptation actions in the next phase of the study. The validation versions include one for validation of SFWSM with historical observations, which omits some newer programs and policies and relies on historical observed data for input[2], and one for validation with a version of HHLSM used for the Water Supply Improvement Program Programmatic Environmental Impact Report. The validation versions are described in the validation section (Section 4). Some important model limitations and input biases that affect validation results (e.g., biases in runoff from the hydrologic models) are described in the main model logic description (Section 3).

1.6 Differences from HHLSM

Key differences between HHLSM and SFWSM are listed in Table 1. The two primary differences are the spatiotemporal resolution and the modeling method. SFWSM includes a higher spatial resolution than HHLSM, with all retail and wholesale customers represented explicitly, and operates at the daily instead of monthly time step. The modeling method is optimization through linear programming, instead of prescriptive rule-based simulation. Another important difference with HHLSM is the logic used to implement drought rationing. The logic used in SFWSM is based on various trigger of Years of Remaining Supply, which is detailed further in Section 3.3.5.

Table 1. Key differences between HHLSM and both planning and validation versions of SFWSM.

Characteristic/component	*HHLSM*	*SFWSM*
Model platform/language	Excel with VBA	Python with Pywr
Model method	Prescriptive rule-based simulation	Optimization (Linear programming)
Time step	Monthly	Daily
SFPUC customers	Five demand centers	36 demand centers (26 wholesale customers; City & County of SF; 9 suburban retail customers)
Seasonal water demand depends on air temperature	Not included	Included

[2] Some input time series, such as inflow to the reservoirs, are simulated time series forced by observed weather conditions.

Demand rationing during drought	Reductions trigger for several total system storage thresholds	Reduction trigger for several Years of Remaining Supply (YRS; Section 3.3.5)
5% minimum demand reduction for SFPUC during drought rationing	Not included	Included
Emergency storage in the Bay Area	Not included	Included (120 days by default; can be adjusted)
Alternative supply (i.e., EBMUD and SCVWD)	Not included	Included
Upper Tuolumne River Ecological Program	Not fully included	Included
Rafting flows	Not included	Included but simplified (i.e., 9000 AF/month uniformly distributed during June, July and August).
Anticipate SJPL maintenance by filling San Antonio with HH water	Not included	Included
Evaporation over the reservoirs depends on the air temperature	Not included	Included
Lower Cherry aqueduct	Included (but never used)	Not included
Transfer from Alameda siphons to SVWTP	Included (but not significant ; ≈2,700 MG/year)	Included only when running simulations that includes water contamination of Hetch Hetchy water

1.7 This document

The purpose of this report is to describe the modeling scope, approach and methods used to simulate operations of the SFPUC water system for the LVTA. Methods are described with enough detail to give the non-modeler and modeler alike a good understanding of the logic that drives the model, but without a detailed description of actual code.

The modeling code, which has been provided to SFPUC independently of this report, in many cases may be more detailed than what is described here in order to achieve simulation objectives. Model code is generally “self-documenting”, such that more detailed descriptions here are unnecessary. Furthermore, code includes programmer comments as needed to explain further some specific code logic.

Source data used in the model is included in this report as needed to document many of the policies that were used to create SFWSM. In many cases, data used is not listed explicitly, but described qualitatively and either depicted graphically (e.g., assumed demand for SFPUC’s suburban retail customers) or referenced as needed.

2 Modeling framework

2.1 Generalized model logic

2.1.1 Theoretical basis

Models for simulating water systems can be developed using strictly "if-then-else" style simulation-based logic, optimization-based simulation logic, or some combination of the two. In a traditional simulation approach, water system operations are represented using logic that is broadly consistent with the way water systems are usually actually operated: with logical rules, specified using "if-then-else" style statements. For example, "if Hetch Hetchy reservoir level is X, then release is Y". HHLSM is set up in such a way. In an optimization-based approach to simulation such as that used here, the "best" water system operations are determined in any given simulation time step based on minimizing the total cost (or maximizing the total benefit) of flow and storage decisions during the time step, subject to various management and physical constraints, as described further below. "Costs" in this context are relative quantities for prioritizing/weighting water allocation operations, and not necessarily dollar values. Optimization may be used for simulation if the costs (benefits) and constraints are set up in such a way that "optimal" operations reflect actual operations. The choice of a relevant objective function and weights (costs) associated with each decision variable allows mimicking the observed operations. The above example if-then logic would still be relevant, but revised as "if Hetch Hetchy reservoir level is X, then maximum release is Y"; desired release is expressed as a parameter to an objective function, while a capacity on that release is defined as an operational constraint, rather than as a mathematical absolute release.

SFWSM uses the second approach, with human-readable code used to define the inputs to a LP optimization model. Such an approach has been extensively used for modeling large water resource systems across the world. Relevant examples are CalSim-II (Draper and Darabzand, 2003), developed by the California Department of Water Resources (DWR), and the well-known WEAP modelling platform (Yates et al., 2005), both of which use a custom programming/function language to describe water system model logic

2.1.2 Linear programming (LP) model

The LP problem in SFWSM is formulated with *Pywr* (Tomlinson et al., 2020), a Python programming language packaged for water system modeling described below. Pywr defines the LP problem most generally as a benefit maximization problem, with an objective to maximize the sum of flow and reservoir storage benefits times the respective relative marginal (per unit volume) benefit of flow and storage. Semantically, however, Pywr uses costs instead of benefits, such that for the Pywr user the problem can be described as a cost minimization problem. Mathematically, the LP problem is defined as an objective function to be minimized:

$$\text{minimize } c = \sum_{n} C_n Q_n + \sum_{r} C_r S_r \tag{1}$$

where c is the objective function value, n is any node representing demand (i.e., water customers, IFR locations, etc.), r is reservoir, C_n is marginal (per-volume) cost of flow through node n, Q_n is flow (i.e., delivery) at node n, C_r is marginal cost of storage at reservoir r, and S_r is end-of-time step storage at reservoir r. Groundwater is considered as a reservoir. Flow and storage are *decision variables* for the optimization problem. Marginal benefit is simply negative cost, such that the more negative the cost parameters C_n and C_r, the more the LP model will try to send water to the associated node or store water, respectively. Constraints are used to represent infrastructure capacities, mass balance through the system, physical processes and system management policies; equations for these constraints are omitted here for brevity. The purpose of the LP optimization model—and, therefore, SFWSM simulation model—is to find the optimal combination of flow (in non-storage nodes) and storage in each time step that minimizes the objective function value subject to physical and operational constraints.

The variables in the equation above, as well as parameters in the LP constraints must be defined in the model. Thus, maximum and minimum flow constraints and associated costs (if any) are assigned to demand centers (i.e., retail and wholesale customers), instream flow requirement (IFR) locations, pipelines, and other facilities. Costs for desired water (e.g., delivery to customers) are defined with negative values (benefits), while costs for undesired water (e.g., spill) are defined with positive values. Reservoirs are also assigned minimum and maximum storage levels and costs of storage. The purpose of Pywr, described below, is to enable defining these values, which may be based on complex operational rules, through an easy to read and easy to write syntax.

The above optimization is solved each time step with no consideration of future time steps. However, the numeric values of the LP constraints can be generated via a script and, therefore, possibly account for other time steps in the past (such as for tracking accumulated flow) or future (such as for hydrologic forecasts). Some LP parameters are also updated every time step to account for changing conditions or operating polices.

One important limitation of linear programming is that, by definition, system behavior is considered linear. This means, for example, that the marginal benefit of water deliveries is the same regardless of the amount delivered. In the case of deliveries, this example is accounted for in real operations through drought reduction policies, whereby demand is reduced, and all reduced demand is met. This is easily achieved in the model. Other cases are not so easily dealt with. Specifically, for operations constraints that depend on reservoir storage, a small change in reservoir storage can result in a dramatic switch from one flow value to another that might not reflect actual practice. Other operations (e.g., hydropower generation) can also exhibit all-or-nothing operations. These issues can be accounted for to some degree through piecewise linearization of otherwise nonlinear behavior. However, this necessarily increases computation time, countering one of the main advantages of linear programming. In SFWSM, piecewise linearization is not used, to prioritize speed and ease the calibration of the various costs through the system.

2.2 Implementation with Pywr

The above general logic is implemented for SFWSM using Pywr[3], a tool designed by Atkins consulting engineers and The University of Manchester and specifically for simulating complex water systems. With Pywr, a water system's infrastructure and operations are described using one or more JSON (JavaScript Object Notation) files or Python code, or some combination of the two. JSON is a human-readable text-based format for describing data in key-value pairs and lists of items. Since JSON is both somewhat more readable of the two code formats and easier to implement using Pywr, the model is as much as possible described using JSON, with more complex policies written in Python. During each time step, Pywr converts JSON/Python code to a lower-level LP problem formulation and passes this formulation to a more generic, computationally efficient LP solver. SFWSM uses GLPK for the LP solver engine, a free and open source LP solver (Free Software Foundation, 2014).

SFWSM itself is run as a single Python script that 1) declares custom Pywr parameters, 2) loads four JSON files into a Pywr model, 3) runs the model, and 4) saves results from the model.

Figure 3 illustrates through the example of a very simple model call with declaration of a simple parameter. In this example, a parameter "demand_2_policy" is declared prior to loading the JSON file "simple_example.json". This parameter is used within the JSON file and, as such, must be declared prior loading the model. Technically, the python class defining the parameter 'demand_2_policy' is called at every time step by the model via the 'load' class method, which returns the output of the 'value' function defined within the class. Major components of the JSON files is given later in the section, and illustrated for "simple_example.json" in Figure 5.

[3] https://www.github.com/pywr/pywr

```python
class demand_2_policy(Parameter):

    def value(self, timestep, scenario_index):

        if self.model.nodes['reservoir_1'].volume > 800:
            demand = 5
        else:
            demand = 0

        return demand

    @classmethod
    def load(cls, model, data):
        return cls(model, **data)

demand_2_policy.register()

m = Model.load('simple_example.json')
m.run()
results = m.to_dataframe()
results.to_csv('../results/Example_Results.csv')
```

Figure 3. Example Pywr script that defines and registers a custom demand policy, loads a JSON file with detailed water system model structure and information, runs the model, and saves the results to a CSV file. The python class is called every time step by the model via the 'load' function and returns the output of the 'value' function.

The four JSON files consist of one JSON file for each of the three SFWSM geographic regions (Upcountry, East Bay, and Peninsula), as well as an additional "master" JSON file describing the timestep, and system-wide information. In addition to these files, comma-separated value (CSV) files are used to store input time series and other required data; these data are imported into SFWSM during model initialization. This general organization is depicted schematically in Figure 4.

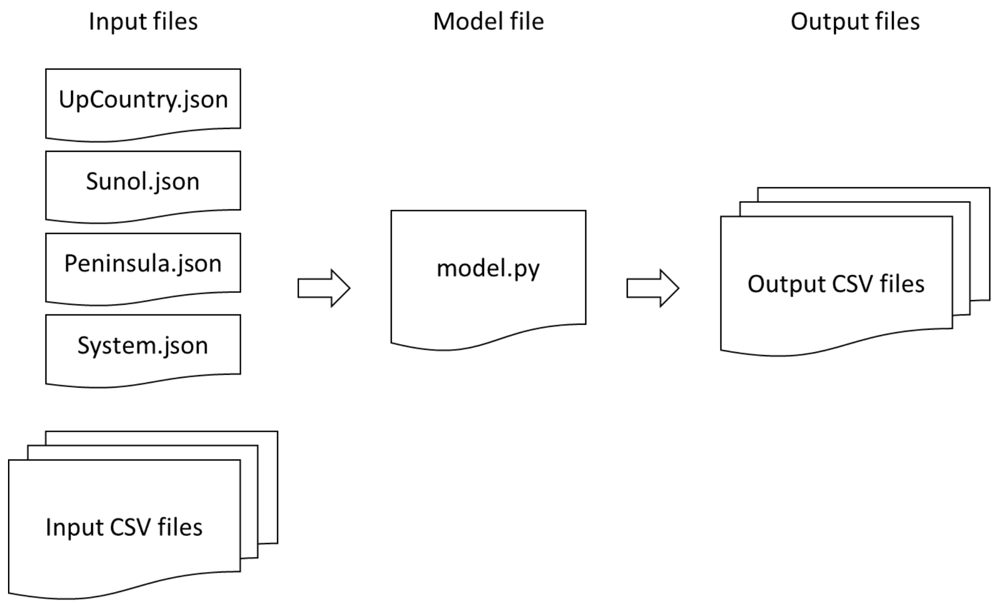

Figure 4. Schematic of the San Francisco Water System Model code and data organization. Filenames are illustrative; actual filenames differ from those shown here.

Several key Pywr concepts are worth noting. With Pywr, a water system is defined as an interconnected set of storage and non-storage *nodes* (e.g., reservoirs and demand centers), as is typical for water supply system models. Storage nodes are the only ones capable of carrying water from one timestep to another. As there is no routing in Pywr, conveyance facilities (e.g., pipelines and tunnels) are also represented as nodes in Pywr. Node connectivity is defined via *edges*. Each node is supplied with information about *parameters*, which vary depending on the node type. The function of parameters is ultimately to define the fundamental parameters used in the mathematical formulation of the LP problem. Since it is inefficient to store all information about system state during the simulation, any value that should be saved as output must be declared via *recorders*.

For illustration of the organization of the basics elements that compose the JSON files, a very simple example water resource system written using the JSON syntax for Pywr is given in Figure 5. Here, only the four main sections are provided; “nodes” declares facilities (storage, pipelines, etc.) or water transfers (e.g., precipitation and evaporation over lakes); “edges” connect the various nodes together; “parameters” give information/constraints about the possible water volume flowing through the specific nodes or water stored in reservoirs; and “recorders” lists the requested model outputs.

```
"nodes": [
    {
        "name": "reservoir_1",
        "type": "storage",
        "max_volume": 1000,
        "initial_volume": 850
    },
    {
        "name": "upstream_basin",
        "type": "catchment",
        "flow": "inflow_to_reservoir_1"
    },
    {
        "name": "demand_1",
        "type": "output",
        "max_flow": 10,
        "cost": -1
    },
    {
        "name": "demand_2",
        "type": "ouput",
        "max_flow": "demand_2_policy",
        "cost": -1
    }
],
"edges": [
    ["upstream_basin", "reservoir_1"],
    ["reservoir_1",    "demand_1"],
    ["reservoir_1",    "demand_2"]
],
"parameters":{
    "inflow_to_reservoir_1": {
        "type": "dataframe",
        "url": "../data/flow/flow_to_reservoir_1.csv",
        "column": "flow_Res_1",
        "index_col": "Date",
        "parse_dates": true
    },
    "demand_2_policy": {
        "type": "demand_2_policy"
    }
},
"recorders": {
    "Reservoir1_storage": {
        "type": "numpyarraystoragerecorder",
        "node": "reservoir_1"
    },
    "Supply_demand_1": {
        "type": "numpyarraynoderecorder",
        "node": "demand_reservoir_1"
    }
}
```

Figure 5. Illustration of the organization of the water system using the JSON format for Pywr.

Water system operations can be described using either Pywr's built-in water system rules specification/syntax, or by customizing each of nodes, parameters, and recorders using additional python scripts. Some representative examples of Pywr's functionality within the above generic organization scheme include:

- Nodes can be created that represent aggregations of other nodes. This is useful, for example, for aggregating all reservoirs to track total system storage. Aggregated nodes cannot be linked to the water system network.
- Virtual nodes can be created to track water in accounting schemes. The Water Bank (Section 3.7.1) is represented as a virtual node. Like aggregated nodes, a virtual storage node cannot be linked directly to the water system network.
- Complex parameters that depend on system state and time can be written using the JSON format (e.g., *control curves*[4], which depend on reservoir storage levels) or written with scripts in Python. Both approaches are used in SFWSM.
- Custom recorders can be created to convert system state into an output parameter. In SFWSM, hydropower generation is tracked this way; flow through powerhouses is converted via recorders to save energy generation as output.

For SFPUC personnel interested in modifying the delivered SFWSM model we recommend following first the Pywr tutorial (https://pywr.github.io/pywr-docs/tutorial.html) and going over the various examples of Pywr models available online, including at:

- https://github.com/pywr/pywr/tree/master/examples
- https://github.com/pywr/pywr/tree/master/tests

3 Specific modeling logic

SFWSM includes representations of infrastructure configurations and capacities, as well as infrastructure operations, driven by management and policy objectives. Both are described generally here.

3.1 Physical system

SFWSM infrastructure is represented as an interconnected network of Pywr nodes, which represent specific facilities (e.g., reservoirs and pipelines), natural inflow catchments (e.g., rivers), and decision points (e.g., flow requirement locations). In addition to physical features, the system, as represented in Pywr, includes a wide range of nodes not typically represented as such that represent physical processes (e.g., evaporation). The system spans the three geographic regions of SFPUC's water system, including the Upcountry, East Bay (Sunol), and Peninsula regions. An overview of

[4] The term "control curve", used in Pywr, is synonymous with "guide curve" or "rule curve", both common industry terms. In SFWSM, a "preferred storage level" acts as a guide curve and is used with a Pywr control curve parameter type for modelling.

the schematic of the system as modeled with Pywr is shown in Figure 6; a larger version of the schematic is in Appendix I.

Table 2 summarizes the facilities included in SFWSM, including their specific Pywr node type. Note that some assets have been either aggregated, in the case of parallel conveyances (pipelines, etc.), or omitted as able to reduce system complexity and minimize computation time. Also, since water quality is not simulated in this model, operations to discharge or treat out-of-specification water is not included in the model.

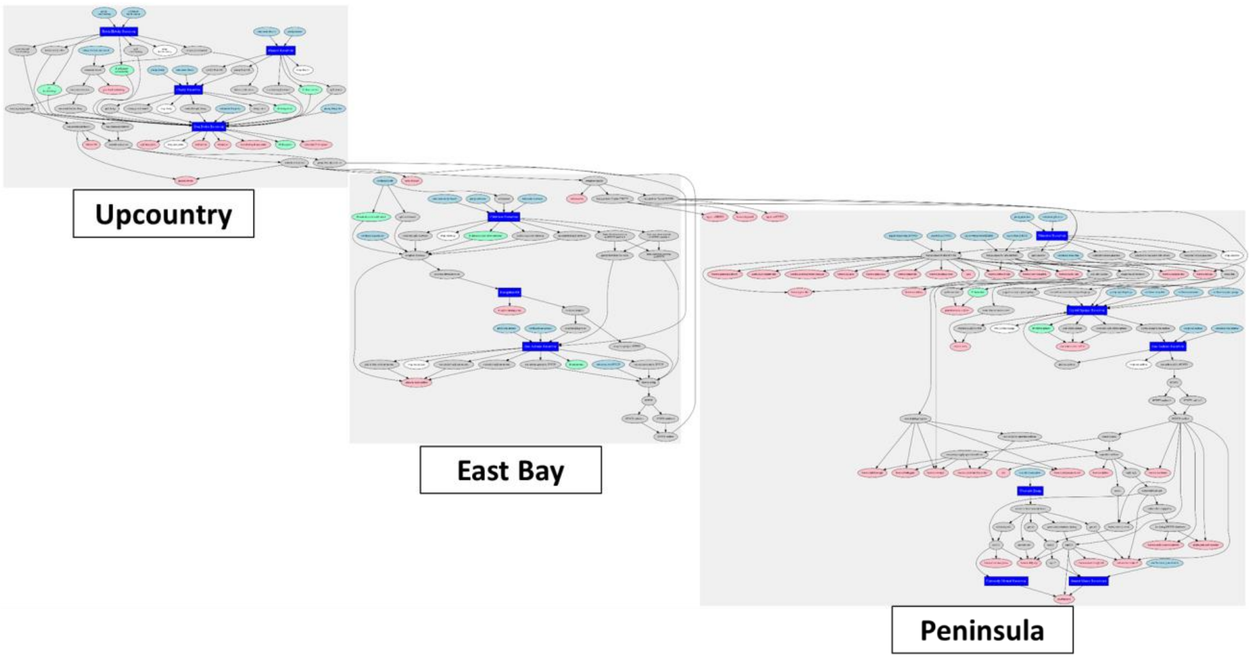

Figure 6. Overview of the schematic of the San Francisco Water System Model as represented in Pywr. Dark blue nodes are reservoirs, light blue nodes are inputs, pink nodes are demand, green nodes are instream flow requirements, and grey nodes are everything else. A larger version of this schematic is in Appendix I. Note also that the version of the schematics presented in this report are obtained for the historical validation version of SFWSM. High resolution pdf for of the schematics for all versions of SFWSM are provided to SFPUC.

Table 2. Infrastructure count and comments by modeled resource class (node or link) and resource type. For links, counts include modeled serial segments, rather than actual parallel asset numbers. Aqueducts and siphons are lumped with pipelines. Counts listed are subject to revision.

System component type	***Comment***	***Pywr node type***
Demand	Includes 27 wholesale customers, the City and County of San Francisco retail customer, other misc. retail customers, and canal diversions for MID and TID.	Output
Storage reservoirs	Includes 4 Upcountry resources (including Don Pedro), 2 East Bay reservoirs, 3 Peninsula reservoirs and 2 City reservoirs (University Mound and Sunset reservoirs)	Storage
Water Bank	Water Bank at Don Pedro Reservoir is represented as a virtual storage node.	Virtual Storage
Watersheds	Includes 4 Upcountry watersheds, 5 East Bay watersheds, and 6 Peninsula watersheds	Catchment
Treatment plants	Includes Harry Tracy and Sunol Valley water treatment plants (WTPs)	Storage
Hydropower stations	Not explicitly represented; hydropower generation calculated from power tunnel flows.	Recorder
Instream flow requirements	Includes instream flow requirements below every major dam or diversion (The actual IFRs that are considered within SFWSM vary with the version. For instance, there is no IFR at Calaveras and ACDD tunnel for the historical validation, while these IFRs are implemented in the HHLSM and Planning versions. See Section 4 for the detail of the considered IFR by each SFWSM version.	Link
Conveyances (pipelines, etc.)	Includes major pipelines, tunnels, aqueducts and siphons, aggregated as needed to improve computational efficiency.	Link

3.2 Costs

Because the model operates on a cost-basis, we note here some general principles related to assigning costs, which indicate the overall priority of water allocations. Table 3 lists costs assigned to each of five major operational categories, including IFR, uncontrolled spill, controlled spill, demand, and evaporation. The values listed indicate that evaporation (for which assigning cost is a modeling technique rather than a value *per se*) is prioritized over all other allocations, followed by IFRs, demand, controlled spill, virtual water and finally uncontrolled spill. In all cases except uncontrolled spill and virtual water, costs are negative, representing a benefit. Deviations from these system-wide costs are described as needed for specific operational needs. For instance, controlled spill can be modeled via a minimum flow constraint that takes positive values when needed, and equals zero otherwise.

Table 3. System-wide costs for major operational categories, from lowest (most valuable) to highest cost.

Category	Cost
IFR	-100,000
Evaporation	-1,000,000
Controlled spill	-100
Demand	-1000
Virtual Water	999
Uncontrolled spill	1,000

3.3 System-wide policies

Operational drivers are management objectives and constraints for the operation of the RWS. Specific modeling approaches for each operational driver are described in this section. In some cases, operational drivers have not been explicitly included in SFWSM. In such cases, the driver is described briefly here, with additional commentary indicating why it was not included. The detail of the modeling approach used depends on the driver, and in some cases a policy is implemented through operations of a specific facility, described elsewhere. References to more detailed modeling descriptions are provided as needed.

Operational objectives are ultimately manifested in the LP model as parameters in a constraint or the objective function. In some instances, the objective is represented in the LP model structure. In other cases, the objective is represented as a function that is calculated to generate the appropriate LP cost or constraint. How each operational objective or constraint affects the LP model is noted.

The list of modeled drivers included here is from the HHLSM model and regular discussions with SFPUC staff.

3.3.1 Retail and wholesale water supply objectives

Water supply is the core business of SFPUC, which supplies water directly to retail customers in the City and County of San Francisco (CCSF) and to regional wholesale customers, collectively organized as the Bay Area Water Supply and Conservation Agency (BAWSCA). The current emphasis of SFWSM is water delivery. Water deliveries goals are to fulfil water demand, as quantified in service contracts (Water Service Agreements) between BAWSCA wholesale customers and SFPUC.

Water deliveries are represented in the LP objective function as decision variables. Delivery is constrained in the Pywr model formulation by a maximum allowable flow constraint for each demand center, representing actual demand. Daily demand for both retail and wholesale customers are specified as a time series of numerical values. The calculation of this demand uses annual and daily time steps to reflect base and seasonal demands, respectively. The method for calculating demand is described below (Section 3.4).

3.3.2 Raker Act

The Raker Act of 1913 is the congressional legislation that authorized the construction of the Hetch Hetchy water system. The Act contains several provisions which affects operation of the system, summarized as follows.*No sale to private profit generating companies*

Operationally, the SFPUC can sell a small percentage of its water to private companies, if it is offset by use of local supplies to ensure those sales are not in excess of local watershed yield. In practice, this means that diversion from local watersheds must equal or exceed delivery to California Water Service (CWS), where "diversion" equals local watershed inflow less spill and managed releases (instream flow requirements and reservoir maintenance releases). CWS is the only for-profit customer of SFPUC. This is not accounted for explicitly in SFWSM, though is checked for during model validation.

Senior right of Modesto and Turlock Irrigation Districts (MID and TID).

These irrigation districts have water rights that are senior to those of 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, District Entitlements consist of the lesser of 1) 2,416 cfs (June 14 through April 14) or 4,066 cfs (April 15 through June 13) and 2) unimpaired flow in the Tuolumne River at La Grange. The implication is that water available to SFPUC from the Tuolumne river (Water Available to the City, or WAC) is any unimpaired flow at La Grange (below Don Pedro) that is greater than District Entitlements.

Operationally, the senior water rights of MID and TID are implemented using a combination of the Water Bank, which offers storage space in Don Pedro to meet District Entitlements if water is diverted in Upcountry reservoirs, and by setting a high cost of MID/TID shortages, subject to drought reductions. Diversions for irrigation do not necessarily equal entitlements. Since actual demand for agricultural water is sensitive to hydrologic conditions, a method was developed to establish a relationship between MID/TID surface water demand and inflow hydrology, in turn a function of climate. Both the MID/TID surface demand estimation method (Section 3.4.1) and Water Bank operations (Section 3.7.1) are described further below.

3.3.3 "One Treatment Plant Running at All Times"

This objective requires that at least one water treatment plant (WTP) (i.e., Harry Tracy or Sunol Valley WTP) is online at all times, so that demand can always be met by two sources. In practice, this means that one treatment plant is off, the other must be on.

This policy is implicitly implemented in SFWSM via a minimum flow constraint for the WTP outflows. If the minimum flow constraint cannot be satisfied, high-cost virtual water (see Section 3.3.10) is supplied just before the WTPs to prevent SFWSM from crashing. The virtual water for WTPs are tracked in SFWSM.

3.3.4 Emergency storage in the Bay Area

To account for an unplanned outage of Hetch Hetchy supply or other failure (e.g., from seismic activity), 120 days of emergency reserve supply must be kept in storage in local (Bay Area) reservoirs. The number of days of storage is a somewhat flexible policy that was never explicitly stated or operated to. The Water Supply Improvement Program (WSIP), which mainly focused on delivery reliability under seismic risk, assumed an interruption of Hetch Hetchy water supply for 90 days[5]. In 2015, before the inspection and repairs of Mountain Tunnel, SFPUC developed an emergency response plan for an unplanned outage of Mountain Tunnel and used 180 days of supply in storage by June 1 as a threshold to evaluate the risk of such failure[6]. For this model, 120 days was selected based on discussions with SFPUC and inspection of historical Bay Area reservoir storage, deliveries and refill operations. The number of days for emergency storage is a parameter that can be changed easily in SFWSM via the JSON model files. Possible values for emergency storage are currently 30, 90, 120, 150 and 180 days, although this could be extended.

This policy is modeled by tracking storage in the Sunol Valley and Peninsula reservoirs. Only San Antonio, Crystal Springs and San Andreas can be refilled using Hetch Hetchy water. If any of the following conditions are met, then water from Hetch Hetchy Reservoir is used to fill San Antonio Reservoir:

1) San Antonio Reservoir is less than 38,000 AF (this threshold value is based on historical observation of storage and refill operations),
2) Sunol Valley reservoirs (San Antonio and Calaveras Reservoirs) together have less than 120 days of supply for South and East Bay demand, or
3) A shut down for maintenance of the San Joaquin Pipelines is scheduled within 50 days

If Peninsula reservoirs (Pilarcitos, San Andreas and Crystal Springs Reservoirs) together have less than 120 days of supply for Peninsula and City demand (exclusive of South Bay demand), then water from the Pulgas diversion is used to fill Crystal Springs Reservoir.

---

[5] PARSONS - CH2MHILL, 2006. Facilities Sizing Report. p.7 “these flows provide the ability to meet 2030 customer demands and maintain sufficient storage in Bay Area reservoirs in the event of an interruption of Tuolumne River supply for up to 3 months.”

[6] Nelson, C. and Dufour, A. to Ritchie S. Effects of Planned and Unplanned Outage of Mountain Tunnel and Hetch Hetchy Supply on the Local System Reservoirs. September 28, 2015. p.10 “At the end of an outage, the RWS is operated to maximize refill of Local System reservoirs. By June 1, which corresponds to the beginning of summer delivery period, the target is 90% chance of having 180 days of supply in storage in case of an unplanned outage of the Hetch Hetchy supply.”

The assumption in the above logic is that the Sunol Valley reservoirs cover emergency reserves for East Bay and South Bay customers, while Peninsula reservoirs cover Peninsula and CCSF customers, based on discussions with SFPUC.

3.3.5 Drought policy

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, 2009b). The WSA defines how shortages may be imposed, including using either voluntary (preferable) or mandatory (if needed) reduction measures. The WSA describes how shortages are allocated between SFPUC and the wholesale customers as a group (Tier 1 split). BAWSCA developed a plan to allocate shortage among wholesale customers (Tier 2 split). Shortages result in percentage use reduction requirements on the part of wholesale customers for shortages up to 20%. For shortages in excess of 20%, the WSA gives SFPUC some discretionary power to decide how to further allocate shortages on a per-drought basis. However, the SFPUC has a water supply goal of not more than 20% rationing. In SFWSM, SFPUC demand is reduced no more than 20%. In this model, the rationing applies to in-city retail customers and suburban retail customers.

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

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

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…*". 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 is called "years of remaining supply" (YRS), and is calculated as:

$$YRS = \frac{Total\ Active\ Storage}{Total\ Delivery}$$

Total active storage is total actual storage less the total of each reservoir's inactive zone (i.e. dead pool volume). Total demand is calculated as:

$$Total\ Delivery = \sum Urban\ Delivery + \sum IFR + \sum Reservoir\ ET$$

where IFR is instream flow requirements lost to the system. Only IFRs that are not recaptured by a reservoir downstream are included in the calculation (i.e., IFRs downstream Stone dam, Crystal springs and Calaveras reservoirs). YRS is calculated once per year, on July 1, based on the observed delivery from the previous year (July 1 of the previous year through June 30 of the current year). Thus, in the model, YRS is an approximate value based on the previous deliveries rather than based on delivery and hydrologic predictions for the year to come. The approach of using YRS is different than HHLSM, which uses total system storage triggers. 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 4, along with Tier 1 splits and the historical total system storage trigger values (in thousands of acre-feet, TAF) that were used to calculate the YRS thresholds. The YSR triggers were developed using HHLSM simulation results for WSIP 2018 with an annual demand of 265 MGD. Both voluntary and mandatory reductions are modeled as demand reductions.

The Tier 1 split for allocating reductions (Step 2) depends on YRS. If YRS ≥ 3.14 (Demand Reduction Factor ≤ 20%), then the Tier 1 split is 63% reduction allocation for SFPUC and 37% for wholesale customers. Otherwise, the split is 62.5% for SF and 37.5% for BAWSCA customers. Tier 2 splits are as listed in Table 5. The splits among urban and sub-urban customers are listed in Table 6.

Table 4. 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	≥ 3.61	0%	N/A (groundwater use)	N/A	1285
Trigger 2	3.41	10%	Voluntary	37/63	1100
Trigger 3	3.14	20%	Mandatory	37/63	900
Trigger 4	1.81	25%	Mandatory	37.5/62.5	500

Table 5. BAWSCA Tier 2 splits for distributing Tier 1 drought allocations.

BAWSCA Customer	Share
ACWD	7.6%
Brisbane GVMID	0.5%
Burlingame	2.9%
CCWD	1.2%
CWS Bear Gulch	6.5%
CWS MPD	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%*

Table 6 Urban and large sub-urban retail customers Tier 2 splits for distributing drought allocations. Contrary to BAWSCA Tier 2, these shares were estimated from historical supply (contribution of retail customer X divided by total retail demand)

Urban and suburban retail customers	Share
San Francisco (CCSF)	94.8326%
llnl_site_300	1.0137%
groveland_community	0.507%
general_electric	0.0413%
town_sunol	0.735%
sfo	1.4716%
golden_natl_cem	0.3192%
nasa	0.7907%
cordilleras	0.0826%
menlo_park_cc	0.2808%
TOTAL	100%

The application of demand reduction differs slightly between wholesale customers and in-city retail customers. As described further below, 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 (base and seasonal demand components are discussed in Section 3.4.2). During the 2012-2016, it was observed that reduction in delivery to San Francisco city was achieved by reduction is base demand instead of seasonal demand. If the reduction of the seasonal demand is not sufficient to match the requested reduction, the base demand is reduced accordingly.

Note that, similar to HHLSM, the above described demand reduction leads to an increase in demand for San Francisco city when system-wide demand is reduced by 10%. To overcome this issue, a minimum 5% demand reduction is accounted for SFPUC urban and large sub-urban retail customers.

3.3.6 Minimize net energy use

Within the broader water supply and other objectives, water should be managed to minimize net energy use. Energy consumption is primarily in pump stations and groundwater pumps, while energy production is from hydropower. In practice, SFPUC uses Peninsula water to meet higher elevation customer and CCSF retail demand first, limiting the use of Hetch Hetchy water in those regions. This policy is not explicitly included in the model as an operational driver, as the immediate goal of the model was to represent historical water system operations. However, at location where both gravity and pumping are available (e.g., between San Antonio and Sunol Valley WTP) a higher cost assigned to pumping (e.g., cost associated with gravity being zero) ensures that gravity is used first, which helps minimize the energy used. Aside from such explicit configurations, it was assumed that existing operational practice already accounts for this policy objective. Other model links have costs to represent the fact that energy consumption through pumps is required to move water through them. This is for instance the case of the Baden valve lot pipeline, or the sunset pipeline 1 whose cost represents the pumping cost from Merced pump station. Note that costs are relative to one another and have been set through a manual calibration to represent at best the SFPUC system operation.

3.3.7 Spill Prevention

Spill represents water lost from the reservoir system for urban water needs and is therefore minimized by SFPUC, particularly in the Bay Area reservoirs. Though spill should be minimized throughout, the risk of spill Upcountry should be greater than in the Bay Area. In particular, the model should try to minimize spill from Bay Area reservoirs when water is imported from Hetch Hetchy. This objective must be coordinated with the objective to keep water in reserve in Bay Area reservoirs for emergencies (as described in section 3.3.4). This objective is accounted for implicitly in SFWSM through the combination of existing preferred storage levels for local reservoirs—which are assumed to take this objective into account—and higher benefits (negative costs) of storage in local reservoirs over up country reservoirs. The latter is also needed for prioritizing emergency storage in the Bay Area.

In any given timestep, this is readily accounted for using the LP approach to simulation: water is valued for all uses other than spill, resulting in an LP model outcome that implicitly minimizes spill by maximizing all other uses first. However, because the LP problem does not have hydrologic foresight, spill could still occur if there is unanticipated excess inflow into reservoirs. Forecasting at several reservoirs is used to account for this in the model, as in actual operations (i.e., Hetch Hetchy, Cherry and Calaveras reservoirs). With forecasting of inflows to reservoirs, SFPUC anticipates potential spill, and can pre-release into other reservoirs or otherwise release water accordingly. Forecasting routines are described as needed below for specific facilities.

3.3.8 Hydropower

Hydropower operations, which include operations of Kirkwood and Holm powerhouses (PH), are modeled to reflect the "Water First, Power Second" policy via Canyon Power Tunnel and Cherry Power Tunnel, respectively. Under this policy, water supply is always prioritized over hydropower. Additionally, instream flow requirements are also prioritized over hydropower (and other uses). However, hydropower is also generated incidentally by default whenever water is released to meet water demand, whether for water supply (Kirkwood PH) or to fill the Water Bank (Holm PH). Hydropower is always generated incidentally at Moccasin Powerhouse. In HHLSM, flows in Canyon and Cherry power tunnels are always actively managed, whether for release for water demand, spill avoidance, or hydropower, as described in Section 3.7.1. Hydropower generation in SFWSM is calculated following the standard formulation of the hydropower equation and accounts for the difference between the reservoir level elevation and powerhouse tailwater elevation for each respective powerhouse. It does not account for hydraulic losses and other efficiency loss due to other factors (i.e., turbine efficiency change with flow rate, ..). As such, the use of the hydropower generation output is not recommended. Instead, the flow through the power tunnels should be considered for assessing change in hydropower generation.

3.3.9 Instream flow requirements

Instream flow requirements (IFRs) represent demand for water in a river, typically below storage or diversion dams. IFRs in the region range from relatively simple fixed flowrates below a diversion dam to relatively complex, with a dependency on a hydrologic indexing scheme. IFRs are represented in SFWSM as Pywr nodes with maximum flow attributes and high negative costs, which ensures that IFRs are prioritized by in SWSM. They are conceptually similar to demand nodes, except that they are non-consumptive.

The IFR modeling method depends on the complexity of the IFR schemes. As with other inputs, IFRs are described as constants, pre-processed time series data, or dynamically in a Python script. For expediency, all pre-processed IFRs were developed as described in HHLSM, rather than the original policies. This resulted in a coarser temporal resolution of one IFR (Calaveras Creek below Calaveras Dam) and omission of ramp rates. Although this results in discrepancies between policy and the model in a couple of cases, the discrepancies are in temporal resolution only, and do not affect overall model accuracy. The original policies are included in specific IFR descriptions in Section 3.7.

3.3.10 Virtual water

Virtual water is high cost water that is not explicitly supplied in the LP model, but that is nonetheless available. Virtual water is used in SFWSM for two overlapping purposes: to prevent model crashes and to provide insight into where additional water may be needed in the future. One drawback for LP optimization problem is that a solution to the optimization problem may not exist (i.e., the problem may be *infeasible*). A common practice to avoid LP infeasibilities is to allow the model to add virtual water at specific locations. A high cost is associated with this virtual water so that it is only used if absolutely necessary. For example, if water cannot be supplied from sources above either one of the WTPs, virtual water is supplied to prevent the model from crashing.

Virtual water can be useful for identifying where additional water may actually be required to alleviate delivery shortages. For instance in the planning version of SFWSM, the EBMUD and SCVWD sources are modeled as virtual water, even though in reality they are used by SFPUC and not technically “virtual”. Virtual water points for emergency supply within the system include:

- EBMUD and SCVWD
- Mountain Tunnel entrance,
- Water treatment plants entrance (SVWTP and HTWTP), which represent critical supply points in the system.

3.4 Demand

3.4.1 Modesto and Turlock Irrigation Districts canal delivery

Irrigation demand in the Modesto Irrigation District (MID) and Turlock Irrigation District (TID) are met by surface water stored in Don Pedro Reservoir and groundwater wells. Irrigation surface water demand for MID and TID is calculated by using the historical daily mean canal diversion adjusted based on an annual linear regression between historical observed diversions and simulated San Joaquin Valley Index (SJVI)[7], modified during drought years based on a drought reduction policy. As accuracy of the main SFPUC water supply system is the modeling priority, representation of MID/TID operations are purposefully kept relatively simple, recognizing that MID/TID operations are buffered from the main SFPUC system on a daily basis through Don Pedro Reservoir only monthly to annual time steps impact the SFPUC system.

The MID/TID drought reduction policy, modeled in the same way as for the New Don Pedro Hydropower FERC relicensing process (Steiner, 2013), is calculated on April 1 from current Don Pedro storage and forecasted April 1st to July 1st inflow. Forecasting the inflow to Don Pedro reservoir requires forecasting the operations of Cherry, Eleanor and Hetch Hetchy Reservoirs. Here, the forecasted inflow to Don Pedro is defined as the difference between the future unimpaired inflow to Don Pedro from April 1st to July 1st and the flows from Cherry, Eleanor and Hetch Hetchy that have been simulated the previous year. MID/TID drought reduction factors are listed in Table 7.

Table 7. MID/TID Drought Reduction Factor (aka: Water Supply Factor).

Don Pedro Storage + Inflow Index (TAF)	MID/TID Reduction Factor
≤1,350	0.60

[7] https://cdec.water.ca.gov/reportapp/javareports?name=WSIHIST

2,000	0.85
9,999	1.00

Observed annual MID/TID canal deliveries for the period 1972 to 2012 were adjusted by accounting for the drought reduction factor. The adjusted canal deliveries were then regressed against the estimated historical SJVI values to obtain a relationship between SJVI and non-drought canal deliveries. Non-drought canal deliveries were then used to scale the corresponding daily time series of normalized (with unit annual volume) mean deliveries.

Regression equations for baseline MID/TID demand were calculated to be:

$$D_{MID} = -7.2 \times SJVI + 319.8$$

and

$$D_{TID} = -18.0 \times SJVI + 611.4$$

where D_{MID} and D_{TID} are annual (i.e., water year) canal deliveries for MID and TID, respectively, each in units of TAF. The data used in these regression equations are shown in Figure 7. There is a negative correlation between irrigation demands and SJVI, indicating that demand for imported water from the Tuolumne River increases when regional water availability declines.

The drought reduction factor is accounted for in SFWSM directly, rather than during pre-processing, as it depends partly on Don Pedro Reservoir storage.

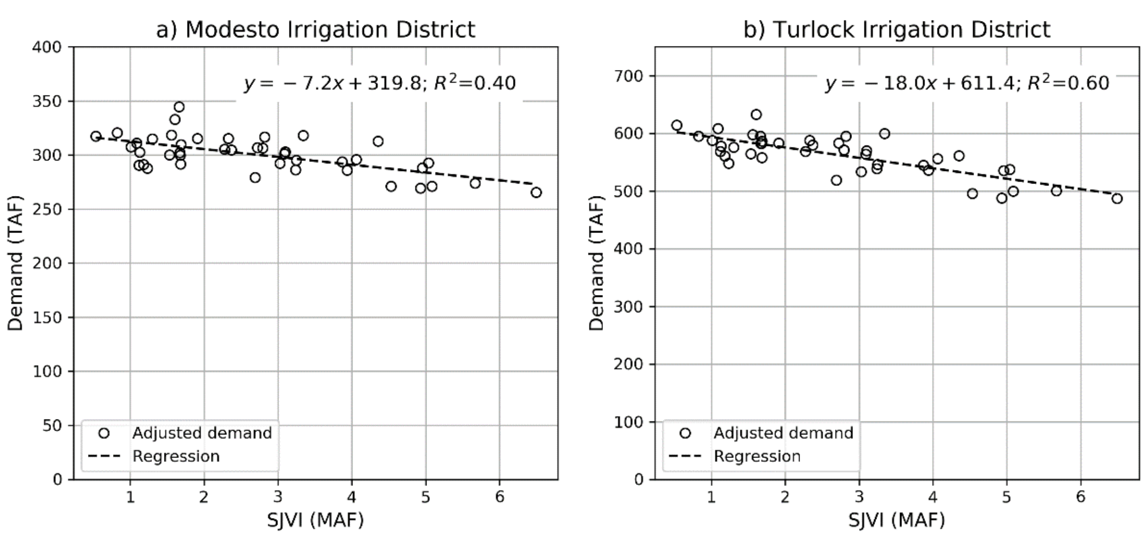

Figure 7. Linear regression between the San Joaquin Valley Water Year Index (SJVI) and each of Modesto Irrigation District (a) and Turlock Irrigation District (b).

3.4.2 Wholesale and retail demand

Note: This section provides a summary of the take-away of the Technical Report 3: Urban Water Demand (HRG TR3, 2020) that is relevant for SFWSM. The demand model described below is for the planning version of SFWSM. The demand used for the historical version of SFWSM is observed deliveries. The demand used for the HHLSM version of SFWSM is detailed in Section 4.2.

The SFPUC provides water to both retail and wholesale customers. Over 2.7 million people within the counties of San Francisco, San Mateo, Santa Clara, Alameda, San Joaquin, and Tuolumne rely entirely or in part on the water supplied by the SFPUC (Figure 8). Approximately two thirds of the SFPUC's water supply is delivered to wholesale customers, and the remaining one third is delivered to retail customers.

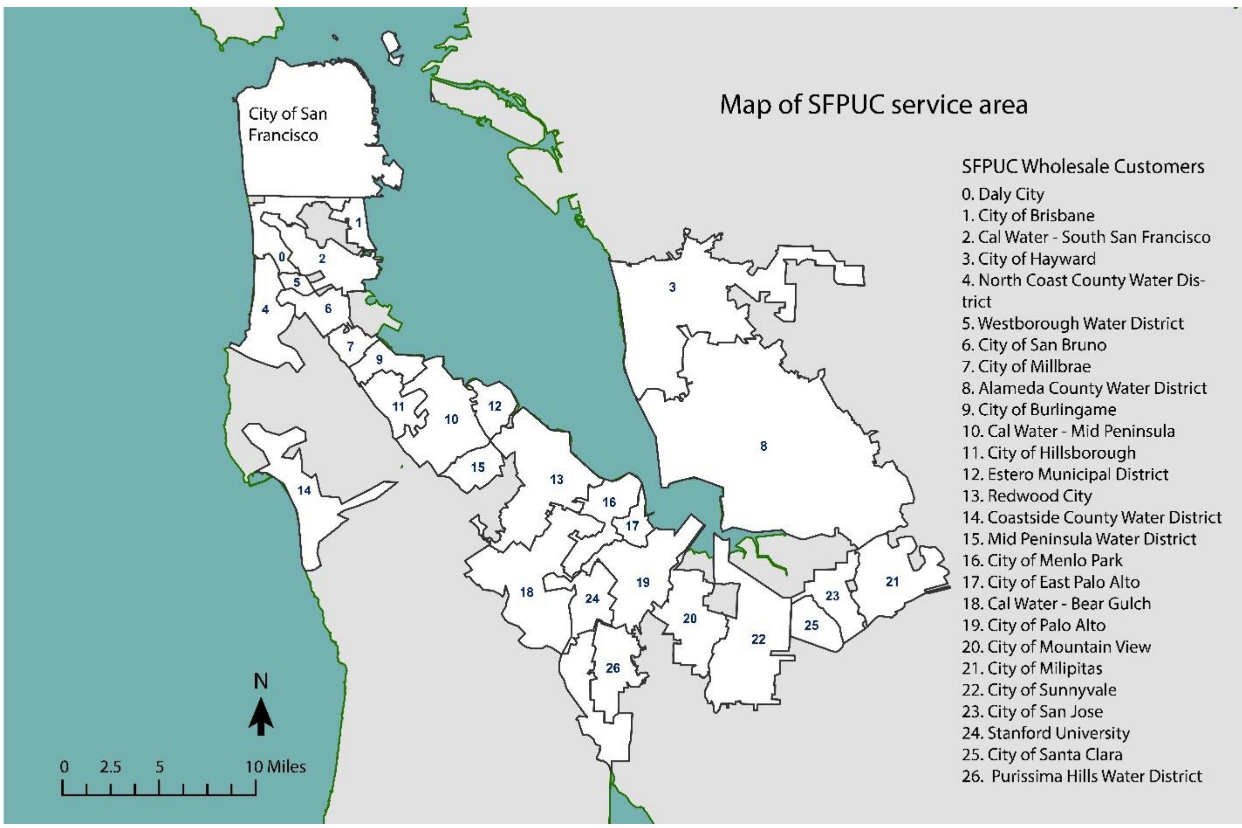

Figure 8: Map of the SFPUC service area (SFPUC, 2016a).

Retail customers include the residents, businesses, and industries located within San Francisco city limits, referred to as the in-city retail service area, as well as a patchwork of smaller non-contiguous customers located outside the City collectively referred to as the suburban retail service area.

The RWS also delivers water to 27 wholesale customers in Alameda, Santa Clara, and San Mateo Counties. The Bay Area Water Supply and Conservation Agency (BAWSCA) represents the interests of the wholesale customers and coordinates their water supply planning. Since 1970, the 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, 2016a).

Water deliveries from SFPUC to retail and wholesale customers are driven by demand on the RWS, for which a demand model was developed, as described in a companion report and summarized here. The demand model is concerned with addressing two questions: 1) How can we take a single point projection of demand and downscale this to the daily timestep? And 2) How can we ensure the time series of demand reflects seasonal fluctuations as well as the influence of future increases in temperature as a result of climate change?

The regression approach provided by Maidment and Miaou (1986) was used to incorporate the influence of temperature at seasonal timescale. This approach relies on decomposing observed water use into base and seasonal components (Figure 9), which represent the portions of total demand that are insensitive and sensitive to air temperature. The base fraction generally represents essential uses such as for drinking, cooking, washing, toilet flushing etc., while the seasonal portion represents non-essential and seasonally variable uses such as for landscaping. Observed fluctuations in the seasonal component of demand are related to observed maximum temperature in the same time step. This observation is used to establish a temperature-demand response function, referred to as a *heat function*. This heat function, shown in Figure 10, can then be applied to future climate realizations to establish a seasonal shape to an annual demand value.

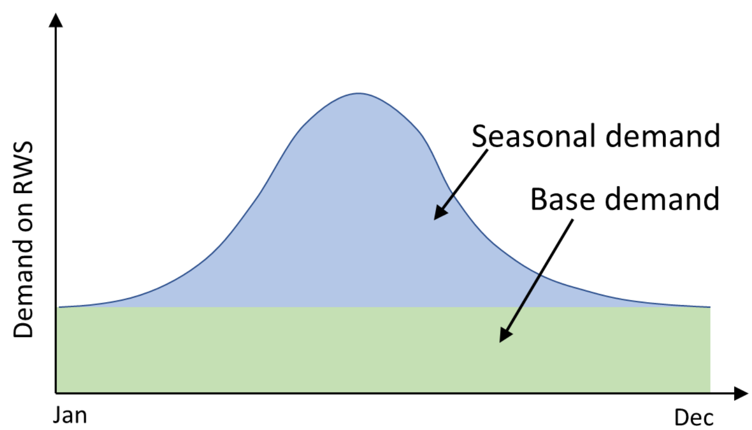

Figure 9. Idealized representation of base and seasonal demand for BAWSCA and San Francisco retail customers. Adapted from Maidment and Miao (1986).

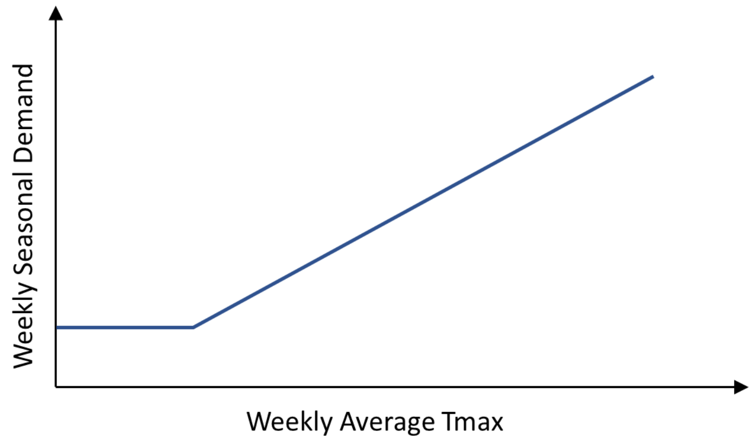

Figure 10. Example heat function representing the response of weekly seasonally demand to weekly average maximum temperature (Tmax).

To establish the base and seasonal components of demand, SFPUC examined weekly delivery data from 2005–2018. The SFPUC share of total demand for each wholesale customer was derived from 2013 to establish the total demand on the RWS (i.e., demand as used in SFWSM). For each customer, the lowest values of demand for the months from December through April are divided by total annual demand and their average across the time frame provides an indication of the base/total demand ratio. As demonstrated in Figure 11, the range of base/total demand ratio across customers is significant, with a low of 0.31 for Hillsborough, a high of 0.92 for the city of San Francisco and a median value of 0.632. These differences reflect differences in water use across regions. The base/total demand ratio for densely population urban areas such as the city of San Francisco is high, where consumers are less likely to have outdoor space and water use is focused on essential uses.

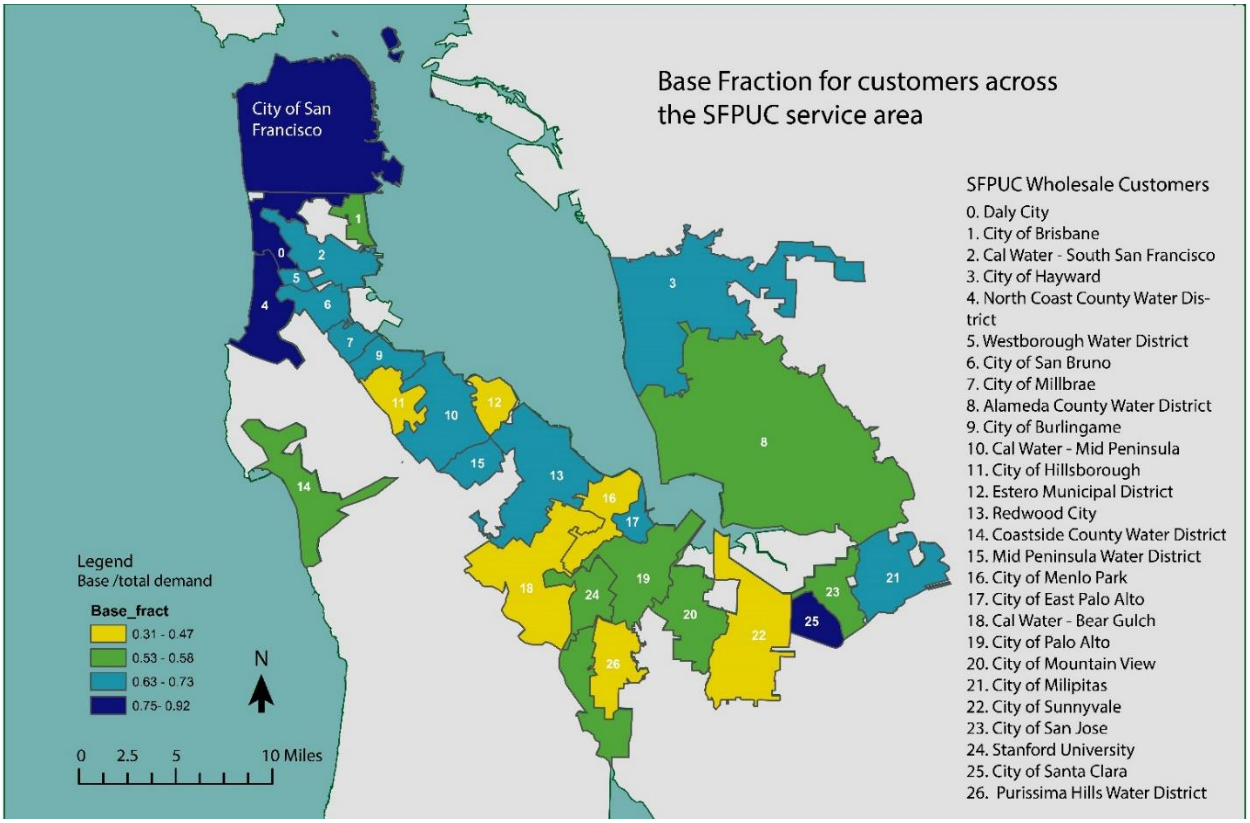

Figure 11: Range of base fraction values across the SFPUC service area.

From the same dataset of weekly deliveries, SFPUC established a heat function or each of the three delivery centers across the SFPUC service area, which represents the seasonal response of customers to changes in temperature. To create this heat function, data points have been normalized by average delivery for their respective week. This is required to ensure the output of the heat function is a unit-less coefficient that can be applied to the disaggregated, seasonal portion of demand based on daily maximum temperature.

The demand model is implemented as a pre-processing step, with results from the demand model used as time series input to SFWSM. The final demand is reduced by a drought reduction factor during drought based on the drought reduction policy described above (Table 4).

One important assumption in the demand model is that demand is only constrained by the drought policy. In actuality, there is a residual "memory" effect during drought recovery periods, whereby demand does not immediately return to pre-drought levels, due to changes in water consumption habits. Consumption changes from increased water use efficiency (more efficient household appliances, landscaping, etc.) that typically occurs during longer droughts is also not included. Excluding these demand effects was based on several discussions with SFPUC staff. The assumption is that long-term demand change scenarios will encapsulate the ranges of demand anticipated with (and without) climate change. Further model refinements can be added as needed, consistent with the approach of the LTVA generally.

3.4.3 Moccasin Fish Hatchery

The Moccasin Fish Hatchery is located just below Moccasin Diversion Dam on Moccasin Creek. A constant demand of 30 cfs (59.5 AF/day) was assumed for the hatchery, based on input from SFPUC staff. Note that the water used by the Fish Hatchery eventually flows to Don Pedro reservoir.

3.5 Hydrology

Hydrologic processes of importance in the system model include surface runoff into reservoirs (inflows), reservoir evaporation, local precipitation at reservoirs, and groundwater recharge in the Peninsula. These processes are summarized here, along with performance issues with important implications for SFWSM.

3.5.1 Surface runoff

Surface runoff is represented in the system model as time series input of flows into reservoirs, which are generated as a pre-processing step with data from the hydrologic models described in the Technical Report 2: Hydrology Modeling Module (HRG TR2, 2021). In the Pywr model, inflows are represented as Catchment nodes, from which a known amount of water enters the system.

Surface runoff in the Upcountry region were generated with the Precipitation Runoff Modeling System (PRMS), a physically-based distributed-parameter hydrologic modeling system developed by the USGS (Markstrom et al. 2015). One hydrologic model was developed for each of the watersheds above Hetch Hetchy and the combined Cherry-Eleanor watershed, as well as for the region between New Don Pedro Dam and the Cherry, Eleanor and O'Shaughnessy Dams (the latter is noted as 'Don Pedro' basin). A PRMS model for Don Pedro watershed model was calibrated by UMass, while the others were provided to by UMass by SFPUC. Surface runoff in the East Bay and Peninsula regions were developed using a version of the Sacramento Soil Moisture Accounting Model (SAC-SMA), a lumped-parameter hydrologic model developed by NOAA, modified to allow the use of distributed parameters (SAC-SMA-DS). Input to each implementation of PRMS and SAC-SMA-DS include i) combination of station or gridded datasets of observed precipitation and temperature variables, and ii) precipitation and temperature data from the weather generator developed for the LTVA, as described in the Technical Report 1: Weather Generator Module (HRG TR1, 2018). The historical validation version of SFWSM uses simulated streamflow using weather forcing i), while the HHLSM validation and planning versions of SFWSM use simulated streamflow using weather forcing ii). When required, pre-processing steps are used to create input data to either PRMS/SACSMA-DS models. Details about these steps are provided in the Technical Report 2: Hydrological Modeling Module (HRG TR2, 2021). Output from these models included surface runoff and evaporation, as well as other hydrologic parameters not used by SFWSM. For the Upcountry region, the simulated streamflow from PRMS have then be bias corrected using a correction method based on an index of precipitation and index of temperature. More details are given in the Technical Report 2 (HRG TR2, 2021).

3.5.2 Evaporation

Evaporation is computed every time step for each reservoir in SFWSM as a function of the area of the respective reservoir at the end of the time step and the potential evaporation (in length units) at the reservoir. Surface area is derived from a volume-area curve, which is provided by SFPUC. Potential evaporation is estimated by using the potential evapotranspiration (PET) output from the hydrology models described in the Technical Report 2 (HRG TR2, 2021). Note that for Don Pedro reservoir, the use of the simulated PET from Don Pedro catchment (i.e., the catchment upstream Don Pedro reservoir but downstream Cherry/Eleanor and Hetch-Hetchy reservoirs) was resulting to an overestimate of the evaporation over the lake. To prevent the reservoir from evaporating too much, the parameter of the PET model used for Hetch-Hetchy catchment was used instead of the calibrated PET model parameter for Don Pedro catchment.

Evaporation is represented in Pywr as output nodes, which allow water to leave the system, with a maximum allowable flow equal to the computed potential evaporation over the reservoirs for the time step. To prevent LP infeasibilities, evaporation over the reservoirs is not represented as a minimum flow constraint out of the reservoir but rather by using a high negative cost (Table 3) to ensure evaporation water leaves the system. SFWSM tracks evaporation deficit that might occur when a reservoir volume drops to its minimum active storage level.

3.5.3 Precipitation

Precipitation is modeled similar to evaporation. Precipitation equals reservoir surface area times precipitation (in length units), with precipitation from the weather generator, as described in the Technical Report 1 (HRG TR1, 2018), and are represented in Pywr as input nodes.

3.5.4 Groundwater recharge

Two conjunctive use schemes—both in the Peninsula—have been implemented by SFPUC, for which groundwater hydrology is potentially important. However, due to the nature of the schemes and the lack of development of a groundwater hydrology model, groundwater recharge was not explicitly modeled. The representation of groundwater recharge is described for each respective scheme in Section 3.7.3.

3.6 Reservoir management

Reservoirs are operated to meet the wide range of SFPUC management objectives. Most of these management objectives are achieved with the LP-based approach by setting water demands (i.e., max flow in Pywr), costs, and so on for non-reservoir nodes, the logic for which is summarized in this section and in Section 3.7. More comprehensive descriptions are provided in the code.

3.6.1 Reservoir preferred storage levels

SFPUC operates all reservoirs with preferred storage levels for reservoirs, which increase during the flood season (approximately November – March) to ensure there is enough space in the reservoir to store higher inflows and avoid uncontrolled spill. For Don Pedro Reservoir, the preferred storage level is also for flood control. As noted by SFPUC (2009), preferred storage levels do not necessarily reflect a storage target, but rather as a "trigger to initiate water movement among the Bay Area reservoirs and into the distribution system". Thus, they act as "guide curves" typically used in reservoir operations generally. These preferred storage levels are defined in HHLSM with monthly values and were carried over to SFWSM directly. Interpolation was used to downscale to preferred storage levels from monthly to daily time steps. Preferred storage levels for each reservoir in the HHLSM validation version of SFWSM are shown in Figure 12, with curves shown in relative terms (i.e., percent of storage capacity). Preferred storage levels are used in SFWSM differently depending on the reservoir in question and vary slightly between model versions. For the SFWSM version to be used for planning, the preferred storage curves are input as time series to allow more flexibility for future use at SFPUC. This could allow, for instance, the addition of temporary reductions of storage capacity during maintenance.

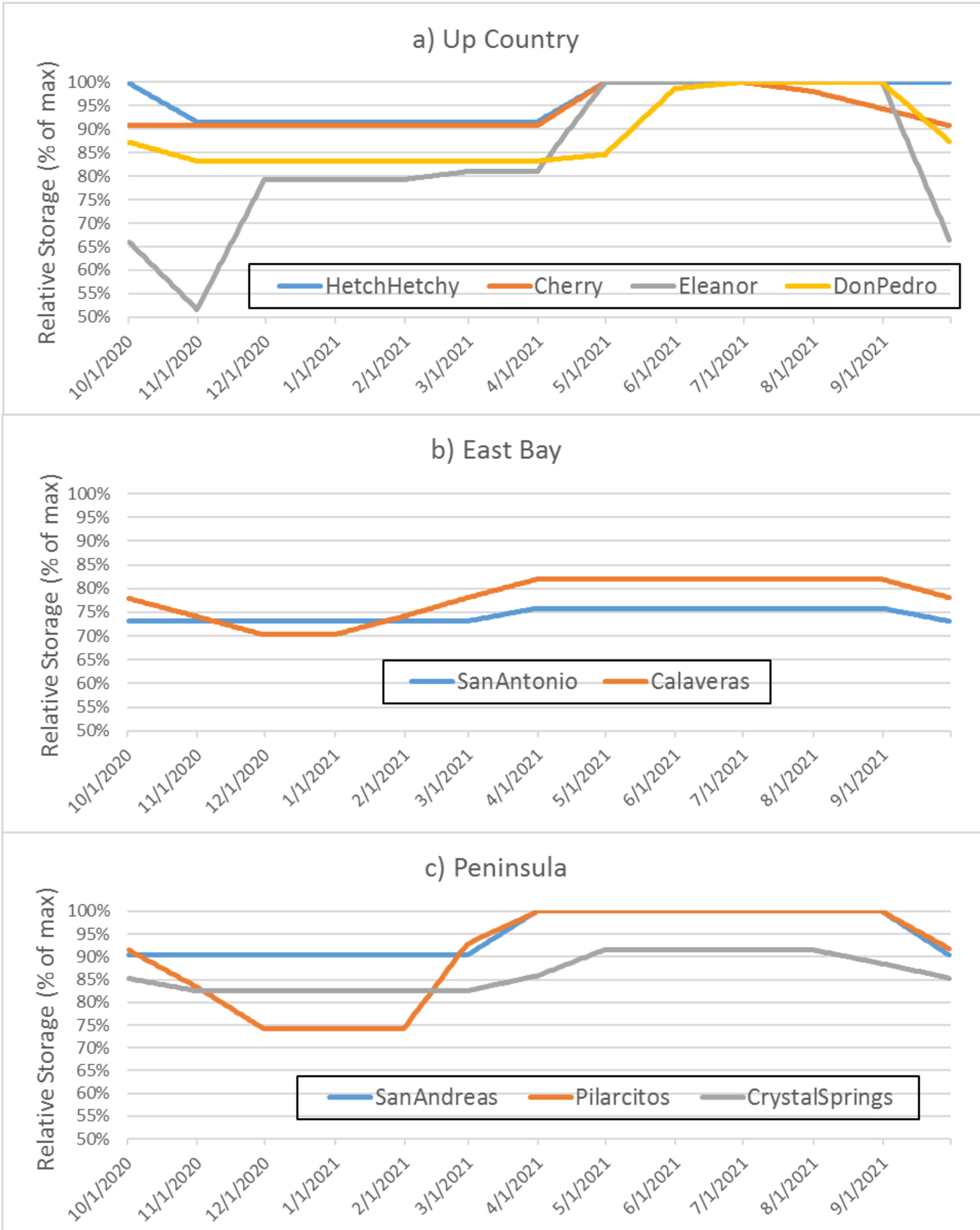

Figure 12. Relative preferred storage levels for reservoirs in each of the a) Upcountry, b) East Bay (Sunol) and c) Peninsula regions. Preferred storage levels are shown for the HHLSM validation version of SFWSM. Note that preferred storage at Calaveras and San Antonio reservoir are significantly lower than 100%. This is because storage capacity for these reservoirs in SFWSM is at the dam crest, which allows to account for a known rating curve for the controlled spill (see Sections 3.6.3).

3.6.2 Reservoir storage costs

In all cases, the cost (value) of reservoir storage is defined partly with preferred storage levels, with cost depending on relative storage in relation to the relative preferred storage level (Table 8). For all reservoirs, storage cost is linearly interpolated between fixed relative preferred storage levels below (0%, 10%, and 50%) and above (100%) the main preferred storage level. These costs have been manually calibrated in order to reproduce the observed storage fluctuation in the historical dataset (Section 5). However, values for some of these costs translate directly to the actual operation of the system by SFPUC. The cost for preferred storage level and above levels is null (zero), which represents the absence of interest of storing water above the preferred storage level. In addition, Table 8 shows that local reservoirs are valued higher than Upcountry reservoirs, which prioritizes the higher storage in local reservoirs over those in the Upcountry. Larger costs are assigned to Sunol Valley reservoirs than to Peninsula reservoirs, ensuring that the latter are prioritized over the former for meeting demand for Peninsula customers since HTWTP cannot supply all customers in South East Bay region.

Table 8. Storage cost value by relative storage for each reservoir in SFWSM.

Region	Reservoir	0%	10%	50%	Preferred storage level value	100%
Upcountry	Hetch Hetchy	-100	-20	-5	0	-0
	Cherry	-100	-75	-50	0	0
	Eleanor	-100	-10	-5	0	0
	Don Pedro	-2 (constant)				
Sunol Valley	Calaveras	-200	-150	-100	0	0
	San Antonio	-150	-100	-50	0	0
Peninsula	San Andreas	-100	-75	-50	0	0
	Pilarcitos	-100	-75	-50	0	0
	Crystal Springs	-100	-75	-50	0	0

3.6.3 Spill

Up to three link nodes represent spill at each reservoir. They are noted as i) *controlled spill*, ii) *uncontrolled spill* and iii) *spill*. '*Controlled spill*' nodes represent the water released through the outlet works prior the reservoirs reach their spillway elevation or storage capacity. '*Uncontrolled spill*' nodes represent the water released through the spillway. As detailed below, only Calaveras and San Antonio reservoirs have such nodes because the required information is missing for the other reservoirs. The '*spill*' nodes represent the spilled water when the reservoirs reach their maximum storage capacity. More detail about each of these nodes is provided below.

Controlled spill

Preferred storage levels are used to trigger controlled spill, though the exact trigger mechanism—and the spill amount—varies by reservoir. Controlled spill refers to the water released through outlet works. For Hetch Hetchy Reservoir, controlled release is a maximum of the guide curve-based release and releases for the Upper Tuolumne River Ecological Program (UTREP, described below), since both represent management of controlled spill. For all other reservoirs except for Pilarcitos and San Andreas, the maximum flow through controlled release depends on the respective reservoir guide curves in a binary manner, with controlled release at a specified capacity when the reservoir level is above the curve, and zero capacity when reservoir level is below the curve (Table 9). Most of these reservoirs have a single controlled spill volume; Calaveras has an interpolated release based on reservoir storage. These are implemented directly in the model JSON files, as they are simple switches.

Forecasting is used to help minimize spill. Forecasting is used with Hetch Hetchy, Cherry, and Eleanor reservoirs to anticipate spill for release under these preferred storage levels. Forecasting is used with Calaveras Reservoir to reduce diversions from ACDD if diversions would lead to spill. No transfer from Pulgas to Crystal Springs is the reservoir is at the preferred storage level. Forecasting is not used with other reservoirs.

Controlled spill nodes are assigned relatively high negative cost to ensure the controlled spill is respected. When there is no conflict with other reservoir operations, the controlled spill flow can also be defined as minimum flow constraint for the LP problem to further ensure the controlled spill is respected.

Table 9. Controlled release value by absolute storage.

Region	Reservoir	Controlled spill (AF/day)	Control type	Controlled spill cost
Upcountry	Hetch Hetchy	Variable (script-based)	Minimum & maximum flow	N/A
	Cherry	8,727	Maximum flow	-100
	Eleanor	1,388	Maximum flow	-100
	Don Pedro	3,000	Maximum flow	-100
Sunol Valley	Calaveras	0 to 1,110.74 (depends on storage level)	Minimum flow	N/A
	San Antonio	635	Maximum flow	-100
Peninsula	San Andreas	N/A		
	Pilarcitos	N/A		
	Crystal Springs	397	Maximum flow	-100

Uncontrolled spill

In SFWSM, uncontrolled spill represents the flow through the spillway. At the time of the creation of SFWSM, the relation between storage level and flow through the spillway was known for Calaveras and San Antonio reservoirs only. As such, a Pywr node *uncontrolled spill* is only implemented at Calaveras and San Antonio reservoirs, although it could be implemented in the future for the other reservoirs with minor modification of the code. No cost is associated with this node. However, both minimum and maximum flow constraints are equal to the flow value obtained from the relationship between spillway flow and storage level.

For Calaveras, SFPUC provided a rating curve between storage filling level and flow. For San Antonio, SFPUC provided a second order polynomial function that gives the flow through the spill as a function of the reservoir elevation (cf. Gitlab issue #84).*Spill*

The *spill* nodes represent the spilled water that does not fall within the definition of either controlled or uncontrolled spill. For Calaveras and San Antonio, *spill* nodes represent water spilled above the dam crest (which should always be 0 under normal system operations / climate). For the other reservoirs, *spill* nodes includes both uncontrolled spill though the spillway and spill above dam crest.

In SFWSM, no maximum constraint is assigned to *spill* nodes and a high positive cost is assigned (i.e., 1000) to ensure that spills are only effective when no other way to release water is able to maintain the reservoirs below their maximum storage capacity.

3.7 Subsystem operations

This section describes the operations of specific facilities, organized by subsystem. Subsystems are generally organized around one or more reservoirs.

All reservoirs receive inflows from contributing catchments (watersheds) and from precipitation; and lose water through evaporation and downstream releases typically consisting physically of a diversion conduit for hydropower generation or delivering water demand, IFRs and spill. In SFWSM, downstream releases are mostly subdivided as needed to align with specific management objectives, even if the same physical conduit is used.

3.7.1 Upcountry

Hetch Hetchy Reservoir

Hetch Hetchy Reservoir (Figure 13) is the keystone reservoir in the RWS. Managed releases from Hetch Hetchy include the Canyon Power Tunnel (for water supply and hydropower), two IFRs, controlled spill, uncontrolled spill, and valve (Hetch Hetchy valve) that is only for emergency releases to meet MID/TID demand when the Water Bank is empty. Controlled spill includes both snowmelt management releases under UTREP (as noted above and described below) and releases to keep storage levels below the Hetch Hetchy maximum preferred storage level (Figure 12). With the exception of releases to the Canyon Power Tunnel, all managed releases subsequently go directly to Don Pedro Reservoir (via the Tuolumne River); Canyon Power Tunnel releases may also go to Don Pedro, via the Kirkwood diversion.

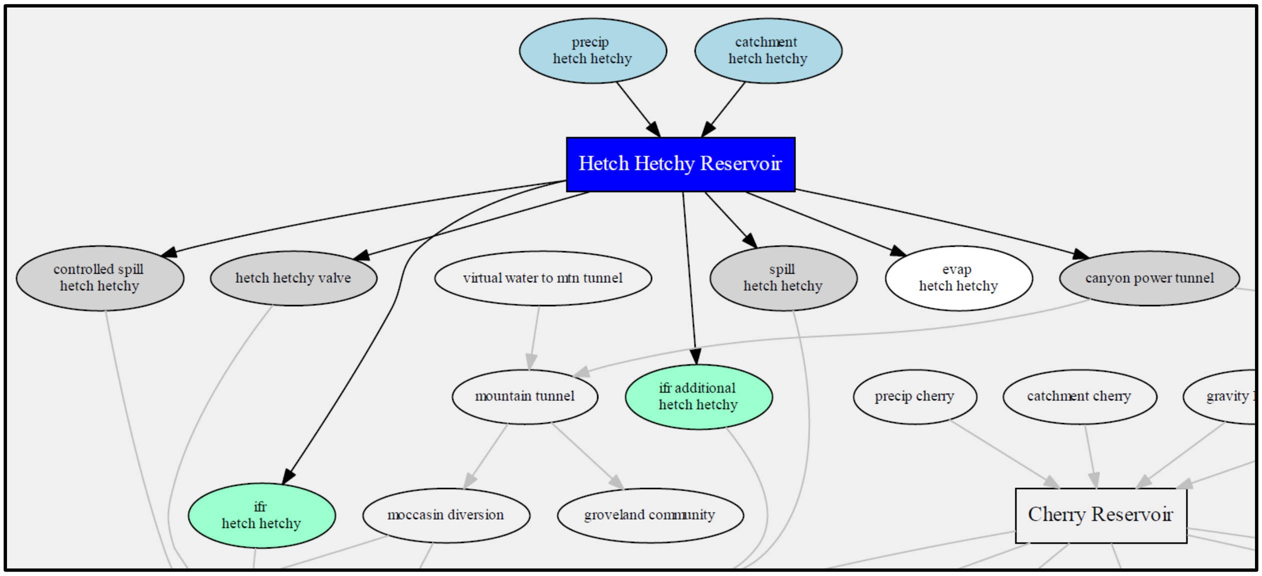

Figure 13. Schematic of flows in to and out from Hetch Hetchy Reservoir.

Controlled spill

Controlled spill from Hetch Hetchy includes both spill management objectives under UTREP (described further down in the section) and normal controlled spill for reservoir management. Each of these are calculated independently, with the actual controlled spill, specified as a minimum instream flow, being the maximum of the reservoir management release and UTREP spill. In contrast to most other reservoirs, which have controlled spill operated in an on/off manner, non-UTREP controlled spill is calculated in a script to include only what is necessary to keep the reservoir level near the Hetch Hetchy preferred storage level. This prevents too much water from being released as controlled spill during the flood period.

Canyon Power Tunnel

Operators of Hetch Hetchy reservoir release water through the Canyon Power Tunnel to send water to the Mountain tunnel/San Joaquin Pipelines to satisfy SFPUC customer demand and to local reservoirs. Water can also be released through the Canyon Power Tunnel, whose capacity is larger than the Mountain tunnel, to Don Pedro reservoir to minimize spill and, in very rare instances, to supply MID/TID water rights directly. The volume of water that exceeds the Mountain tunnel capacity is diverted via the Kirkwood diversion and eventually flows into Don Pedro reservoir. However, being a “water first” organization, SFPUC aims at limiting such water transfers from Hetch Hetchy to Don Pedro reservoir. Several rules constrain the operation of the Canyon Power Tunnel. In SFWSM, three attributes control the operation of the Canyon Power Tunnel, represented as a Pywr Node: a maximum flow constraint, a minimum flow constraint and the cost associated with the node.

The maximum flow through the tunnel, which in SFWSM equals 2,776 AF/day, is reduced to zero when Hetch Hetchy storage is lower than 40,000 AF.

A minimum flow constraint is used to avoid or reduce spill from February 1st to June 30th. A set of conditions must be met for SFWSM to trigger the minimum flow constraint. If the below condition is met, the capacity of the Canyon Power Tunnel (i.e., 2,776 AF/day) is used as a minimum flow constraint. Otherwise, the minimum flow constraint is set to 0.

First, the Hetch Hetchy reservoir storage must be larger than 150 TAF. This storage threshold was set after discussion with SFPUC personnel to prevent the model, which forecasts inflow to the reservoir with perfect foresight, for emptying the reservoir to minimize spill. Second, the total forecasted net inflow to the Hetch Hetchy reservoir must be larger than the available space in the reservoir (defined as the difference between the preferred storage on July 1st and the current storage). The total forecasted net inflow is defined as the inflow to Hetch Hetchy Reservoir during this period reduced by the expected IFRs below the reservoir and a climatologically adjusted demand to the San Joaquin Pipelines (estimated from historical flows through the pipelines during this period). If the two above conditions are met, the next condition to check depends on the water bank storage. If water bank storage is lower than 550 TAF, the minimum flow constraint is set to 2,776 AF/day. The logic here is that the water release to Don Pedro Water bank through the Canyon Power tunnel could be used to fill Hetch Hetchy reservoir at a later time. On the other hand, if water bank storage is larger than 550 TAF, any water release from Hetch Hetchy to the bank is likely to be spilled at water bank. This is only acceptable if the forecasted Water Available to the City (WAC) is larger than the available space in the Upcountry region plus the forecast water demand. The former is the sum of the available spaces at Hetch Hetchy, Cherry, Eleanor and Water Bank reservoirs. The latter is the sum of the demand to the San Joaquin Pipelines plus the evaporative demand over the Upcountry reservoirs.

When the current time step is after July 1st or prior February 1st the Canyon Power Tunnel can still be operated at flow rates larger than the Mountain tunnel to avoid spill. In such a case, a short-term forecast (20 days) of net inflow into Hetch Hetchy is considered. When the volume of inflows to Hetch Hetchy is larger than available storage, which accounts for current reservoir storage, preferred storage level, instream flow requirements and total system water demand, the cost attribute of the Canyon Power Tunnel node is set to a value highly negative, which gives a strong incentive to the LP optimization to use the canyon at full capacity. Otherwise, the cost associated to the Canyon Power Tunnel node is set to zero by default.

IFR in the Tuolumne River below O'Shaughnessy Dam

There are two flow requirements in the Tuolumne River below O'Shaughnessy Dam, based on various agreements between SFPUC and the U.S. Department of Interior (SFPUC, 2009a):

1. *Base flow requirements* – This is the base IFR below O'Shaughnessy Dam. There are three release requirements, corresponding to three water year types: A, B and C (Table 10). The year types are defined by a combination of cumulative precipitation at Hetch Hetchy (Jan-Jun) and reservoir storage (Jul-Aug).

2. *920/64 rule* – This rule stipulates that whenever flow through Canyon Power Tunnel equals or exceeds 920 cfs in year types A and B, an additional 64 cfs should also be released into the Tuolumne River.

The above rules are modeled using two respective Pywr nodes with differing maximum flows based on scripted logic implementing the rules. Base flow requirements were modeled directly using the original requirements (Table 10). The 920/64 rule is modeled by specifying the requirement based on the previous day's release.

Table 10. Base flow requirements below O'Shaughnessy Dam (SFPUC, 2009a).

	Year Type A		Year Type B		Year Type C
Month	Release (cfs)	Criteria[a, b]	Release (cfs)	Criteria[a, b]	Release (cfs)
January	50	8.80"	40	6.10"	35
February	60	14.00"	50	9.50"	35
March	60	18.60"	50	14.20"	35
April	75	23.00"	65	18.00"	35
May	100	26.60"	80	19.50"	50
June	125	28.45"	110	21.25"	75
July	125	575,000 acre-feet	110	390,000 acre-feet	75
August	125	640,000 acre-feet	110	400,000 acre-feet	75
September 1-14	100		80		75
September 15-30	80		65		50
October	60		50		35
November	60		50		35
December	50		40		35

[a] Precipitation indicator in inches is cumulative, measured at Hetch Hetchy Reservoir, starting October 1.
[b] Runoff indicator in acre-feet is the calculated inflow into Hetch Hetchy Reservoir commencing on the previous October 1.

Upper Tuolumne River Ecological Program (UTREP)

One goal of Upper Tuolumne River Ecological Program (UTREP) is to manage spill from Hetch Hetchy reservoir in a way that is ecologically beneficial, where "spill" is water in excess of other management priorities. The general approach to achieve UTREP spill management objectives is to release water in a controlled manner as an ecologically beneficial hydrograph whose total volume equals the anticipated spill. UTREP releases are managed via the Hetch Hetchy valve.

Several "template" hydrographs have been created for this purpose: depending on the anticipated spill, a different template hydrograph is used as a basis for defining the final hydrograph to be released. Each hydrograph consists of functional flows and flow constraints (e.g., ramping rate constraints) intended to achieve specific ecological functions or avoid ecological harm. Template hydrographs are selected based on forecasted spill threshold values, with the hydrograph modified to account for spill in excess of the threshold value. Hydrographs are further modified in real-time, depending on actual hydrologic conditions and reservoir storage. For instance, UTREP releases may be curtailed if reservoir storage falls too low. UTREP releases are generally about 30-40 days long. The initiation of UTREP releases varies, depending on forecasted conditions and reservoir operation objectives. Releases under the UTREP, while providing ecological benefits, are not considered as instream flow requirements *per se*, but rather a re-shaping of water through time that would otherwise be spilled.

An algorithm to model UTREP release targets was developed from the quantitative UTREP guidance found in (SFPUC, 2014), in consultation with SFPUC staff. Within Pywr, UTREP releases are via a dedicated node below Hetch Hetchy, which has minimum and maximum flows equal to the UTREP calculated release target. Generally, the algorithm to calculate UTREP releases is based on the following principles:

1) Base UTREP requirements are from template hydrographs pre-generated based on Table 21 in SFPUC (2014), as updated and provided by SFPUC staff.
2) UTREP spill initiation begins:
   a) when the reservoir will exceed 300 TAF in the next seven days,
   b) when reservoir storage is above a specified threshold,
   c) no earlier than April 1, and
   d) no later than May 15
3) Spill volume is calculated on April 1, based on assumed perfect forecast of inflows through July 1 (the target Hetch Hetchy refill date) less storage space, power tunnel capacity, and base instream flow requirements (including the 920/64 rule). Forecasted spill is calculated using this logic with a daily time step subroutine for simplified Hetch Hetchy operations applied for April 1 through July 1.
4) For forecasted spill above thresholds:
   a) For the four lowest thresholds and the 104,000 AF threshold, the hydrograph is extended in time at the last fixed-flow shelf.
   b) During other years, the template hydrograph is scaled linearly toward the template hydrograph with the next higher threshold.
5) UTREP releases are curtailed when reservoir storage levels decrease to below several specified thresholds, based on buffer coefficients established for each threshold; UTREP releases are stopped completely if the reservoir storage decreases to below a third storage threshold.

Hydrographs were adjusted slightly to ensure their total volume equals their associated spill threshold, as the algorithm did not work without this alignment. Resulting adjusted template hydrographs are shown in Figure 14. Buffer coefficients (Table 11) were developed through trial-and-error to approximate operations. During model validation, it was found that the algorithm did not perfectly represent ideal UTREP operations. However, the algorithm was not further refined, as imperfect modeled operations were found to generally reflect actual challenges in operating Hetch Hetchy Reservoir for UTREP.

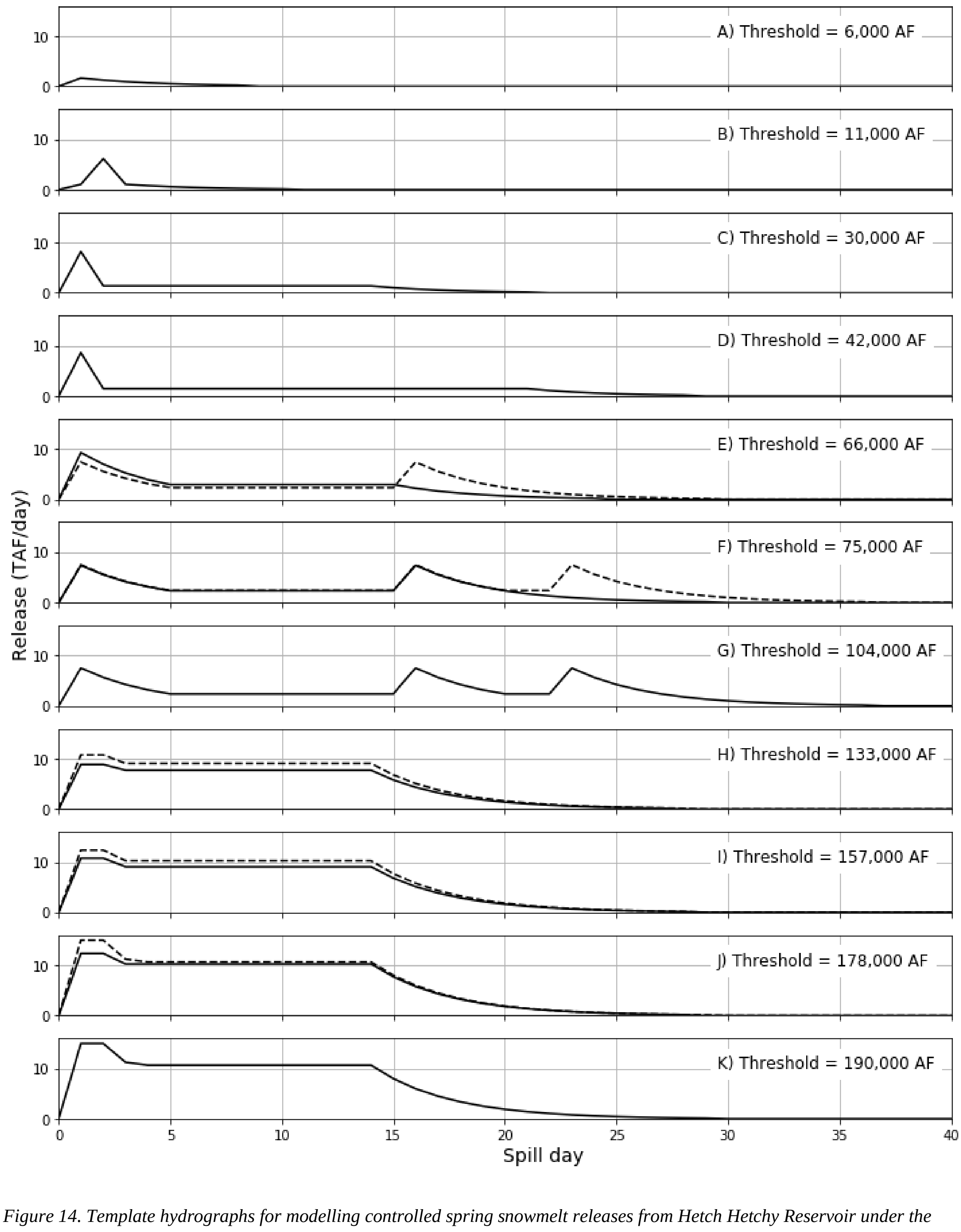

Figure 14. Template hydrographs for modelling controlled spring snowmelt releases from Hetch Hetchy Reservoir under the Upper Tuolumne River Ecological Program (UTREP). Extra spill is allocated either to the next template hydrograph (dashed lines) or to extensions of main hydrograph shelves (extensions not shown).

Table 11. Storage thresholds and buffer coefficients to curtail releases Upper Tuolumne River Ecological Program (UTREP).

Storage threshold (TAF)	Buffer coefficient
≤ 200	0
210	0.25
225	0.5
250	0.75
360	1.0

Lake Eleanor and Cherry Lake

Lake Eleanor and Cherry Lake (also called Eleanor Reservoir and Cherry Reservoir, respectively) are operated conjunctively, whereby water is transferred from Lake Eleanor to Cherry Lake by gravity flow or pumping via a connecting pipeline (Figure 15). Releases from Eleanor thus consist of pumping/gravity flow releases, an IFR, controlled spill, and uncontrolled spill. The pumping/gravity flows go to Cherry Lake, while other releases continue to Don Pedro Reservoir. Cherry Lake, therefore, receives inflow from the pumping/gravity flows, in addition to normal inflows. Outflows from Cherry Lake include controlled spill, uncontrolled spill and a single IFR, each of which continue to Don Pedro Reservoir.

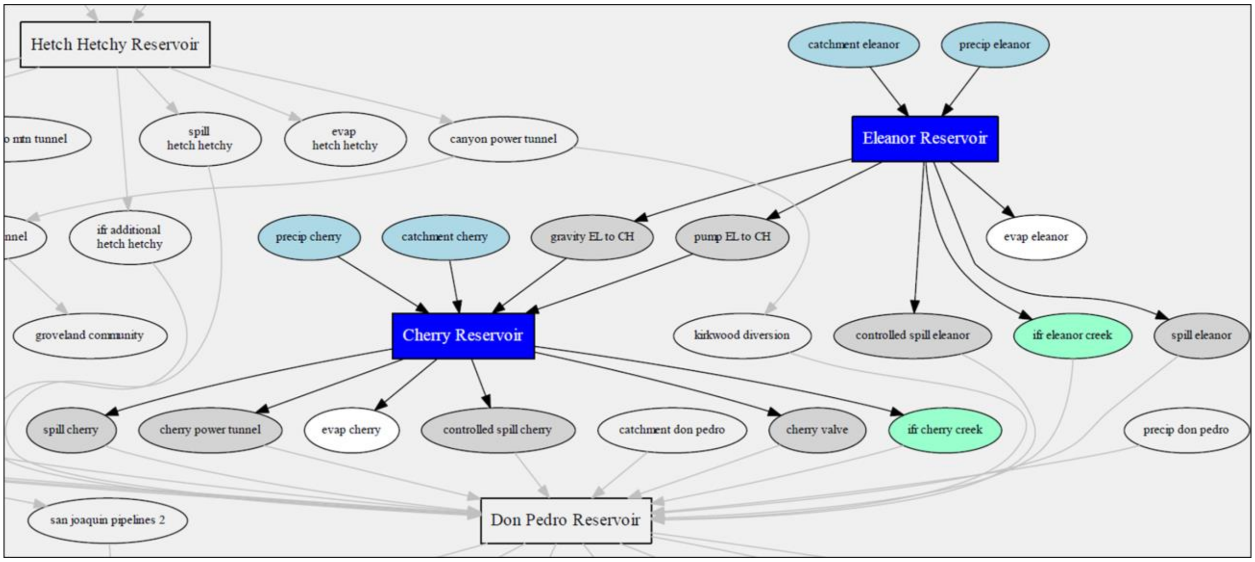

Figure 15. Schematic of flows in to and out from both Lake Eleanor (Eleanor Reservoir) and Cherry Lake (Cherry Reservoir).

Controlled spill

The controlled spill for both Lake Eleanor and Cherry Lake are implemented with a node with maximum flow constraint in a Pywr node below each respective reservoir which either zero if storage is below the preferred storage level, or takes a positive value if storage gets above the preferred storage. Controlled spill from Cherry and Eleanor, if any, equals 8,727 AF or 1,388 AF respectively.

Eleanor-Cherry transfers

The purpose of the Eleanor-Cherry transfer is to minimize total spill from the two reservoirs, while meeting other management objectives. Eleanor-Cherry transfers are modeled by connecting Lake Eleanor with Cherry Lake with two independent nodes, one each for pumping and gravity flow, with a required flow specified for each. Transfer generally occurs when spill is expected to occur in the next 90 days based on an assumed 90-day perfect forecast and accounting for current storage and instream flow requirements below Lake Eleanor. Gravity flow is used when net head is greater than zero, and when Lake Eleanor equals or exceeds either 17,900 AF (Jul-Sep) or 5,000 AF (all other months). Otherwise, pumping is used. The volume of transferred water depends of the head difference and results from an interpolation of the "head difference–flow" curves provided by SFPUC. Note that no pumping is not allowed after January 1st till Eleanor spills more than 50 cfs. Note also that transfers do not consider storage or potential spill from Cherry Lake.

Cherry Power Tunnel

The Cherry Power Tunnel, with a capacity of 1,924 AF/day, is operated using similar logic as the Canyon Power Tunnel, described above, including ceasing operations if Cherry storage is below 40,000 AF. The operation of the Cherry Power Tunnel from the Canyon Power Turnnel from two aspects. First, a minimum flow constraint of 9,000 AF/day is set during June, July and August months to supports rafting activities below the dam. Second, outside the February through June period, a minimum flow constraint is set to supply with Kirkwood and Moccasin power plants a target daily load (i.e., 2576.2 MWh). The residual load to be supplied by Holm power plant (i.e., Cherry Lake) is estimated from the average generation from Kirkwood and Moccasin power plants during the previous 5 days. The residual load is only supplied if Cherry reservoir storage is larger than 100 TAF and water bank is lower than 550 TAF.

In HHLSM, the use of the Cherry Power Tunnel is set to null (zero) if the Water Bank storage is greater than 90% of the maximum water bank account storage for power release. This operation is not accounted for in SFWSM for two reasons. First, the minimum flow constraint associated with the power tunnel already accounts for Water Bank via the WAC. Second, the implementation of the above threshold was found to result in an increase spill at Cherry when Water Bank was higher than 90% of its storage capacity.

IFR in Cherry Creek below Cherry Lake (Lake Lloyd)

The IFR below Cherry Lake is a minimum flow requirement that varies by time of year, with 5 cfs (9.9 AF/day) during October – June and 15.5 cfs (30.7 AF/day) from July – September. This is modeled as a Pywr link node with a pre-processed time series input for maximum flow and high value (negative cost).

IFR in Eleanor Creek below Lake Eleanor

The IFR below Lake Eleanor is a minimum flow requirement that varies by time of year and whether or not water is pumped from Lake Eleanor to Cherry Lake. Minimum release values, modeled as a Pywr link node with script-based maximum flow definition and high value (negative cost), are depicted in Table 12.

Table 12. Minimum releases in Eleanor Creek below Lake Eleanor (SFPUC, 2009a).

	With Pumping [a]		Gravity Flow Without Pumping [a]	
Month	(cfs)	(acre-feet)	(cfs)	(acre-feet)
October	10	615	5	307
November	5	298	5	298
December	5	307	5	307
January	5	307	5	307
February	5	278	5	278
March	10	615	5	307
April 1-14	10	278	5	139
April 15-30	20	635	5	159
May	20	1,230	5	307
June	20	1,190	5	298
July	20	1,230	16	953
August	20	1,230	16	953
September 1-15	20	595	16	461
September 16-30	10	298	16	461
Total		9,106		5,535

[a] The agreement for the operation of the Eleanor-Cherry Tunnel and Pumping Plant calls for different fishery release schedules below Eleanor Dam depending on whether or not the pumping plant is used.

Water Bank

SFPUC water rights for diverting from the Tuolumne River are junior to MID/TID. Based on prior agreement with the Districts, SFPUC operates a water accounting scheme called the Don Pedro Water Bank, whereby water released by SFPUC to Don Pedro Reservoir is accumulated as a credit for later use by the Districts. Subtractions from the Water Bank are the District Entitlement. The Water Bank has a maximum allowable storage of 570 TAF, plus 1/2 of the available storage space if there is sufficient flood storage space (this is known as the “bubble account”). This arrangement provides SFPUC some flexibility in operations, as it allows them to choose when to release water to fulfill Water Bank obligations. However, if the Water Bank is depleted, SFPUC is obligated to respect the actual District Entitlements, and may be required to curtail its own abstractions from the Tuolumne River. SFPUC’s operational objective is therefore to keep the Water Bank as full as possible from excess supply during wet years and to draw down the bank by releasing less into it during dry years, resulting in more stable and predictable operations of its reservoirs. These operations are depicted in Figure 16, with modeled behavior using inflow time series realizations generated for the LTVA. Refill occurs during wetter years (e.g., 2045 and 2050 in Figure 16a) and drawdown during drier years (e.g., 2044, 2047 and 2048 in Figure 16a).

The Water Bank is modeled as a Pywr *virtual reservoir* that tracks accumulation and depletion of water in exactly the same way as actual operations. Accumulation consists of a summation of all inflows to Don Pedro Reservoir. Depletion consists of the District Entitlements and evaporation. Evaporation from the Water Bank is defined as the fraction of the evaporation over Don Pedro in regards with the ratio between Water Bank storage and Don Pedro storage. Note that the additional flood encroachment space available during the end of the flood season is not included in the model, as this space is ephemeral and was deemed insignificant for LTVA purposes by SFPUC staff.

If the Water Bank storage, decreases significantly during a drought, water from Cherry, Eleanor and Hetch Hetchy Reservoirs can be requested by SFWSM to ensure supply of the district right. In SFWSM, the cost associated with the Cherry Power Tunnel takes a large negative value when the Water Bank storage is below 35,000 AF, increasing the incentive for the LP model to send water to the Water Bank. If the Water Bank storage drops to below 10,000 AF, outflow from the valve at Cherry Reservoir is used to complement the remaining balance and satisfy the district right. The capacity of the valve at Cherry has been estimated from historical observations to 7,000 AF/day. If there is still a remaining balance or if water available from Cherry is not sufficient, water from Eleanor is used following the same logic, with a valve capacity of 4,000 AF/day. The same logic is also applied for Hetch Hetchy, with a valve capacity of 18,000 AF/day. Note that SFWSM calls in priority water from Cherry, then Eleanor and eventually, if necessary, water from Hetch Hetchy. These drought operations are demonstrated with two drought periods in Figure 16, derived from two climate realizations developed for the LVTA. In one drought (Figure 16a), valve releases are not used, while in the other (Figure 16b), the Cherry valve is used. While typically refill occurs via Holm PH, Kirkwood PH may also be used to fill the Water Bank.

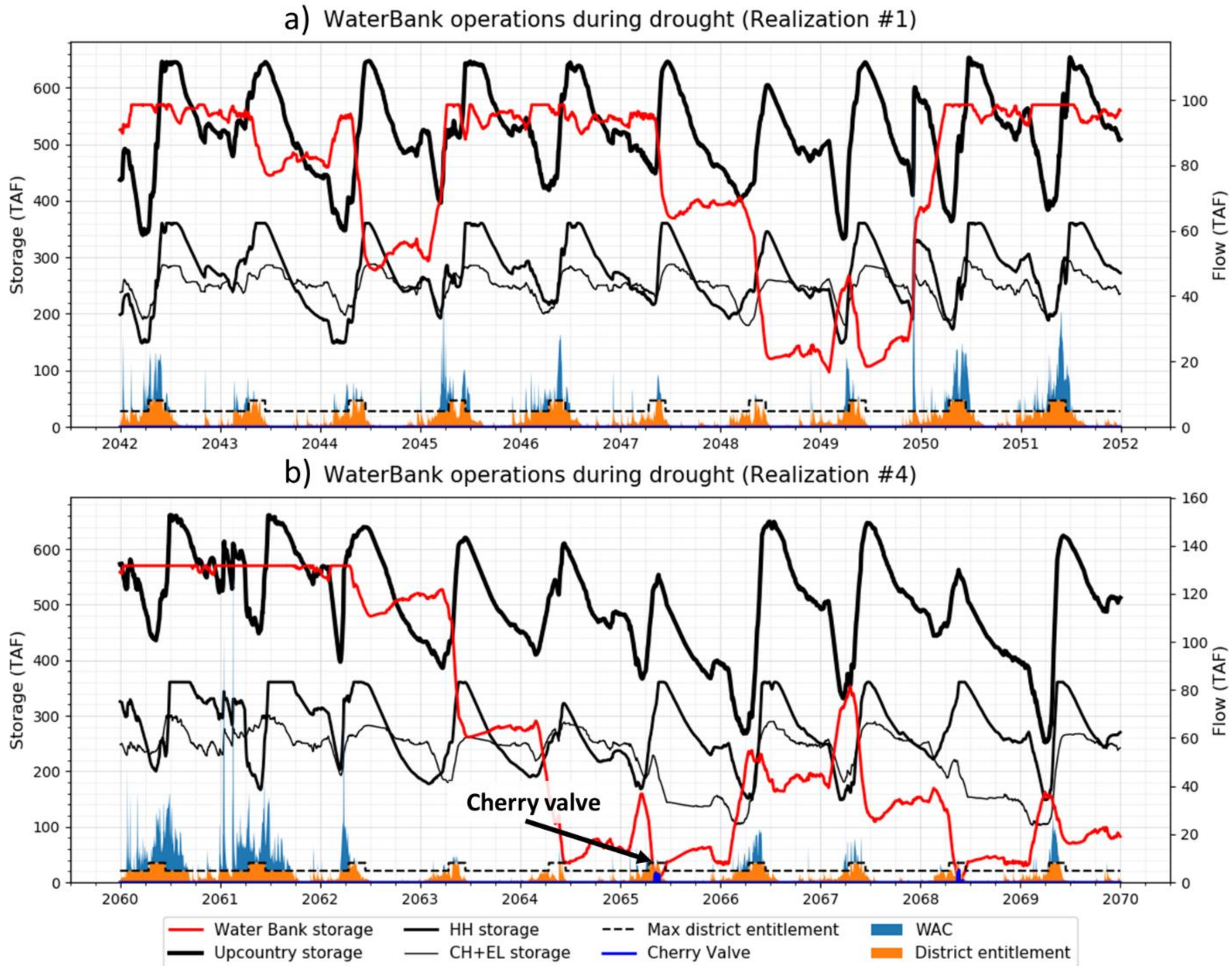

Figure 16. Example Don Pedro Water Bank operations during drought for two climate realizations (historical climate conditions). Upcountry Storage includes Hetch Hetchy Reservoir, Lake Eleanor, and Cherry Lake. WAC is Water Available to the City. Climate realizations are described by UMass (2018).

Don Pedro Reservoir

Don Pedro Reservoir (Figure 17), a major multi-purpose reservoir in California, receives inflows from each of the discrete outflows from respective upstream reservoirs (IFRs, controlled spill, and uncontrolled spill for each of Hetch Hetchy, Eleanor and Cherry reservoirs), as well as from the Cherry Power Tunnel, Kirkwood diversion below the Canyon Power Tunnel, and Moccasin Fish Hatchery. Outflows include diversions to each of MID and TID, in addition to an IFR, controlled spill, and uncontrolled spill. Since Don Pedro is near the system boundary of the SFPUC RWS, each of the outflows from Don Pedro are also outflows from SFWSM as a whole. Note that MID and TID irrigation canal diversions are described in the Demand section above (Section 3.4).

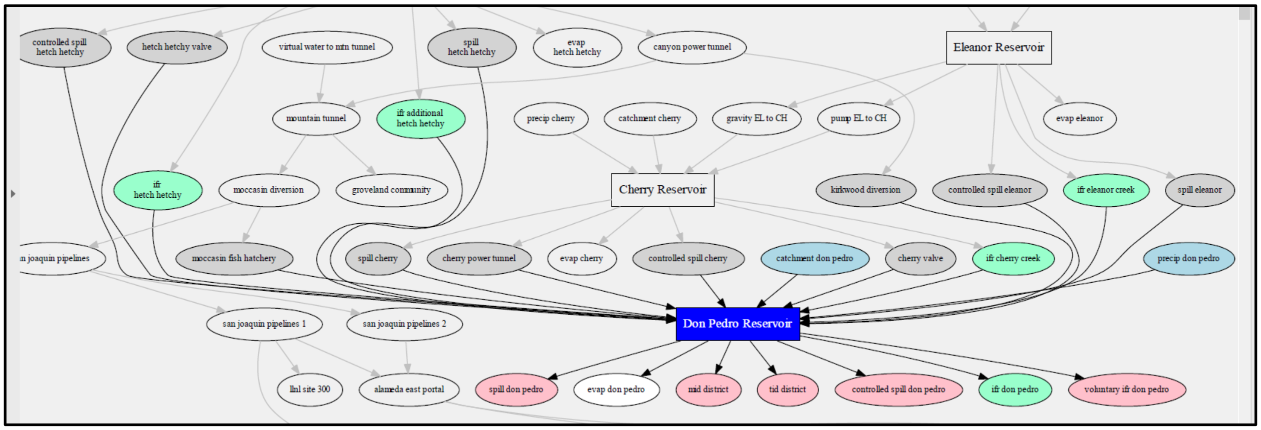

Figure 17. Schematic of flows in to and out from Don Pedro Reservoir.

Controlled spill

The controlled spill for Don Pedro Reservoir is implemented using a minimum flow requirement triggered by the preferred storage level. Controlled spill, when triggered, is a constant 9,000 cfs (17,851 AF) when spill is required. This flow value corresponds to the capacity of a powerhouse located downstream (Modesto). Unlike other Upcountry reservoirs, controlled spill for Don Pedro does not use a forecast for operations.

IFR in Tuolumne River at La Grange (below Don Pedro Reservoir)

The license issued by the Federal Energy Regulatory Commission (FERC) to operate the New Don Pedro Hydroelectric Project includes instream flow requirements at La Grange (below Don Pedro Reservoir and diversions to Modesto and Turlock Irrigation Districts' diversions) that depend on the SJVI. The license defines 10 year type classifications, listed in Table 13, which are used in determining the actual flow requirements, listed in Table 14. Flow requirements include an annual volume, which is allocated between base flows and seasonal attraction and out migration pulse flows (Table 14). For SJVI values between FERC license breakpoints, annual volumes are linearly interpolated between those listed.

These FERC-mandated flow requirements are modeled as a Pywr output node (water downstream of this node is outside the system boundary) with a maximum flow and high value. As with other IFRs, these flow requirements are pre-processed, since flow requirements do not depend on system state. Flow requirements are calculated exactly as with HHLSM, with the SJVI calculated using perfect foresight. Like in HHLSM, additional annual water calculated through interpolation is allocated to the out-migration pulse flows, proportioned between April and May. Attraction pulse flow during wet years are not accounted for by SFWSM.

Table 13. Year type classifications for instream flows in the Tuolumne River at La Grange as mandated by the FERC license for the New Don Pedro Hydroelectric Project.

FERC Year Type Classification	San Joaquin Valley Hydrologic Classification 60-20-20 Index (1,000 AF)
Critical and Below	<1,500
Median Critical	1,500
Intermediate Critical / Dry	2,000
Median Dry	2,200
Intermediate Dry / Below Normal	2,400
Median Below Normal	2,700
Intermediate Below Normal / Above Normal	3,100
Median Above Normal	3,100
Intermediate Above Normal / Wet	3,100
Median Wet / Maximum	3,100

Table 14. Instream flow requirements in the Tuolumne River at La Grange as mandated by the FERC license for the New Don Pedro Hydroelectric Project.

	FERC Year Type Classificatoin						
Period	Critical and Below	Median Critical	Intermed Critical / Dry	Median Dry	Intermed Dry / Below Normal	Median Below Normal	Intermed Below Normal / Above Normal and Above
Annual Volume (acre-feet)	94,000	103,000	117,016	127,507	142,502	165,002	300,923
October 1 – 15	100	100	150	150	180	200	300
Attraction Pulse Flow (acre-feet)	None	None	None	None	1,676	1,736	5,950
October 16 - May 31	150	150	150	150	180	175	300
Out migration Pulse Flow (acre-feet)	11,091	20,091	32,619	37,060	35,920	60,027	89,882
June 1 – September 30	50	50	50	75	75	75	250
Units: cfs unless otherwise noted.							

San Joaquin Pipelines

The San Joaquin Pipelines (SJPL) are comprised of a set of three parallel pipelines that form the longest segment of the conveyance system carrying water from the Upcountry to the East Bay. In SFWSM, SJPLs are modeled as two pipelines that aggregate all conveyances from the Moccasin Diversion Dam to Alameda East Portal in the Sunol Valley. The first SJPL (SJPL A[8]) has a maximum capacity of 900 AF/day (288 MGD). To avoid high frequency fluctuations in the pipeline flow, which is a common side-effect of LP-based simulation models, as discussed in Section 2.1.2 SJPLA maximum flow constraint is kept constant for a minimum of 21 days. The minimum flow for SJPLA is defined as the minimum value between 245 AF/day (80 MGD), which is the SFPUC minimum operating flow through the SJPL, and half of the maximum flow defined by the 21-day flow rule (i.e., there is a maximum ramp down constraint of 50%). A new maximum flow constraint is set at the end of each 21-day window. There is also a maximum ramp up

[8] To prevent confusion between the real name of the pipelines and the name of the two pipelines in SFWSM are noted SJPLA and SJPLB, while in the code they are named 'san_joaquin_pipelines_1' and 'san_joaquin_pipelines_2', although they do not correspond to the actual pipelines 1 and 2.

constraint of 500 AF/day from the previous 21-day period flow rate. No cost is associated with SJPLA.

The second SJPL (SJPLB) has a dynamic maximum capacity and a high cost, which deprioritizes flow. SJPLB maximum capacity ranges from 76 AF/day (25 MGD) to 715 AF/day (233 MGD) so that total capacity (i.e., SJPLA+SJPLB) equals 960 AF/day (313 MGD). This approach, developed in consultation with SFPUC, ensures that SJPL is not used at capacity when Bay Area sources are available, whih is consistent with operating practice. For example, if Bay Area reservoirs are below the emergency storage target then SJPL B could be positive flow to refill the Bay Area reservoirs. Another example is during summer days with high demand, SJPL B could be positive even when SJPL A is at capacity.

3.7.2 East Bay (Sunol Valley)

Calaveras Reservoir

Calaveras Reservoir (Figure 18) captures water from several local catchments in the Alameda Creek basin, including the Calaveras Creek and Arroyo Hondo catchments. Additionally, water is diverted from Alameda Creek itself via the Alameda Creek Diversion Dam (ACDD) and associated tunnel. Outflows from Calaveras Reservoir include the typical IFR, controlled spill, uncontrolled spill and spill above dam crest, each of which contribute to the "recapture diversion", a Pywr link added to SFWSM to allow IFRs below the ACDD to be recaptured via the Alameda Creek Recapture Project (ACRP), described below. Additionally, water may be released from Calaveras Reservoir to either the Sunol Valley WTP (SVWTP) or to San Antonio Reservoir (described below) via gravity flow. Both of these first pass through an intermediary Pywr link representing the actual release conduit.

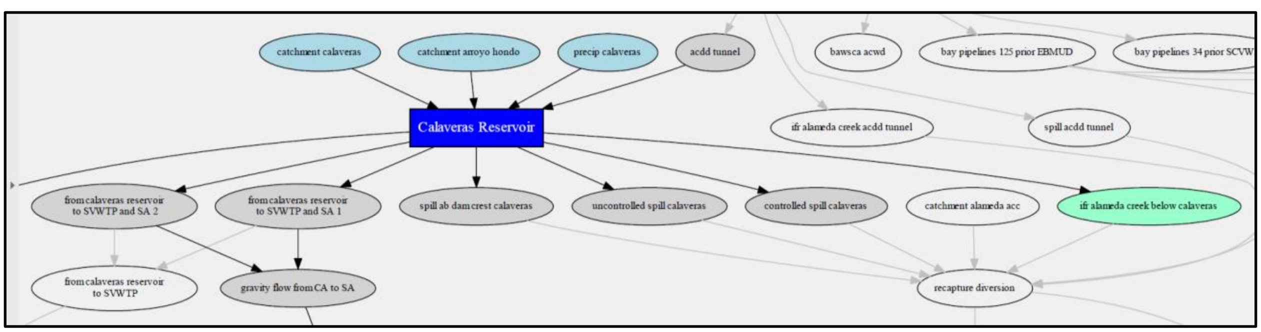

Figure 18. Schematic of flows in to and out from Calaveras Reservoir.

<u>Controlled spill</u>

Calaveras Reservoir spill is managed following the logic described in Section 3.6.3.

National Marine Fisheries Service Biological Opinion

In 2011, the National Marine Fisheries Service (NMFS) issued a Biological Opinion (BO) to the U.S. Army Corps of Engineers for Calaveras Dam upgrade work that included, among other things, recommended minimum instream flow requirements and ramping rates in Calaveras Creek below Calaveras Reservoir and minimum releases below Alameda Creek Diversion Dam (ACDD). These proposals, developed with SFPUC input and described further below, are included in SFWSM for instream flow requirements below the respective dams.

These BO instream flow requirements have prompted the development of the Alameda Creek Recapture Project, described below.

IFR in Alameda Creek below Alameda Creek Diversion Dam

Note: This operation is currently included in the HHLSM version of SFWSM.

The NMFS BO includes a minimum flow (called a *minimum bypass* flow in the BIOP) of 30 cfs (59.5 AF/day) below ACDD from December 1 through March 31. This is represented in SFWSM as a constant; ACDD diversions are disabled April through November. The ACDD tunnel is not activated in the validation version of SFWSM.

IFR in Calaveras Creek below Calaveras Reservoir

Note: This operation is currently included in the HHLSM version of SFWSM.

The proposed minimum instream flow schedule for Calaveras Creek below Calaveras Reservoir, listed in Table 15, includes flows for Dry and Normal/Wet years, which are calculated based on cumulative flows for the water year in Arroyo Hondo (a tributary to Calaveras Creek that flows directly into Calaveras Reservoir). The proposed ramping rates—including both ramp up and ramp down rates—also depend on water year, as listed in Table 16. The minimum instream flows are included in SFWSM as a pre-processed time series input. As noted above, ramping rates were not included. Because of a positive bias in the East Bay hydrologic model during the fall season, the dry schedule (B) from January 1 to April 30 is much less frequent than historical.

Table 15. Minimum instream flow schedule for Calaveras Creek below Calaveras Reservoir as proposed by the National Marine Fisheries Service and implemented in the San Francisco Water System Model.

Flow Schedule Decision Date	Flow Schedule Application period	Dry (Schedule B)		Normal/Wet (Schedule A)	
		Cumulative Arroyo Hondo flows for water year classification (MG)[1]	Flow release (cfs)	Cumulative Arroyo Hondo flows for water year classification (MG)	Flow release (cfs)
N/A	October	N/A	7	N/A	7[2]
N/A	Nov.01 – Dec. 31	N/A	5	N/A	5
Dec. 29	Jan. 01 – Apr.30	≤ 360	10[2]	> 360	12[2]
April 30	May 01 – Sept. 30	≤ 7,246	7	> 7,246	12

Notes:
[1] MG = million gallons
[2] Flows would be ramped as shown in the daily schedule in Table 3.

Table 16. Maximum ramping rates for Calaveras Creek below Calaveras Reservoir as proposed by the National Marine Fisheries Service and implemented in the San Francisco Water System Model.

Dates	Dry[1] (Schedule B)	Normal/Wet[2] (Schedule A)
10/1-10/2	7	9 (ramping down)
10/3 to 10/31	7	7
11/1 to 12/29	5	5
12/30	5	7 (ramping up)
12/31	7 (ramping up)	10 (ramping up)
1/1 to 3/31	10	12
4/1 to 4/30	10	12
5/1 to 9/30	7	12

Notes:
[1] The threshold value for dry (Schedule B) and normal/wet years (Schedule A) is 60 percent exceedance probability. Sixty percent of the time, cumulative flows in Arroyo Hondo would be higher than the dry year thresholds identified in Table 2. The "dry" schedule would apply to 40 percent of all months.
[2] Normal/wet schedule would apply to 60 percent of all months.

San Antonio Reservoir

San Antonio reservoir (Figure 19) includes gravity flow from Calaveras Reservoir, Pumped flow from Alameda East Portal (i.e., from the Upcountry region), and, optionally, from the Recapture Pit (Pit #2), a part of the Alameda Creek Recapture Project (described below). In addition to controlled and uncontrolled spill, water from San Antonio Reservoir can be released via either pump or gravity to the Sunol Valley WTP.

Reservoir storage management and controlled spill

Through a license agreement with the California State Water Resources Control Board (SWRCB), no more than 15,300 AF can be diverted within from San Antonio Creek within a Water Year. San Antonio Reservoir must be operated to account for this, as reservoir storage is considered a diversion under California water law. This is achieved with an accounting of storage through the year. Once the 15,300 AF limit is reached, no further storage is allowed, and no additional inflows can be impounded. However, though noted here, this is not explicitly implemented in SFWSM. Instead, San Antonio Reservoir level management is modeled with controlled spill using a preferred storage level trigger, as for other reservoirs.

IFR below San Antonio Reservoir

Note that there is currently no IFR below San Antonio Reservoir. A narrative with a new IFR below San Antonio reservoir has been investigated though the LTVA. Description of this narratives is found in the LTVA final report (HRG LTVA, 2021).

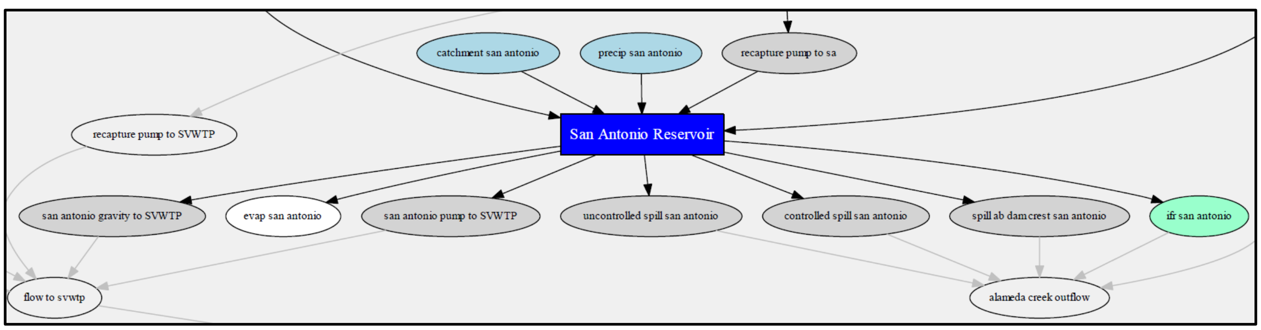

Figure 19. Schematic of flows in to and out from San Antonio Reservoir and the Recapture Pit (Pit #2), a part of the Alameda Creek Recapture Project.

Alameda Creek Recapture Project

SFPUC has a pre-1914 right to divert water from Alameda and Calaveras Creeks, meaning it does not need a permit from the SWRCB for these diversions. One component of the 2008 SFPUC Water System Improvement Program (WSIP) entails improving environmental flows below Calaveras Dam and the Alameda Creek Diversion Dam (ACDD), potentially impacting SFPUC’s water rights. To account for this, SFPUC will implement a project called the Alameda Creek Recapture Project (ACRP), whereby it will construct infrastructure to recapture water below Calaveras Dam and ACDD, thereby enabling it to continue to use its water rights (SFPUC, 2016b). Specifically, SFPUC will re-purpose a depleted quarry (Pit #2) below the confluence of Alameda and San Antonio Creeks. The quarry is hydrologically connected to Alameda Creek, such that water from Alameda Creek naturally refills the quarry. SFPUC will pump water from the filled quarry to San Antonio Reservoir.

The ACRP is modeled using a range of physical constraints and operational objectives, developed in consultation with SFPUC and described as follows, with ACRP flow path configuration around Pit #2 as shown in Figure 19 above.

Infiltration from Alameda Creek to Pit #2—the potentially recapturable water—equals total inflow from upstream (release/spill from ACDD and Calaveras Reservoir, plus local accretions) or 17 cfs, whichever is less.

The amount of water that can be recaptured (pumped from Pit #2) depends on an accumulated recapturable flow credit for flow that is uncapturable due to a maximum pumping capacity. The flow credit is limited by 1) a recapturable daily flow limit and 2) the amount of free storage space SFPUC would otherwise use to store water released for IFRs.

The recapturable daily flow limit depends on whether or not Calaveras Reservoir spills. If Calaveras Reservoir does not spill, the recapturable daily flow is the total of IFR below Calaveras Reservoir and ACDD. Otherwise, recapturable flow is zero.

The flow credit accumulation in each time step is limited by the available storage space in Calaveras reservoir, calculated as storage at 756.2 feet (i.e., 93,717 AF) less actual storage.

From Pit #2, water may be pumped into the system. Since water can only flow in one direction to/from San Antonio Reservoir, water from Pit #2 is pumped either to San Antonio Reservoir or to the Sunol Valley WTP.

If San Antonio Reservoir is being refilled and its storage is less than 90% of capacity, then water is pumped to San Antonio reservoir, subject to the recapturable flow credit constraint, recapture water rights, and the pumping policy of pumping at a constant rate for a minimum of seven days.

Water is pumped to the Sunol Valley WTP subject to a capacity of 30 cfs, and only: July through June, if recapturable rights exceed 416 AF, if San Antonio Reservoir storage is at 90% of storage capacity or higher. Pumping to the Sunol Valley WTP is subject to the same goal of a constant rate for a minimum of seven days.

Pit #2, like other reservoirs generally, has a minimum storage and preferred maximum storage. If, based on physical constraints and operational objectives, Pit #2 storage exceeds the preferred maximum storage, then a "mining pump" is used to pump excess storage from the pit. This excess water is lost from the system, and is considered as spill (i.e., it does not contribute to accumulated flow credits).

3.7.3 Peninsula

All reservoir preferred storage levels in the Peninsula are operated with controlled spill as described above, so no further description of preferred storage level is provided here.

Pilarcitos Reservoir

Pilarcitos Reservoir (Figure 20) releases include uncontrolled spill (i.e., spill), controlled release to Stone Dam and a diversion to Crystal Springs Reservoir. Note that when the reservoir gets below 2,274 AF, the controlled release to Stone Dam is no longer possible. However, the IFR at Stone Dam can still be provided by Pilarcitos reservoir thanks to a temporary siphons with a capacity of 20 cfs. Spill from Pilarcitos continues to Stone Dam, from which water is released as an IFR, uncontrolled spill, delivery to the BAWSCA Coastside County Water District (CCWD), or a diversion to Crystal Springs Reservoir. CCWD can also be supplied with water from Crystal Springs Reservoir via pumping. To avoid pumping costs, SFWSM prioritizes supply from Pilarcitos Reservoir by assigning a positive cost to the connection between Crystal Springs Reservoir and the CCWD node.

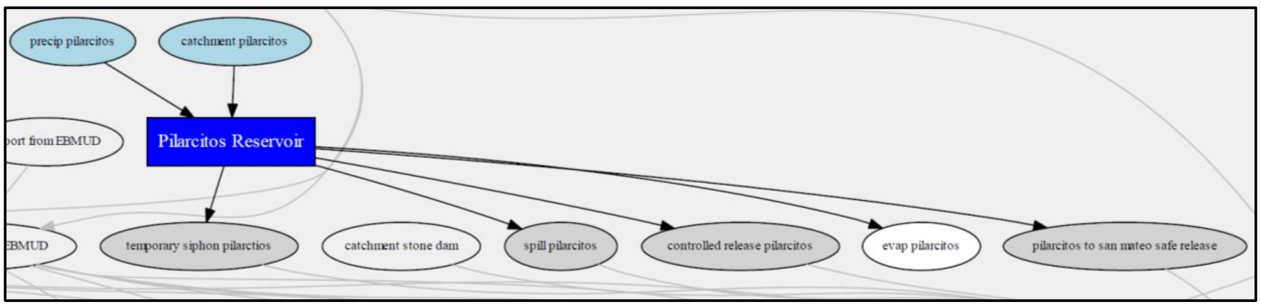

Figure 20. Schematic of flows in to and out from Pilarcitos Reservoir.

<u>IFR in Pilarcitos Creek below Pilarcitos Reservoir</u>

There is no official flow requirement in Pilarcitos Creek below Pilarcitos Reservoir. However, SFPUC nonetheless releases a constant 1.5 cfs (2.99 AF/day) from Stone Dam. This was modeled as a constant maximum flow in Pilarcitos Creek below Stone Dam, with a high negative cost.

San Andreas and Crystal Springs Reservoirs

San Andreas and Crystal Springs Reservoirs (Figure 21) are operated jointly, with spill from San Andreas going to Crystal Springs via San Mateo Creek and pumping from Crystal Springs to San Andreas. Water enters the urban delivery system from the Peninsula reservoirs via the Harry Tracy WTP (HTWTP). San Andreas Reservoir is used as the storage facility immediately preceding HTWTP. Thus, water is pumped from Crystal Springs to San Andreas before entering the urban delivery system. Outflow from San Andreas therefore includes spill to Crystal Springs and diversion to HTWTP. Outflows from HTWTP include several BAWSCA customers, as well as delivery system connection points. Inflows to Crystal Springs Reservoir includes the Pulgas Tunnel, which delivers water from the East Bay, diversions from Pilarcitos Reservoir and Stone Dam, several local catchments, and spill from San Andreas Reservoir. Outflows from Crystal Springs includes an IFR, controlled spill, uncontrolled spill, diversion (pumped) to San Andreas Reservoir, and the pumped diversion to CCWD noted above.

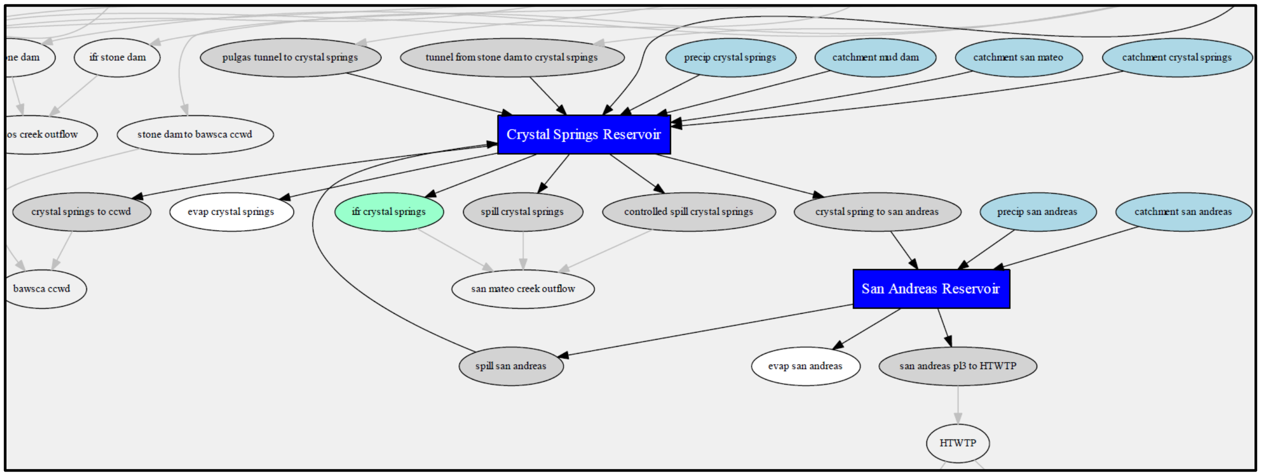

Figure 21. Schematic of flows in to and out from San Andreas and Crystal Springs Reservoirs.

IFR in San Mateo Creek below Crystal Springs Reservoir

Note: This operation is not included in the historical validation version of SFWSM.

Instream flow requirements in San Mateo Creek below Crystal Springs Reservoir consist of a base minimum flow in all months and flow ramping restrictions on specific flow level transition days. There are two flow schedules—Schedule A and Schedule B—which are determined every January 11 by the cumulated San Mateo Creek Precipitation Index (SMCPI), a local precipitation index calculated as the average of four rainfall stations within the San Mateo Creek watershed. SMCPI calculations result in three water year types; Schedule A is used in Normal/Wet years (cumulative SMCPI on January 11 > 10.3 inches), while Schedule B is used in Dry years (≤10.3 inches). Flow schedules are listed in Table 17. The change in flow schedule occurs on January 12 based on January 11 cumulative SMCPI.

As with other IFRs, this IFR is modeled with a Pywr node for controlled spill and was pre-processed following the above logic, with SMCPI values calculated using precipitation output from the weather generator. As noted above, ramping rates were not modeled.

Table 17. Instream flow requirements for San Mateo Creek below Crystal Springs Reservoir.

Flow schedule decision date	Flow schedule application period	Dry (Schedule B)			Normal/Wet (Schedule A)		
		Cumulated precipitation index for water-year classification (in)	Flow ramping schedule (cfs)	Flow requirement (cfs)	Cumulated precipitation index for water-year classification (in)	Flow ramping schedule (cfs)	Flow requirement (cfs)
N/A	Oct 1 – Dec 14	N/A	N/A	3	N/A	N/A	3
N/A	Dec 15 – Jan 12	N/A	N/A	5	N/A	N/A	5
Jan. 12[2]	Jan. 15 – Mar 15	≤ 10.3	Jan 13: 5 Jan 14: 7	10	> 10.3	Jan 13: 7 Jan 14: 12	17
N/A	Mar 16 – Mar 30	N/A	N/A	8	N/A	Mar 16-17: 15 Mar 18-19: 12	10
N/A	Mar 31 – Apr 30	N/A	N/A	5	N/A	Mar 31-Apr 1: 7	5
N/A	May 1 – Sep 30	N/A	N/A	3	N/A	N/A	3

From: 'Corps – opinion SJPL LCSDI CSSA 2010-10-29.pdf': Table 1, pg. 16-17

San Francisco Groundwater Project

The San Francisco Groundwater Project is an in-city project that provides the city with a year-round supply. The production capacity is theoretically 4 MGD but is limited in practice to 1 MGD. This is represented in the model as a Pywr input node with a maximum flow of 1 MGD (3.1 AF/day). No cost is associated with groundwater.

Regional Groundwater Storage and Recovery

Under a program called the Regional Groundwater Storage and Recovery (GSR), SFPUC and two regional wholesale customers (CWS South San Francisco and Daly City) have agreed to operate a 75,000 AF (24,439 MGD) groundwater storage account using the Westside groundwater basin in the Peninsula. Under this scheme, both SFPUC and the wholesale customers pump from groundwater storage whenever drought reductions occur ("take years"), before all other drought reduction actions. During non-drought years ("put years") pumping is ceased to allow natural recharge of the basin, with SFPUC providing in-lieu deliveries to the wholesale customers to ameliorate impacts of not pumping. Withdrawals are limited by pumping capacities, with withdrawal capacities listed in Table 18.

The GSR program is modeled with a reservoir, with a capacity of 75,000 AF, that is refilled with a Pywr input node and is connected to participating customers (Daly City, California Water Service Company South San Francisco and San Bruno). Recharge, defined with a script, is set at zero if YRS is less than 3.6 or if demand reduction factor is less than 1; otherwise, recharge is 5.0 MGD (15 AF/day). The Westside basin is connected to each connection point with the system first via a single link that limits total withdrawals then by connection-specific links that limit withdrawals (via pumping) for each connection. The maximum total withdrawal is 7.2 MGD (26.5 AF/day).

Table 18. Maximum withdrawal capacity for each connection to the Westside groundwater basin under the Regional Groundwater Storage and Recovery program.

Connection / customer	Withdrawal capacity (MGD)	Withdrawal capacity (AF/day)
SFPUC transmission pipelines		
Sunset Pipeline	0.48	1.47
Crystal Springs Pipeline 2	0.55	1.7
San Andreas Pipeline 2	3.39	10.4
Wholesale customers		
CWS South San Francisco	0.62	1.9
Daly City	1.3	3.99

San Francisco Westside Recycled Water

Recycled water in San Francisco city (2 MGD) is accounted for in SFWSM as a Pywr Input node to the city of San Francisco.

4 Model version variations

As noted above, three versions of SFWSM have been created: one for validation with historical observations, one for validation with HHLSM, and one for planning (i.e., for use with the LTVA analysis). This section summarizes key differences between these versions, though technical implementation details are found in model code rather than here. Table 19 lists key differences between the three versions of the model.

Table 19. Key differences between 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'Shaughnessey Dam	YES	YES	YES
IFR below Cherry Valley Dam	YES	YES	YES
IFR below Lake Eleanor Dam	YES	YES	YES
IFR below San Antonio Dam	NO	NO	NO
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

4.1 Historical validation version

To model suburban retail demand for historical validation, some customers have a constant demand, while for others historical mean daily demand is used. The specific methods for each suburban retail demand customer are listed in Table 20. Constant demand values are also listed in Table 20, while mean daily demand values are shown graphically in Figure 22.

Table 20. Method for modeling suburban retail customer demand for the historical validation version of SFWSM.

Region	Customer	Demand method	Demand (AF/day)
Upcountry	Groveland Community Services District	Constant	1.47
	Lawrence Livermore National Laboratory (LLNL) Site 300	Constant	2.46
East Bay	Town of Sunol	Constant	1.78
	General Electric	Historical daily mean	Variable
Peninsula	Golden Gate National Cemetery	Historical daily mean	Variable
	Menlo Park Country Club	Historical daily mean	Variable
	NASA	Historical daily mean	Variable
	SFO	Historical daily mean	Variable
	Cordilleras Mutual Water Company	Constant	0.02

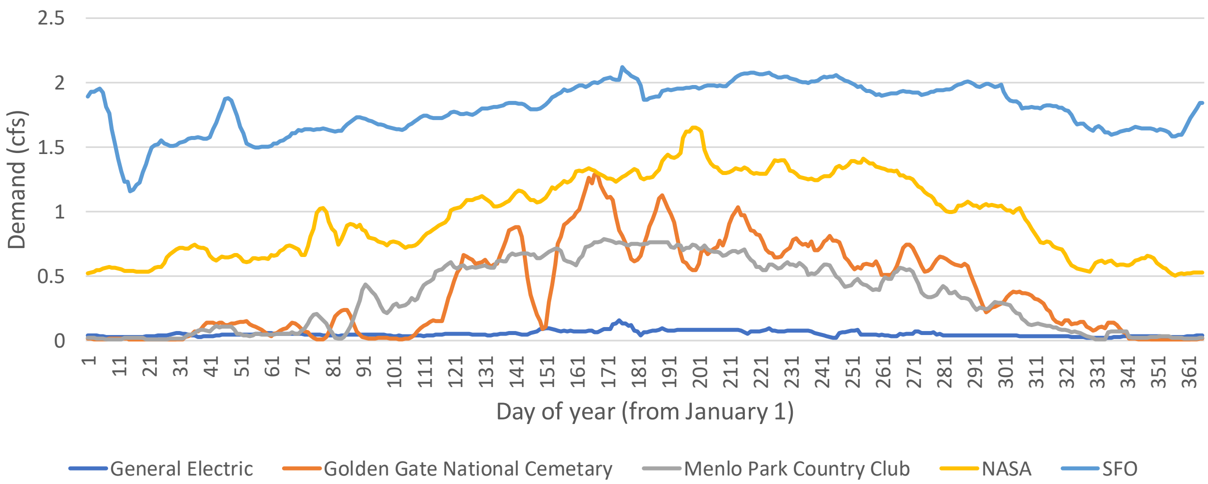

Figure 22. Daily demand for SFPUC retail customers for the historical validation version of SFWSM, with demand based on historical daily mean demand. Demand is repeated every year.

4.2 HHLSM validation version

Similar to the historical validation, a version of SFWSM was developed through discussions between UMass and SFPUC staff for validation with HHLSM. Modifications were made to match, as closely as possible, the “WSIP PEIR 2018” version of HHLSM, which includes an annual average system-wide water demand of 265 MGD as a major assumption. Comparisons were made with the Real Hydrology (RH) version of HHLSM.

In order to validate the SFWSM model against the HHLSM model, a demand time series for each wholesale and retail customer needs to be established that provides a reasonable comparison to the demand time series used by HHLSM. The HHLSM model splits demand according to three delivery centres: Peninsula, East Bay, and City of San Francisco as well as retail customers (see Table 21 for breakdown of demand across delivery centres in HHLSM). Demand is held constant on an inter annual basis and is modelled at the monthly time step. The SFWSM model has a finer spatial resolution, including all wholesale and retail customers as distinct demand nodes and runs at the daily time step.

Table 21 Mean annual demand to HHLSM service areas in the HHLSM validation version of SFWSM

Delivery Centre	**Annual demand**	
	AF	**mgd**
Cordilleras	0.02	0.01
Golden Gate Bridge Cemetery	0.77	0.25
Menlo Park	0.68	0.22
NASA	1.92	0.62
San Francisco Airport	3.56	1.16
Town of Sunol	1.78	0.58
Groveland	1.23	0.40
LLNL_site_3000	2.46	0.80
General Electric	0.10	0.03
RETAIL TOTAL	**12.51**	**4.08**
Peninsula	130.73	42.6
South and East Bay	440.37	143.5
City of San Francisco	229.00	74.62
TOTAL	**812.62**	**264.79**

The following process was applied in order to reconcile the spatial and temporal differences between the input demand timeseries used by HHLSM and that required by SFWSM.

1. The average demand for each day of the year for the base demand scenario from the SFWSM model is established. This timeseries is then normalised by dividing all values by the average annual daily demand across all days of the year to create the timeseries shown in Figure 23.

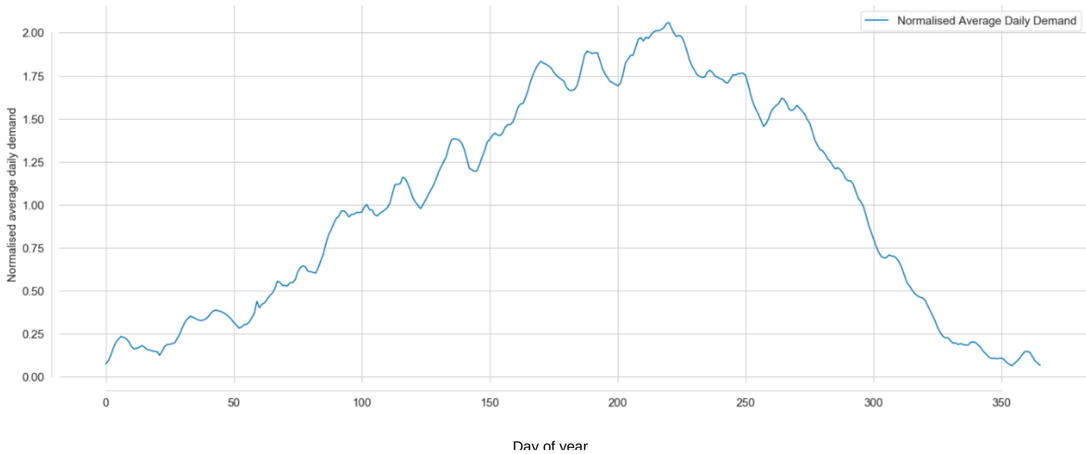

Figure 23 *Normalized average demand for each day of the year*

2. Divide HHLSM daily demand equally amongst customers belonging to that delivery centre. Table 22provides an overview of which wholesale customers belong to which delivery centre. In reality there is significant variation in the level of demand amongst customers within delivery centres, however for the purposes of the HHLSM validation the priority was to ensure that SFWSM demand aligned with HHLSM demand at the delivery centre scale.

Table 22 Overview of wholesale customers and their corresponding region and base fraction, which describes the proportion of total demand that make up the base demand use (essential uses).

Wholesale Customer	Region	Base Fraction
BAWSCA_ACWD	SEB	0.55
BAWSCA_BRISBANE_GVMID	PEN	0.56
BAWSCA_BURLINGAME	PEN	0.64
BAWSCA_CWS_BEAR_GULCH	SEB	0.40
BAWSCA_CWS_MPD	PEN	0.65
BAWSCA_CWS_SOUTH_SF	PEN	0.71
BAWSCA_CCWD	PEN	0.58
BAWSCA_DALY_CITY	PEN	0.83
BAWSCA_EAST_PALO_ALTO	SEB	0.71
BAWSCA_ESTERO_MID_FOSTER_CITY	PEN	0.45
BAWSCA_HAYWARD	SEB	0.67
BAWSCA_HILLSBOROUGH	PEN	0.31
BAWSCA_MENLO_PARK	SEB	0.47
BAWSCA_MID_PENINSULA_WD	PEN	0.65
BAWSCA_MILLBRAE	PEN	0.66

BAWSCA_MILPITAS	SEB	0.63
BAWSCA_MOUNTAIN_VIEW	SEB	0.55
BAWSCA_NORTH_COAST_COUNTY_WD	PEN	0.75
BAWSCA_PALO_ALTO	SEB	0.53
BAWSCA_PURISSIMA_HILLS_WD	SEB	0.40
BAWSCA_REDWOOD_CITY	SEB	0.65
BAWSCA_SAN_BRUNO	PEN	0.63
BAWSCA_SANTA_CLARA	SEB	0.76
BAWSCA_SAN_JOSE	SEB	0.56
BAWSCA_STANFORD	SEB	0.58
BAWSCA_SUNNYVALE	SEB	0.44
BAWSCA_WESTBOROUGH_WD	PEN	0.73
SAN_FRANCISCO	CCSF	0.92

3. For each day in the time series, multiply average daily demand for each customer by the corresponding normalized average demand established in Step 1.
4. Divide daily demand into base and seasonal components according to observed seasonal demand fractions provided in Table 22.

5 Validation

Two validations were performed with SFWSM: one with historical operations ("historical validation"), and one with the "WSIP PEIR 2018" version of HHLSM ("HHLSM validation").

5.1 Validation with historical operations

This section includes a series of representative validation figures with brief qualitative assessments of model performance compared with historical operations. The assessment covers the most important operations across the region, system-wide operations and selected operations in each region.

5.1.1 System-wide

Comparisons of major operations system-wide are shown in Figure 24, including total system storage (Figure 24a), total storage in each region (Figure 24b-d), flow in San Joaquin Pipelines (Figure 24e), and outflow from each WTP (Figure 24f). 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, as discussed above, SFWSM results in peak storage less than observed. Simulated Upcountry storage is significantly lower than observed during the first months of 2010, which can be attributed to the use of perfect foresight forecast to operate the Canyon and Cherry power tunnels (cf. discussion below). Storage otherwise generally follows observed historical trends.

Generally, flows in SFWSM, which are shown in Figure 24 using a 20-day moving average, 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[9], which results in discrete jumps from one state to another. Flow through the SJPL and outflow from the treatment plants

[9] In this section, several daily time series are discussed. Most of the time series present some high frequency variability due the nature of the simulation technique (using linear programming). These high fluctuations have been reduced for some link nodes (e.g., SJPL) by defining dynamically appropriate minimum and maximum flow constraints to the LP problem. This is not discussed further. It is = important to note, however, these high frequency fluctuations are not expected to influence the LTVA.

are rather well represented. Discussions that are more detailed are provided later in this section.

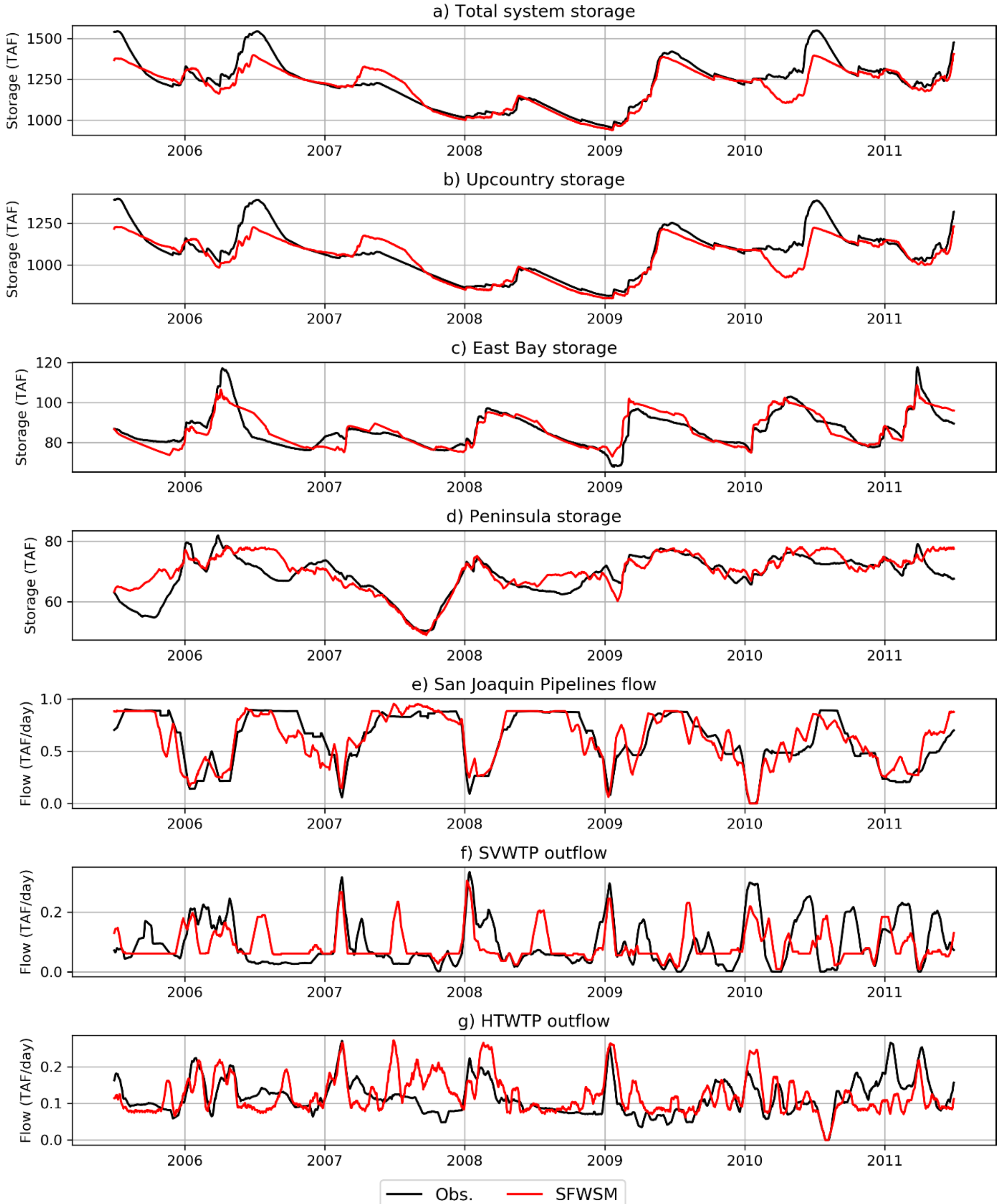

Figure 24. Observed and simulated major operations system-wide. Operations with flow are with a 20-day moving average.

Figure 25 compares observed and simulated daily storage for each reservoir in SFWSM, including the Water Bank, while Figure 26 compares mean daily storage for each reservoir. Most reservoirs are reasonably well simulated. This is especially true for the Upcountry reservoirs and the Water Bank. Again, it is noted that the high storage portion of the Water Bank was not considered important to represent in SFWSM. The use of perfect foresight forecasts to operate the power tunnels at Hetch Hetchy and Cherry leads to more aggressive operations in a sense that the reservoirs empty more than they have in reality. This is especially the case during the years 2009, 2010 and to some extend 2011 for Cherry Lake. For the purposes of the LVTA, 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. We note a negative bias for the year 2007 for Pilarcitos annual minimum storage. Although the cause of this has not been determined, Pilarcitos is a small reservoir and identified biases (i.e., 500 to 1,000 AF), though relatively large within the reservoir, are relatively small within the water system.

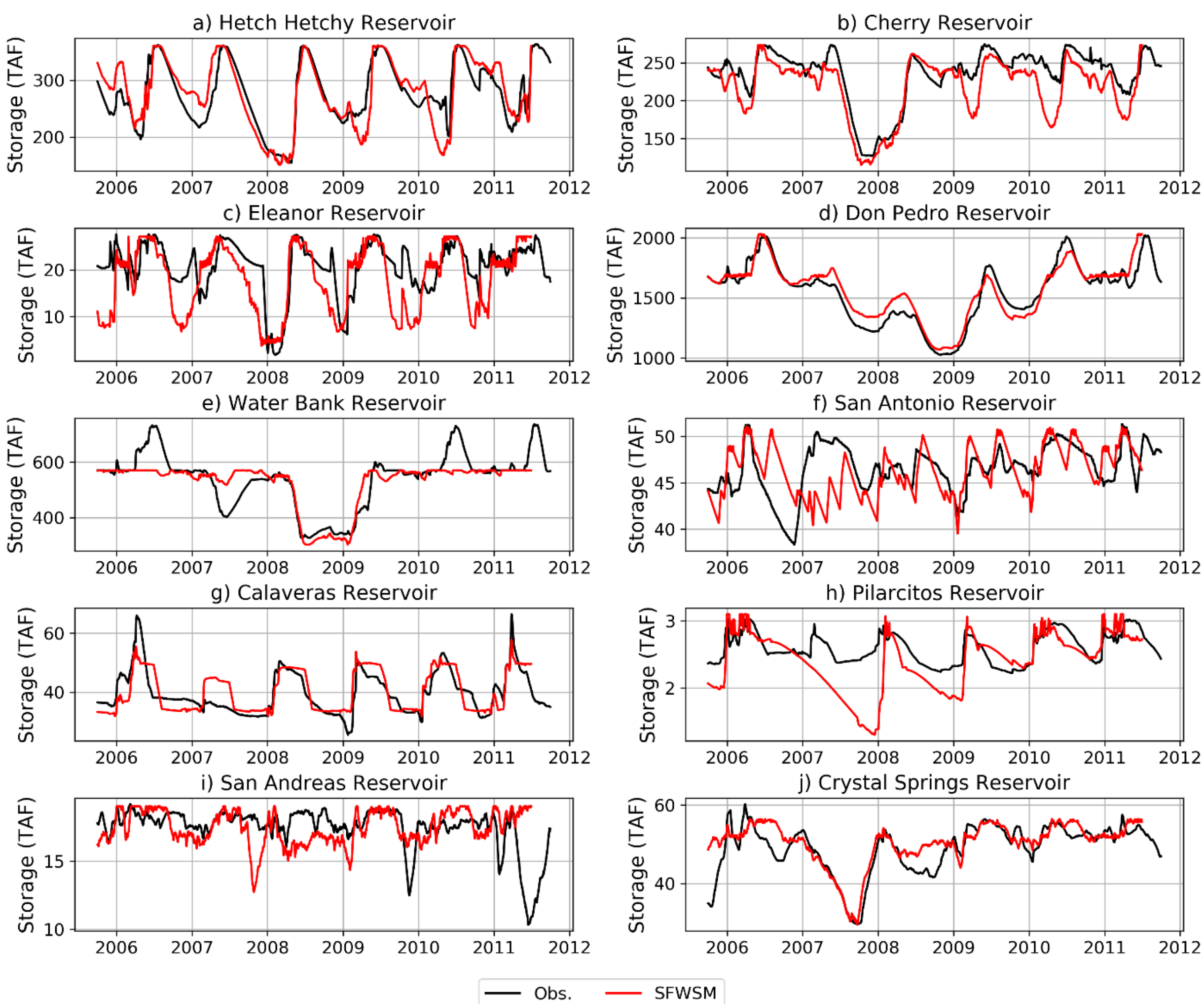

Figure 25. Observed and simulated daily storage for all reservoirs, including the Don Pedro Water Bank.

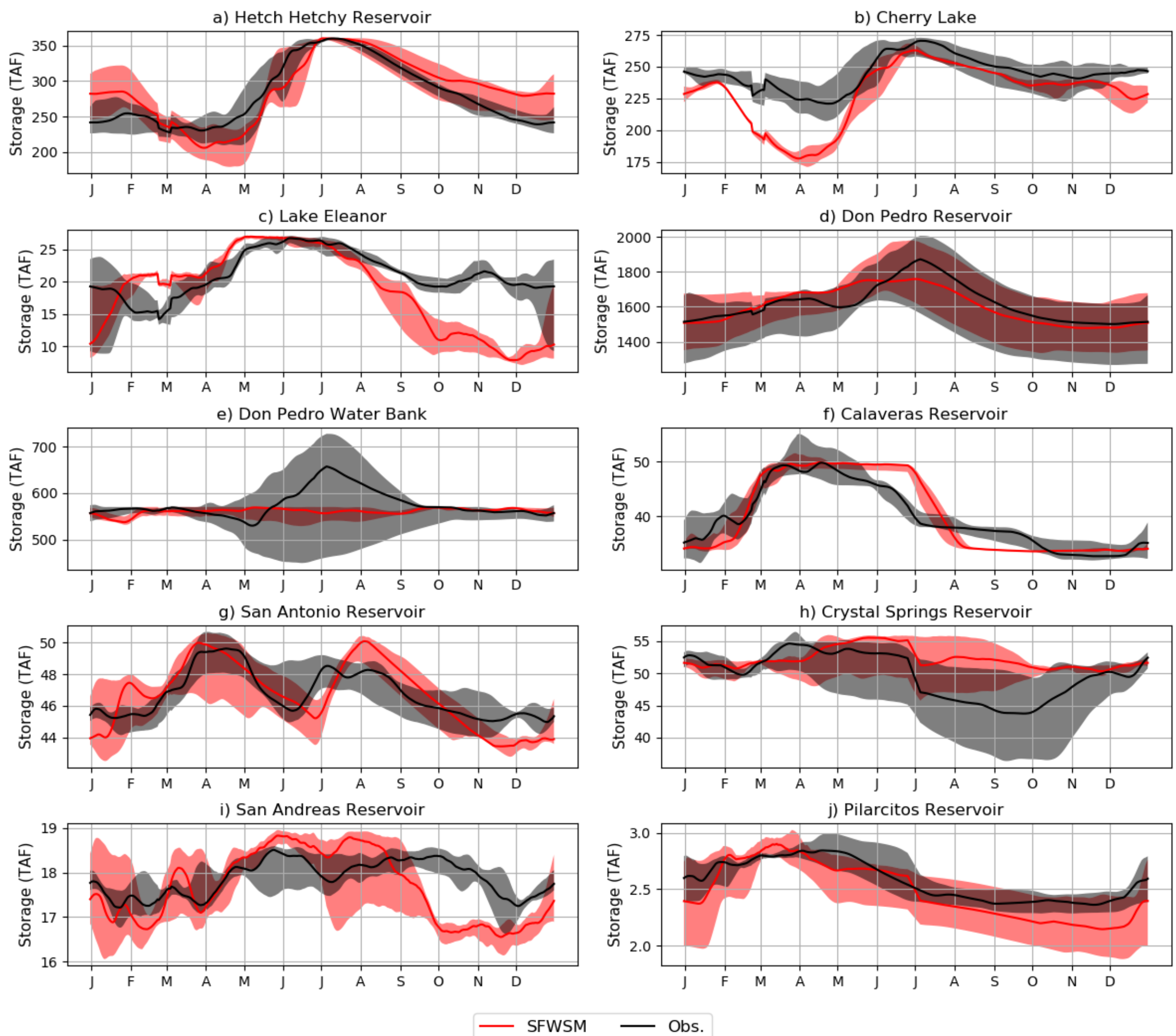

Figure 26. Observed and simulated mean daily storage for all reservoirs, including the Don Pedro Water Bank.

Observed and simulated flows through the San Joaquin Pipelines are shown in Figure 27. Here, the greater fluctuations in SFWSM are more apparent than in the 20-day moving average shown in Figure 24. However, on a mean daily (Figure 27b) and mean annual (Figure 27c) basis, trends follow historical operational patterns. The short-term fluctuations are not likely to have any significant influence in the LTVA context. The inter-annual variability of total flow through the

SJPL are well represented.

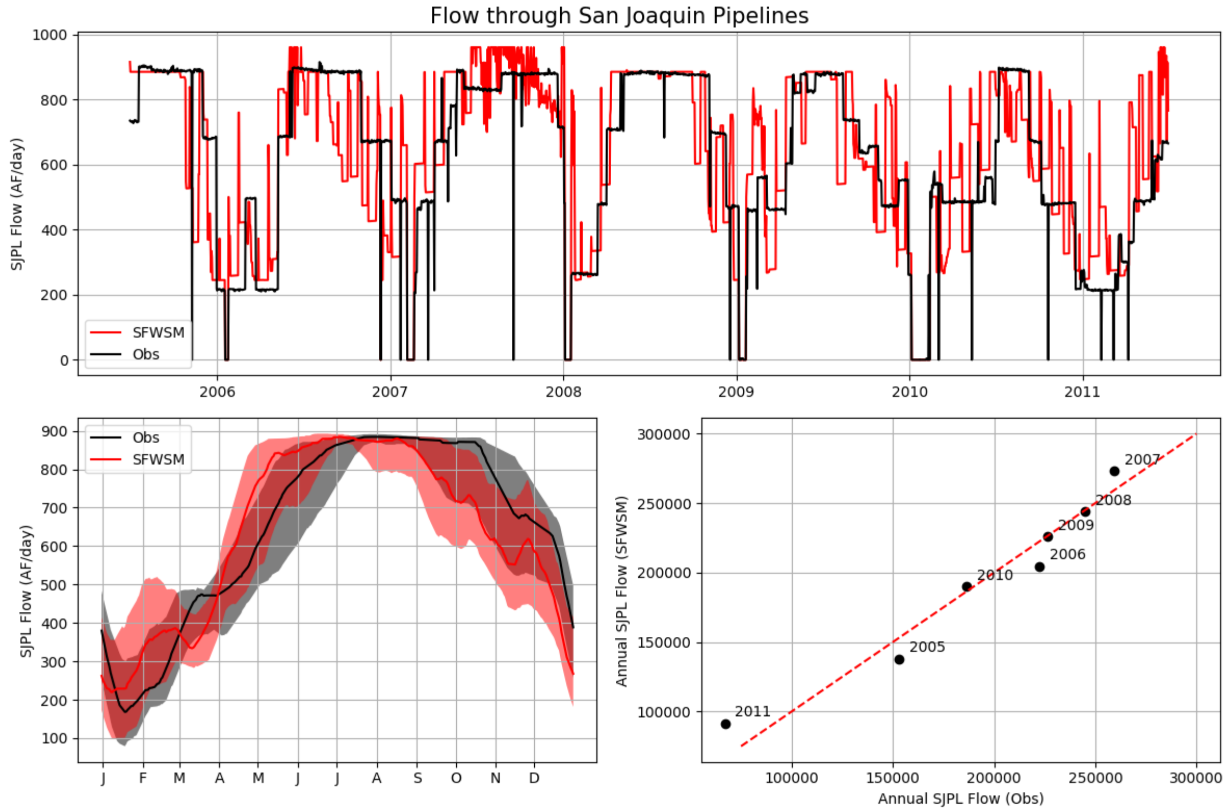

Figure 27. Comparison of historical observed and SFWSM simulated flow in the San Joaquin Pipelines (SJPL). The top plot shows daily time series for the simulated (red) and observed (black) flow through the SJPL. The bottom left plot shows the medians (bold curves) and the deviation between inter-quartiles smoothed over a 30-day moving windows. The bottom right plot show the scatter plot of the annual flow (calendar year).

Finally, Figure 28 compares observed and simulated water deliveries—which are to meet demand based directly on historical deliveries, so these should be the same. No virtual water is needed to meet demand..

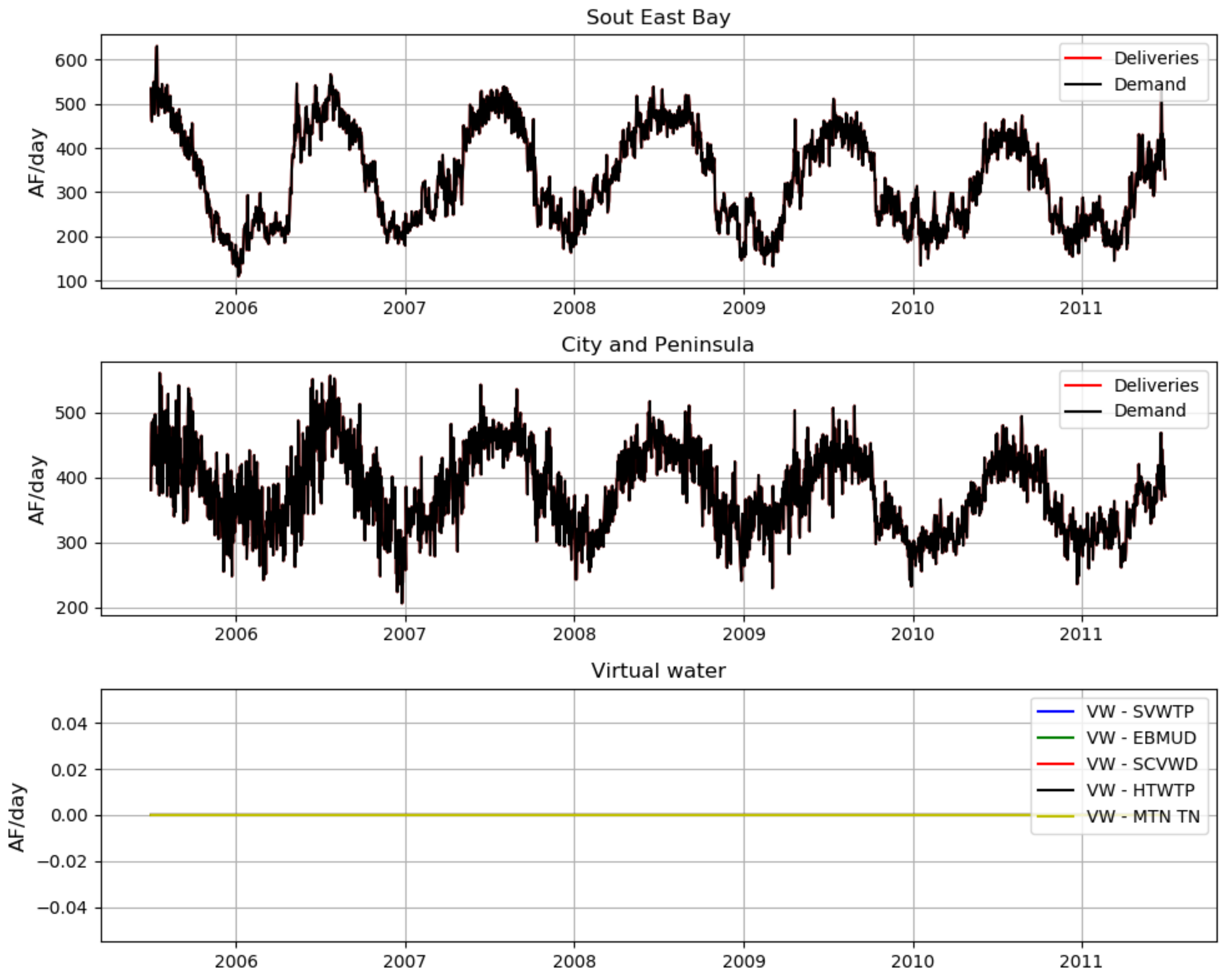

Figure 28. Comparison of historical deliveries and SFWSM simulated deliveries with historical deliveries as input demand in each of the South East Bay (a) and the City and Peninsula (b), and "virtual water" (c) needed to supply historical deliveries.

5.1.2 Up-country

Water Bank operations

Water Bank operations, including Water Bank storage and related flows, are shown in Figure 29. Simulated operations generally track observed operations well. We note a negative bias in the water bank balance during spring and summer seasons of the years 2006 and 2011. These discrepancies follow from the absence of the bubble account in the Don Pedro Water Bank, which makes the Water Bank spill. Including the bubble account could be considered in any improvements to SFWSM. A more detailed view of Water Bank operations is shown in the Appendix III.

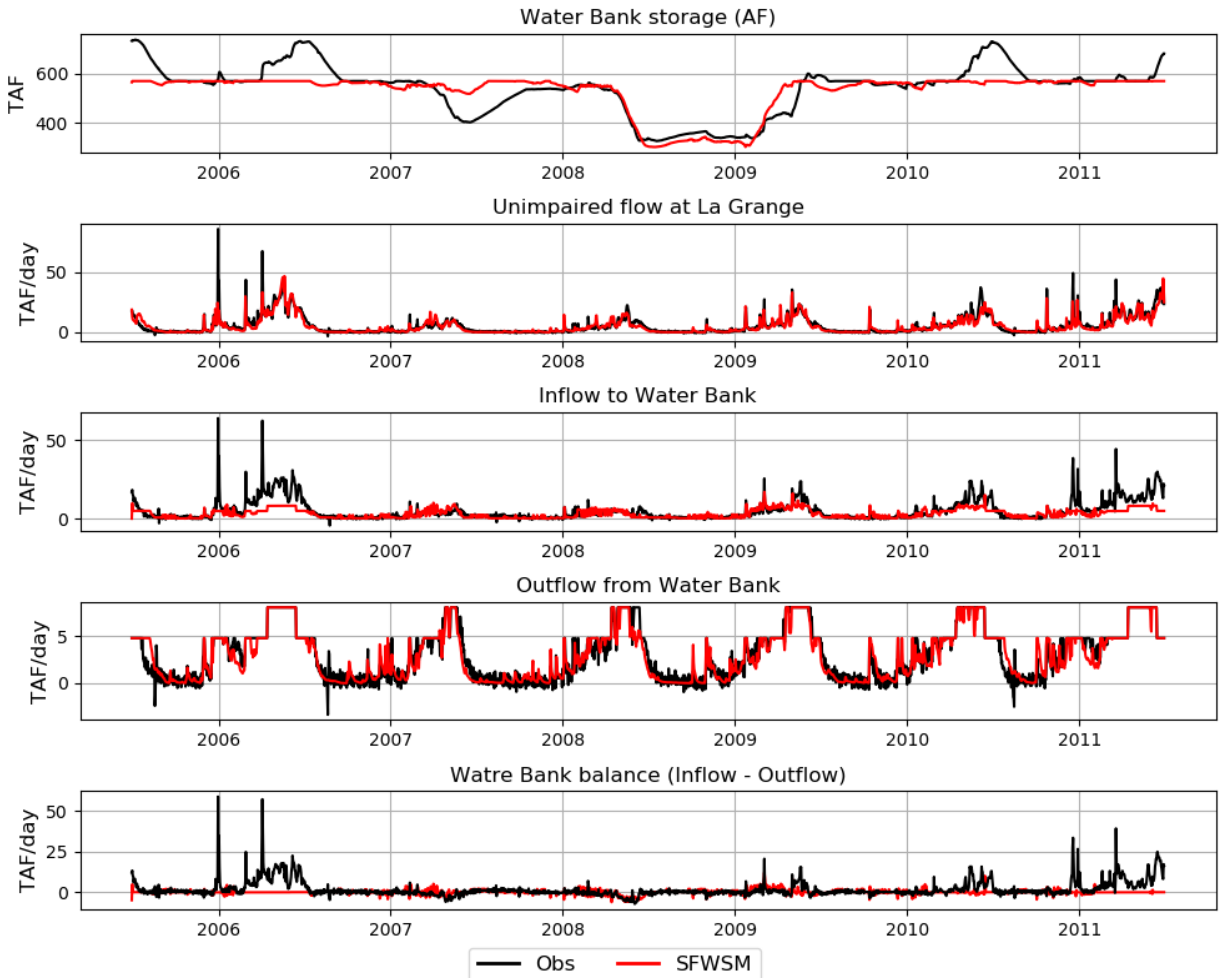

Figure 29. Water mass balance at the Don Pedro Water Bank, including: storage (a), unregulated flow at Lagrange (b), inflow (c), outflow (d) and balance (e).

5.1.3 Water Transfers in the Sunol Valley

Reminder: The prime objective of the comparison between the historical operation of the RWS and historical version of SFWSM is to validate the logic implemented within the planning version of the SFWSM model. As such, it was decided to somewhat bend the planning version of SFWSM to represent the historical operations, rather than creating a model with the only purpose to model the historical operations. As a results, some historical operations are hardly reproducible, especially in the Sunol Valley region where the operations during construction of the new Calaveras reservoir are not necessarily representative of the way SFPUC operates nowadays.

An important water transfer within the Sunol Valley is the water flow from Calaveras Reservoir to the Sunol Valley Water Treatment Plant (SVWTP). Figure 30 illustrates this transfer. In regards with the inter-annual variability (Figure 30, bottom right), we note that SFWSM underestimates that transfer, especially for the years 2006, 2008 and 2010. This underestimate results from the implemented logic that consists at sending water to both the treatment plant and San Antonio reservoir when the Calaveras reservoir is forecasted to spill. Given the reduced storage at Calaveras for the historical period, Calaveras storage is very quickly near to its preferred storage level (i.e., 50,000 from February through June; 34,480 AF otherwise), which results into a strong a frequent incentive for transferring water to SVWTP and San Antonio, which can be observed Figure 30 through Figure 32 with a peak flow starting in July, which is the time when Calaveras storage must be decrease from 50,000 to 34,480 AF.

Given the constraint on Calaveras storage during the construction period, the above-described logic leads to too much water sent to San Antonio and not enough water sent directly to SVWTP (i.e., when compared with the historical operations during this period). During the manual calibration of SFWSM, various costs associated with the transfer from Calaveras to San Antonio have been tested to prioritize the transfer to SVWTP rather than San Antonio. When doing so, the required outflows to reduce Calaveras from 50 TAF to nearly 34 TAF were sent directly to SVWTP, which led to a significant overestimate of the yield of the treatment plant. In other words, SFWSM tries to use San Antonio as a buffer to smoothed out the flow at the entrance of Calaveras reservoir when the latter is near full, which is what it is expected from the implemented logic. The reason why SFWSM does that too frequently is due to the limited storage at Calaveras. For this reason, it has been decided to keep the logic as it is, even if the operations within the Sunol Valley system are not representative of the operations during the construction.

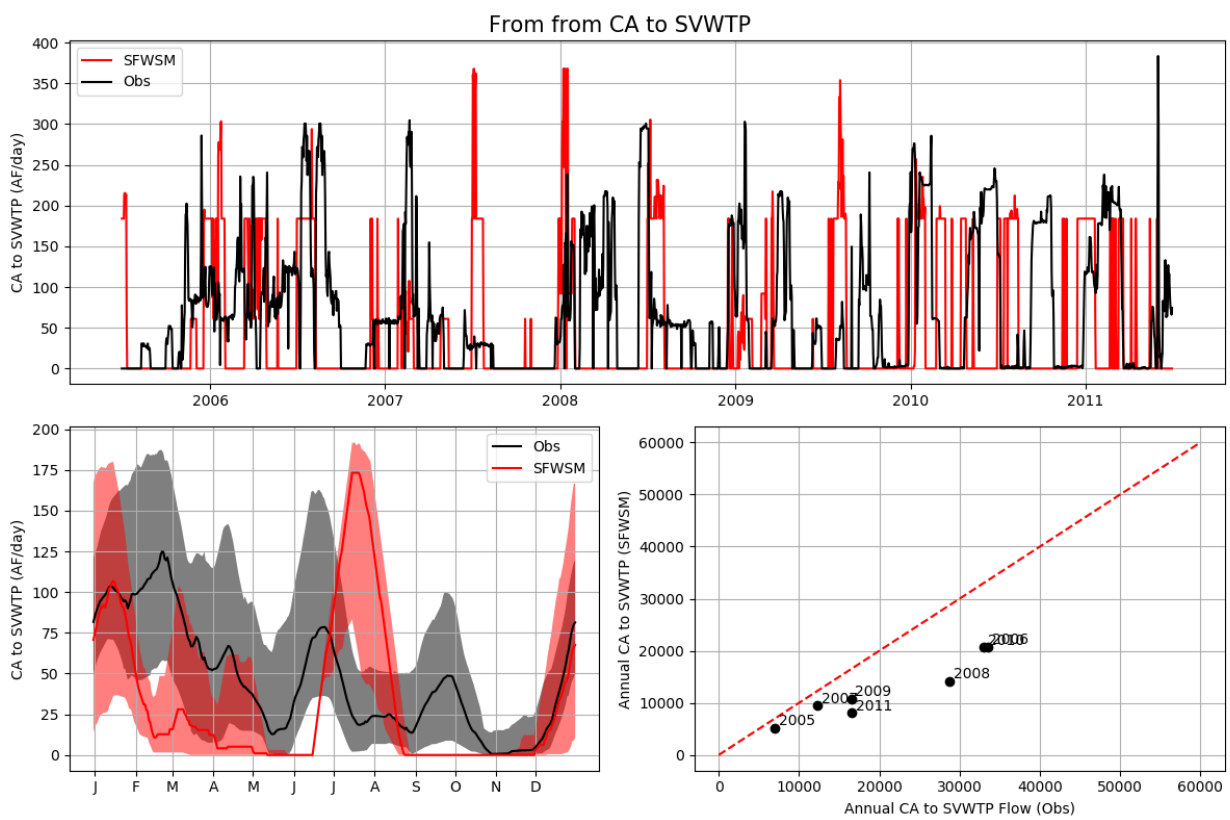

Figure 30: Flow from Calaveras to the Sunol Valley Water Treatment Plant.

Figure 31 summarizes the transfers within Sunol Valley. Note that observations are missing in 2006 and 2008 for the transfer by gravity from Calaveras to San Antonio (Figure 31 bottom left). A large water transfer from Hetch Hetchy to San Antonio (Figure 31, bottom right) explains the jumps in storage at San Antonio (Figure 31, top right). These water transfers from Hetch Hetchy to San Antonio take place following the emergency storage policy described in Section 21. Based on the observed record, it appears this policy was not in place during the construction of Calaveras.

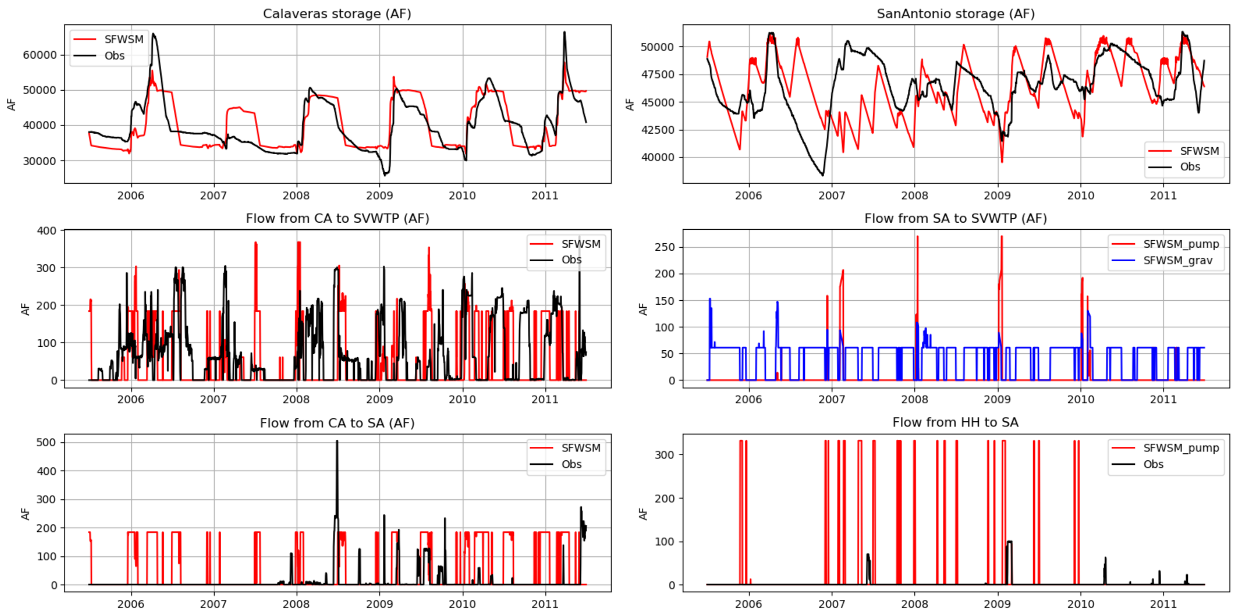

Figure 31 Outlook of the water transfer in the Sunol Valley.

Figure 32 illustrates the outflow from the SVWTP. The seasonal pattern of the outflow from the treatment plant is similar to the transfer from Calaveras to SVWTP discussed in Figure 30. The difference between the two figures comes mostly from the water release via gravity from San Antonio that often provides the minimum water flow requested by the treatment plant (blue curve on Figure 31). At the annual scale, the yield of the water treatment plant is fairly well reproduced, with largest underestimate occurring during the years 2010 and 2011 (Figure 32, bottom right). Besides these minor anomalies, mostly due the construction of the new Calaveras reservoir and the frequent shutdowns of SVWTP during these years, the representation of the operation of SVWTP in SFWSM is good.

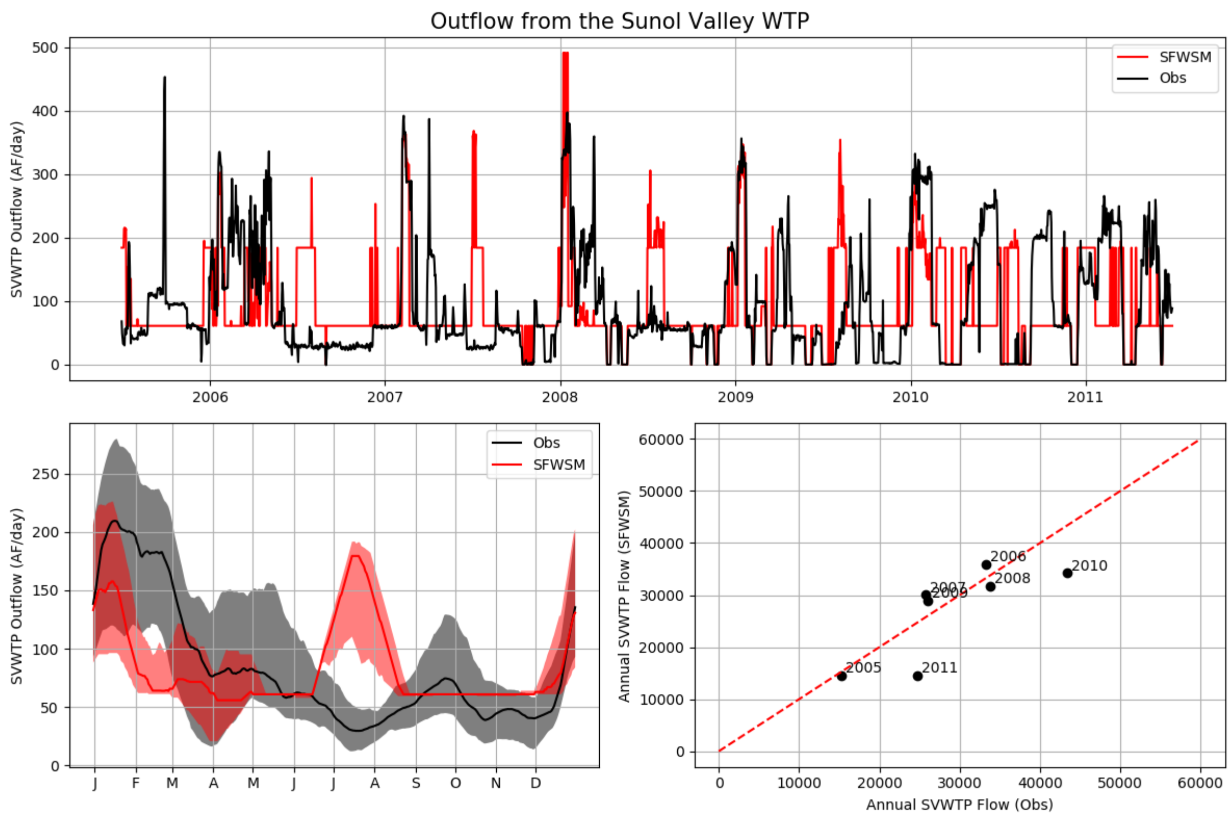

Figure 32 Outflow from the Sunol Valley Water Treatment Plant.

5.1.4 Water Transfers in the Peninsula

An important operation in the Peninsula region consists at ensuring a continuous flow of water at the entrance of the Harry Tracy Water Treatment Plant (HTWTP), which delivers water to the Peninsula. HTWTP receives water from San Andreas Reservoir, which receives most of its inflow via pumping water from Crystal Springs Reservoir. Both San Andreas and Crystal Springs reservoirs have storage capacities that are significantly larger than their natural annual inflow. To ensure the flow to HTWTP, water from either Sunol Valley reservoirs or Hetch Hetchy Reservoir is used to feed Crystal Spring via the Pulgas diversion.

Figure 33 illustrates the outflow from HTWTP. Both inter annual variability and seasonal pattern are fairly well reproduced by SFWSM. There is a positive bias in the annual outflow during the year 2007 (Figure 33, bottom right) that is due to a significant overestimation of the flow at the treatment plant during the fall (Figure 33, top), which causes the bump in the seasonal pattern in this season (Figure 33, bottom left). This large overestimation of the flow in SFWSM results from the reduction in storage at Crystal Spring reservoir (cf. Figure 25) due to maintenance (communication with SFPUC personal). This maintenance is accounted for in SFWSM via a modified preferred storage curve for this year. To decrease the water level in Crystal Spring reservoir, SFWSM sends a much larger volume of water than usual to HTWTP during this year. Besides the anomaly during the year 2007, simulation of the HTWTP is good for the LTVA.

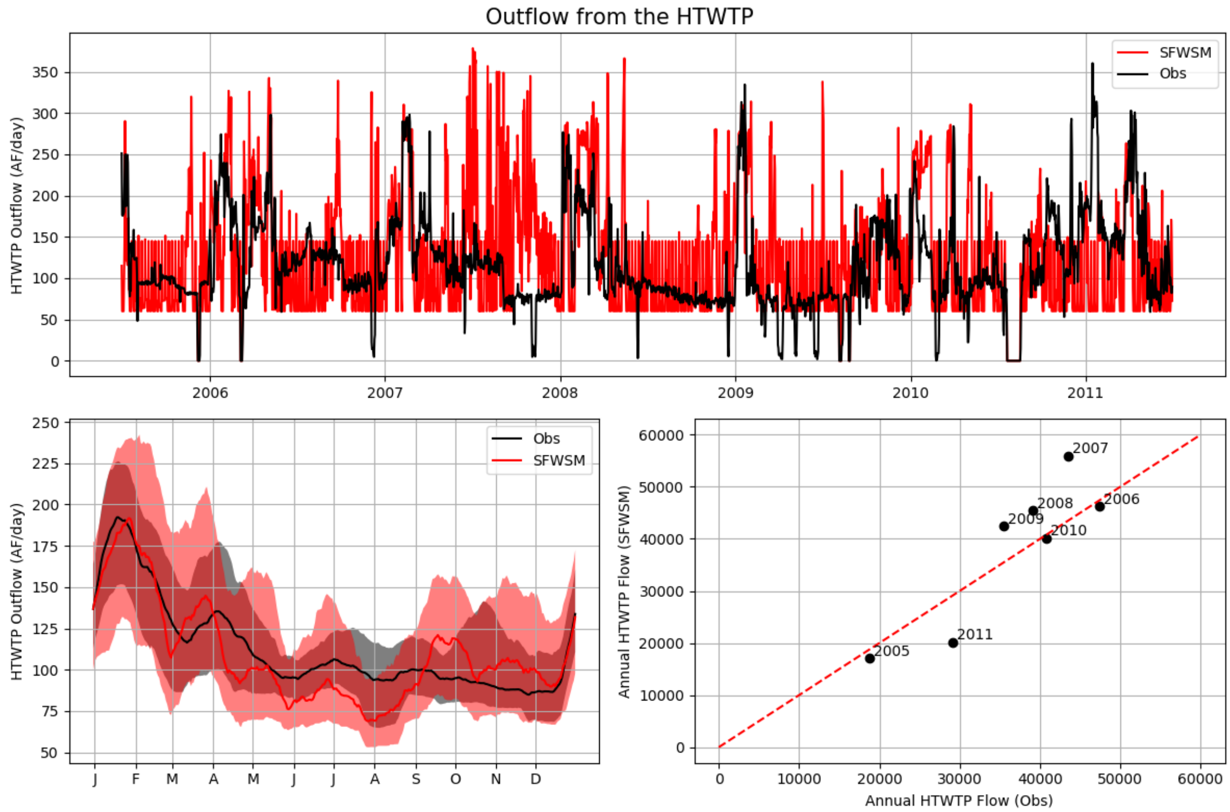

Figure 33 Outflow from the Harry Tracy Water Treatment Plant.

Figure 34 shows transfers from Crystal Spring reservoir to San Andreas reservoir. Beside the high-frequency due to the LP-solver, The interannual variability is well reproduced, except for the year 2007, when SFWSM overestimates the transfers from Crystal Spring to San Andreas for the reason mentioned above.

Figure 35 illustrates the transfers from Pulgas diversion to Crystal Spring reservoir. The seasonal pattern is not well reproduced as a positive bias is observed during spring and a negative bias is observed during the fall. There are several possible causes for this misrepresentation. For instance, any discrepancies between observed and simulated streamflow to Peninsula reservoirs would affect the need for water from the upstream part of the RWS via the Pulgas diversion. Another example is the emergency storage policy that is accounted for in SFWSM, although it is not certain whether this has been actually implemented in the historical operations. Finally, the Pulgas diversion shown on Figure 35 is the only link discussed in this section that is not located directly downstream of a reservoir. The flow through Pulgas diversion is therefore influenced by any biases and mismatches that could take place further up in the system.

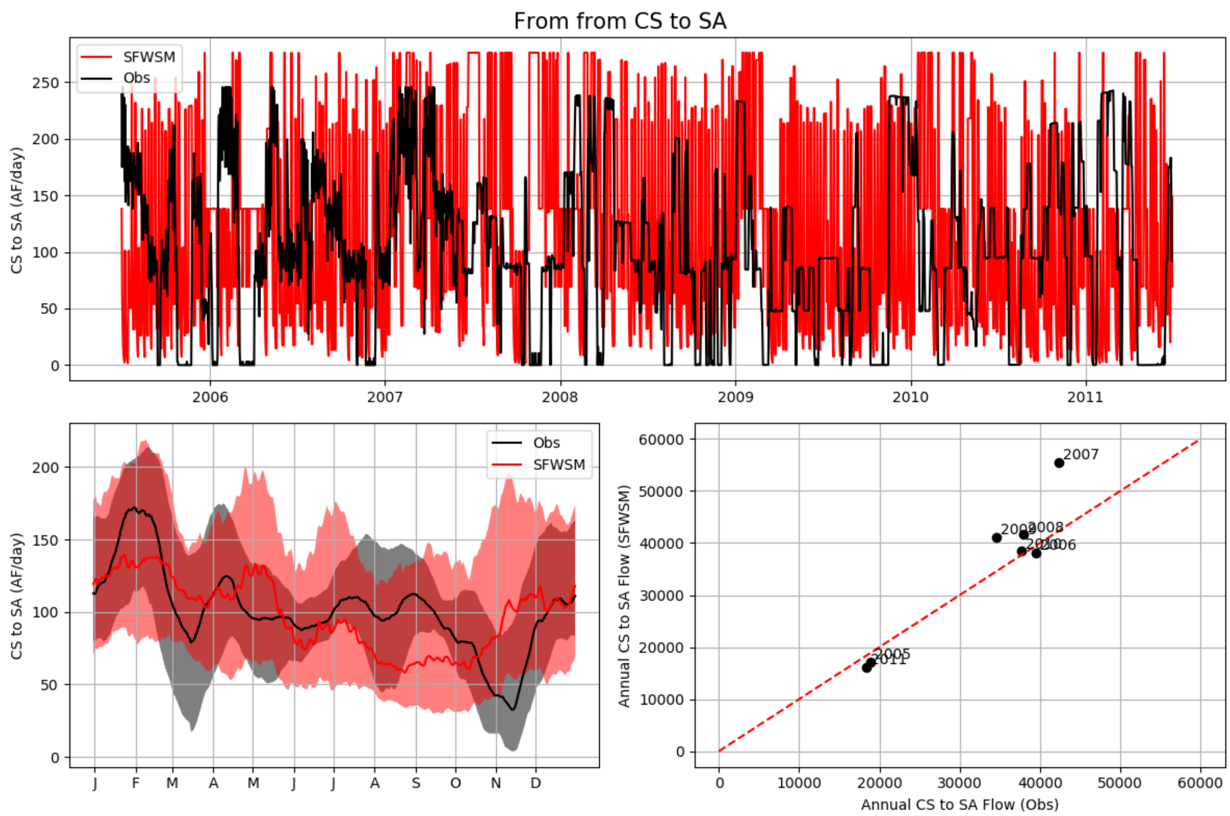

Figure 34 Water transfer from Crystal Spring to San Andreas.

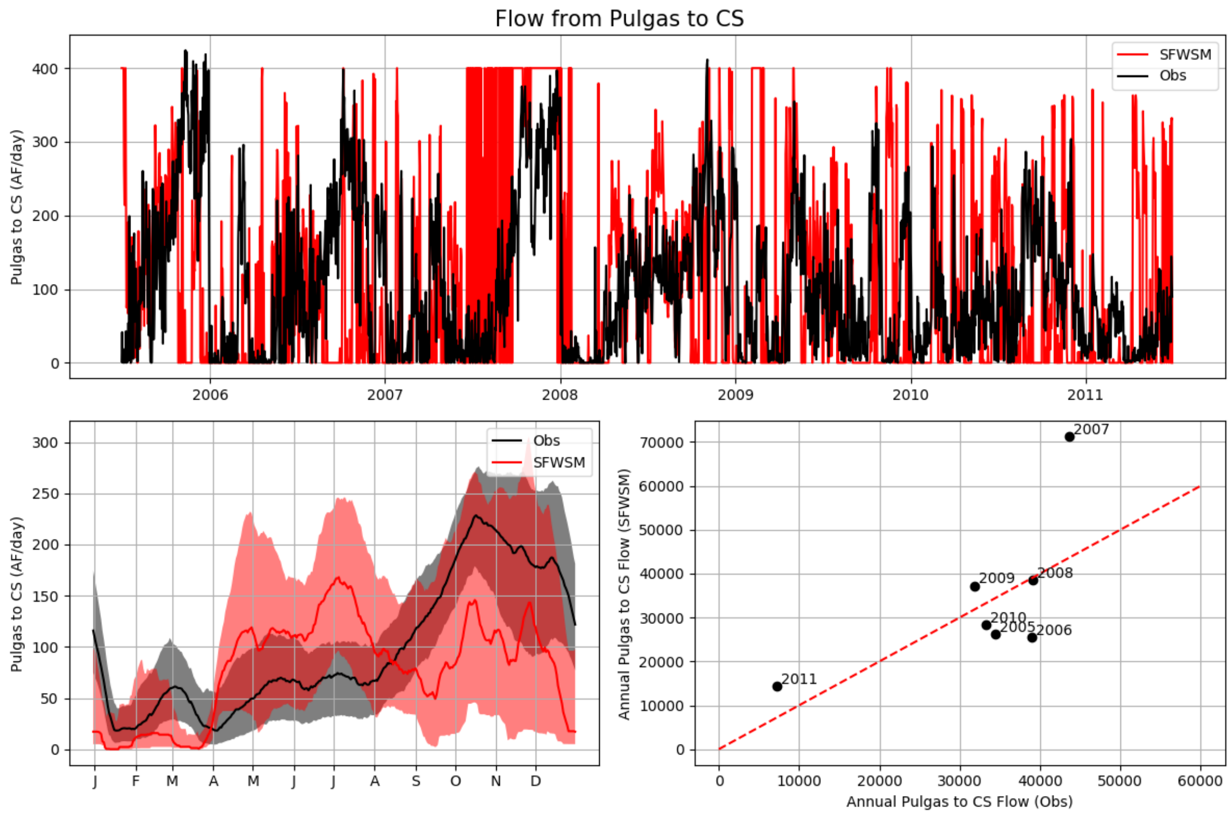

Figure 35 Water transfer from the Pulgas diversion to Crystal Spring

5.1.5 Hydropower Generation in Upcountry

Figure 36 and Figure 37 show hydropower operations at Kirkwood (Figure 36) and Holm (Figure 37) powerhouses. Observed power generation is not available. As such, only the water releases through the power tunnels (power draft) are discussed here, although SFWSM also outputs power generation. Note also that observed power draft at Kirkwood powerhouse is only available from the end of 2009, which significantly limits the comparison with SFWSM simulation to this period.

Figure 36 and Figure 37 highlight the effect on the regulation of the fluctuation in the SJPL on the water release pattern through the Canyon Power Tunnel. Both seasonal and inter-annual variabilities of the power draft in Cherry and Canyon power tunnels are fairly well represented in SFWSM. However, there is a peak in the power draft during February in both Cherry and Canyon power tunnels that is absent in the historical record (this is especially true for the Canyon Power Tunnel). These peaks are like due the minimum flow constraints associated with the tunnels. As noted above, the minimum flow constraints in Cherry and Canyon power tunnels use perfect foresight forecasts for WAC and for the net inflows to the reservoirs and compares it to the available storage. The forecast procedure starts in February, which may explain the excess generation at Kirkwood powerhouse during this month for the year with large forecasted WAC and/or large forecasted net inflow to Hetch Hetchy reservoir.

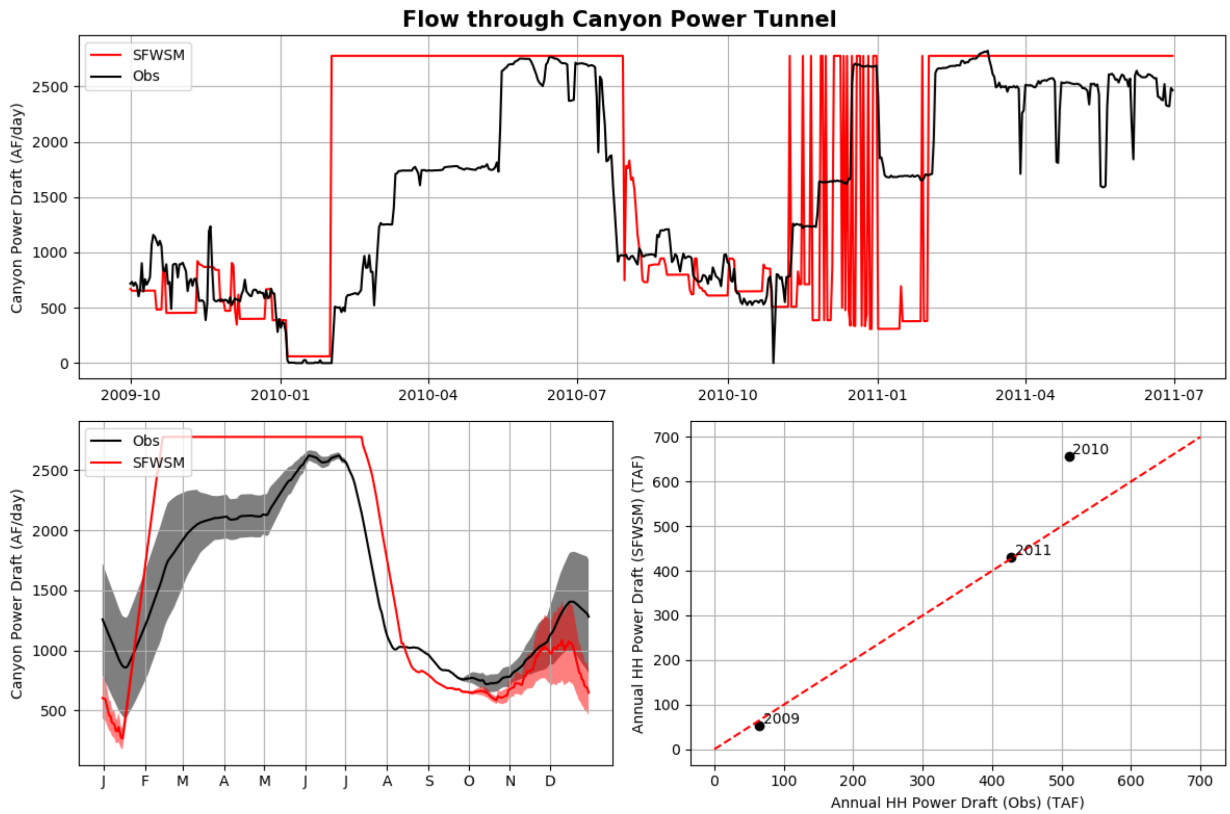

Figure 36 Flow through the Canyon Power Tunnel

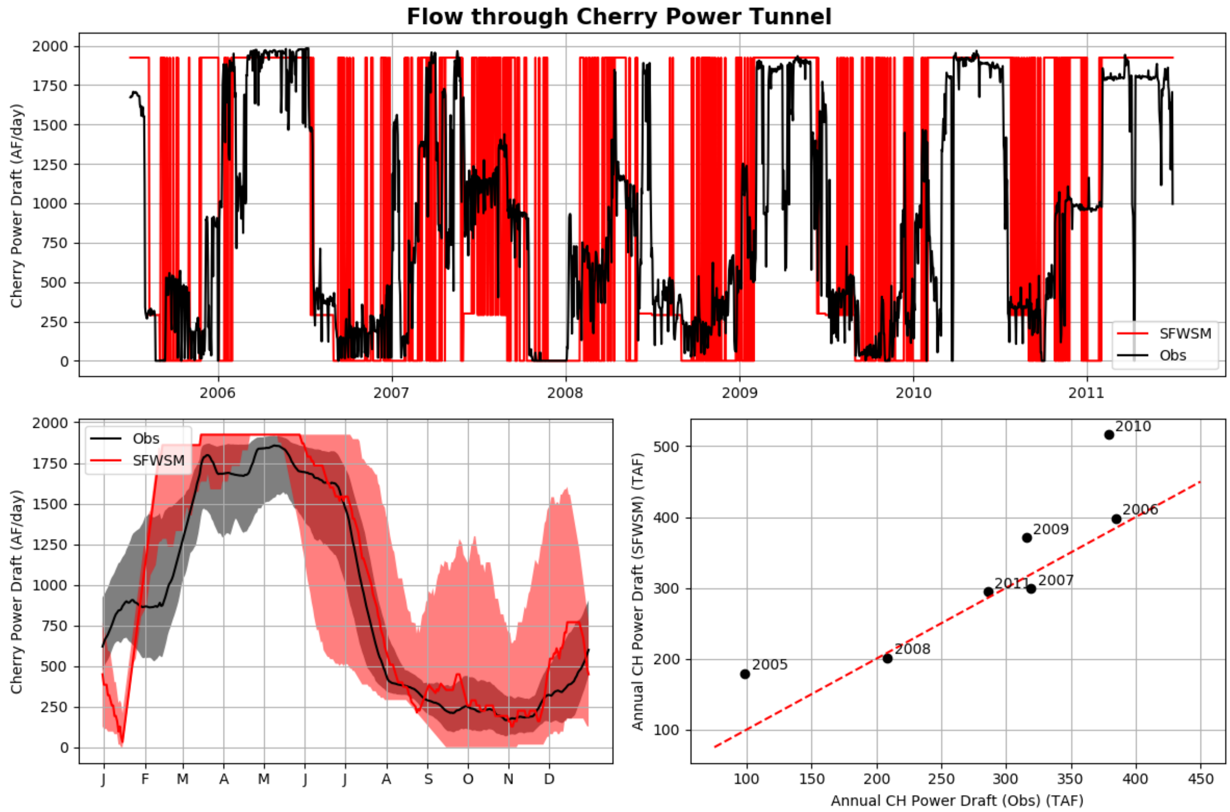

Figure 37 Flow through the Cherry Power Tunnel

5.1.6 Conclusion

SFWSM is skillful at representing system-wide variables and reservoir storage fluctuations of the most important infrastructure. It successfully delivers the demand with no deficit (i.e., no virtual water is needed), despite some challenging periods with maintenance of the different water treatment plants and conveyances. SFWSM also represents reasonably well the water flow through key components of the system such as SJPL and the Sunol Valley and Harry Tracy water treatment plants. It represents fairly well the power draft through the power tunnels at Kirkwood and Holm powerhouses, although an overestimation of flow in winter can occur during the year with large forecasted WAC and/or net inflow to the Hetch Hetchy and Cherry reservoirs

Overall, the performance of SFWSM at mimicking the operations of the WRS is good and no major issue has been detected that would prevent a reasonable validation with HHLSM, described in the following section.

5.2 Validation with HHLSM

The HHLSM validation version cannot be directly compared with HHLSM, due to a mismatch in modeling requirements related to inflow hydrology. There is no sufficiently accurate daily time step dataset of inflows that spans a sufficiently long time period as HHLSM, which uses a reconstructed 91-year monthly time step inflow dataset. On the other hand, it is not possible to easily convert hydrologic output generated for SFWSM for use in HHLSM. Because of this discrepancy, monthly averages were instead compared between the previously-run HHLSM model provided by SFPUC and output from the HHLSM validation version of SFWSM run with nine (9) climate realizations from the LTVA weather generator assuming historical climate characteristics.

In the comparisons that follow, the spread of averages across all nine realizations are considered. However, in some instances subsets of output are compared to assess behavior under similar—though not exact—hydrologic conditions. Of specific interest is the need to compare behavior under more extreme drought conditions. In addition to mean monthly comparisons, representative time series are compared graphically, and some quantitative metrics are compared between HHLSM and the nine realizations.

Importantly, the purpose of this validation is not necessarily to show that the HHLSM validation version of SFWSM mimics HHLSM exactly, as HHLSM is not considered better, but rather to show that more important operations are consistent and that any significant differences can be explained where needed.

We also note that in the comparison figures, the system component names found in HHLSM are used.

5.2.1 System-wide

Reservoir storage

Comparisons of aggregated reservoir storage across the RWS are shown in Figure 38, while disaggregated storage is shown in Figure 39. The Upcountry region behaves broadly consistently between SFWSM and HHLSM. Hetch Hetchy Reservoir fills slightly earlier in HHLSM. Cherry Lake storage peaks slightly higher in HHLSM, while Lake Eleanor storage remains higher longer in HHLSM compared to SFWSM. The Water Bank account peaks significantly higher in HHLSM compared to SFWSM, due specifically to the lack of the Water Bank bubble account in SFWSM, as noted above. Collectively, these result in similar storage level in the Upcountry Hetch Hetchy system, except during May month (Figure 38b). The discrepancy between HHLSM and SFWSM during May months likely results from the use of perfect foresight forecasts that tends operate the power tunnel more aggressively than SFPUC does in reality, leading to lower storage levels during that period. SFWSM also exhibits a much greater variability in storage levels, especially when looking at the minimum and maximum simulated values, which likely results from both the use of perfect foresight forecasts, and more extremes hydro-meteorological conditions within the nine realizations than in the HHLSM inputs.

Reservoirs in both the East Bay and Peninsula are generally fuller, on average, in HHLSM. The main reason for this is that the reservoir preferred storage levels are more strictly adhered to in HHLSM. This is particularly apparent in all three reservoirs in the East Bay region—especially

Calaveras reservoir—(Figure 38h-i). A possible explanation is the difference in temporal resolution used by HHLSM and SFWSM. Arroyo-Hondo catchment, which feds Calaveras reservoir, is subject to flash flood events, particularly challenging but with limited impact on the operation when averaged at monthly time step. SFWSM operates Calaveras to release water to either the treatment plant or through the valve (i.e., controlled spill) using 7-day lead time perfect forecast to reduce the uncontrolled spill, which may explain the lower levels, on average. In addition, SFWSM is much more actively managed, based on system needs rather than strictly preferred reservoir storage, resulting in a much greater variation in reservoir storage. SFWSM was not modified to operate reservoirs to more strictly follow the preferred storage levels as in HHLSM.

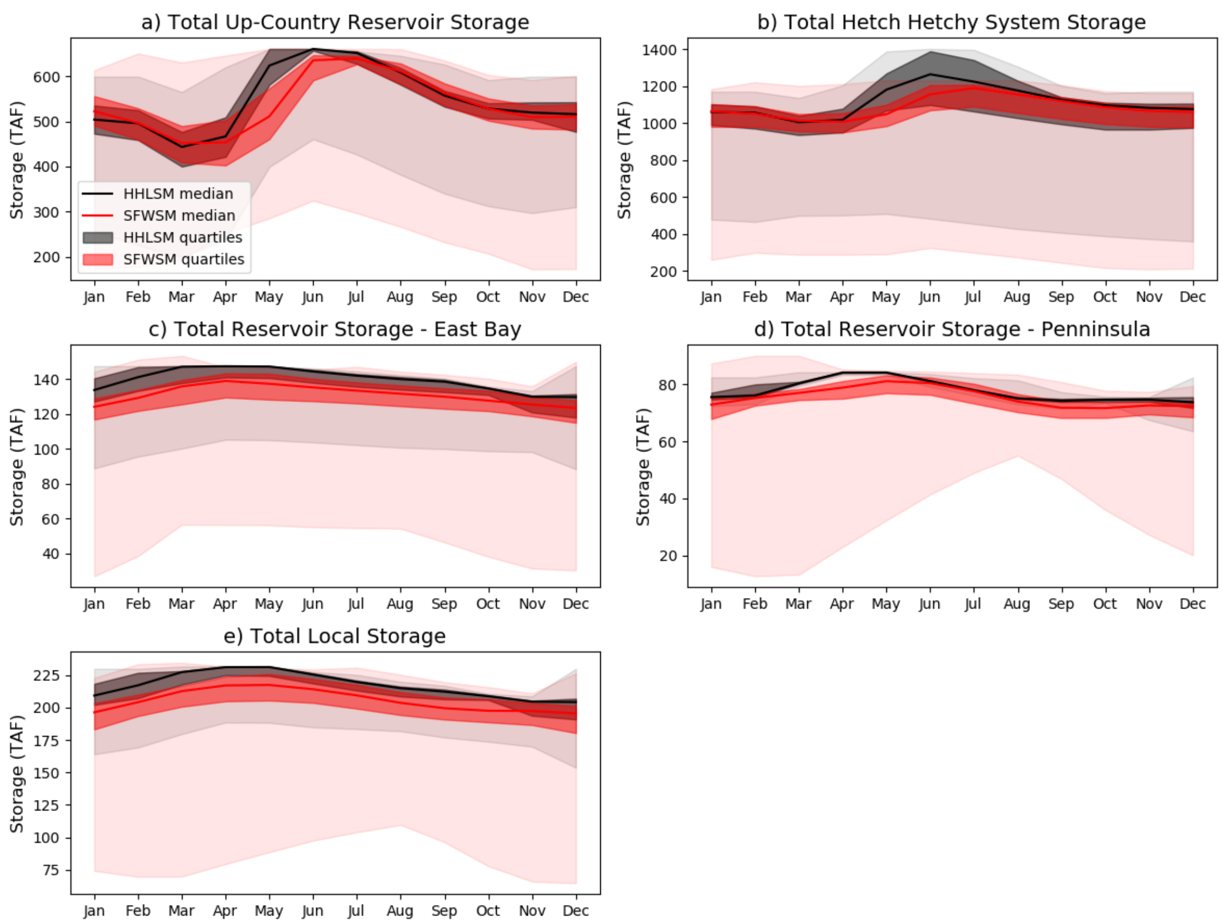

Figure 38. Comparison monthly aggregated reservoir storage for HHLSM and SFWSM (9 climate realizations). The dark shaded areas show the deviation between the inter-quartiles. The light shaded areas show the deviation between the maximum and minimum simulated values across the nine realizations.

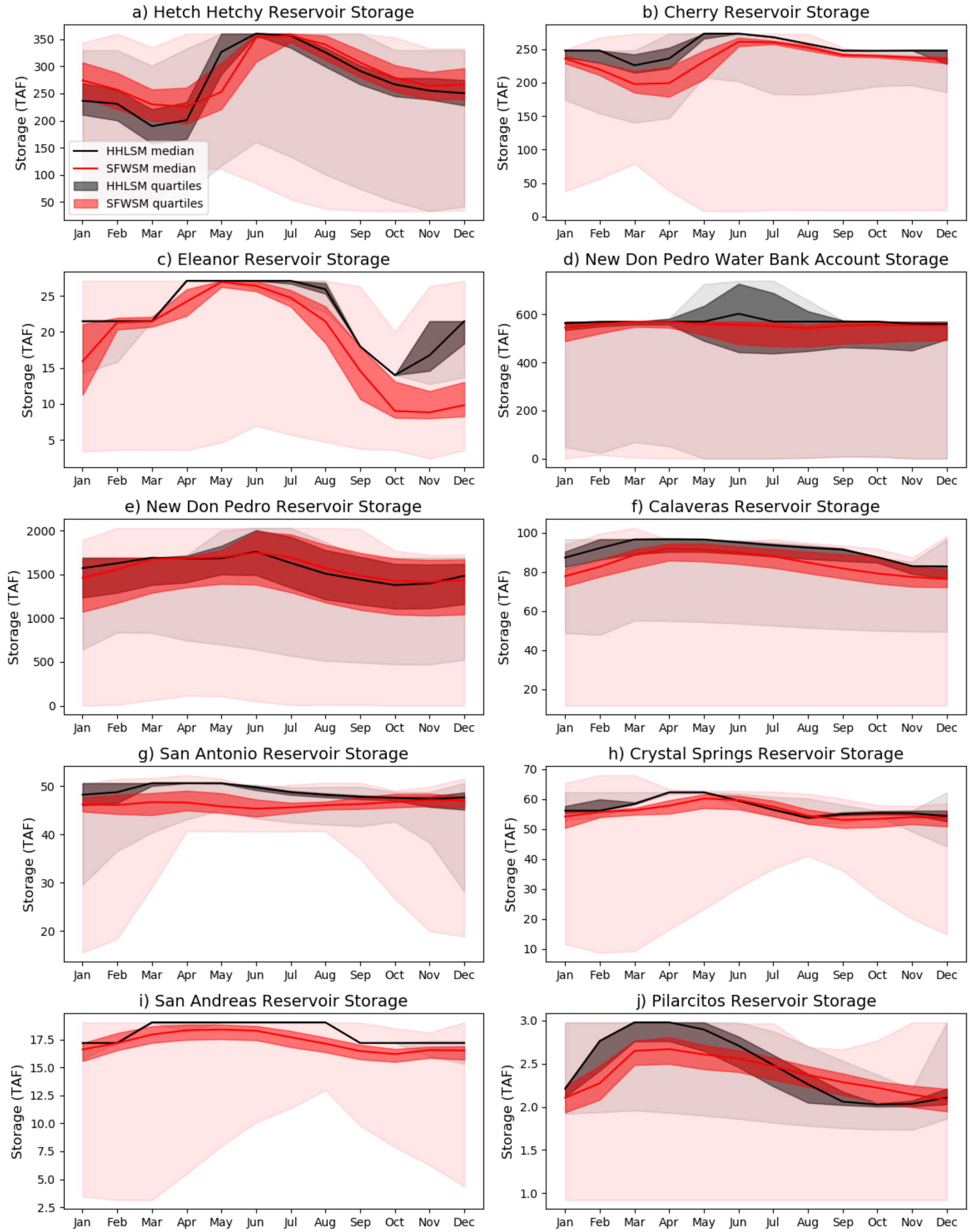

Figure 39. Comparison monthly reservoir storage for HHLSM and SFWSM (9 climate realizations). The dark shaded areas show the deviation between the inter-quartiles. The light shaded areas show the deviation between the maximum and minimum simulated values across the nine realizations.

San Joaquin Pipelines

Figure 40 shows flow in the San Joaquin Pipelines (SJPL) in both HHLSM and SFWSM. Operations are broadly consistent between the two models, though the deviation between the maximum and minimum monthly flow values are larger in SFWSM. As with reservoir operations, we attribute this greater variability in SFWSM as a more detailed response to a more complex system represented by SFWSM and to hydro-meteorological conditions that are more challenging within the nine realizations than in the HHLSM inputs.

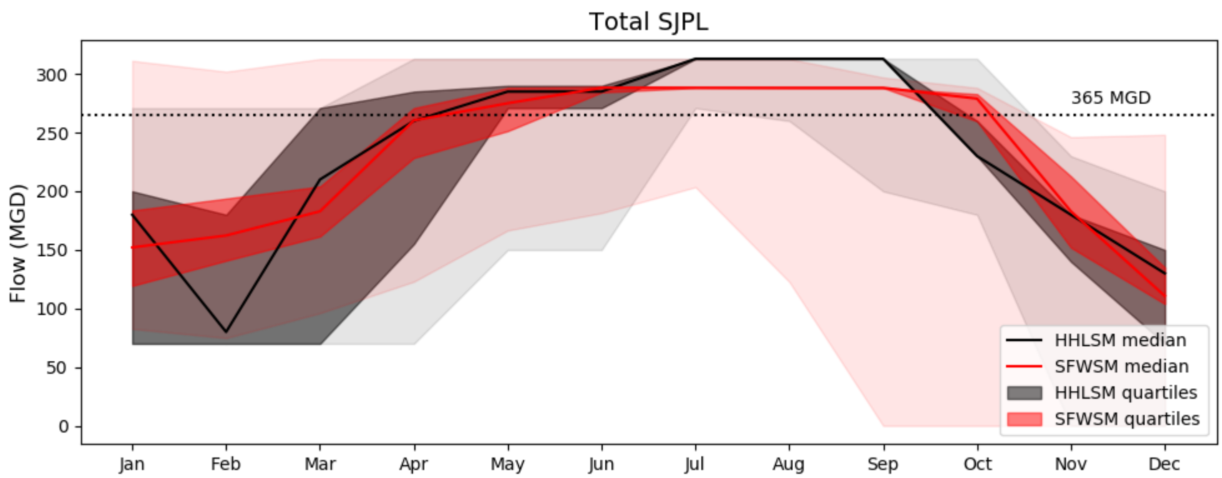

Figure 40. Comparison monthly San Joaquin Pipeline (SJPL) flow for HHLSM and SFWSM (9 climate realizations). The dark shaded areas show the deviation between the inter-quartiles. The light shaded areas show the deviation between the maximum and minimum simulated values across the nine realizations.

Years of Remaining Supply

Years of Remaining Supply (YRS) is a critical metric used in SFWSM to trigger different drought rationing measures, as discussed above. Table 23 includes a comparison of basic statistical parameters (across all climate realizations), including mean, standard deviation, minimum, and maximum. Generally, YRS is represented similarly in both models, with SFWSM having slightly less variability on average, although some realization shows larger variability, which is the case for the eight realization, which also has a lower minimum YRS than HHLSM does (i.e., 1.81 vs. 2.1, which is the simulated values by HHLSM after the 1987-1992 drought).

Table 23. Comparison of statistical parameters for Years of Remaining Supply (YRS) between HHLSM and SFWSM for each LTVA climate realization

	Mean	Std. Dev.	Median	Min.	Max.
HHLSM	3.98	0.57	4.12	2.10	4.50
SFWSM R1	4.03	0.43	4.17	2.70	4.83
SFWSM R2	3.95	0.41	4.12	2.73	4.50
SFWSM R3	3.95	0.57	4.13	2.50	5.20
SFWSM R4	3.94	0.51	4.17	2.63	4.80

SFWSM R5	4.02	0.45	4.17	2.42	5.06
SFWSM R6	4.03	0.37	4.16	2.98	5.13
SFWSM R7	3.99	0.44	4.13	2.69	5.23
SFWSM R8	3.95	0.62	4.14	1.81	5.28
SFWSM R9	3.99	0.41	4.16	2.68	4.71
SFWSM mean	3.98	0.47	4.15	2.57	4.97

Reservoir spill

Spill (including both controlled and uncontrolled SFWSM spill) is compared generally for a subset of reservoirs in Figure 41 and Figure 42, which show mean monthly spill and monthly spill quantiles, respectively, for HHLSM and SFWSM. Spill for Hetch Hetchy, Don Pedro and San Antonio reservoirs are generally consistent between the models. Spill for Calaveras is significantly higher in SFWSM than in HHLSM. The latter is due to the high bias in runoff into Calaveras Reservoir, as noted above.

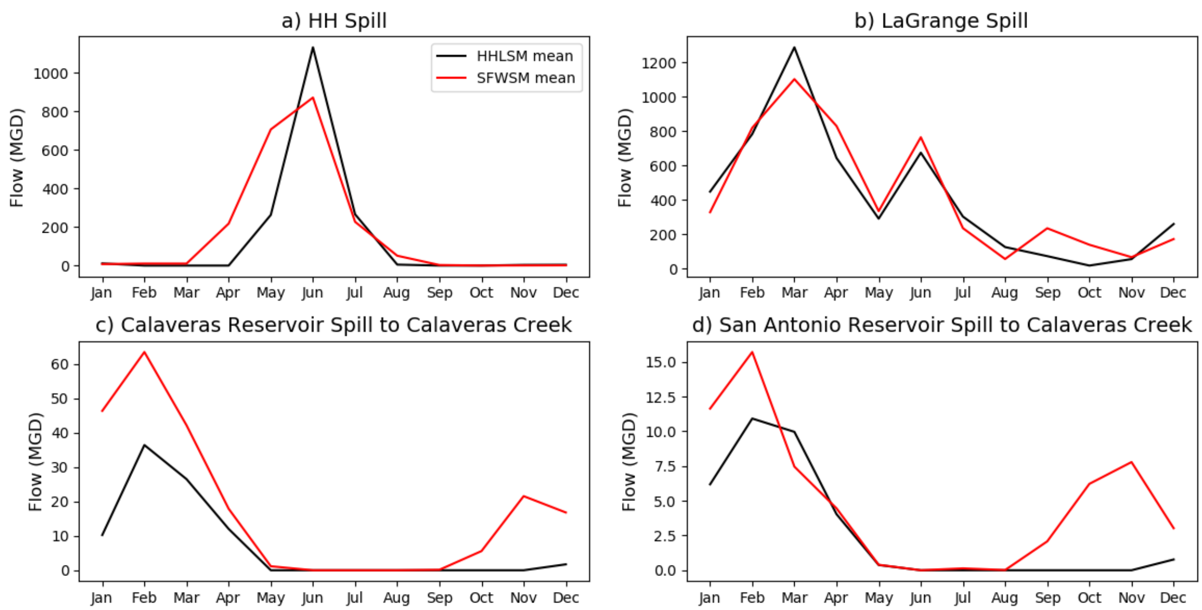

Figure 41. Comparison of mean monthly reservoir spill for HHLSM and SFWSM (9 climate realizations) for selected reservoirs.

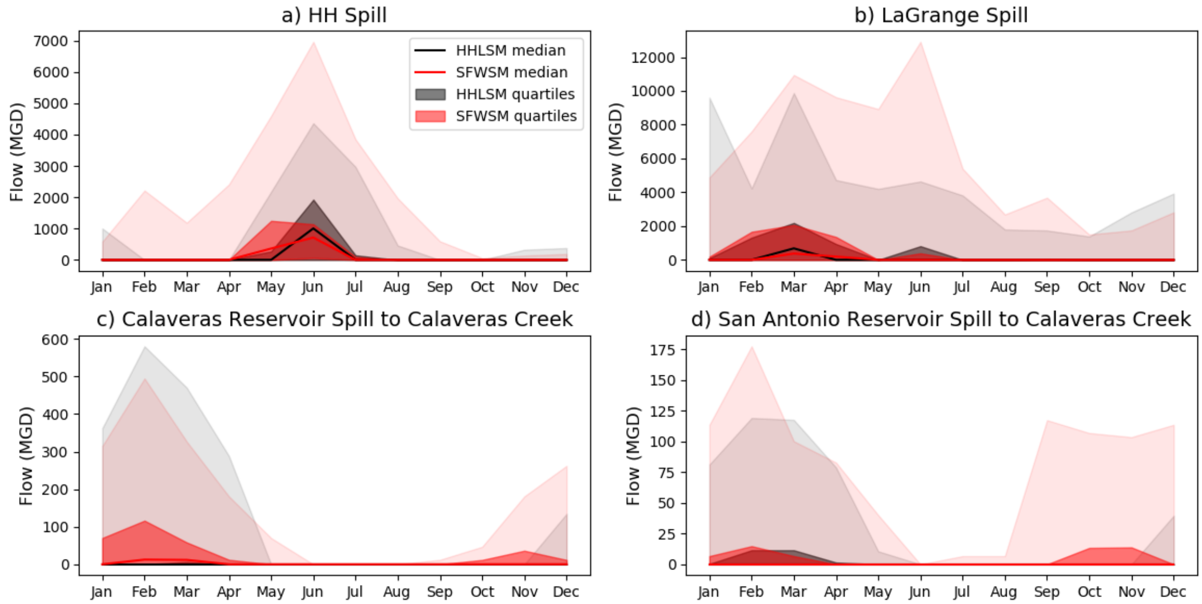

Figure 42. Comparison of monthly reservoir spill for HHLSM and SFWSM (9 climate realizations) for selected reservoirs.

Virtual Water

No virtual water is required but for the ninth realization. The virtual water is required to supply water during the summer 2063 in South East bay region. The overall deficit equals 2895 AF. Figure 43 illustrates the operations in the Upcountry region during the drought leading to need of virtual water in South East Bay region. The requirement for virtual water in 2063 follows the driest year from the ensemble of nine realization. As illustrated in Figure 43, the year 2062 not only has no WAC, the inflow at La Grange are almost null too. During this exceptional dry water year, Hetch Hetchy reservoir draws down quickly to its deadpool. Early 2063, local reservoirs are solicited much more they usually are to supply the demand in the Bay area. Early summer 2063, Calaveras runs dry and reaches deadpool (not shown). The water transfer from San Antonio to the Sunol Valley treatment, although used at full capacity (i.e., 270 AF/day) is not enough to deliver the demand in South East Bay, which leads to deficit (i.e, virtual water). Note that no deficit in observed in the Peninsula region since Crystal Springs did not run dry during this event.

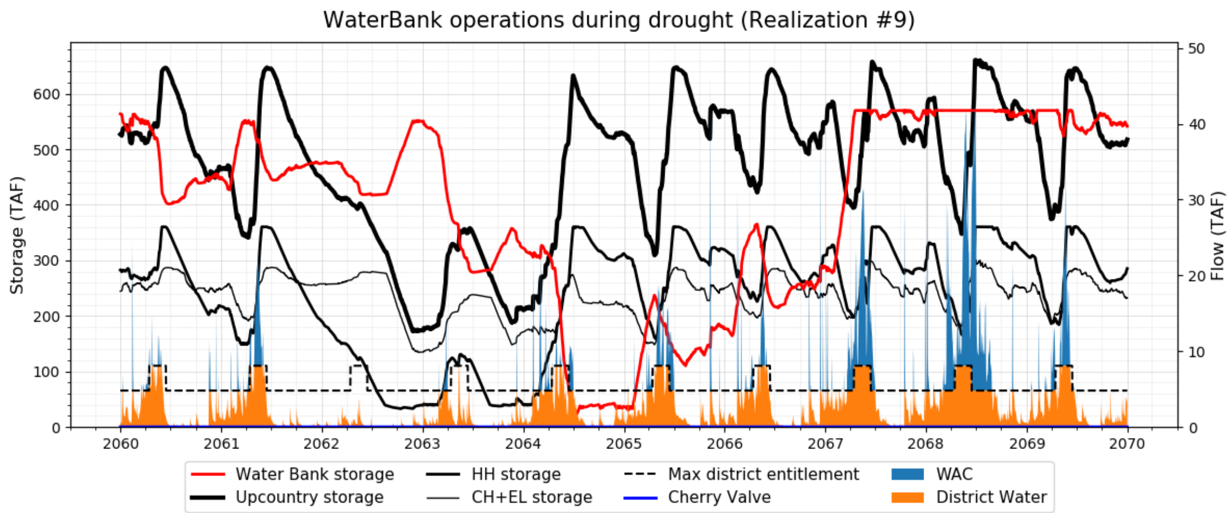

Figure 43 Operation of Upcountry reservoirs during drought for the ninth climate realization (historical climate conditions). Upcountry Storage includes Hetch Hetchy Reservoir, Lake Eleanor, and Cherry Lake. WAC is Water Available to the City.

5.2.2 Conclusions

The HHLSM validation version of SFWSM generally represents operations in a manner consistent with HHLSM, particularly for some of the more important operations such as reservoir storage and drought rationing triggers. In some specific cases, operations between SFWSM and HHLSM diverge, though in ways that are known and understood. Specifically, operations of local reservoirs were not modeled in exactly the same way, with HHLSM more strictly modeled to follow reservoir preferred storage levels. In fact, the local reservoirs were not actually operated this way, based on the historical record.

Finally, we note that this is only a partial comparison of the HHLSM validation version of SFWSM with HHLSM. The purpose here is again to compare overall behavior, without a thorough diagnostic. A more comprehensive comparison of the two models could help advance understanding of the differences between the two models. However, results from this general assessment generally indicates that such a more thorough comparison may not be warranted, given the model reasonably well represents historical operations. Based on both the historical comparison above and this comparison with HHLSM, we again conclude that SFWSM reasonably well represents operations of the Hetch Hetchy RWS and is good for performing further studies for long term planning. Furthermore, it can readily be improved as needed for future studies beyond the immediate purpose of the LTVA.

References

Draper, A., Darabzand, S., 2003. CalSim II Simulation of Historical SWP-CVP Operations (Report). California Department of Water Resources.

Free Software Foundation, 2014. GNU Linear Programming Kit Reference Manual for GLPK Version 4.55 (DRAFT, August 2014).

Maidment, D.R., Miaou, S.-P., 1986. Daily Water Use in Nine Cities. Water Resour. Res. 22, 845–851. https://doi.org/10.1029/WR022i006p00845

Markstrom, S.L., Regan, R.S., Hay, L.E., Viger, R.., Webb, R.M.., Payn, R.A., LaFontaine, J.H., 2015. Modeling Techniques, book 6. chap. B7, in: PRMS-IV, the Precipitation-Runoff Modeling System,, Version 4, Techniques and Methods. U.S Geological Survey, U.S. Geological Survey, Reston, Virginia, p. 169.

SFPUC, 2016a. 2015 URBAN WATER MANAGEMENT PLAN for the City and County of San Francisco [Public Review Draft]. SFPUC, San Francisco.

SFPUC, 2016b. Alameda Creek Recapture Project Fact Sheet.

SFPUC, 2014. O'Shaughnessy Dam Instream Flow Management Plan – Stakeholder Draft. San Francisco Public Utilities Commission, San Francisco.

SFPUC, 2009a. HH/LSM Documentation Volume 1: Water Supply System Modeling Report. SFPUC.

SFPUC, 2009b. WATER SUPPLY AGREEMENT between THE CITY AND COUNTY OF SAN FRANCISCO and WHOLESALE CUSTOMERS in ALAMEDA COUNTY, SAN MATEO COUNTY AND SANTA CLARA COUNTY.

Steiner, D., 2013. Project Operations/Water Balance Model (DON PEDRO PROJECT FERC NO. 2299 No. FERC No. 2299).

Tomlinson, J.E., Arnott, J.H., Harou, J.J., 2020. A water resource simulator in Python. Environ. Model. Softw. 126, 104635. https://doi.org/10.1016/j.envsoft.2020.104635

HRG TR1 (2018). Technical Report 1: Weather Generator Module. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.

HRG TR2 (2021). Technical Report 2: Hydrologic Modeling Module. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.

HRG TR3 (2020). Technical Report 3: Urban Water Demand. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts.

HRG LTVA (2021). Long Term Vulnerability Assessment and Adaptation Plan for the SFPUC Water Enterprise – Phase 1. Hydrosystems Research Group, University of Massachusetts, Amherst, Amherst, Massachusetts

Yates, D., Sieber, J., Purkey, D., Huber-Lee, A., 2005. WEAP21 – A Demand-, Priority-, and Preference-Driven Water Planning Model: Part 1– Model Characteristics. Water Int. 30, 487–500.