https://doi.org/10.7289/V5/TM-SWFSC-597

NOAA Technical Memorandum NMFS

MARCH 2018

CALIBRATION AND VALIDATION OF LINKED WATER TEMPERATURE MODELS FOR THE SHASTA RESERVOIR AND THE SACRAMENTO RIVER FROM 2000 TO 2015

Miles E. Daniels[1,2], Vamsi K. Sridharan[1,2], Sara N. John[1,2], and Eric M. Danner[2]

[1] University of California, Santa Cruz, Institute of Marine Sciences

[2] NOAA Fisheries, SWFSC Fisheries Ecology Division
110 McAllister Way, Santa Cruz, CA 95060

NOAA-TM-NMFS-SWFSC-597

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Southwest Fisheries Science Center

About the NOAA Technical Memorandum series

The National Oceanic and Atmospheric Administration (NOAA), organized in 1970, has evolved into an agency which establishes national policies and manages and conserves our oceanic, coastal, and atmospheric resources. An organizational element within NOAA, the Office of Fisheries is responsible for fisheries policy and the direction of the National Marine Fisheries Service (NMFS).

In addition to its formal publications, the NMFS uses the NOAA Technical Memorandum series to issue informal scientific and technical publications when complete formal review and editorial processing are not appropriate or feasible. Documents within this series, however, reflect sound professional work and may be referenced in the formal scientific and technical literature.

SWFSC Technical Memorandums are accessible online at the SWFSC web site.
(http://swfsc.noaa.gov)

Print copies are available from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.
(http://www.ntis.gov)

Recommended citation

Daniels, Miles E., Vamsi K. Sridharan, Sara N. John, and Eric M. Danner. 2018. Calibration and validation of linked water temperature models for the Shasta Reservoir and the Sacramento River from 2000 to 2015. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-597.
https://doi.org/10.7289/V5/TM-SWFSC-597

Calibration and Validation of Linked Water Temperature Models for the Shasta Reservoir and the Sacramento River from 2000 to 2015

Miles E. Daniels[1], Vamsi K. Sridharan[1], Sara N. John[1], and Eric M. Danner[2]

[1] University of California, Santa Cruz, Institute of Marine Sciences, Affiliated with: Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration

[2] Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration

Table of Contents

Abstract: Mechanistic-based water temperature models simulating how environmental and operational conditions affect water temperature dynamics in the Shasta/Sacramento system can aid water resource management in the region. This report outlines the process of linking two water temperature models (an upstream reservoir model to a downstream river model), including calibration and validation from 2000 to 2015 and the procedures used to run the models operationally. Validation results (even years) indicated the reservoir model matched observed vertical temperature distribution and discharge temperature well, with a RMSE on the order of 1° C. When ran with known upstream boundary conditions, the river model accurately matched daily average river temperatures, with a RMSE near 0.5° C. Running the models in series (i.e. the reservoir model supplying boundary conditions to the river model) resulted in the RMSE of daily average river temperatures to increase to approximately 1° C. Both the reservoir and river models tended to perform poorer from May to October during the temperature management season and predict warmer temperatures than observed. While further model refinement is needed, the linked model framework represents a useful tool to evaluate temperature dynamics of the system under past and future environmental and operational conditions.

1. Introduction

Dams and reservoirs fundamentally alter the flow and temperature of water within a watershed, and these changes can have important consequences for aquatic habitats downstream. Managers have the ability to control the timing, volume, and in some systems, the temperature of the water released. As a result, the ability to predict the thermal conditions of reservoirs and downstream rivers based on operations is important to both water and wildlife managers. When temperature targets are set for downstream locations, they can constrain other utilities of reservoirs, such as timing and volume of water released for downstream agricultural, industrial, or municipal uses. Additionally, when water resources are limited, such as during droughts, environmental and human uses downstream of dams both tend to rely more heavily on reservoir water, which increases the challenge of meeting multiple use objectives.

Shasta Reservoir is the largest reservoir by volume in the state of California (Figure 1). It is a major source of water for the Central Valley Project, which is connected to the State Water Project; combined they account for approximately 13% of California's dedicated urban and agricultural water supply (*1*). Prior to the construction of Shasta Dam in the 1940s, Sacramento River winter-run Chinook salmon (*Oncorhynchus tshawytscha*) migrated past the current location of Shasta Dam and into the upper reaches of the Sacramento River, McCloud River, and other waterways to spawn (*2*). Since the construction of Shasta and Keswick Dams, winter-run Chinook are forced to spawn below Keswick Dam (the after-bay of Shasta) in the Sacramento River, where they have been reliant on reservoir discharge to provide enough cold-water habitat for early life stage development (i.e. egg incubation).

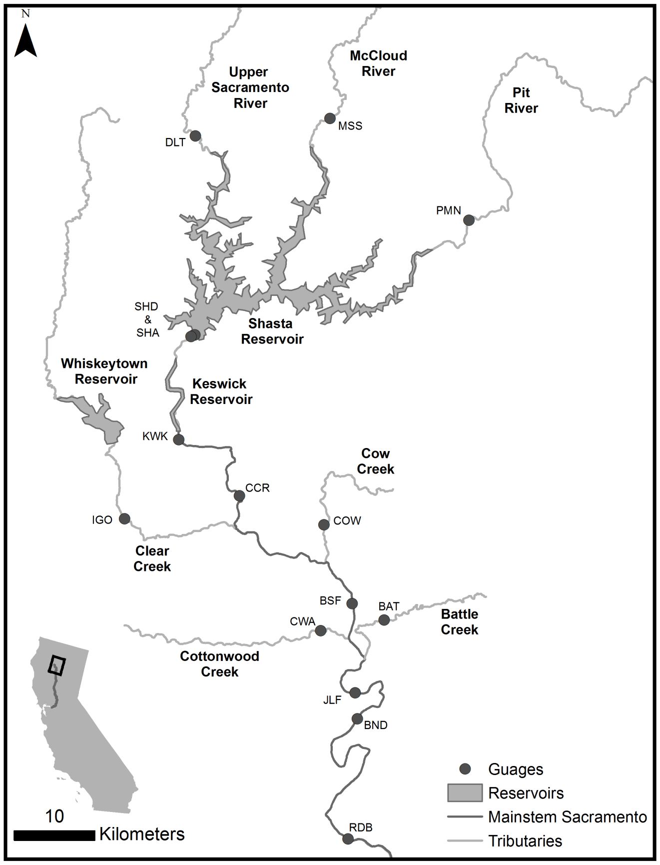

Figure 1: Study site with primary reservoirs and tributary inputs into the main stem of the Sacramento River shown. Specific gauge locations are labeled with the identifier used on the California Data Exchange Center website (cdec.water.ca.gov).

To mitigate the adverse effects of warm water temperatures on winter-run Chinook spawning, egg incubation, and fry emergence, the U.S. Bureau of Reclamation (USBR) installed a temperature control device (TCD) in 1997 on Shasta Dam to allow for selective withdrawal from the thermally stratified reservoir. By opening gates at different elevations on the TCD, managers can access colder, deeper water during summer months, while maintaining power generation capabilities (*3*). This improved flexibility of cold-water management provided by the TCD has added to the need for additional tools to evaluate the temperature dynamics of the system and model multiple operating scenarios to ensure enough cold-water for salmon while supplying water for additional uses downstream.

Mechanistic-based simulation models of water temperature provide a method to understand factors affecting thermal dynamics in reservoirs and rivers, and can inform the planning of cold-water resource allocation. Understanding how a particular model works, assumptions behind the model, the data used to make predictions, and the level of agreement between model predictions and observed conditions are important steps in the appropriate application of a model (*4*).

There is a need to generate operational forecasts to aid in Shasta Reservoir achieving downstream temperature targets for informed cold-water management. The goals of this report are twofold. First, to describe the process of linking two water temperature models (a reservoir model and a downstream river model) which are used to generate model predictions that are then pushed to a decision support tool website for water temperature management in the Shasta/Sacramento system. Description of the website, named the Central Valley Temperature Mapping and Prediction (CVTEMP), is not the focus of this report, but output from the reservoir and river models can be viewed at the following web address: http://oceanview.pfeg.noaa.gov/CVTEMP/. The second goal of this report is to evaluate the ability of the models to accurately predict water temperature through a calibration and validation process. The work described here is ongoing, and this report describes the current operational state of the modeling effort.

This document is divided into eight primary sections: 1) introduction, 2) background on water temperature modeling, 3) outline of the processes involved with linking the models to run operationally for the CVTEMP website, 4-7) description of each model including calibration and validation performance, and 8) conclusions.

2. Mechanistic approaches to modeling water temperature

Simulating temperature dynamics in reservoirs and rivers involves accounting for the various inputs and outputs of thermal energy, which cause the temperature of a parcel of water to either increase or decrease. Predominantly, there are four pathways thermal energy can take in the water column. Energy can be exchanged (1) at the air-water interface and (2) at the bed-water interface (*5*). Energy can also be (3) redistributed vertically or (4) gained or lost horizontally via advection and mixing processes (*6*). To account for these inputs and outputs of thermal energy (i.e. fluxes), simulation models typically use a heat budget approach (*5*).While the exact form of the heat budget may differ between models, there is a general conceptual form of the heat budget and its components for reservoirs and rivers (Figure 2). The full details on the methods of simulating thermal energy dynamics for the two water temperature models used in this work are beyond the scope of this report, and are described elsewhere (*7, 8*).

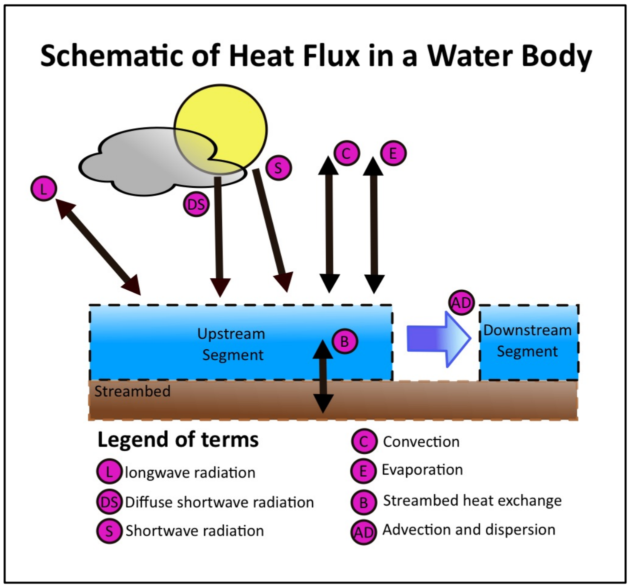

Figure 2: Conceptual diagram of the various inputs and outputs of heat considered to be the primary factors affecting heat flux in a water body.

3. Linked operational modeling framework

Two primary water temperature models were used to link reservoir temperature to river temperature: CE-QUAL-W2 (W2) for the reservoir (*8*), to simulate the hydraulic and thermal dynamics of Shasta Reservoir; and the River Assessment for Forecasting Temperate (RAFT) model (*7*), to simulate the hydraulic and thermal dynamics of the Sacramento River. This section outlines the general process of linking W2 and RAFT. All processes and computations were performed within the MATLAB (*9*) and R (*10*) modeling environments.

3.1 Model set up and data acquisition

When running in an operational setting, the linked models were set up to be updated daily over a user specified time period (both start and end date, hereafter “simulation window”). All required model inputs for W2 and RAFT were specified for the simulation window, which involved generating both hindcast and forecast sets of input files consisting of operational conditions of reservoirs, hydrological inputs, and meteorological forcing conditions. The time period from simulation start date to model run date is referred to as the hindcast period, while the period from model run date to simulation end date is referred to as the forecast period. As each day passed, observed/updated model input data were downloaded via web services or updated manually and ingested by W2 and RAFT. With this approach, the models used observed/updated conditions of the system when available in an effort to improve the skill of the forecast going forward.

To update model inputs automatically via web services, download modules were developed in R and MATLAB to access servers. Input data were accessed via the California Data Exchange Center (CDEC) (http://cdec.water.ca.gov/), NOAA’s Operational Model Archive and Distribution System (NOMADS) (http://nomads.ncep.noaa.gov/), and the California Nevada River Forecast Center (CNRFC) (http://www.cnrfc.noaa.gov/) websites. Within the modules were additional processes to check for obvious input data errors dependent on the measurement unit (e.g. water temperature > 100° C). The download modules filled input data gaps using an algorithm based on singular spectrum analysis (*11*), which maintained dominant trends and periodicity evident in the original input data (see Figure 3 for example of gap filling air temperature).

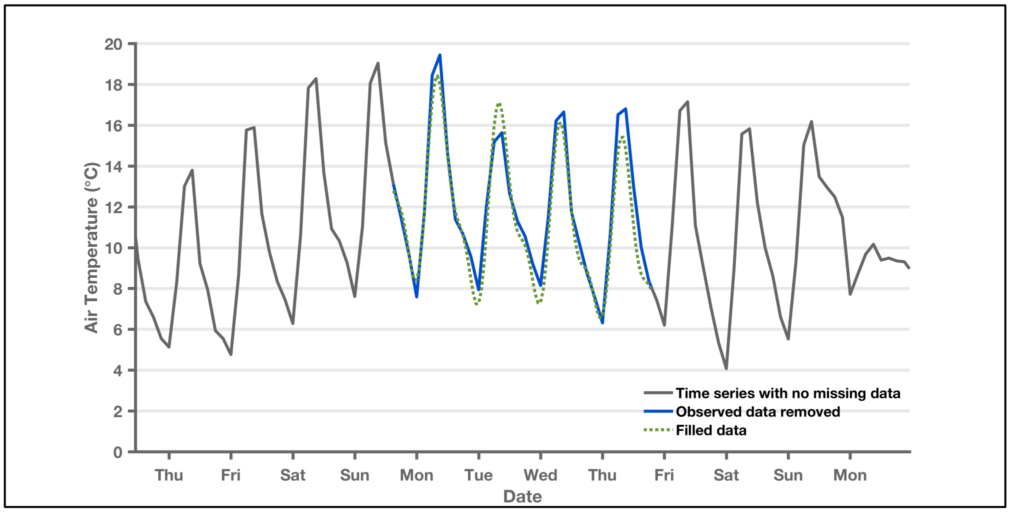

Figure 3: Example of gap filling procedure where approximately 4 days of air temperature data were intentionally removed and filled to demonstrate the ability of the procedure to generate a continuous time series of model input.

3.2 <u>Running the models</u>

There are four primary steps taken to run the models on a daily basis:

- Update operational conditions (reservoirs)
- Update hydrological data (reservoirs and rivers)
- Update meteorological forcing data (reservoirs and rivers)
- Format inputs and run models

Additional details are provided in subsequent sections.

3.2.1 <u>Update operational conditions of reservoirs</u>

The operational inputs required to run the W2 and RAFT models were discharge volumes and discharge temperatures for Shasta and Keswick Reservoirs, and Shasta TCD operations. These were provided by USBR, who operate Shasta and Keswick Reservoirs. Files were received in two formats: a Data Storage System (DSS) file, which is a file system developed by the US Army Corps of Engineers Hydrologic Engineering Center (HEC) and is compatible with HEC developed models, and a log file, which is a text file containing the operational conditions of the TCD at Shasta Reservoir.

Each time new files were received, the contents of the DSS file were exported to a Microsoft Excel file and saved in the root directory along with the log file. At the start of each simulation, a module was used to identify the files with the most recent creation date and those files were used to generate a time series of forecasted inputs relating to reservoir operations. Therefore, as USBR made adjustments to their forecasted operations of Shasta and Keswick Reservoirs, the models reflected these adjustments.

For the hindcast period, observed data for Shasta and Keswick discharge were gathered via the download module outlined previously. For hindcast TCD operations, there was no web service available providing machine-readable data. Therefore, the TCD operations were manually adjusted when they differed from the planned operations using information from records (USBR notification e-mails).

3.2.2 Update hydrological data

At locations that experience high hydrologic influx, W2 and RAFT require specified discharge and temperature inputs (Figure 1).

For the hindcast period, discharge volume inputs were based on gauge data. For discharge temperature, gauge data were used when available, otherwise historical data from 2000-2015 were used to generate monthly or daily 10^{th}, 50^{th}, and 90^{th} percentile estimates of discharge temperature for a given location. Gauges without near real-time discharge temperature data included CWA, BAT, and COW.

During the forecast period, input data were generated from historical conditions or data provided by the CNRFC. Specifically, CNRFC data were used to assign discharge volumes for three of the primary inputs into Shasta Reservoir (gauges DLT, MSS, and PMN) at the 10^{th}, 50^{th}, and 90^{th} percentile levels using a data download module that accessed the Advanced Hydrologic Prediction Services (AHPS) product. For all other forecasted inputs, historical data from 2000-2015 were used to generate monthly or daily 10^{th}, 50^{th}, and 90^{th} percentile estimates of discharge volume and temperature.

3.2.3 Update meteorological forcing data

Both W2 and RAFT require meteorological forcing inputs, such as air temperature and wind speed, to simulate water temperature dynamics. To generate these inputs over the spatial domain of Shasta Reservoir and the Sacramento River, regional and global weather models were used that produce gridded meteorology products. As not all required meteorological inputs were available from one product, they were gathered from three different products: 1) the North American Region Reanalysis (NARR), 2) the Global Forecast System (GFS), and 3) the bias-corrected and spatially downscaled Global Ensemble Forecast System (GEFS). A download module was used to access NOAA's Operational Model Archive and Distribution System (NOMADS) and download the GFS and GEFS products on a daily basis, while the NARR product was downloaded at six-month intervals as it is a reanalysis product.

NARR weather products are produced at three-hour intervals over the spatial domain of North America at a gridded resolution of ~ 32 km, but are only produced retrospectively every six months. GFS products are produced at three-hour intervals over the entire Earth at a gridded resolution of 0.5 degrees, and GEFS products are produced at six-hour intervals and downscaled to ~ 0.023 degrees. The GFS 0.5 degree product produces forecasts 10 days out and the GEFS produces forecasts 16 days out. GEFS produces an ensemble of simulations and provides 10^{th}, 50^{th}, and 90^{th} percentile forecasts.

While GEFS meteorology was already bias-corrected, NARR was not. Three-hourly NARR data were bias-corrected using the empirical quantile mapping method (EQM) (*12*). To bias-correct, EQM maps the empirical cumulative distributions function (ECDF) of a model's time series for a meteorological parameter to the ECDF of a ground observation station. Specifically, each member of the model ECDF is adjusted by the difference between the model and observed ECDF for a particular percentile (see Figure 4 for an example). The ground observation station data used to bias-correct the NARR output were obtained from the Durham and Davis meteorological stations ran by the California Irrigation Management System (http://www.cimis.water.ca.gov/) as these had the longest, most complete record of any stations nearby. For each NARR grid, the nearest ground station was used for bias-correction. Meteorological parameters that were bias-corrected included air temperature, relative humidity, and wind speed. Bias-correction occurred on a monthly basis, such that the historical record of conditions for a given month were used to adjust the 3-hour NARR data.

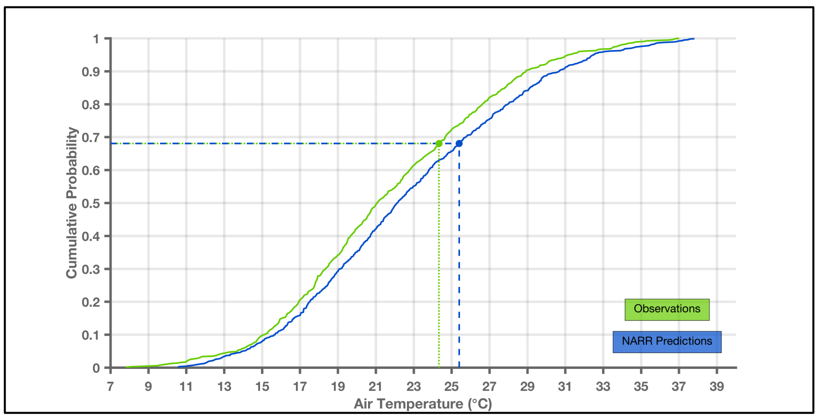

Figure 4: Example of bias-correcting meteorological data using the empirical quantile mapping method, showing NARR air temperature prediction being corrected by approximately -1° C.

During the hindcast period, both W2 and RAFT used meteorology variables produced from GFS (downward solar radiation and cloud cover) and the 50th percentile estimate from GEFS (air temperature, wind speed, wind direction, and a vapor pressure term) models at the 00Z hour mark. The 00Z hour mark are the initialized states of the weather/climate models and therefore are taken as the best estimate of the model domain's state at the point in time the forecast was generated.

In the forecast period, the models simulated meteorology 10 days (GFS) and 16 days (GEFS) out were used, with the 10th, 50th, and 90th percentile forecast used for GEFS variables. Following 16 days, an ensemble of the bias-corrected NARR product was used to account for the uncertainty in long range meteorology (i.e. > 16 day). A 25-member ensemble representing the meteorology from 2000-2015 was used. Therefore, each W2 and RAFT simulation was composed of 3 ensemble simulations during the 16 day forecast period and of 25 ensemble simulations during the > 16 day forecast period to represent the uncertainty in long range meteorology.

3.2.4 Format model inputs and run models

Once all operational, hydrological, and meteorological data were updated, the models were run in series, starting with W2 and ending with RAFT.

To set up a W2 run, modules were called which generated input files formatted for W2. Modules were grouped into four main tasks: 1) build hourly time series of meteorology, 2) build daily time series of inflowing discharge and temperature, 3) initialize reservoir with starting water elevation and vertical temperature profile, and 4) build daily time series of discharge elevation and volume at the downstream boundary condition (i.e. TCD at Shasta Dam). The executable of the W2 model was then called via command line to run the model.

Upon completion of the W2 simulation, the output of predicted daily release temperature from Shasta Dam was adjusted to account for the heat transfer expected to occur during the travel time through Keswick Reservoir to Keswick Dam (~ 17 km). However, information on the bathymetry of Keswick Reservoir (the after-bay of Shasta Reservoir and the upstream boundary of the Sacramento River) was lacking to generate a hydraulic model to simulate water discharge and temperature dynamics. Therefore, an autoregressive (AR) model, run in the R modeling environment in a modular form, was used. Further details of the AR model are provided in section 5.

RAFT modules were then called to set up the input files. Modules were grouped into 3 main tasks: 1) build 15-minute time series of meteorology, 2) build 15-minute time series of inflowing discharge volume and temperature, and 3) initialize the river with starting discharge volume and temperature. After the input files were created, the RAFT model was run.

Upon completion of all tasks, folders containing model inputs and outputs for W2 and RAFT were generated and named with the run date for archive purposes and to assess modeling results in graphical and tabular format with outputs pushed to the CVTEMP website.

The next three sections describe the W2, ARIMA, and RAFT models in greater detail.

4. Reservoir model

W2 is an open source hydrodynamic model developed by the United States Army Corps of Engineers with source code maintained and updated by Portland State University (*8*). The model discretizes a reservoir along its longitudinal axis into a series of multi-layered water columns (Figure 5) and simulates the mass, momentum, and thermal dynamics in the longitudinal and vertical directions, while averaging across the lateral direction in each grid cell. W2 has been shown to be well suited to model the thermal dynamics of stratified systems and has previously been applied to Shasta Reservoir to examine effects of TCD operations on cold-pool dynamics (*13, 14*).

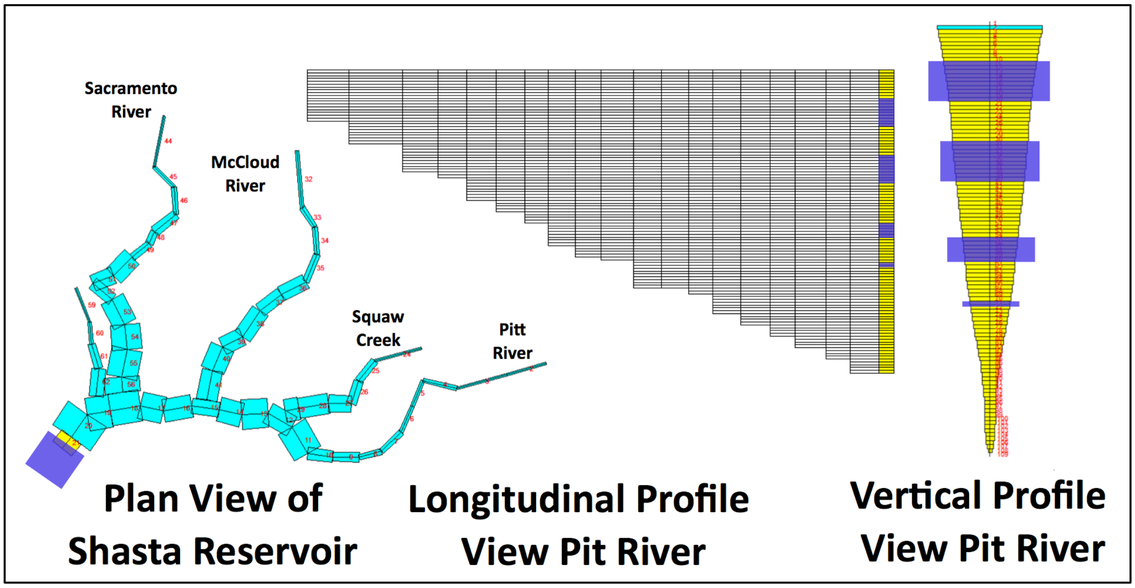

Figure 5: Geometry of the W2 model simulating water discharge and temperature in Shasta Reservoir, with location of the dam in plan view, and depths of the TCD gates in longitudinal (main branch of Pit River displayed) and vertical (at the downstream boundary of the dam) profile view denoted in purple. Note, the side gate does not technically have a width relative to other TCD gates and therefore a thin line is used to represent the elevation of the gate.

4.1 Model inputs

For inputs, W2 requires a geometry file to describe the physical domain of Shasta Reservoir, meteorological forcing terms to drive the energy flux calculations at the water surface and channel bed, observed vertical profile data to define initial temperature conditions of the reservoir, and upstream and downstream boundary conditions, such as inflow volume and temperature at primary locations where tributaries enter Shasta, and the discharge volume of the dam. The version of the W2 model used in this report (version 3.72) of Shasta Reservoir also has outflow depth specified to simulate the selective withdrawal (*8*) capabilities of the TCD at Shasta Dam.

4.1.1 Reservoir geometry

The geometry file is the foundation for the computations required to simulate discharge and temperature dynamics throughout a reservoir. In W2, geometry of a waterbody is represented as a series of branches, made of segments and layers. Geometric data are composed of four primary pieces of information: 1) longitudinal and 2) vertical grid resolution, 3) lateral cross-sectional width, and 4) channel slope.

An existing geometry file of Shasta Reservoir was used (supplied by Dr. Laurel Saito, personal communications) (Figure 5). The vertical grid resolution of the model was set at 1.5 m for the entire domain, while the longitudinal grid resolution was variable and ranged between approximately 1,000 and 5,000 meters. The reservoir was composed of five primary branches that represented the five primary physical branches of Shasta Reservoir. In all, the W2 model of Shasta Reservoir was composed of 63 longitudinal segments and 109 vertical layers representing nearly 7,000 grid cells.

4.1.2 Meteorology

Meteorological inputs required to run W2 included air and dew point temperature, wind speed and direction, cloud cover, and solar radiation (optionally entered as a measured variable). Only one meteorology input file is specified for the entire W2 model domain. Therefore, while gridded meteorological products were available, only the grid closest to Shasta Dam was used to supply meteorological forcing to Shasta Reservoir as this encompassed the majority of the reservoirs' surface area.

As previously described, when in operational mode, GFS and GEFS weather products were used during the hindcast period. During the forecast period, a 3-member ensemble (10th, 50th, and 90th percentile) of GEFS products were used 16 days out and GFS products used 10 days out from the model run date, followed by a 25-member ensemble of NARR products representing meteorology from 2000 to 2015 to complete the time series and account for uncertainty in long range meteorology (i.e. > 16 days out). All meteorological products were temporally downscaled to hourly data using linear interpolation before being used in the W2 model.

4.1.3 Initial reservoir temperature and storage

To Initialize the W2 model required specifying the water temperature of the entire model domain and storage at the starting time of the simulation. The model can be initialized with a constant temperature, or one that varies vertically and/or longitudinally. Observed vertical temperature profiles (collected by USBR with water surface elevation defined) near the face of Shasta Dam were used to initialize the model domain assuming there was no significant longitudinal gradient in temperature (*15*). To automate the initialization process, a module was incorporated that identified the observed vertical profile nearest in time (either before or after) to the model start date. Observed data (typically at ~7.5 meter intervals) were then linearly interpolated to match the vertical grid resolution of the W2 model (1.5 meters) and adjusted to the proper W2 input format. Linear interpolation was considered valid, as the model was initialized typically in February or March when running in operational mode and in January (when nearly isothermal conditions were present in the reservoir) when being calibrated and validated.

4.1.4 Upstream boundary conditions

There were four primary simulated inputs of water discharge and temperature into Shasta Reservoir (Sacramento River, McCloud River, Pit River, and Squaw Creek). In the W2 model, the upstream boundary conditions of water discharge and temperature can be defined as point or nonpoint sources and therefore distributed over the domain of a branch or at a particular location (i.e. at the upstream segment). The upstream inputs of water discharge and temperature were simulated as point sources at the upstream segments of their respective branches.

When in hindcast mode, data on discharge volume and temperature for Sacramento (DLT), McCloud (MSS), and Pit River (PMN) were gathered from available gauge data via CDEC (Figure 3) using the download scripts previously described. Inputs were formatted as mean daily values.

For Pit River, stage rather than discharge volume was available on a real-time basis via CDEC. Therefore, a stage-discharge relationship was used to estimate Pit River discharge (Figure 6).

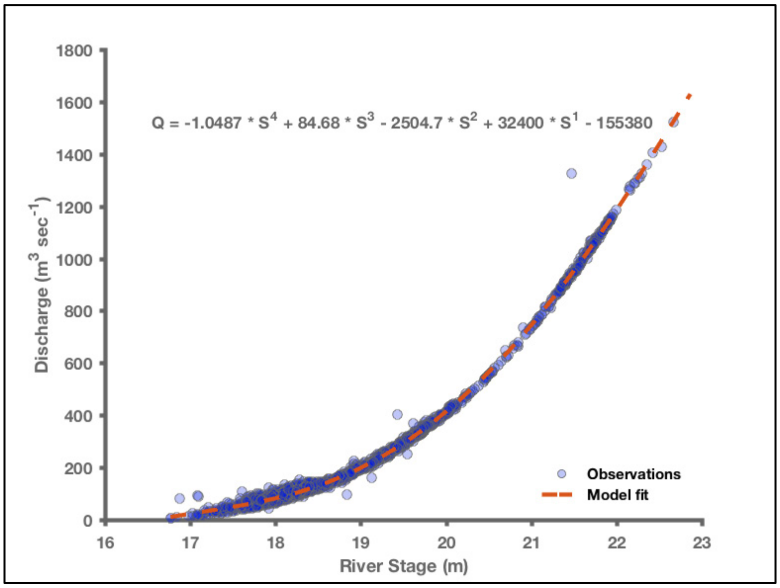

Figure 6: Stage to discharge relationship for Pit River generated from data spanning 1993 to 2004, with the best fit equation shown, where *Q* is discharge and *S* is river stage.

Similarly, real-time data were not available for Squaw Creek. Therefore, Squaw Creek discharge was estimated using a linear regression-based relationship with observed discharge at McCloud River, a neighboring watershed (Figure 7).

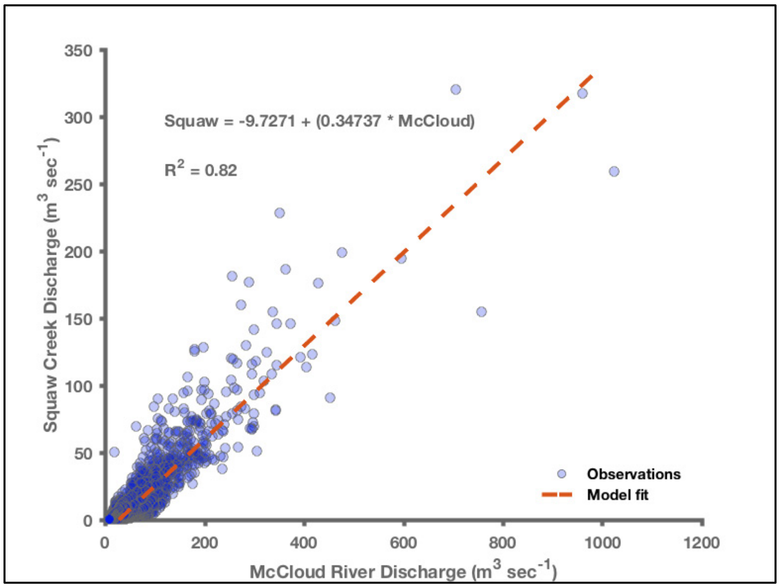

Figure 7: Relationship between McCloud River and Squaw Creek discharge from data spanning 1944 to 1966, with best fit equation shown.

Squaw Creek discharge water temperatures were assigned the same values as observed in the McCloud River, its neighboring watershed, as was done in Saito (*16*).

To account for additional un-gauged inputs and outputs (e.g. seepage and evaporation) and ensure a water balance for Shasta Reservoir, the total inflow volumes were adjusted to equal the USBR estimated total reservoir inflow on a daily basis (available via CDEC at the SHA gauge) using accretion/depletion estimates. The differences between the gauged inflows and USBR's estimate of inflow were distributed among the four simulated inputs based on the historical contributions of total discharge (i.e. Pit = 57.8%, McCloud = 26.8%, Sacramento = 13.3%, and Squaw = 2.1%) as done previously by Saito (*16*). Additionally, as USBR's estimate accounted for losses due to evaporation, the evaporation term was turned off and not directly simulated in the W2 model runs. However, this only affected the water balance and the effect of evaporative cooling was still simulated in the model.

When running W2 in forecast mode, CNRFC's Advanced Hydrologic Prediction Services products were used to set inflowing discharge for Sacramento, McCloud, and Pit Rivers, which were updated daily. Inflowing water temperature was based on the historical range observed from 2000 until 2015. For each forecasted input, the daily 50th percentile was used for a given simulation. The same approach was used to simulate Squaw Creek inflowing discharge and temperature during both the hindcast and the forecast periods.

4.1.5 Downstream boundary conditions

Shasta Reservoir's downstream boundary conditions included the TCD, bypass outlets, and a floodgate. The TCD is composed of four levels/elevations, while there are three levels/elevations of bypass outlets. The use of the floodgate and bypass outlets has been rare since 2000, therefore the text in this report is devoted to describing the use of the TCD (See Figure 8 for a conceptual schematic of the TCD at Shasta Dam).

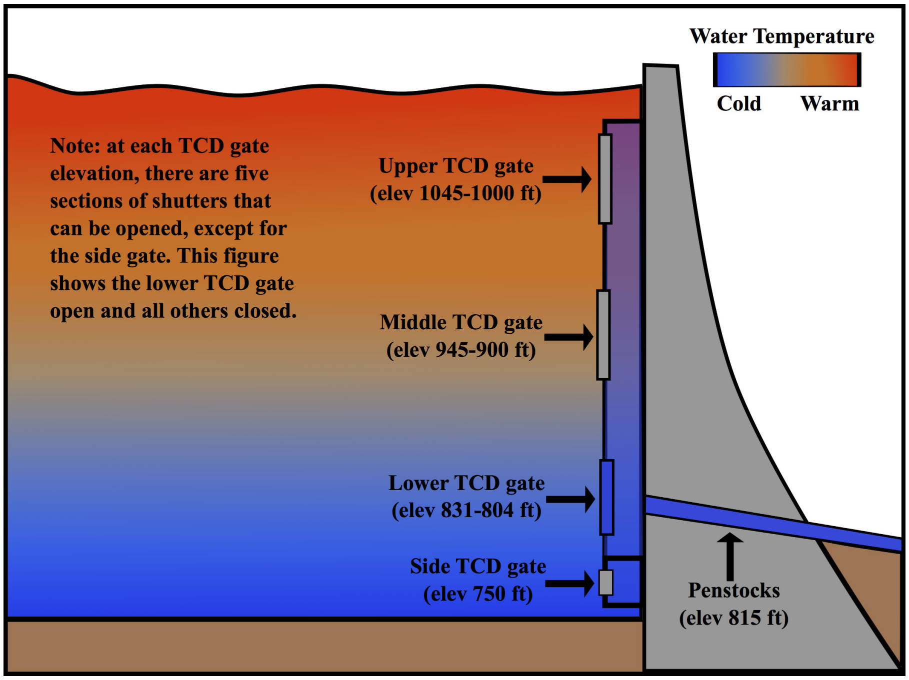

Figure 8: Conceptual schematic of the temperature control device (TCD) on Shasta Dam showing the locations and elevations of intake gates with elevations relative to mean sea level.

Previous studies have indicated the potential for water to leak at various elevations of the TCD at Shasta Dam and potentially affect reservoir release temperatures (*17*). To assess the need to include leakage in the W2 model, a pilot modeling study was conducted using two sets of simulations, those adjusting for leakage and those ignoring leakage. To parameterize TCD leakage in the W2 model, previously published reports by Resource Management Associates (RMA) were used that indicated leakage occurred over seven primary elevations of the TCD (*17*). The results of the reports were adapted to the structure of the W2 model, such that each of the seven leakage zones was represented as a point sink in the model (point sinks described in next paragraph). Mean elevations of the seven leakage zones outlined in the RMA report were used to set point sink elevations in the W2 model. Pilot study results (example of one year shown in Figure 9) revealed adjusting for leakage improved the fit between observed and W2 predicted release temperature (tail water) at Shasta Dam. Therefore, leakage was explicitly modeled in W2.

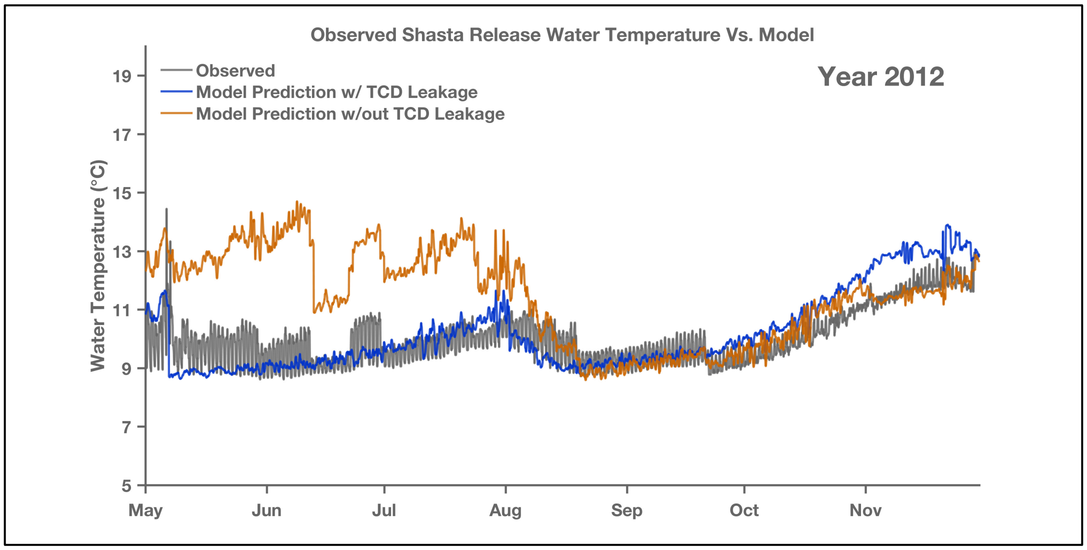

Figure 9: Comparison of observed (grey line) and W2 models with (blue line) and without (orange line) simulating leakage around the TCD for the year 2012.

In W2, the downstream boundary conditions at a dam can be specified in multiple ways, such as over weirs/spillways or through gates and other outlet structures. The outlet structures on Shasta Dam (the TCD and bypass outlets) were simulated as point sinks in the W2 model using W2's selective withdrawal algorithm (*8, 18*). The selective withdrawal algorithm estimates the vertical extent of the withdrawal zone and which parcels of water from the different vertical layers would be withdrawn based on water density, the strength of the vertical stratification, the pressure head, and discharge at the withdrawal location. For each TCD and bypass level, the mean elevation of the outlet was used to assign the elevation of the point sink, while the model internally calculated pressure head at the outlet location. In the current version of the W2 model for Shasta Dam, discharge is an input rather than internally calculated in the model. Prescribing discharge at each outlet structure is described below and required additional steps that were dependent on if W2 was ran in hindcast or forecast mode.

When in hindcast mode, total discharge from Shasta Dam was gauged and volumes of water passed through the bypass outlets and floodgates were known. However, the volume of discharge from specific levels of the TCD was not directly measured. Therefore, a module was developed to estimate and assign discharge to each section of the TCD using three primary pieces of information (total discharge data from Shasta Dam, gate operations of the TCD, and reservoir elevation).

Daily discharge data from Shasta Dam were obtained from CDEC using the download module. The log file of TCD operations provided by USBR was adjusted if observed conditions differed from planned operations, and a time series of gate operations was generated. Lastly, an algorithm (pseudocode shown in Figure 10) was used to estimate the discharge through a given TCD level and account for leakage through the TCD. The algorithm used the surface water elevation in the reservoir, number of gates open at a particular elevation, the fraction a given TCD level that was submerged, three parameters that adjusted the fraction of total discharge between two TCD levels when more than one level was in use, and the volume of water assumed to leak through the TCD.

```
for d in start date to end date

Q_Total_Leak_Z1(d) = Leak_Coeficient_Z1 * Q_Total(d) * Elevation_Weight_Z1(d)
Q_Total_Leak_Z2(d) = Leak_Coeficient_Z2 * Q_Total(d) * Elevation_Weight_Z2(d)
Q_Total_Leak_Z3(d) = Leak_Coeficient_Z3 * Q_Total(d) * Elevation_Weight_Z3(d)
Q_Total_Leak_Z4(d) = Leak_Coeficient_Z4 * Q_Total(d) * Elevation_Weight_Z4(d)
Q_Total_Leak_Z5(d) = Leak_Coeficient_Z5 * Q_Total(d) * Elevation_Weight_Z5(d)
Q_Total_Leak_Z6(d) = Leak_Coeficient_Z6 * Q_Total(d) * Elevation_Weight_Z6(d)
Q_Total_Leak_Z7(d) = Leak_Coeficient_Z7 * Q_Total(d) * Elevation_Weight_Z7(d)

Q_Total_Leak(d) = Q_Total_Leak_Z1(d) + Q_Total_Leak_Z2(d) + Q_Total_Leak_Z3(d)
   + Q_Total_Leak_Z4(d) + Q_Total_Leak_Z5(d) + Q_Total_Leak_Z6(d) + Q_Total_Leak_Z7(d)

if Upper(d) = 1 & Middle(d) = 0 & Lower(d) = 0 & Side(d) = 0

   Q_Upper(d) = Q_Total(d) - Q_Total_Leak(d)

else if Upper(d) = 1 & Middle(d) = 1 & Lower(d) = 0 & Side(d) = 0

   Elevation_Weight(d) = [Lake_Elevation(d) - Upper_Elevation]/100
   Percent_Upper = [Upper_Open(d) / Total_Open(d)] * Elevation_Weight(d) * Adjust_Upper
   Q_Upper(d) = [Q_Total(d) - Q_Total_Leak(d)] * Percent_Upper(d)
   Q_Middle(d) = [Q_Total(d) - Q_Total_Leak(d)] * [1 - Percent_Upper(d)]

else if Upper(d) = 0 & Middle(d) = 1 & Lower(d) = 0 & Side(d) = 0

   Q_Middle(d) = Q_Total(d) - Q_Total_Leak(d)

else if Upper(d) = 0 & Middle(d) = 1 & Lower(d) = 1 & Side(d) = 0

   Percent_Middle = [Middle_Open(d) / Total_Open(d)] * Adjust_Middle
   Q_Middle(d) = [Q_Total(d) - Q_Total_Leak(d)] * Percent_Middle(d)
   Q_Lower(d)  = [Q_Total(d) - Q_Total_Leak(d)] * [1 - Percent_Middle(d)]

else if Upper(d) = 0 & Middle(d) = 0 & Lower(d) = 1 & Side(d) = 0

   Q_Lower(d) = Q_Total(d) - Q_Total_Leak(d)

else if Upper(d) = 0 & Middle(d) = 0 & Lower(d) = 1 & Side(d) = 1

   Percent_Lower = [Lower_Open(d) / Total_Open(d)] * Adjust_Lower
   Q_Lowe(d) = [Q_Total(d) - Q_Total_Leak(d)] * Percent_Lower(d)
   Q_Side(d)  = [Q_Total(d) - Q_Total_Leak(d)] * [1 - Percent_Lower(d)]

else if Upper(d) = 0 & Middle(d) = 0 & Lower(d) = 0 & Side(d) = 1

   Q_Side(d) = Q_Total(d) - Q_Total_Leak(d)

end
```

Figure 10: Conceptual representation of the algorithm used to assign discharge and leakage through the TCD on Shasta Dam

The same approach was used for both hindcast and forecast mode, with the exception that the proposed discharge from Shasta Dam and TCD gate operations from the USBR supplied DSS and log files were used for forecasts.

4.2 <u>W2 model outputs</u>

W2 provided estimates of thermal flux at the surface layer, estimates of water temperature at each grid location in the model domain, as well as discharge water temperature for each outlet structure and a volume weighted discharge temperature for the entire set of outlet structures in use. For the purposes of this project, the focus was on outputs related to reservoir discharge weighted release temperature as well as the vertical distribution of water temperature in the reservoir near the downstream boundary (i.e. Shasta Dam). The model output included a time series of simulated vertical temperature distributions in Shasta Reservoir during hindcast and forecast periods (Figure 11).

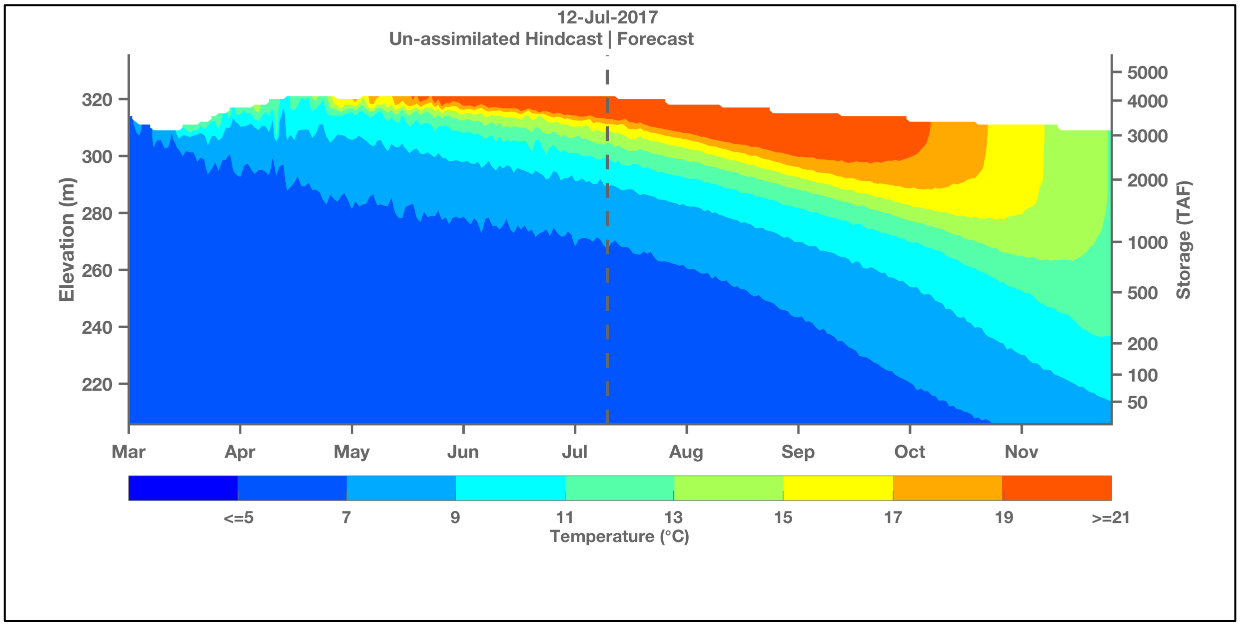

Figure 11: Time series of simulated vertical temperature distribution produced by the W2 model over a hindcast and forecast during 2017. Both reservoir elevation (m) and storage in thousand acre feet (TAF) are displayed on the y-axes, with color denoting water temperature (°C).

4.3 Model calibration and validation

The W2 model was calibrated to odd years from 2000 to 2015 and validated with even years over the same time period. This time period of calibration and validation was chosen as the TCD was operational post 1997 and because input data quality and availability were suboptimal prior to 2000. Over this time period, the distribution of water year types (as classified by California Department of Water Resources: http://cdec.water.ca.gov/cgi-progs/iodir/WSIHIST) was similar between calibration and validation years, with calibration and validation years having three and two years of either wet or above normal years respectively, and five and six years of below normal, dry, or critical years respectively. Each calibration and validation period represented a full calendar year (i.e. simulation was initialized on January 1st and ended on December 31st). During calibration, observed upstream and downstream boundary conditions were used along with bias-corrected NARR meteorology, as archived GFS and GEFS data before 2015 were not acquired.

For each year, model predictions of vertical temperature profiles and discharge release temperature were compared to observed values, with the absolute difference between observations and predictions used as the basis for calibration. Observed vertical temperature profiles were available monthly to bi-weekly depending on year and season, while observed release temperature was available on a daily basis. The model was simultaneously calibrated to reservoir vertical temperature profile and reservoir release temperature as there is utility in predicting both cold-pool resources within Shasta Reservoir and in predicting release temperature to meet temperature compliance downstream.

Before beginning calibration, the primary model parameters in W2 that were assumed to affect thermal dynamics as well as those in the custom modules that assigned discharge through the TCD (see Table 1 for list of W2 calibration parameters and descriptions) were identified. In all, seven model parameters were selected to adjust for thermal calibration purposes.

Table 1: Parameters used to calibrate the W2 model to Shasta Reservoir. Note: parameters not native to the W2 model relating to assigning discharge through the TCD are also described in Figure 10.

Parameter ID	Native to W2 model	Description
BETA	YES	Percent of solar radiation absorbed at the surface layer
EXH20	YES	Light extinction coefficient of water which controls the attenuation rate of solar radiation due to water
WSC	YES	Wind sheltering coefficient which controls how input wind speed is adjusted, which in turn affects thermal mixing
Adjust_Upper	NO	Parameter for the upper TCD gate, which adjusts how much of the total discharge is coming from the upper versus middle gate when both in use
Adjust_Middle	NO	Parameter for the middle TCD gate, which adjusts how much of the total discharge is coming from the middle versus lower gate when both in use
Adjust_Lower	NO	Parameter for the lower TCD gate, which adjusts how much of the total discharge is coming from the lower versus side gate when both in use
Leak_Coefficient_Z6	NO	Parameter for the leakage estimate, which adjusts how much of the total leakage discharge is coming from leakage zone 6 (i.e. at elevation of penstocks, which is assumed to have the largest leakage rate)

The calibration procedure used a probabilistic-based approach to maximize the likelihood of model predictions matching observations. Specifically, assumptions were made about the distribution of each parameter being calibrated (i.e. its prior probability density function (PDF)) by assuming uniform PDFs. To set the bounds of uniform PDFs for each parameter being calibrated, prior knowledge was used in the form of the W2 user manual (i.e. default parameter values and ranges), site specific information about Shasta Reservoir (*17*), and the authors' judgments as to how the TCD functioned to draw water from the reservoir. Next a likelihood function was developed which represented the authors' belief about the variation between model predictions and observations (i.e. error) (*19*) and assumed this variation was normally distributed:

$$Error = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where, x is the absolute difference between the observed value and model predicted values, μ is the mean in variation and was assumed to be zero, and σ is the standard deviation from the mean in variation and was assumed to be one.

With the prior PDFs and likelihood function defined, a Metropolis Markov Chain Monte Carlo (MCMC) routine (*20, 21*) was used to jointly sample the parameter space of the seven calibration parameters and calculate posterior PDFs. Metropolis MCMC routines can be generally described as a method to approximate the distribution of multi-dimensional parameter space. The process begins by randomly assigning starting parameter sets conditional on the bounds supplied by the modeler. In the following iterations, the algorithm randomly (i.e. Monte Carlo property) searches the parameter space and accepts a move to the new location if the move improves the likelihood function compared to the current location (i.e. a Markov Chain property). To better ensure that the algorithm does not get stuck in a local minimum or maximum likelihood space, sometimes a move that does not improve the likelihood function is accepted (i.e. a Metropolis property). The theory is that as iterations increase, the multi-dimensional parameter space is better approximated.

Each MCMC routine for each calibration year was comprised of three chains, each with 500 iterations and resulting in 12,000 simulations in total. As each 1-year model run took approximately ten minutes to complete, the calibration routine was run in parallel, to reduce the calibration time from months to about one week. Parameter convergence was assessed using the Gelman and Rubin diagnostic from the BOA package in R (*22-24*). With this approach, the most likely parameter values were identified based on maximum likelihood principles and the uncertainty around each point estimate was quantified (*19*).

To validate the calibrated model and evaluate how well the W2 model could predict vertical temperature profiles and release temperatures in the non-calibrated years, a series of W2 model simulations were run for even years from 2000-2015. Specifically, simulations were run using known upstream and downstream boundary conditions and NARR meteorology for a given year, while using a Monte Carlo based sample of the calibration parameters' posterior PDFs, which resulted in more samples being drawn from parameter sets which minimized the difference between observed and modeled reservoir conditions. A total of 600 simulations (200 samples from each chain) were completed for each year.

4.3.1 Calibration and validation results

Results from the calibration routine (Figure 12) for the seven W2 parameters revealed some parameters converged to a well-defined range, while others had considerable uncertainty (i.e. posterior density nearly equal across the prior's range). For example, the parameter value of one for the light extinction coefficient appears to be much more likely compared to a value of four (at the upper end of the prior PDF), while all parameter values for the middle TCD gate parameter seem to be equally likely across the prior PDF.

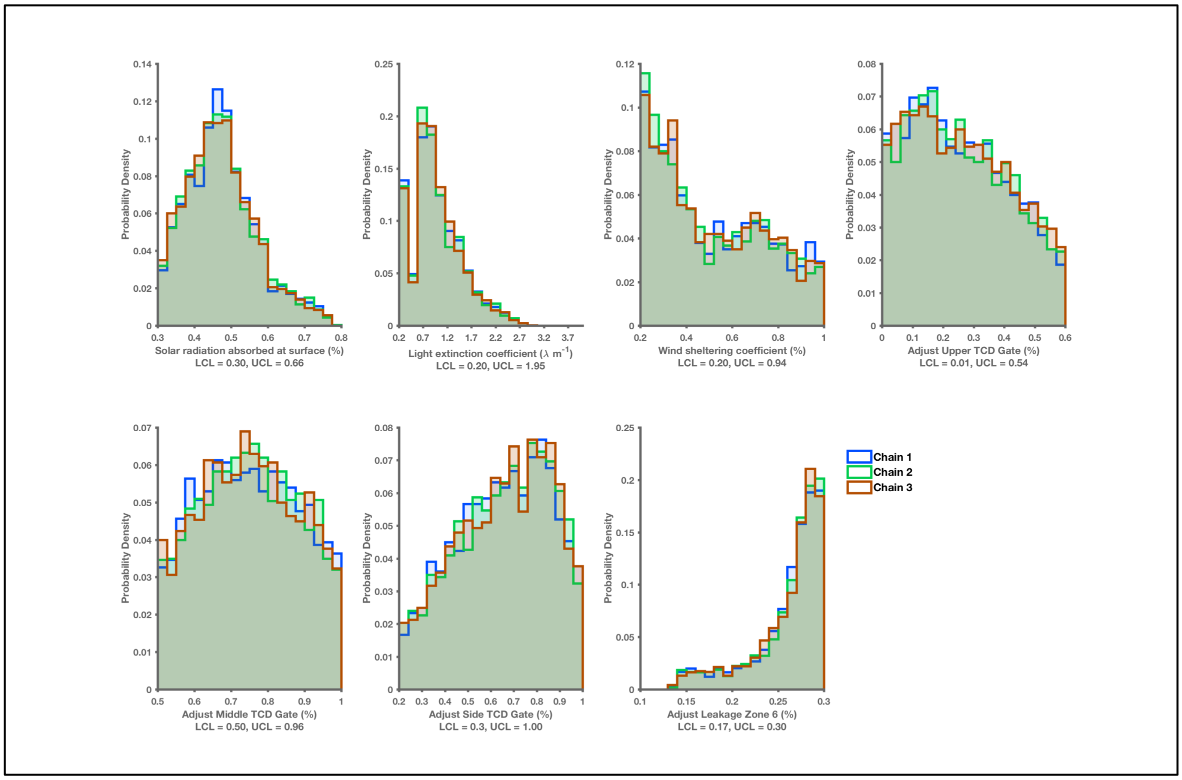

Figure 12: Posterior (three chains of different color) PDFs for the calibration parameters of the W2 model of Shasta Reservoir based on comparing observed vertical temperature profiles and discharge temperature to model predictions over odd years from 2000-2015. Lower and upper credible levels (LCL and UCL) at the 95th percentile level are displayed for each parameter. Note, x-axis bounds represent the prior PDF bounds of calibration parameters.

For an example of model skill/performance (measured as root mean square error (RMSE) and bias), the W2 output of vertical temperature profiles during 2008 indicated the model can capture the vertical distribution in temperature over the course of one calendar year (Figure 13). In this example, W2 has a tendency to estimate colder temperatures than observed at most elevations with an RMSE less than 1° C for the majority of the year, while tending to overestimate temperatures in the deepest sections of the reservoir by a few tenths of a degree. Similar plots for each year used in the validation procedure are provided in Appendix A.

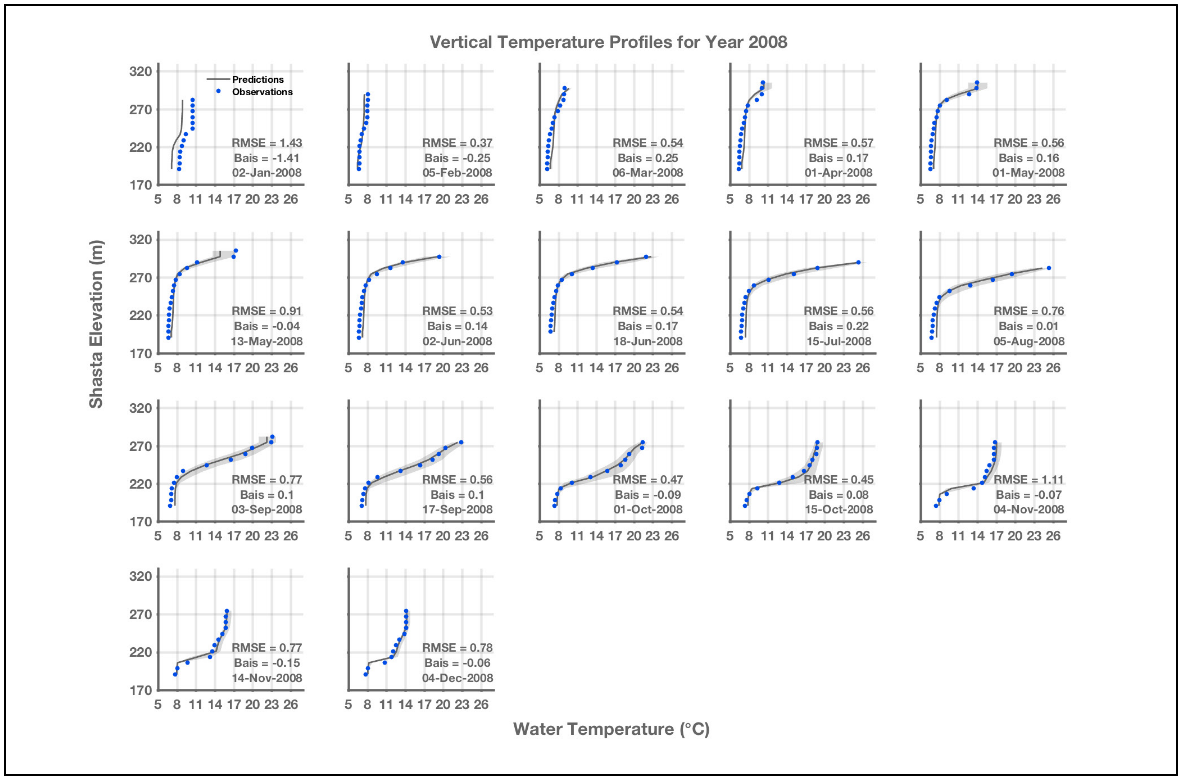

Figure 13: Observed (blue dots) versus W2 predicted (grey line) vertical temperature profile near the face of Shasta Dam for year 2008. The 50th percentile of the W2 predictions are displayed as a solid line with 5th and 95th percentile predictions shown as a shaded region. Goodness-of-fit metrics are based on the 50th percentile model output.

Model validation results collapsed by year for vertical temperature profiles indicated model performance (i.e. RMSE and bias) varied in time and space, but was generally on the order of < 1° C (Figure 14). Figure 14 shows results stratified by month and by water temperature (i.e. temperature ≤ 10° C used as a cutoff point for cold-pool resources and all temperatures in Shasta Reservoir). Across all temperatures, the model simulated warmer water than observed on average with a mean bias of approximately 0.1° C. This bias was slightly greater (0.2° C) when only considering temperature ≤ 10° C. However, mean RMSE was greater when examining all water temperatures (approximately 0.9° C) compared to temperature ≤ 10° C (0.7° C).

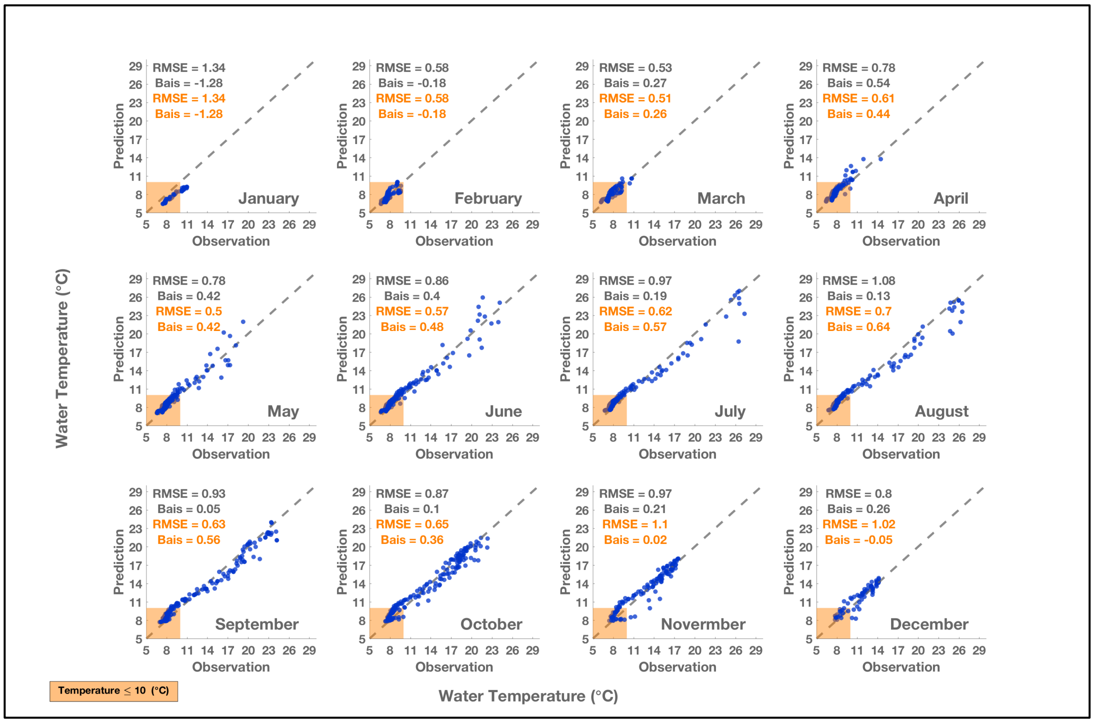

Figure 14: Observed profile temperatures plotted against W2 predicted profile temperatures stratified by month and temperature (all-temperatures and temperature ≤ 10° C) for all simulations during even years from 2000-2015 at the 50th percentile level, with RMSE and bias displayed for each stratum.

Comparison of observed versus W2 predicted Shasta Dam discharge temperature (Figure 15) indicated the model adequately predicted the trends of discharge temperature over time, but typically predicted warmer release temperature on the order of 0.5° C. Similar to vertical profiles, predicted discharge temperature tended to be upwardly biased compared to observed data with bias being greater in May to October during the temperature management season (mean of 0.4° C) compared to the entire year (mean of 0.2° C). Total error was similar between times of year (mean RMSE of 0.8° C for entire year and 0.5° C from May to October).

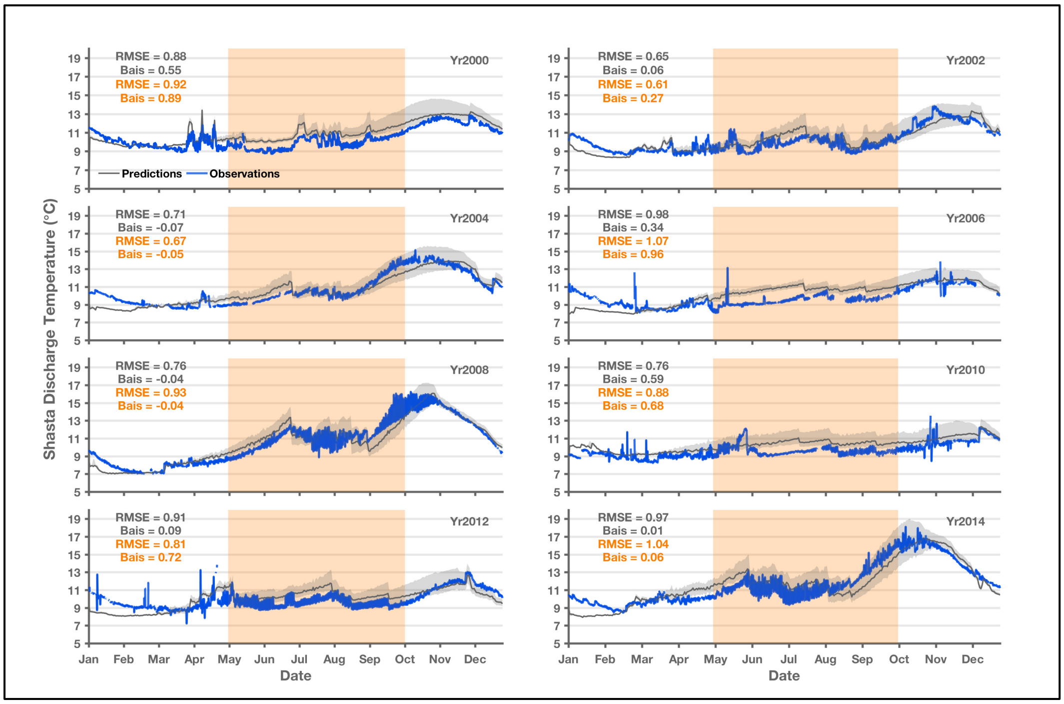

Figure 15: Observed (blue line) versus W2 predicted (grey line) Shasta Dam daily discharge temperature during model validation years from 2000-2015. Predicted temperatures are the result of 600 separate simulations with the 50th percentile (solid grey line) and the 5th and 95th percentiles (shaded region). Results are stratified by year and time of year: all-time versus temperature management season from May-October (shaded orange region). RMSE and bias are displayed for each stratum and based on the 50th percentile model results.

Comparison of observed versus W2 predicted Shasta cold-pool elevation and storage (Figure 16) indicated the model tended to under predict cold-pool, with greater uncertainty in the years 2012 and 2014. Predicted storage tended to be downwardly biased compared to observed data (mean bias of -100 thousand acre feet (TAF)) over the course of a year, and slightly less during the temperature management season from May to October (-60 TAF). Total error varied by year. All-time mean RMSE was ~ 300 TAF. RMSE between May and October was slightly less, at ~ 200 TAF. Lastly, RMSE for cold-pool storage at the end of September, which has been a commonly used metric in water resource management for Shasta Reservoir, was often < 100 TAF.

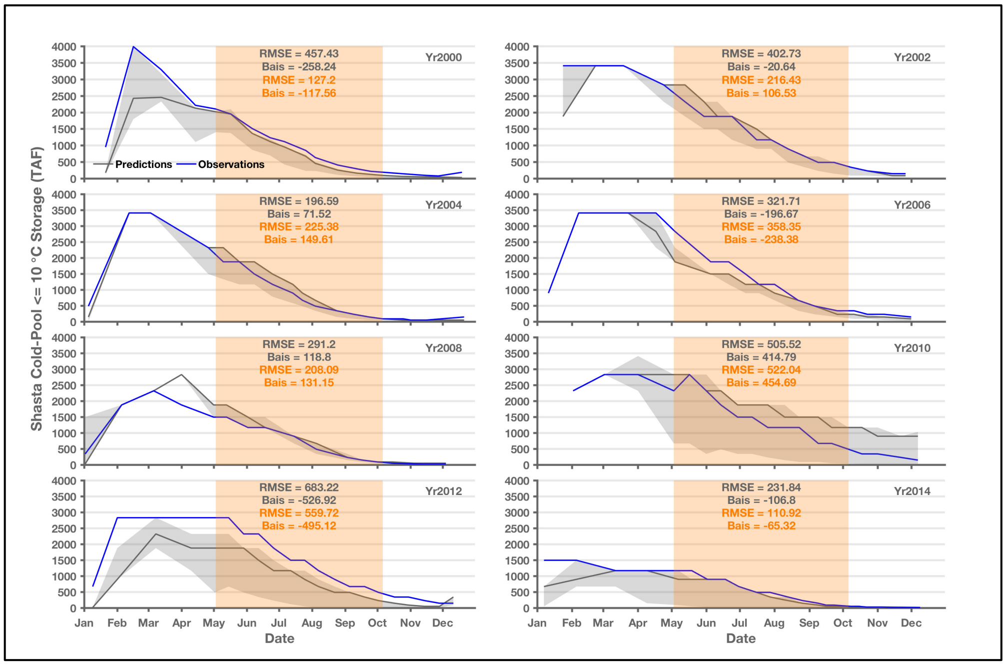

Figure 16: Observed (blue line) versus W2 predicted (grey line) Shasta cold-pool elevation during model validation years from 2000-2015. Predicted cold-pool elevations are the result of 600 separate simulations with the 50th percentile (solid grey line) and the 5th and 95th percentiles (shaded region). Results are stratified by year and time of year: all-time versus temperature management season from May-October (shaded orange region). RMSE and bias are displayed for each stratum and based on the 50th percentile model results for storage in thousand acre feet (TAF).

5. AR model

Due to the lack of geometry data for Keswick Reservoir it was not possible to generate a hydrodynamic model, so a simplified version of an autoregressive integrated moving average (ARIMA) model was used to predict the expected temperature change from Shasta Dam to Keswick Dam. ARIMA models are most often used to forecast time series data and have been described elsewhere (*25*). In general, there are three components to an ARIMA model: the autoregressive component (p) that relates to temporal lags in the time series, the differencing component (d) that relates to stationarity of the time series, and the moving average component (q) that relates to the autocorrelation of errors in the time series. Together, these parameters make up an ARIMA (p,d,q) model.

For this project, an ARIMA (1,0,0) model was used, as the monthly time series data were stationary as indicated by a significant P-value for an augmented Dickey-Fuller test, and did not require differencing, and had no apparent significant autocorrelations in the errors. Therefore, the model was reduced to an autoregressive model (AR). The model was calibrated to even years from 2000 to 2014 and validated during odd years over the same time period.

5.1 AR calibration and validation results

Model validation results for the AR model simulating the temperature change from Shasta Dam to Keswick Dam indicated the model nearly always predicted warmer water compared to observed values (Figure 17). Bias of the AR model was upward and tended to be higher during temperature management from May to October (mean of 0.5° C) compared to the entire year (mean of 0.3° C). Model error was consistently ≤ 1° C and was typically greater during May to October (mean RMSE of 0.7° C) compared to the entire year (mean RMSE of 0.6° C).

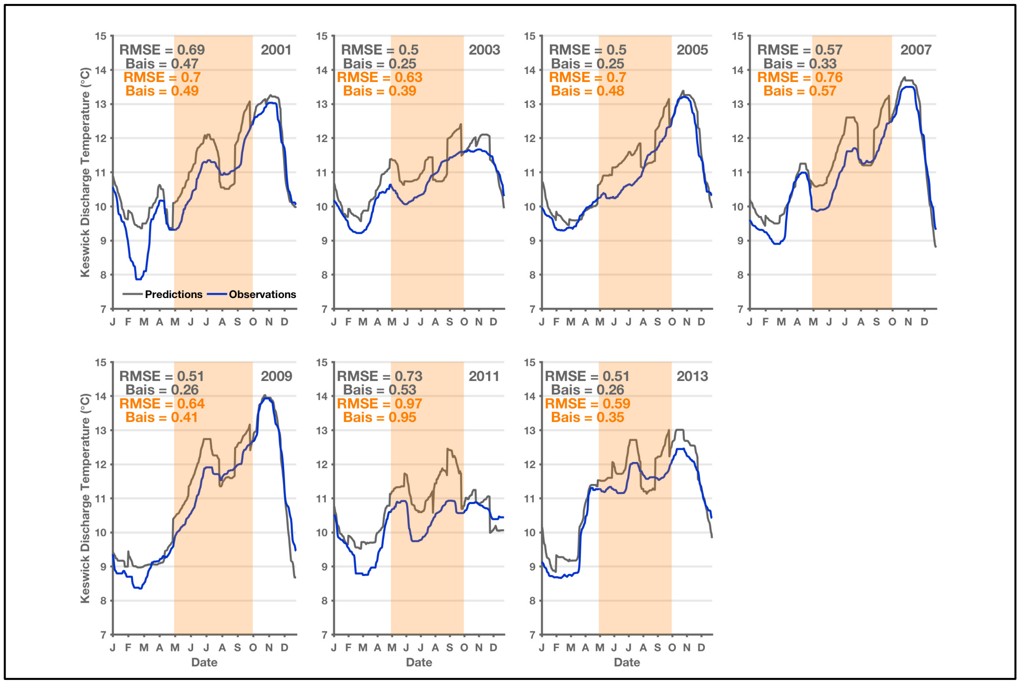

Figure 17: Observed (blue line) versus AR predicted (grey line) daily Keswick Dam release temperature for validation years. Results are stratified by year and time of year: all-time versus temperature management season from May-October (shaded orange region). RMSE and bias are displayed for each stratum.

6. River model

RAFT was developed by NOAA and with funding from the National Aeronautics and Space Administration (NASA) to simulate thermal dynamics in the Sacramento River (*7*). The model is one-dimensional and discretizes a river into a series of longitudinal segments and simulates thermal dynamics in the longitudinal direction, while averaging across the lateral and vertical directions for a given grid cell. Natively, RAFT has a spatial resolution of 2 km and a temporal resolution 15 minutes, but these dimensions are adjustable. The native spatial resolution of 2 km was used for model simulations in this report, as tests reducing the resolution to < 2 km resulted in similar model output, but at greater computational cost. For this report, during calibration and validation, the RAFT time step was adjusted to ten minutes to ensure the model solution was stable during winter flows which can generate high channel velocities (i.e. that the Courant condition was met and a simulated water parcel would not move greater than one grid cell within one time step). RAFT is well suited to simulate the thermal dynamics of well-mixed systems and has previously been applied to the Sacramento River. Additional details of RAFT can be found elsewhere (*7*).

6.1 Model inputs

For inputs, RAFT requires a geometry file to describe the physical domain of the Sacramento River, meteorological forcing terms to drive the energy flux calculations, an observed longitudinal profile of Sacramento River discharge and temperature to define initial conditions of the river along the longitudinal grid, and boundary conditions, such as inflow volume and temperature at the upstream segment (i.e. Keswick Reservoir) as well as at locations of significant tributary inputs.

6.1.1 River geometry

In RAFT, the geometry of a river is represented as a series of longitudinal segments and each segment is described by channel characteristics, such as channel width, depth, and slope. The domain of RAFT presented in this report begins at Keswick gauge approximately 2 km below Keswick Reservoir and extends approximately 100 km downstream to the city of Red Bluff, represented by 50 grid cells.

Methods to develop RAFT geometry have been described elsewhere (*7*). Briefly, observed cross-sectional data for the model domain were collected by California Department of Water Resources. Survey locations were spaced approximately 500 meters apart and imported into the Hydrologic Engineering Center's River Assessment System (HEC-RAS). In HEC-RAS, a series of steady flow simulations (varying from ~ 50 $m^3\ sec^{-1}$ to ~ 5700 $m^3\ sec^{-1}$) were run to generate rating curves for each of the main channel parameters for each cross-section. Cross-sectional rating curves were then converted to a lookup-table and interpolated to the RAFT grid and represented a computationally efficient method to assign channel geometry for a given discharge in the model domain (*7*).

The model formulation of RAFT differed slightly from the formulation presented in (*7*). The differences primarily included simplifications to the hydrologic routing, the data assimilation technique, and the solution scheme of the advection-dispersion equation for river temperature. First, the Muskingum-Cunge flow routing model and Kalman filtering based data assimilation for flow computations have been replaced by a simple nearest neighbor interpolation between the flows (or their forecasted values) available at different grid points. The second difference involved solving the advection-dispersion equation. The hybrid Eulerian-Lagrangian two-step computation of the material derivative (see e.g., (*26*)) of temperature has been replaced by an explicit Forward Euler scheme. Similarly, the implicit Crank-Nicolson time stepping scheme for dispersion has been replaced by a more straightforward explicit central differencing spatial discretization scheme. These changes necessitate the adherence of the chosen time step size to the Courant-Freidrichs-Levy stability condition (see e.g., (*27*)). Any reduction in accuracy due to these computational simplifications is small compared to the error associated with the meteorological input data (Figure 5 in (*7*)).

6.1.2 Initial river conditions

For each RAFT simulation, the model domain required initialization to a starting discharge and temperature value. For discharge and temperature, the model was initialized using gauge stations along the model grid. Similar to W2, a module was incorporated that searched for the observed discharge and temperature data points nearest in time to the model start date. Observed data were then interpolated using a nearest neighbor method to match the longitudinal grid resolution of the RAFT model.

6.1.3 Boundary conditions

As a one-dimensional model, there is a single upstream boundary condition in RAFT, which is the gauge station (KWK) approximately 2 km downstream from Keswick Reservoir. Additionally, there are four primary tributaries that enter the main stem of the Sacramento River incorporated into RAFT (Figure 1). Observed data for each gauge station (upstream and tributary inputs) were obtained using the download modules previously described.

When run in hindcast mode, observed gauge data were used for all boundary conditions when available. For discharge volume, all boundary conditions had available data via CDEC. For inflowing water temperature, only the upstream boundary and one tributary (IGO gauge) were available for download via CDEC. All other tributary temperature inputs were based on the historical range observed from 2000 until 2015. For each input based on historical data, the monthly 50th percentile was used for a given simulation.

In forecast mode, the upstream temperature boundary condition was daily Shasta Dam release temperature simulated with the W2 model, adjusted for thermal effects expected in Keswick Reservoir using the AR model and temporally downscaled to a 15-minute time series with linear interpolation. The upstream discharge boundary condition was the proposed Keswick operations provided by USBR. Tributary inputs during forecast mode were assigned the monthly 50th percentile values of discharge and temperature calculated over the time period of 2000-2015.

6.1.4 Meteorology

The minimum meteorological inputs required to run RAFT include air temperature, a vapor pressure term such as relative humidity, wind speed, and solar radiation. RAFT used the same weather/climate products as the W2 model (i.e. NARR, GFS, and GEFS) and the same process of using different products for different time periods (i.e. during hindcast vs. forecast period). Therefore, as with W2, there were three ensembles when GFS/GEFS was used and 25 ensembles with varying meteorology when NARR was being used.

Unlike W2, which does not allow meteorology to vary spatially over the model domain, the meteorology of RAFT was spatially explicit, such that each grid cell had a separate meteorological time series. To generate the spatially explicit time series of meteorology, gridded linear interpolation was used for each weather/climate product. Then, for each of the 50 grid cells in the RAFT domain, the nearest meteorology grid was chosen. Lastly, all meteorological products were temporally downscaled from three-hour data to 15-minute data using linear interpolation before being used in the RAFT model.

6.2 Data assimilation

When in hindcast mode, RAFT has a data assimilation component to ingest observed information about water temperature along the model domain. Further details can be found elsewhere (*28*) and (*7*), but the primary goal of data assimilation is to update a model state to align with an observation in order to improve the simulation going forward. Specifically, RAFT uses a version of the Ensemble Kalman Filter to assimilate data (*28*). Briefly, the Ensemble Kalman Filter works by perturbing model states stochastically to generate an ensemble of model states which are weighted by a Kalman Gain matrix (*28*). The ensembles are then used to produce an updated model state along with estimates of uncertainty around the new model state. Model states are perturbed based on assumptions about the error structure of the model and observations, which are input by the modeler (*28*). For the simulations presented in this report, model error was assumed to be 0.3° C and observational error of gauged data was assumed to be 1° C. Figure 18 provides an example of the RAFT model state variable of water temperature being updated to more closely align with observed water temperatures over a period of four days.

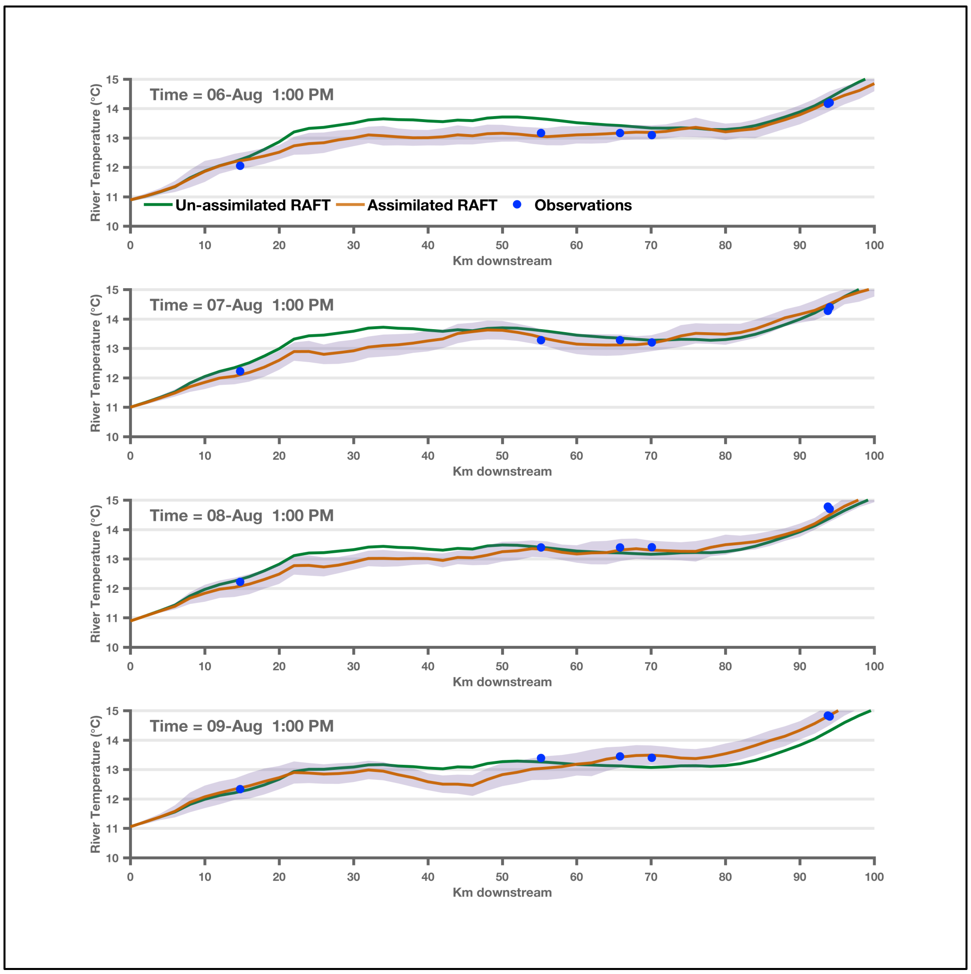

Figure 18: Example of data assimilation in RAFT using the Ensemble Kalman Filter over a period of four days at 1PM showing longitudinal temperature in the Sacramento River with (red line with 95th percentile confidence interval in shaded region) and without (green line) data assimilation. Observations (blue dots) correspond to the river gauges downstream from KWK in Figure 1.

6.3 Model outputs

As outputs, RAFT provided estimates of thermal flux at the air-water and bed-water interface and estimates of water temperature at each grid location in the model domain. For the purposes of this project, outputs relating to river water temperature across the model domain were of primary interest. An example of output from the model is shown in Figure 19, which depicts a time series of longitudinal temperature during the hindcast and forecast periods.

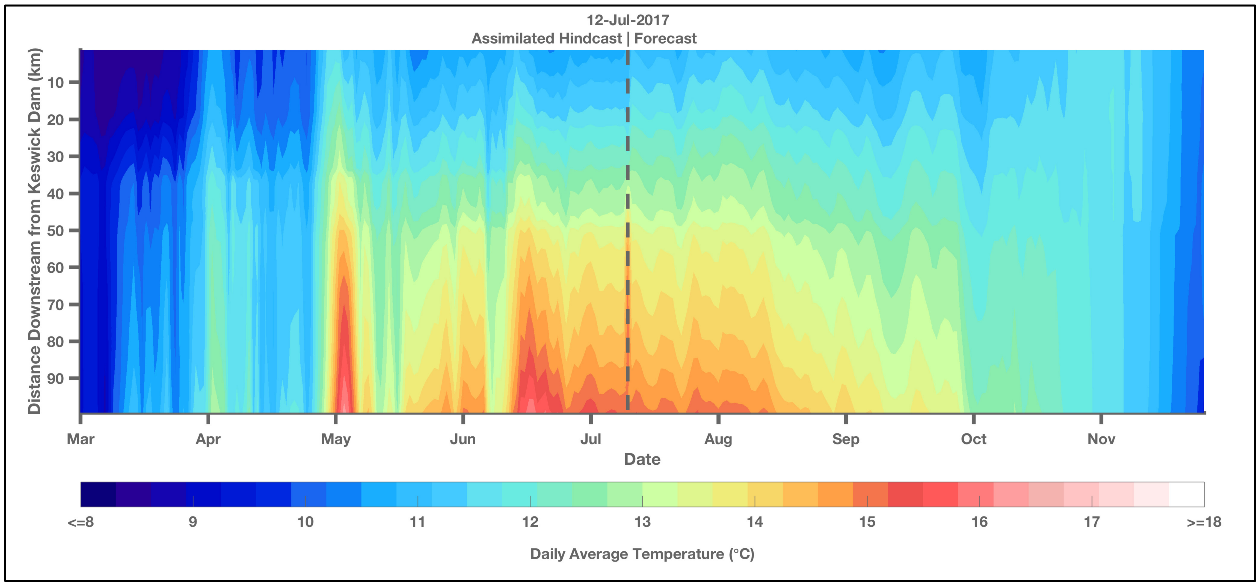

Figure 19: Time series of simulated longitudinal temperature produced by the RAFT model over a hindcast and forecast during 2017. Distance (km) downstream from Keswick Reservoir is displayed on the y-axis (note the axis is reversed), with color denoting water temperature (°C).

6.4 Model calibration and validation

The RAFT model was calibrated to odd years from 2000 to 2015 and validated with even years over the same time period. This time period of calibration and validation was chosen to be consistent with the calibration time period of the W2 model. Each calibration and validation period represented a full calendar year (i.e. simulation was started on January 1st and ended on December 31st) and was run in parallel as with the W2 calibration. For each year, model predictions of longitudinal temperature profiles were compared to observed values along the Sacramento River at gauging stations (CCR, BSF, JLF, BDN, RDB). During calibration, observed boundary conditions were used along with bias-corrected NARR meteorology.

Before beginning calibration, model parameters in RAFT were identified that affected thermal dynamics, which may be expected to vary by region. Three model parameters were selected to adjust for thermal calibration purposes. These included two parameters related to evaporative heat flux from wind and one parameter related to heat exchange at the bed-water interface. Specifically, RAFT uses the following wind function in estimation of evaporative flux:

$$f_{wind} = \alpha + \beta * WS$$

where α and β are unitless parameters, and WS is wind speed at 10 meters above ground (m sec^{-1}) (*7*).

The parameter relating to heat exchange at the bed-water interface is the depth of the conduction layer represented in the following equation:

$$\theta_{bed} = \frac{K_{cl} * (T_{bed} - T_{water})}{D_{cl}/2}$$

where θ_{bed} is the heat exchange/flux of the bed channel, K_{cl} is the thermal conductivity of the sediment conduction layer (15.9 W m^{-1} °C^{-1}), T_{bed} is the temperature of the bed (°C), T_{water} is the temperature of the water (°C), and D_{cl} is the depth of the conduction layer (m).

The details about the specifics of how the parameters relate to heat dynamics in RAFT are described elsewhere (*7*).

Similar to the W2 calibration, a probabilistic-based approach was used to maximize the likelihood of model predictions matching observations assuming uniform PDFs. Bounds of uniform PDFs for each calibrated parameter were set using prior knowledge in the form of the reported values by other water temperature models (*29*) and using site specific information about the Sacramento River. Similar to W2 calibrations described in Section 4.3, the variation between model predictions and observations was assumed to be normally distributed, with a mean of zero and standard deviation of one. Each MCMC routine for a given year was comprised of three chains, each of 1,000 iterations. As in W2 calibration, convergence of parameters was confirmed using the Gelman and Rubin convergence diagnostic from the BOA package in R (*22-24*).

To validate the calibrated model and evaluate how well the RAFT model could predict longitudinal temperature profiles in the non-calibrated years, a series of RAFT model simulations were run. As with W2, simulations were run using known boundary conditions and NARR meteorology for a given year, while using a Monte Carlo based sample of the calibration parameters based posterior PDFs. A total of 600 simulations (200 samples from each chain ran in parallel) were completed for each even year between 2000 and 2015.

6.4.1 Calibration and validation results

Results from the calibration routine for the three parameters used in the RAFT model (Figure 20) indicated a low degree of uncertainty for two parameters and a high degree of uncertainty for one parameter under assumed prior PDFs. Specifically, the parameters relating to the wind function were better constrained over the parameter space examined compared to the parameter relating to depth of the thermal conduction layer between the water column and stream bed.

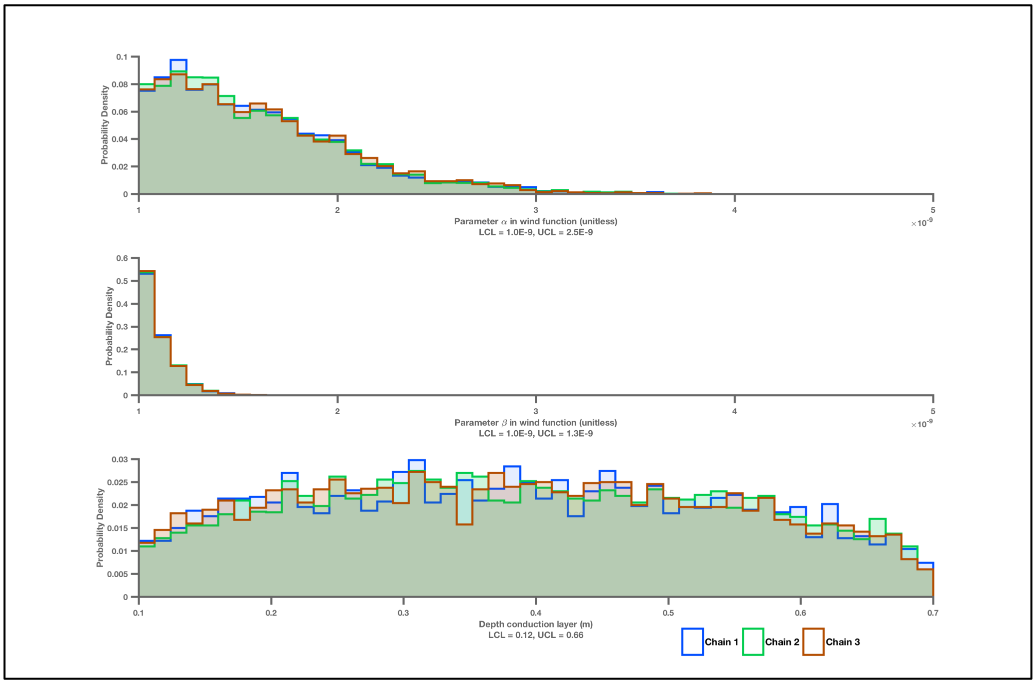

Figure 20: Posterior (three chains of different color) PDFs for the calibration parameters of the RAFT model of Sacramento River from Keswick Reservoir to Red Bluff based on comparing observed longitudinal temperature profiles to model predictions over odd years from 2000-2015. Lower and upper credible levels (LCL and UCL) at the 95th percentile level are displayed for each parameter.

As an example of model skill during August, Figure 21 shows how RAFT captured the diel fluctuations in water temperatures approximately 40 km downstream of the boundary conditions at KWK. Over this example, RAFT predictions typically estimated warmer peak daily temperatures with an RMSE often near 0.5° C. Note, the fit statistics for year 2014 are a result of an observational error in the gauge data that was not captured during screening of data and therefore not an adequate assessment of RAFT for that location and time.

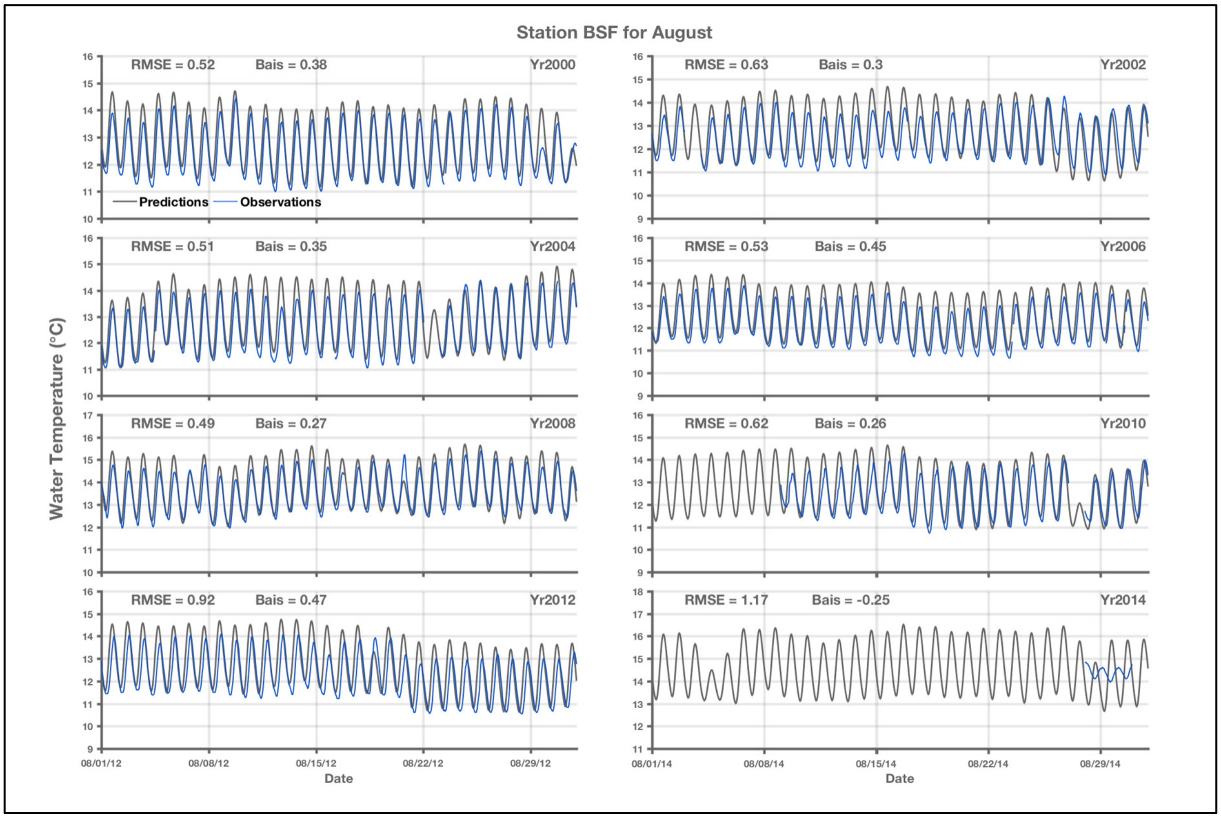

Figure 21: Observed (blue line) versus RAFT predicted (grey line) hourly temperatures at Balls Ferry (BSF) gauge during the month of August for 2000 to 2014. Predicted temperatures are the result of 600 separate simulations with the 50th percentile displayed as a solid grey line. RMSE and bias are displayed and based on the 50th percentile model results. Note there was a gauge error in 2014.

Collapsing RAFT validation results by gauge station and stratifying by month for longitudinal temperature profiles indicated that when considering all months and gauge locations, RAFT generally predicted colder temperatures (mean bias of -0.1° C) with a mean RMSE of 0.7° C (Figure 22). Error varied by season. During the temperature management months (i.e. May to October), RAFT tended to predict warmer temperatures, and during winter months RAFT tended to predict colder temperatures. Additionally, when examining RAFT output stratified by gauge stations (Figures in appendix B), RAFT predictions at the Clear Creek gauge location typically had the lowest RMSE and bias of the five sites examined, which is expected, as this is the closest site to the upstream boundary conditions.

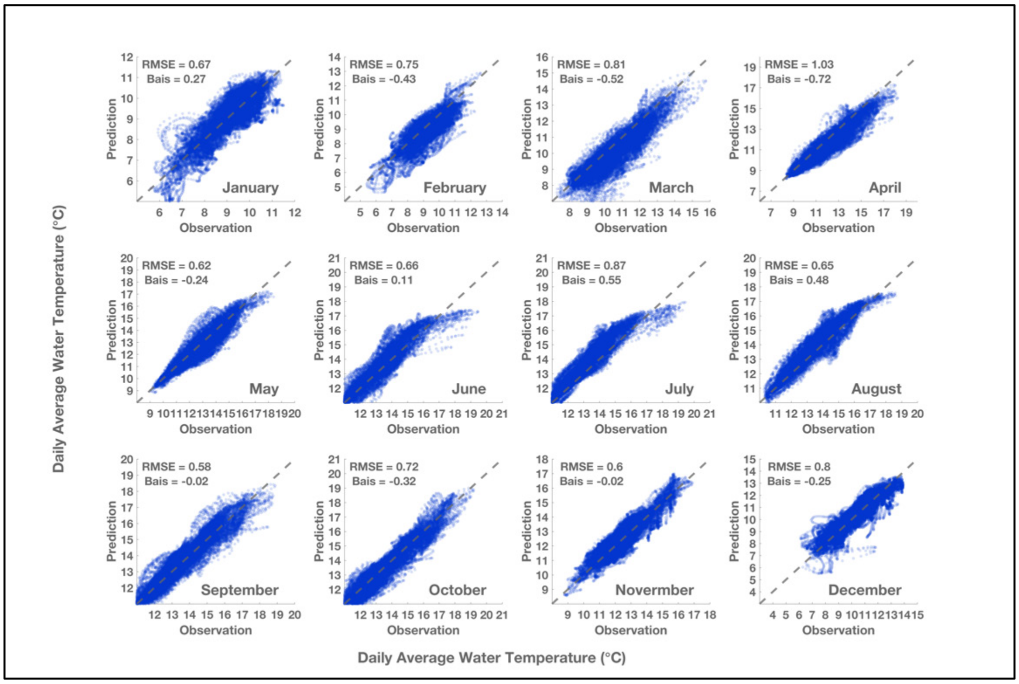

Figure 22: Observed longitudinal river temperatures plotted against RAFT predicted temperatures stratified by month during even years from 2000-2015 at the 50th percentile level from the 600 validation simulations, with RMSE and bias displayed for each stratum.

7. Evaluation of linked models

To evaluate overall model error when the W2 reservoir model, the AR model, and the RAFT river model were linked, the models were run in series during even years from 2000-2015. An example of these results is provided in Figure 23 comparing predicted water temperature at the Balls Ferry (BSF) gauge during August in validation years from 2000-2015. The model results for this time and gauge location indicated RAFT consistently predicted warmer water temperature on the order of 0.8° C. Comparing the same time period and location when RAFT was run with known upstream boundary conditions indicated that, for this example, RMSE and bias both increased by 0.4° C when using the linked models (note error statistics exclude the faulty gauge data of year 2014).

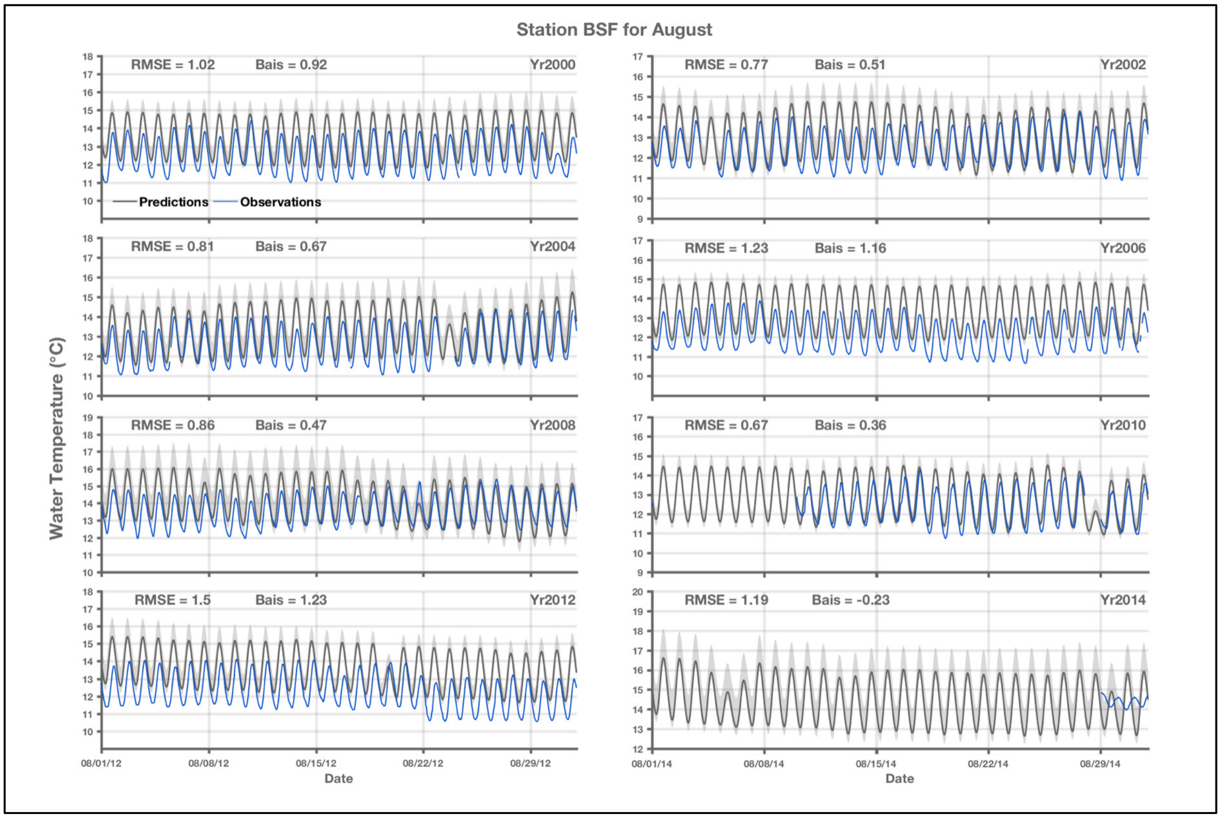

Figure 23: Observed (blue line) versus RAFT predicted (grey line) hourly temperature at Balls Ferry (BSF) gauge during the month of August for even years from 2000-2015 using W2 and AR output as the upstream boundary condition for RAFT simulations. Predicted temperatures are the result of 600 separate simulations with the 50th percentile displayed as a solid grey line and the 5th and 95th percentiles as the shaded region. RMSE and bias are displayed and based on the 50th percentile model results.

Collapsing the results by gauge station and year to assess model performance by month (Figure 24) indicated the linked models on average had an RMSE of 1° C and typically predicted warmer water temperatures (mean bias of 0.4° C) than observed. Error and bias were both typically greatest during summer months from June to October, with some model predictions being nearly 4° C warmer than observed.

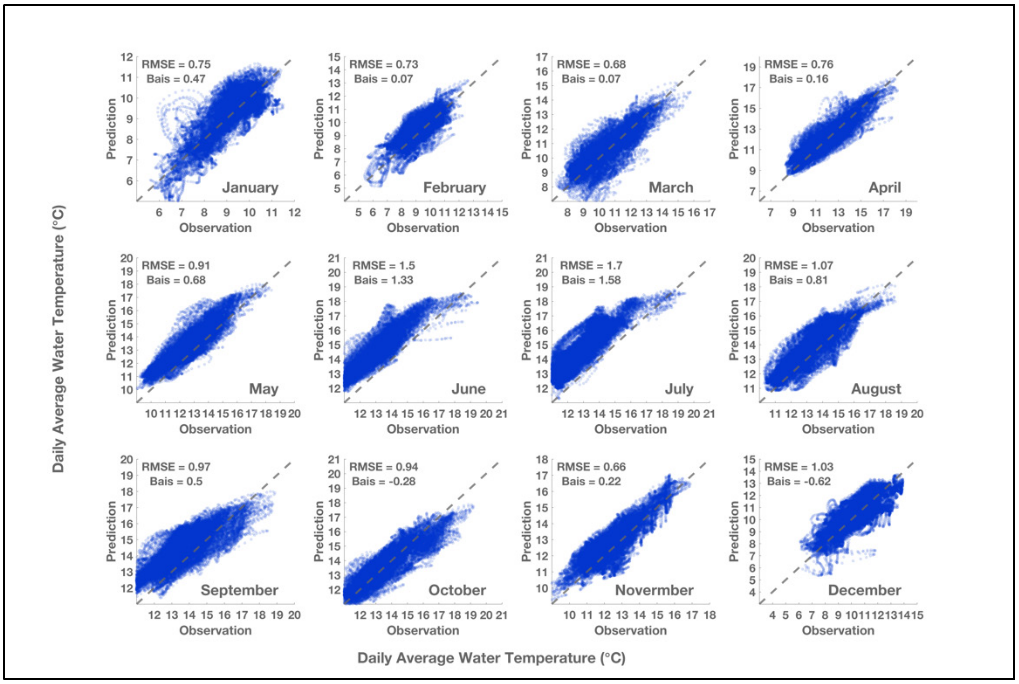

Figure 24: Observed longitudinal river temperatures plotted against RAFT predicted temperatures stratified by month during even years from 2000-2015 and using W2 and AR output as the upstream boundary condition for RAFT. RMSE and bias displayed for each stratum based on the 50th percentile estimates from 600 simulations.

8. Conclusions

The calibration and validation of the W2, AR, and RAFT models in this report revealed some key insights in relation to using the W2 and RAFT models to aid in water temperature management in the Shasta/Sacramento system.

As a tool to characterize the vertical distribution of temperature in Shasta Reservoir, the W2 model tended to have better skill in predicting the epilimnion and metalimnion layers of the reservoir compared to the hypolimnion. Specifically, the model often overestimated the temperature of water in the deepest sections of the reservoir, typically on the order of < 1° C (Appendix A). This finding may have implications for long-term seasonal operations of Shasta Reservoir that depend on cold water in the deeper layers of the reservoir to be accessed later in the temperature management season, especially if the model is run over multiple years. However, because some cold water in Shasta Reservoir is too deep to be accessible by the TCD (i.e. the dead-pool), the overestimation of water temperature in the deepest sections may not have a large impact on the overall management of cold-water.

While the W2 simulations often under-predicted cold-pool resources (water ≤ 10° C) in Shasta Reservoir during the course of a year, predictions of end of September (EOS) cold-pool storage aligned well and were often within 100 TAF or less of observed values. This indicated that the current version of the model would be expected to aid in evaluating seasonal operational scenarios of the Shasta Reservoir to manage EOS cold-pool resources and provide a conservative estimate as the model is downwardly biased.

Predictions of discharge temperature indicated the W2 model provides utility as a tool to aid in seasonal management of Shasta Reservoir release temperatures. The model accurately matched the temporal trends of temperature as well as the magnitude of discharge temperature on an order of 1° C when considering the 50th percentile of model predictions. The 5th and 95th percentiles generally encompassed the observed temperature values and provided a valuable assessment of uncertainty around discharge temperatures later in the year, which were shown to vary by as much as 5° C. Times when the range of model predictions did not encompass observed values after extensive adjustment of model parameters provides justification to further explore additional factors in the model that need consideration for calibration. For example, post-hoc analysis comparing years with high model skill (such as 2008) to those with lower relative skill (such as 2012) may provide valuable insights into factors related to these differences. Additional empirical data collected in field studies on reservoir and TCD dynamics may also be needed to better resolve reasons for discrepancies between years.

Examining the parameters calibrated in the W2 model itself also provided insights into how the reservoir model responded to meteorological forcing. The wind speed related calibration parameter (i.e. the wind sheltering coefficient or WSC parameter in Table 1) was observed to vary from 0.2 to 0.94 over the 95% credible level, with values less than 0.5 more likely than those greater than 0.5 (Figure 12). The WSC parameter dampens the effect of hydrodynamic mixing of water in a reservoir with lower values resulting in a greater level of dampening. Previous reports have found that values are typically between 0.5 and 1 (*8*). While it is unknown if other models were calibrated over the range of values explored in this calibration routine (i.e. if other model calibration parameter sets used values less than 0.5 for the WSC), finding a WSC value less than 0.5 indicates that reducing the level of wind-driven mixing improves the fit between observations and predictions of vertical temperature distribution in Shasta Reservoir as well as discharge temperature.

The calibration of how water was portioned to leak through the TCD also revealed some insights into reservoir dynamics. The parameter controlling the volume of water leaking through the TCD near the penstocks of Shasta Reservoir (i.e. parameter Leak_Coefficient_Z6 in Table 1) was found to improve model fit when increased above the value reported in a previous calibration of a one-dimensional model of Shasta Reservoir, where a coefficient of 0.1373 was used (*17*). In the present calibration, the parameter value varied from 0.17 to 0.3 over the 95% credible level, and was most likely ~ 0.29. While there are surely differences between the one-dimensional model and the two-dimensional W2 model calibrated in this report, this finding indicated that model fit was improved when a greater volume of cold water near the penstocks intake was assumed to leak through the TCD. This finding, while difficult to empirically validate, may have important implications for cold-water management that should be further assessed, as this type of leakage may reduce the efficiency of the TCD to conserve cold water.

While the statistical AR model was usually able to estimate Keswick Reservoir release temperatures based on Shasta release temperatures within 1° C of accuracy, the AR model lacks some capabilities of a physically based model. For example, in addition to Shasta Reservoir, Whiskeytown and Trinity Reservoirs also discharge into Keswick Reservoir via the Spring Creek Tunnel. Being able to more explicitly account for the discharge volumes and temperature variations into Keswick Reservoir with a physical model may become more important under scenarios where the fraction of water from Spring Creek Tunnel versus Shasta

Reservoir differs from historical trends and thus results in non-stationarity and a poorer statistical model fit. As an example, while a fraction of Keswick discharge has historically originated from Trinity Reservoir via the Spring Creek Tunnel, concerns about salmonids in the Trinity Watershed could result in reduced inputs to Keswick Reservoir from Spring Creek in the future.

The calibration and validation of RAFT confirmed its utility as a tool to assess water temperature impacts downstream from Shasta Reservoir as shown elsewhere (*7*) with error typically being less < 1° C over the model domain when using known boundary conditions and assessing daily average water temperatures. However, temperatures were often overestimated during June-July across sites and years, which has been a peak time for winter-run Chinook spawning according to redd aerial survey data (*30*). Future work is needed to evaluate potential factors that systematically result in overestimation of water temperature in summer months using the RAFT model, such as incorporating additional dynamics that affect river water temperature, such as hyporheic flow and riparian shading.

When the models were used in series and under known boundary conditions, modeled daily average temperatures often aligned with observed temperatures on the order of 1° C over a period of a year from below Keswick Reservoir to the city of Red Bluff (See Figure 25 for example at site BSF and Appendix C for additional sites). While error and uncertainty often increased when using the models in series compared to only using RAFT with known boundary conditions (Figure 23 vs. Figure 21), using the models in series for future analysis allows for a more comprehensive modeling framework of the system and more in-depth analysis of different strategies to achieve both reservoir and downstream temperature targets. Currently, the majority of the error propagation occurs from the W2-AR model providing upstream boundary conditions of temperature for the RAFT model (Figure 25). Therefore, it is recommended that future efforts be focused on reducing error in predicting Shasta Reservoir discharge temperatures as a starting point for RAFT.

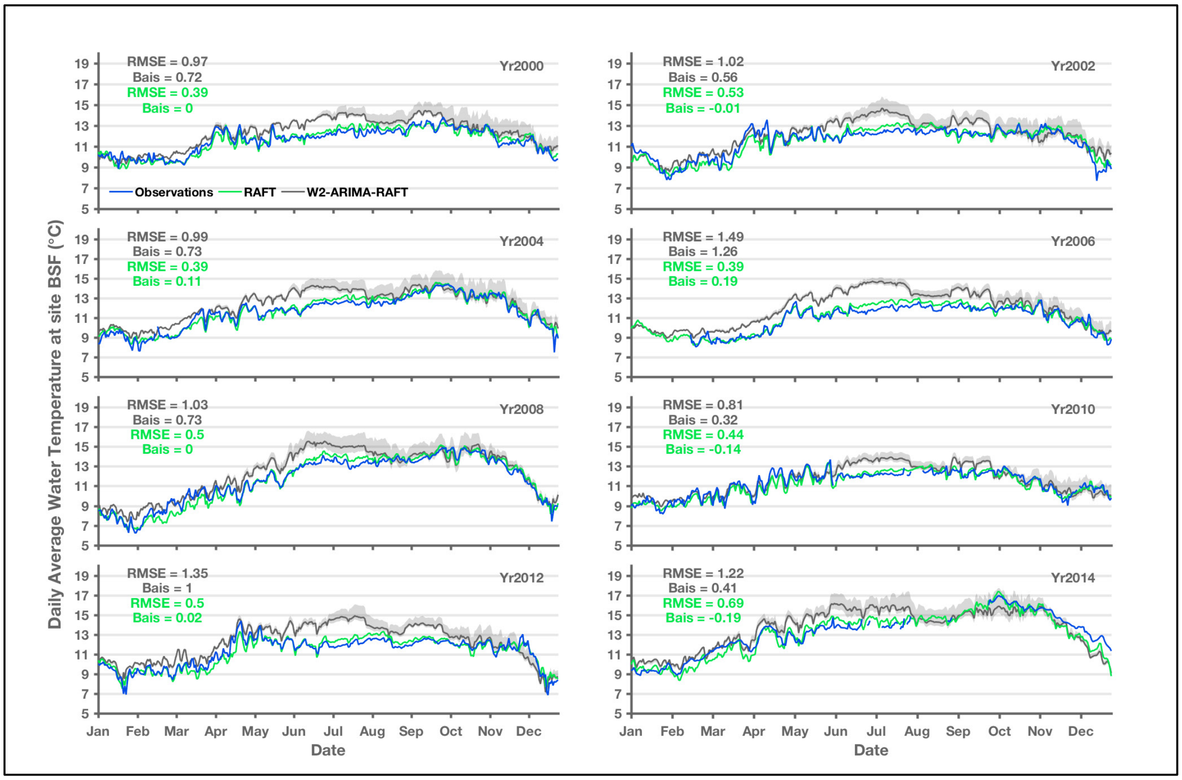

Figure 25: Time series of daily average observed river temperatures at gauging station BSF (blue line), along with RAFT predictions with known boundary conditions (green line), and RAFT predictions with boundary conditions supplied by the W2-AR models (grey line with 95th percentile confidence interval in shaded region) stratified by validation years from 2000-2015. RMSE and bias displayed for each stratum are based on the 50th percentile estimates from 600 simulations.

In an operational/planning setting, use of the models should be made with the understanding that additional error added from uncertain conditions not present in a validation setting where boundary conditions are known, may inflate model error. For example, there are three primary components that produce additional uncertainty in long range forecasting for the Shasta/Sacramento system that are not represented in the validation of these models presented: 1) reservoir operations, 2) meteorology, and 3) hydrology. While there will always be uncertainty in the forecasting process associated with these inputs, further analysis may help to better understand and quantify this uncertainty. For example, an in-depth global sensitivity analysis exploring how much of the model error can be attributed to these sources and others in space and time would provide insight into how best to use these models in an operational context with imperfect knowledge of how the TCD at Shasta Dam is operated for a given season.

Ultimately, a primary function of utilizing the W2 and RAFT models described in this report is to predict daily average and other temperature metrics at specific locations or along reaches of the Sacramento River under varying operational and environmental scenarios in the Shasta/Sacramento system. Temperature predictions can then be used with other information to minimize the exposure of endangered winter-run Chinook salmon populations below Shasta Reservoir to harmful water temperatures (*31*). The findings from this calibration and validation report provide a robust assessment of the models' skill to aid management of cold-water resources. While improvements should be made (including further calibration/validation) to reduce model error, especially when error is greatest during the summer, the models in their current state do enable assessment of expected water temperature in Shasta Reservoir and the Sacramento River under varying operations and environmental conditions.

References

1. USBR, "Water Supply and Yield Study " (United States Department of the Interior; Bureau of Reclamation, Mid-Pacific Region, 2008).
2. NMFS, "CENTRAL VALLEY RECOVERY DOMAIN: 5-Year Review, Summary and Evaluation of Sacramento River Winter-run Chinook Salmon ESU," (National Marine Fisheries Service, Southwest Region, Long Beach, CA, 2011).
3. USBR, "Shasta Lake Water Resources Investigation," (United States Department of the Interior; Bureau of Reclamation, Mid-Pacific Region, 2006).
4. O. Balci, in *Simulation Conference Proceedings, 1994. Winter*. (IEEE, 1994), pp. 215-220.
5. S. C. Chapra, *Surface water-quality modeling*. (Waveland press, 2008).
6. Z. Xing, D. A. Fong, E. Yat-Man Lo, S. G. Monismith, Thermal structure and variability of a shallow tropical reservoir. *Limnol. Oceanogr.* **59**, 115-128 (2014).
7. A. Pike *et al.*, Forecasting river temperatures in real time using a stochastic dynamics approach. *Water Resources Research* **49**, 5168-5182 (2013).
8. T. M. Cole, S. A. Wells, CE-QUAL-W2: A two-dimensional, laterally averaged, hydrodynamic and water quality model, version 3.5. (2006).
9. I. MathWorks. (Natick, Massachusetts, United States, 2016).
10. R Development Core Team. (R Foundation for Statistical Computing, Vienna, Austria, 2017).
11. J. B. Elsner, A. A. Tsonis, *Singular spectrum analysis: a new tool in time series analysis*. (Springer Science & Business Media, 2013).
12. J. Chen, F. P. Brissette, D. Chaumont, M. Braun, Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two North American river basins. *J. Hydrol.* **479**, 200-214 (2013).
13. R. B. Hanna, L. Saito, J. M. Bartholow, J. Sandelin, Results of simulated temperature control device operations on in-reservoir and discharge water temperatures using CE-QUAL-W2. *Lake Reserv. Manage.* **15**, 87-102 (1999).
14. J. Bartholow, R. Hanna, L. Saito, D. Lieberman, M. Horn, Simulated limnological effects of the Shasta Lake temperature control device. *Environ. Manage.* **27**, 609-626 (2001).
15. P. Yang *et al.*, Observations of vertical eddy diffusivities in a shallow tropical reservoir. *Journal of Hydro-environment Research* **9**, 441-451 (2015).
16. L. Saito, Colorado State University, Fort Collins, CO (1999).
17. RMA, "Upper Sacramento River Water Quality Modeling with HEC-5Q: Model Calibration and Validation," (Resource Management Associates, Inc., Suisun City, California, 2003).
18. I. R. Wood, Extensions to the theory of selective withdrawal. *Journal of Fluid Mechanics* **448**, 315-333 (2001).
19. K. Beven, *Environmental modelling: an uncertain future?* , (Taylor & Francis, 2008).
20. G. H. Givens, J. A. Hoeting, *Computational statistics*. (John Wiley & Sons, 2012), vol. 708.
21. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087 (1953).
22. S. P. Brooks, A. Gelman, General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* **7**, 434-455 (1998).

23. A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple sequences. *Stat. Sci.*, 457-472 (1992).
24. B. J. Smith, boa: an R package for MCMC output convergence assessment and posterior inference. *J Stat. Softw.* **21**, 1-37 (2007).
25. T. C. Mills, *Time series techniques for economists*. (Cambridge University Press, 1991).
26. P. Kundu, I. Cohen, Fluid mechanics. 2004. *Elsevier Academic Press, San Diego). Two-and three-dimensional self-sustained flow oscillations* **307**, 471-476 (2008).
27. J. H. Ferziger, M. Peric, *Computational methods for fluid dynamics*. (Springer Science & Business Media, 2012).
28. G. Evensen, *Data assimilation: the ensemble Kalman filter*. (Springer Science & Business Media, 2009).
29. M. Boyd, B. Kasper, "Analytical Methods for Dynamic Open Channel Heat and Mass Transfer: Methodology for heat source model Version 7.0.," (2003).
30. D. Killam, Sacramento River winter-run Chinook salmon escapement survey April–September 2004. *Calif. Dept. Fish and Game SRSSAP Tech. Rep*, 06-01 (2005).
31. B. T. Martin *et al.*, Phenomenological vs. biophysical models of thermal stress in aquatic eggs. *Ecol. Lett.* **20**, 50-59 (2017).

Appendix A: Vertical temperature profiles of Shasta Reservoir

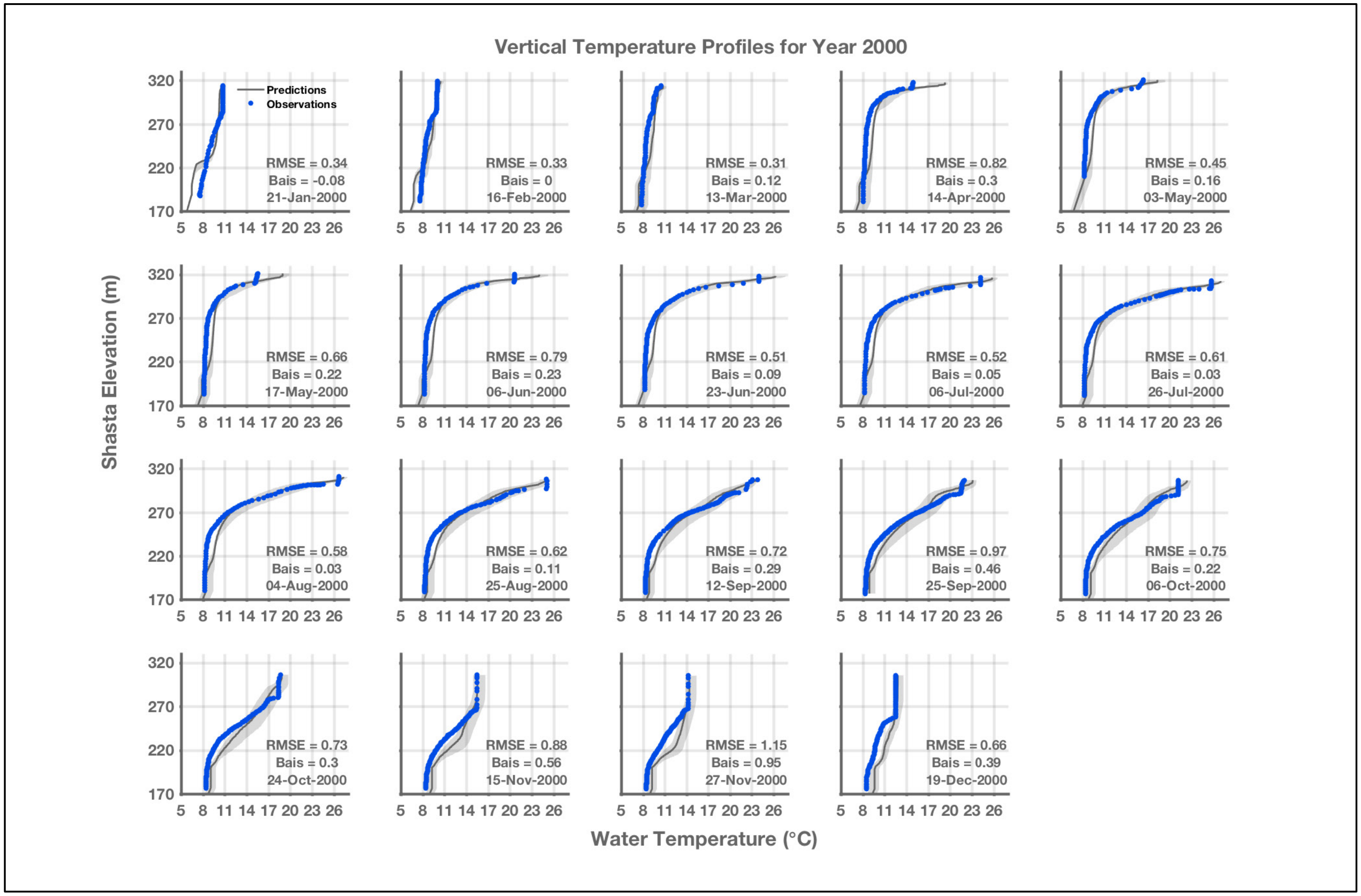

Figure 1-A: Observed (blue dots) versus W2 predicted (grey line) vertical temperature profile near the face of Shasta Dam for year 2000. The 50th percentile of the W2 predictions are displayed as a solid line with 5th and 95th percentile predictions shown as a shaded region. Goodness-of-fit metrics are based on the 50th percentile model output.

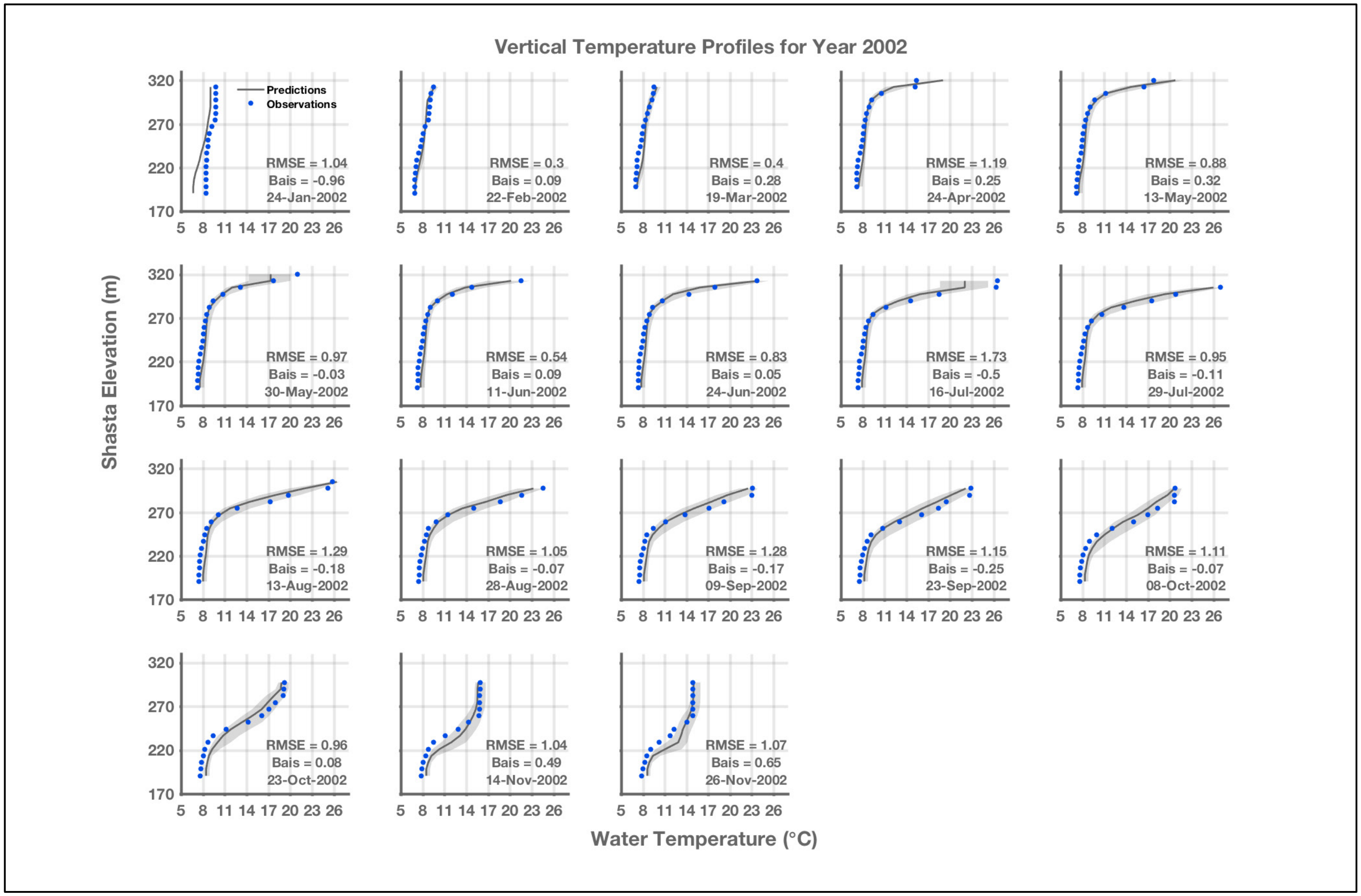

Figure 2-A: Observed (blue dots) versus W2 predicted (grey line) vertical temperature profile near the face of Shasta Dam for year 2002. The 50th percentile of the W2 predictions are displayed as a solid line with 5th and 95th percentile predictions shown as a shaded region. Goodness-of-fit metrics are based on the 50th percentile model output.

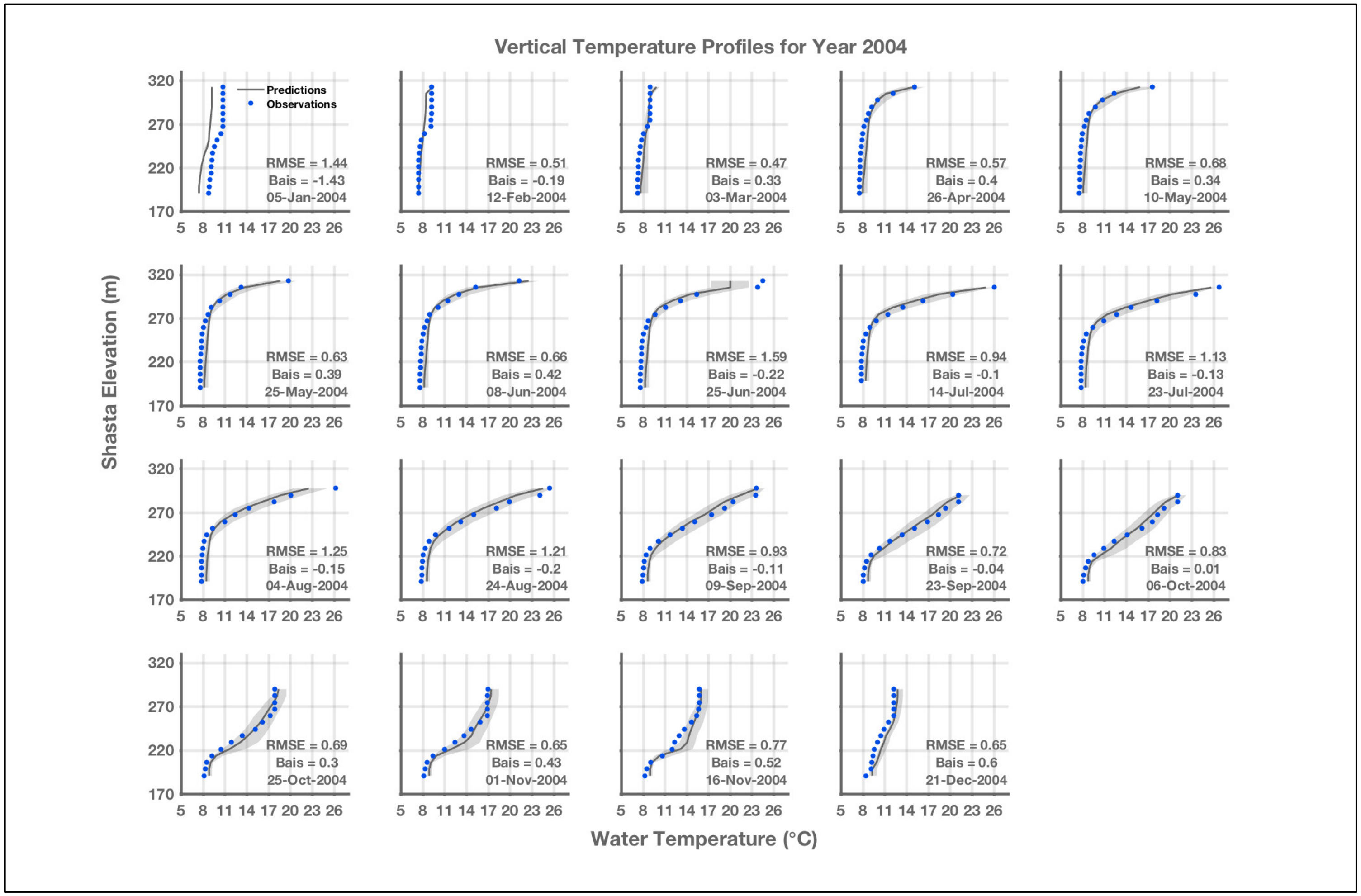

Figure 3-A: Observed (blue dots) versus W2 predicted (grey line) vertical temperature profile near the face of Shasta Dam for year 2004. The 50th percentile of the W2 predictions are displayed as a solid line with 5th and 95th percentile predictions shown as a shaded region. Goodness-of-fit metrics are based on the 50th percentile model output.

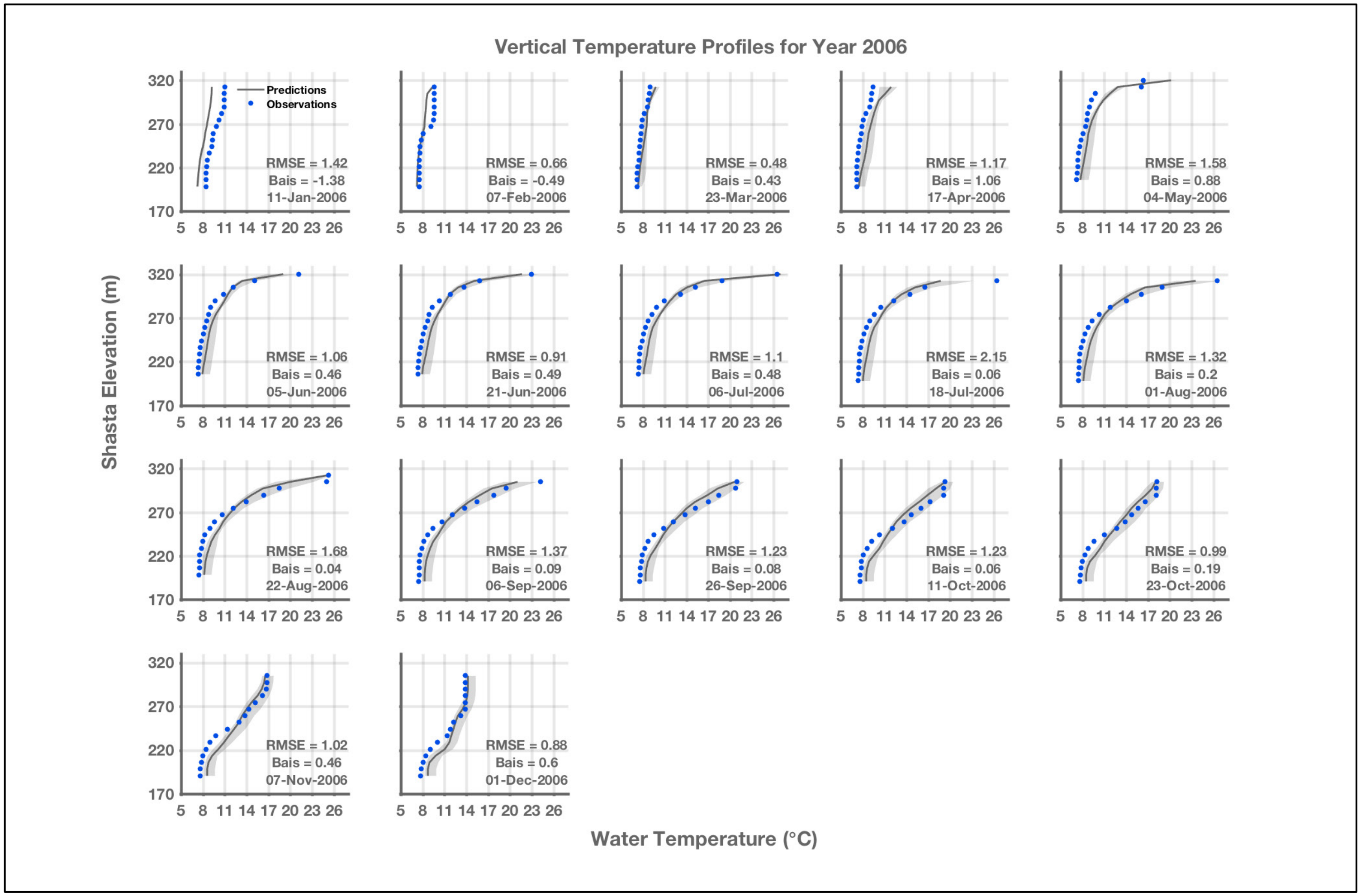

Figure 4-A: Observed (blue dots) versus W2 predicted (grey line) vertical temperature profile near the face of Shasta Dam for year 2006. The 50th percentile of the W2 predictions are displayed as a solid line with 5th and 95th percentile predictions shown as a shaded region. Goodness-of-fit metrics are based on the 50th percentile model output.

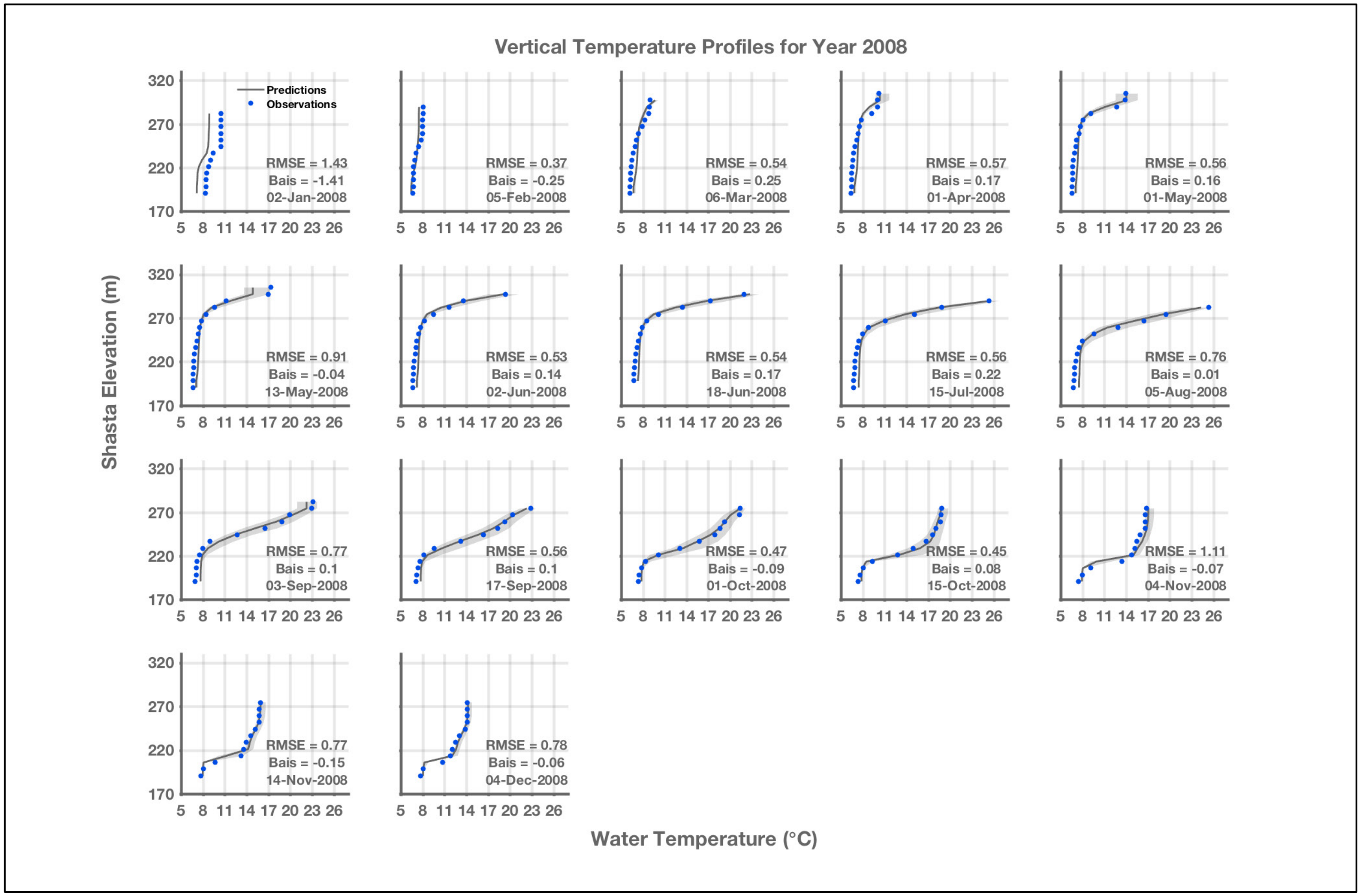

Figure 5-A: Observed (blue dots) versus W2 predicted (grey line) vertical temperature profile near the face of Shasta Dam for year 2008. The 50th percentile of the W2 predictions are displayed as a solid line with 5th and 95th percentile predictions shown as a shaded region. Goodness-of-fit metrics are based on the 50th percentile model output.

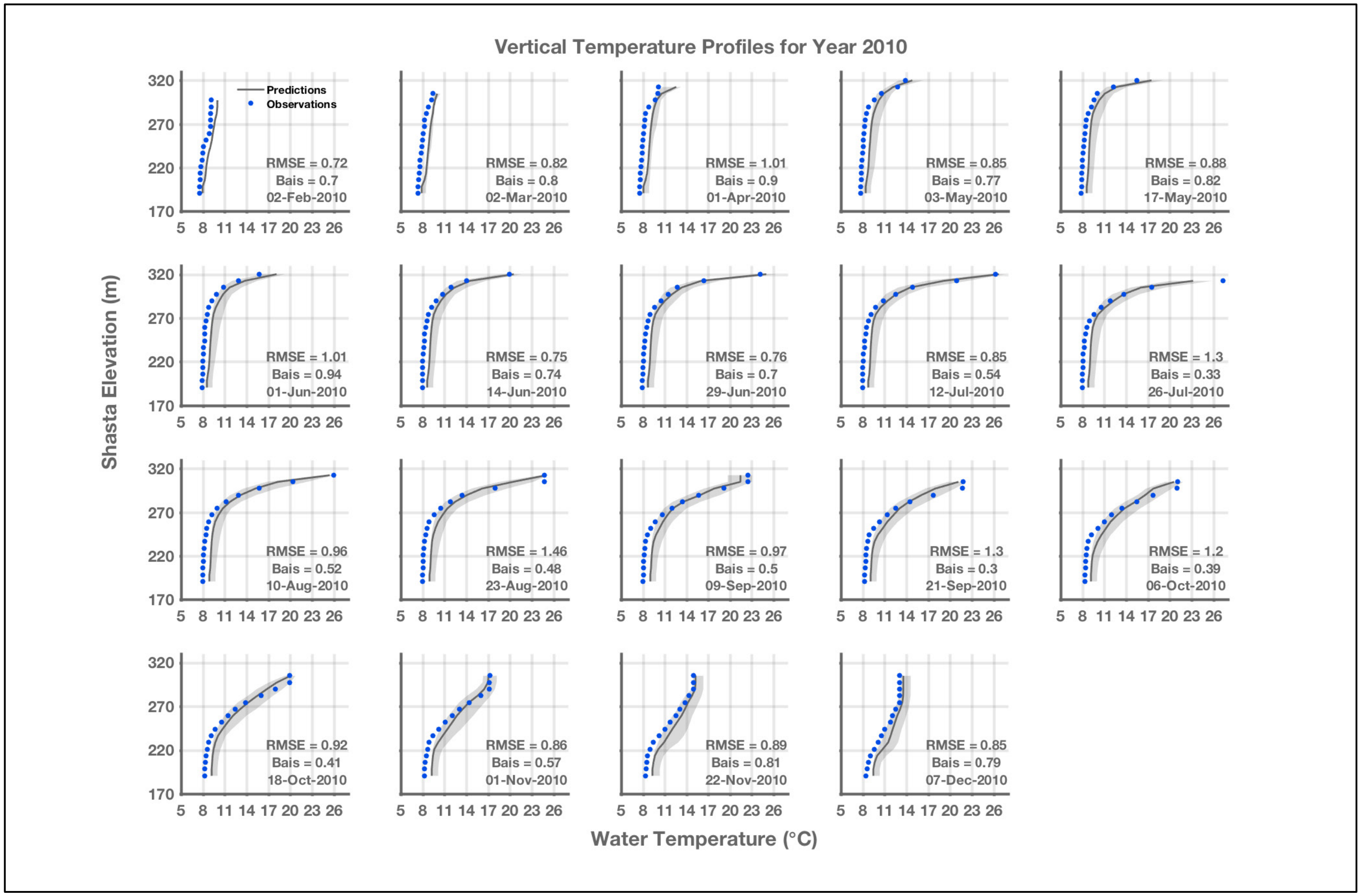

Figure 6-A: Observed (blue dots) versus W2 predicted (grey line) vertical temperature profile near the face of Shasta Dam for year 2010. The 50th percentile of the W2 predictions are displayed as a solid line with 5th and 95th percentile predictions shown as a shaded region. Goodness-of-fit metrics are based on the 50th percentile model output.

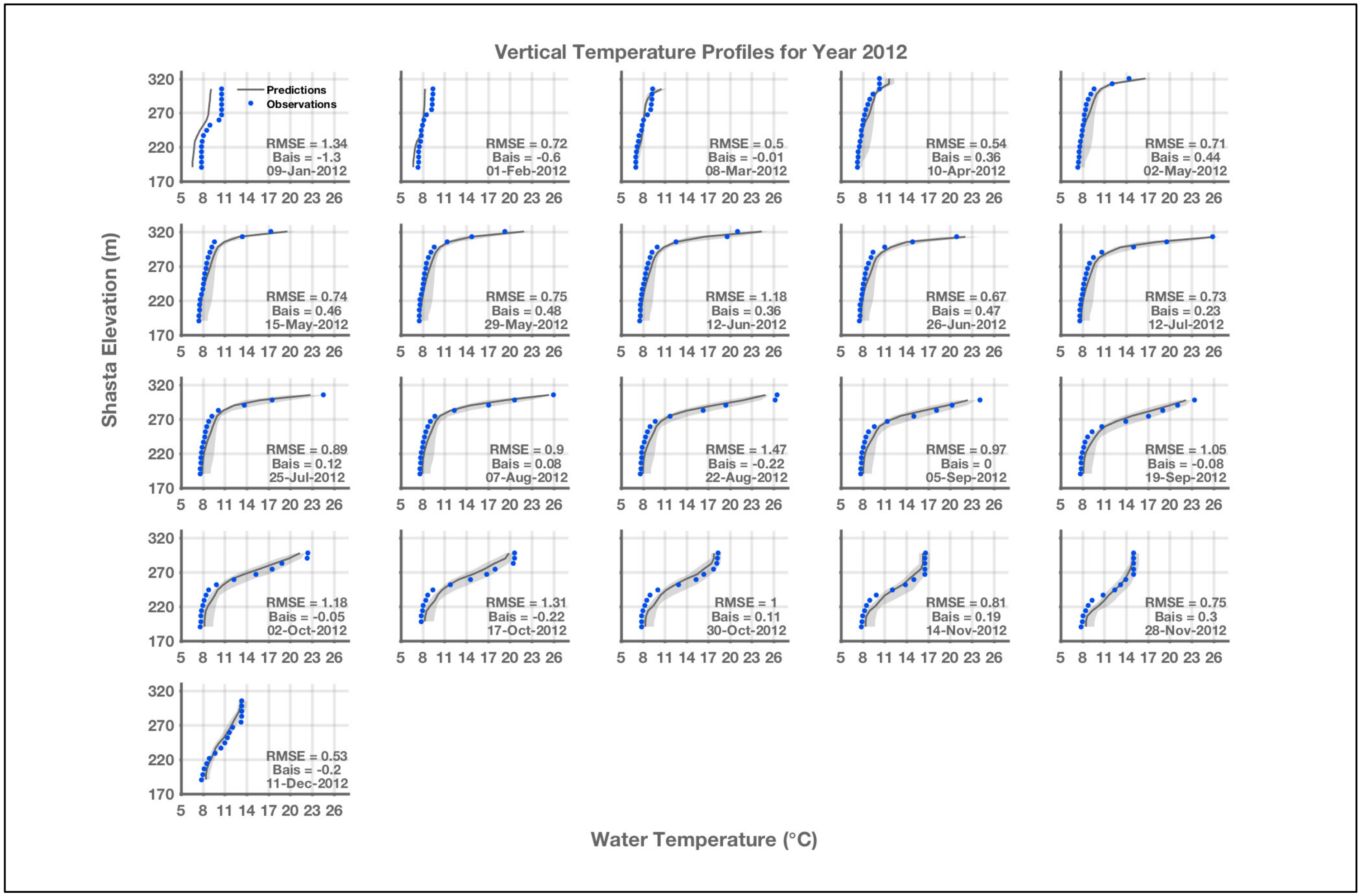

Figure 7-A: Observed (blue dots) versus W2 predicted (grey line) vertical temperature profile near the face of Shasta Dam for year 2012. The 50th percentile of the W2 predictions are displayed as a solid line with 5th and 95th percentile predictions shown as a shaded region. Goodness-of-fit metrics are based on the 50th percentile model output.

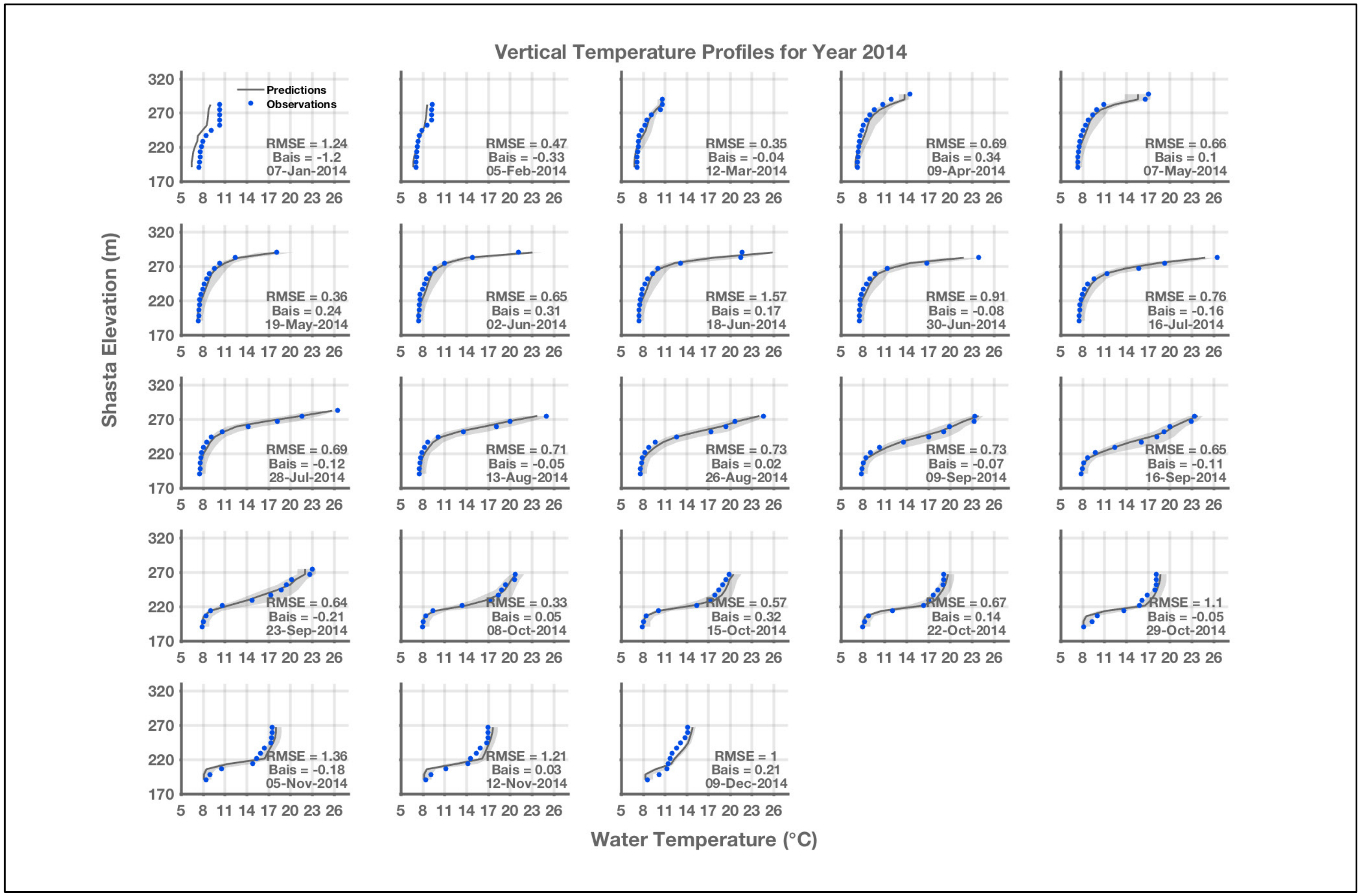

Figure 8-A: Observed (blue dots) versus W2 predicted (grey line) vertical temperature profile near the face of Shasta Dam for year 2014. The 50th percentile of the W2 predictions are displayed as a solid line with 5th and 95th percentile predictions shown as a shaded region. Goodness-of-fit metrics are based on the 50th percentile model output.

Appendix B: RAFT Validation 1:1 relationship by gauge station

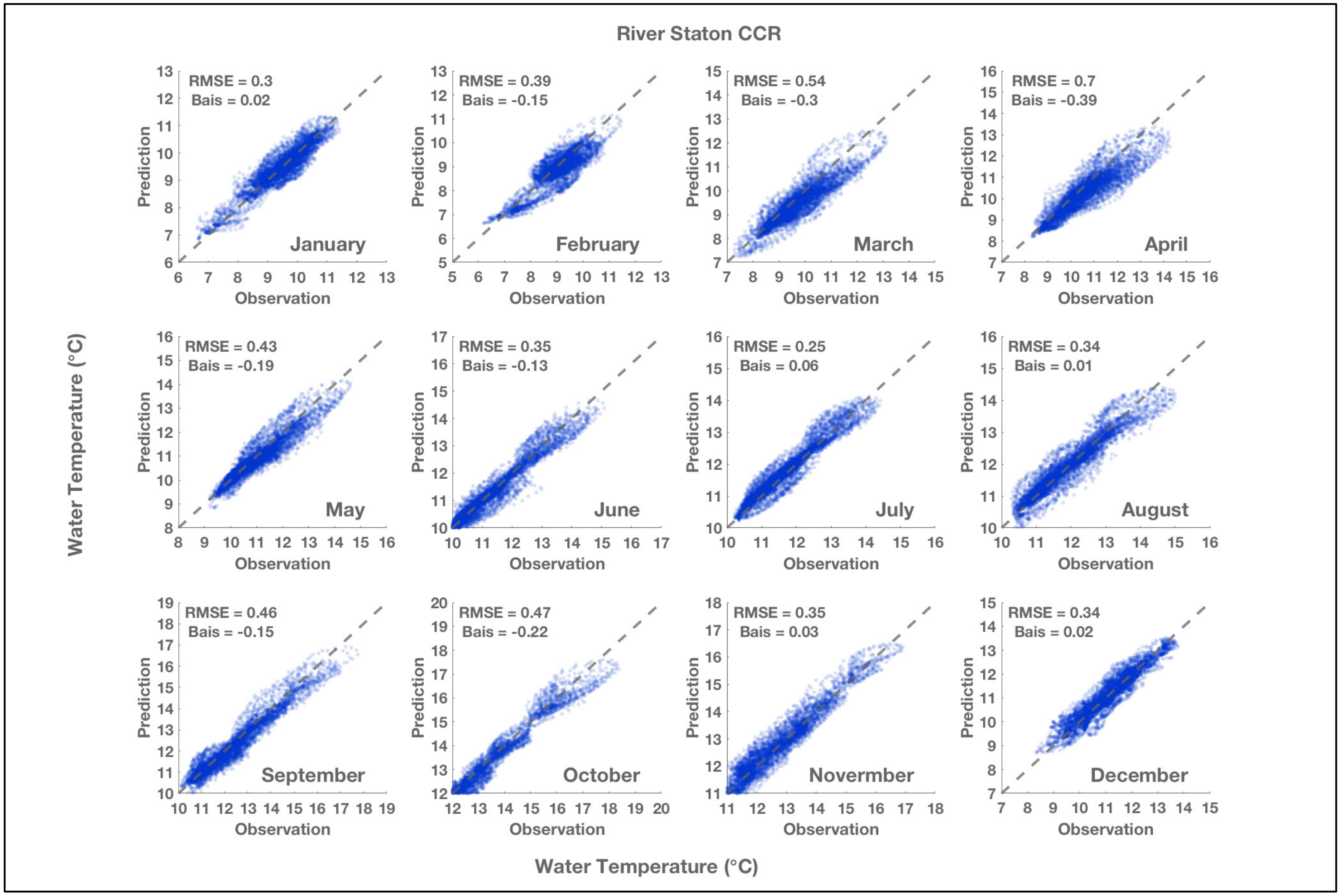

Figure 1-B: Observed river temperatures at gauge station CCR plotted against RAFT predicted temperatures stratified by month during even years from 2000-2015 at the 50th percentile level from the 600 validation simulations, with RMSE and bias displayed for each stratum.

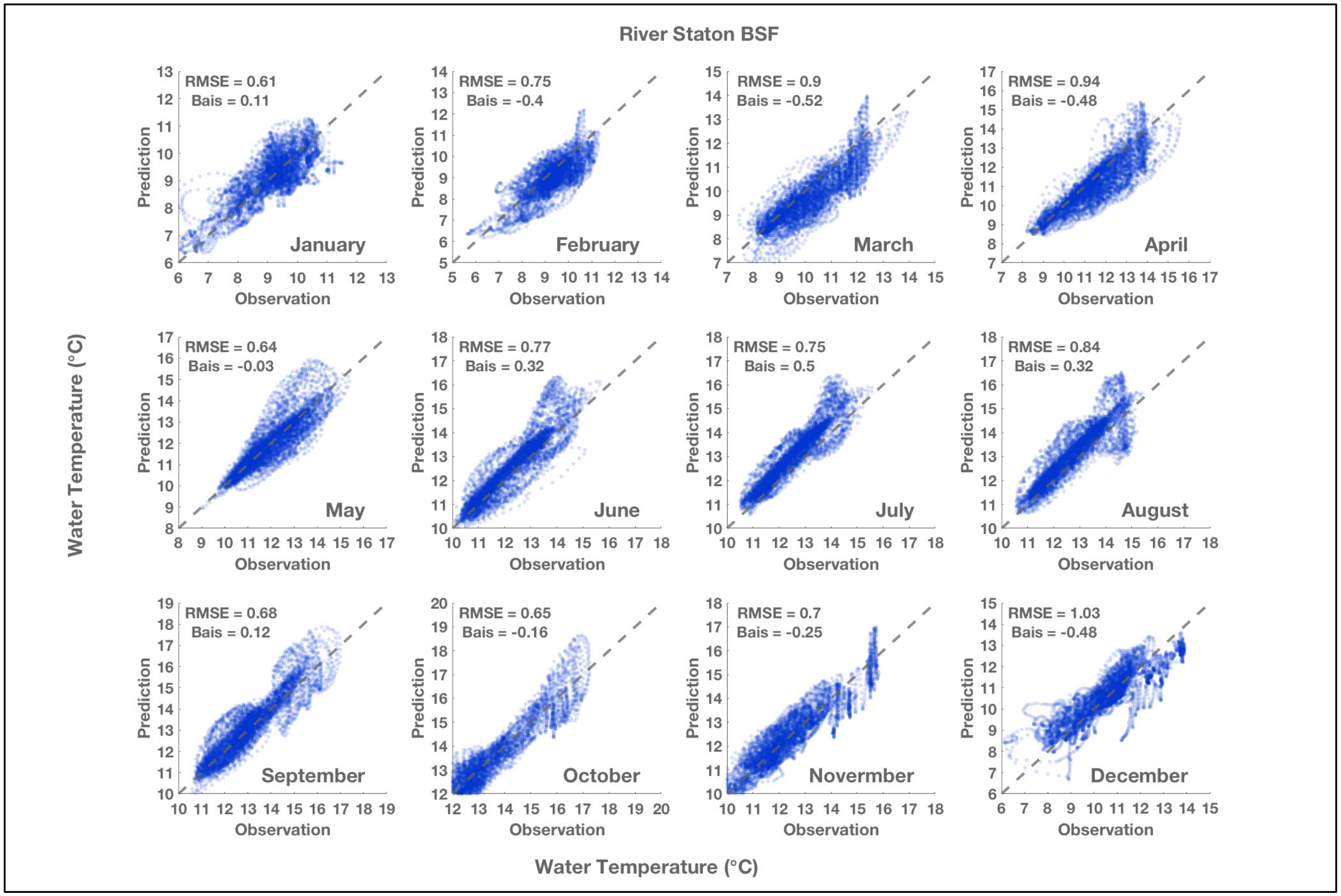

Figure 2-B: Observed river temperatures at gauge station BSF plotted against RAFT predicted temperatures stratified by month during even years from 2000-2015 at the 50th percentile level from the 600 validation simulations, with RMSE and bias displayed for each stratum.

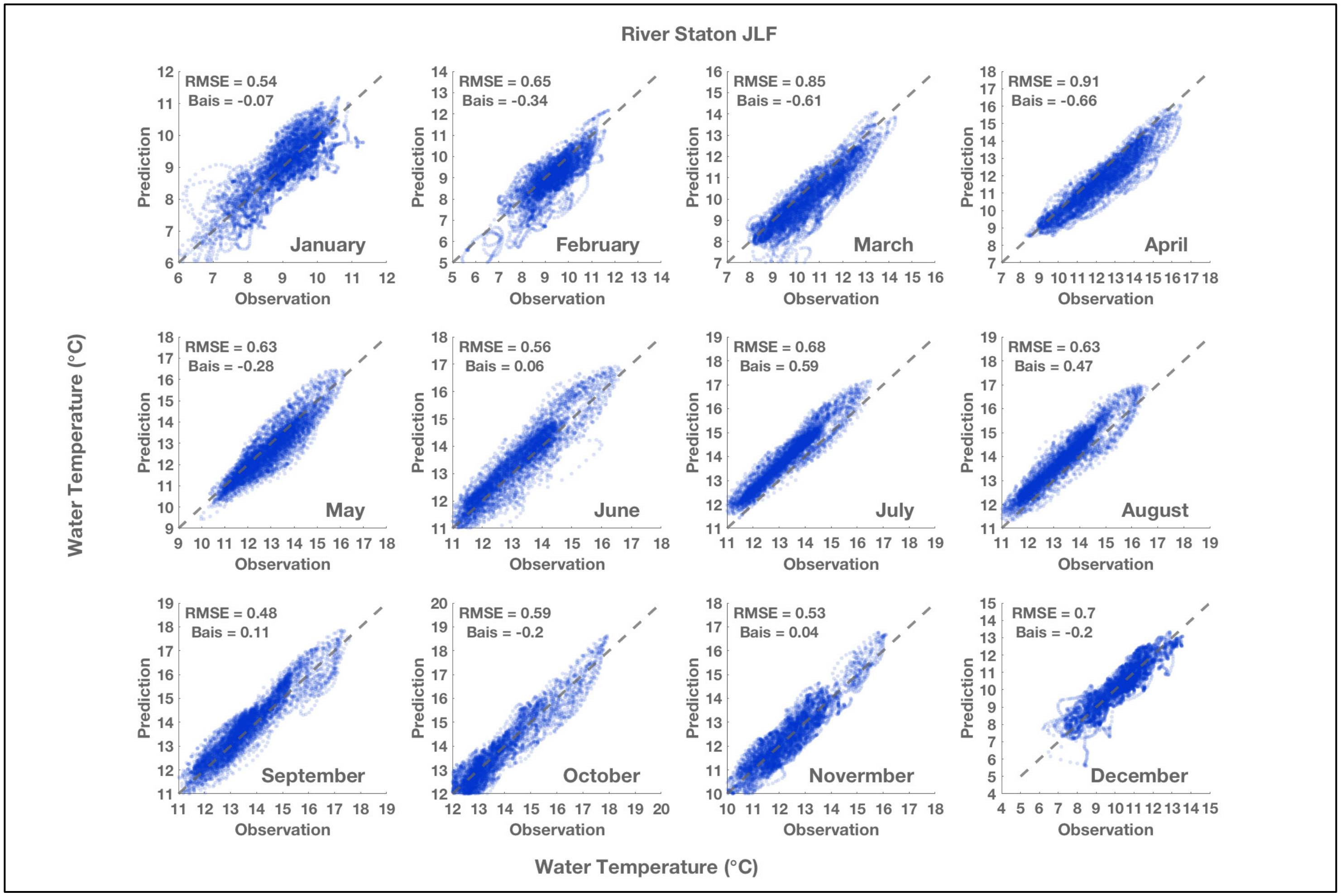

Figure 3-B: Observed river temperatures at gauge station JLF plotted against RAFT predicted temperatures stratified by month during even years from 2000-2015 at the 50th percentile level from the 600 validation simulations, with RMSE and bias displayed for each stratum.

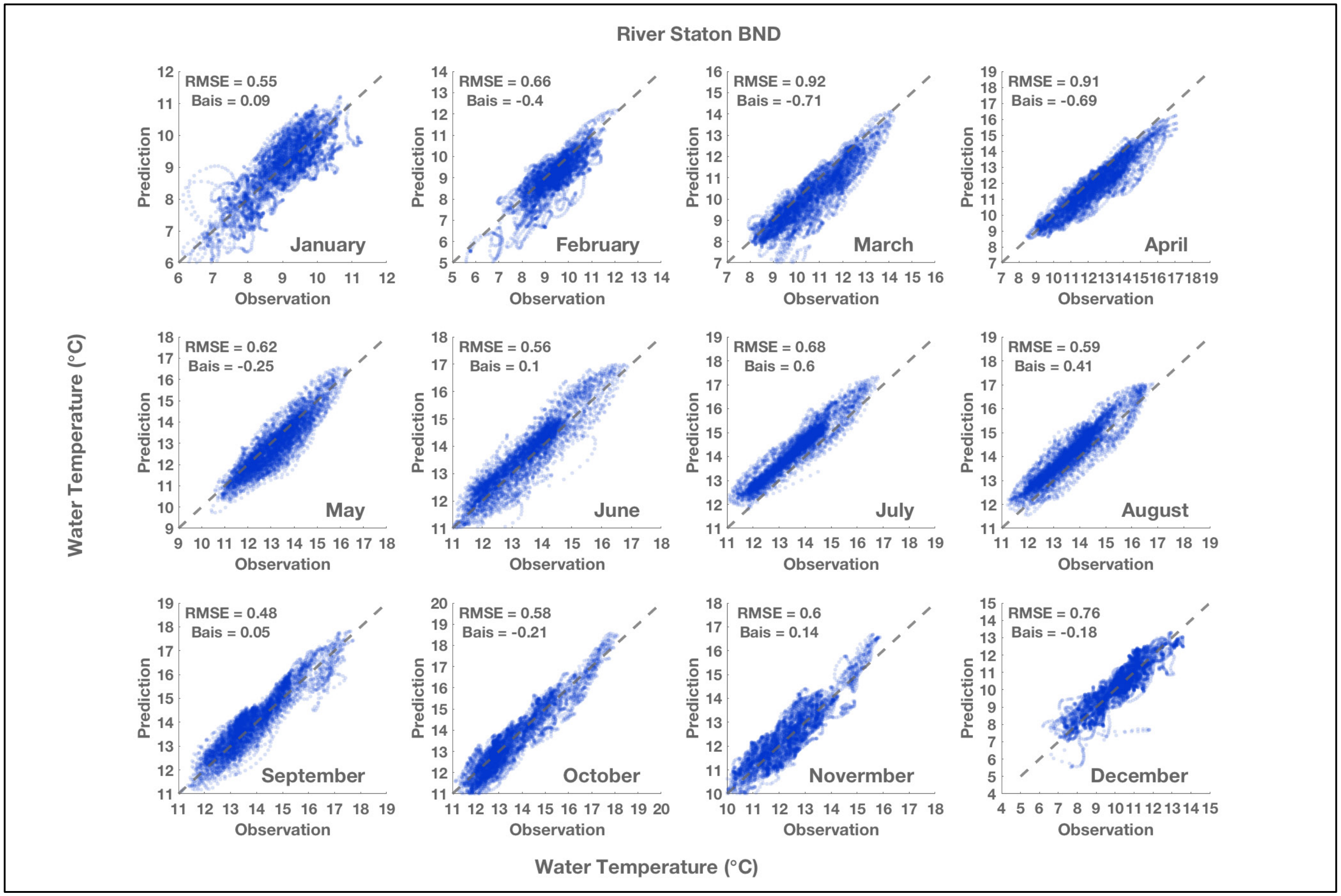

Figure 4-B: Observed river temperatures at gauge station BND plotted against RAFT predicted temperatures stratified by month during even years from 2000-2015 at the 50th percentile level from the 600 validation simulations, with RMSE and bias displayed for each stratum.

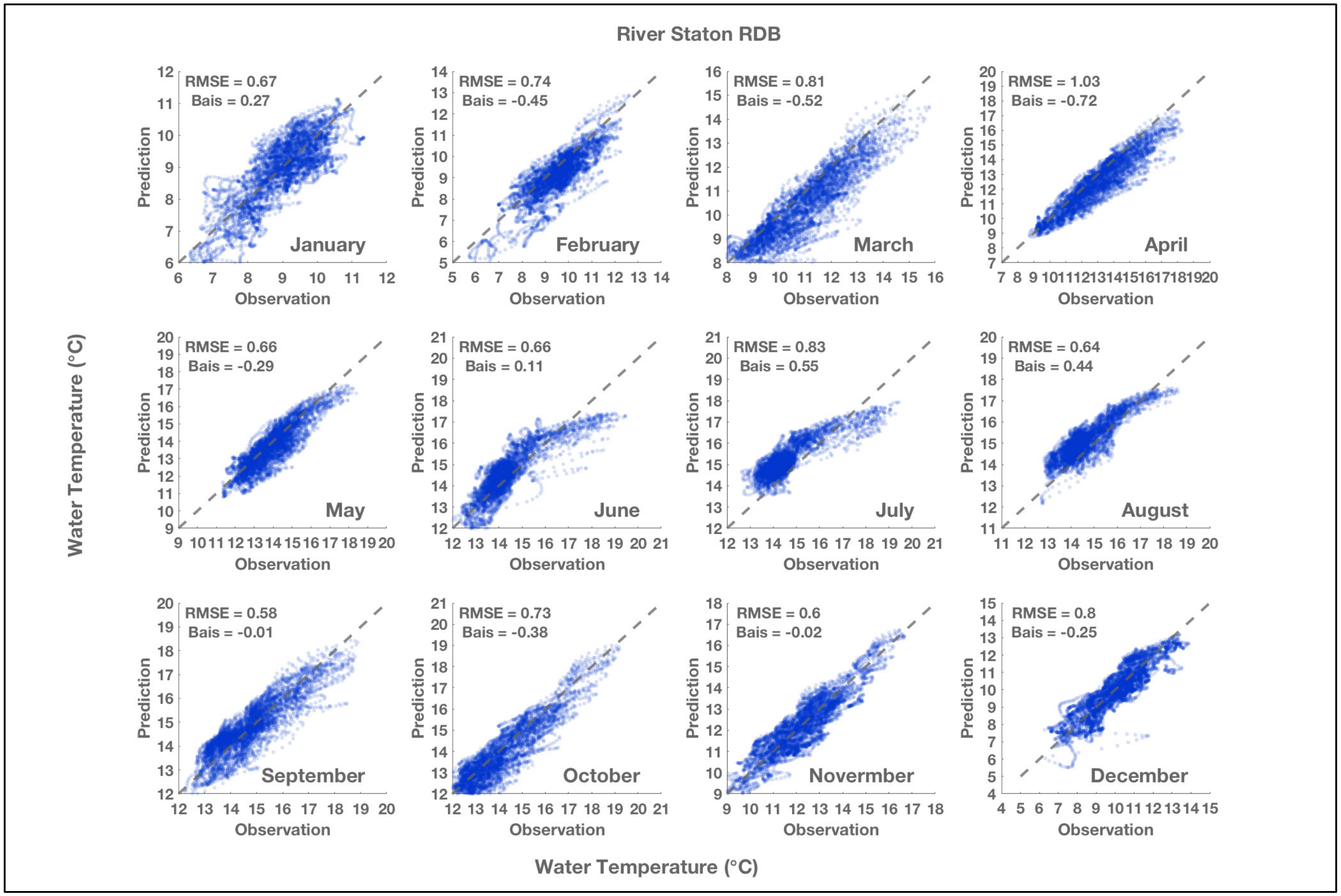

Figure 5-B: Observed river temperatures at gauge station RDB plotted against RAFT predicted temperatures stratified by month during even years from 2000-2015 at the 50th percentile level from the 600 validation simulations, with RMSE and bias displayed for each stratum.

Appendix C: W2-AR-RAFT daily average temperature plots by gauge station

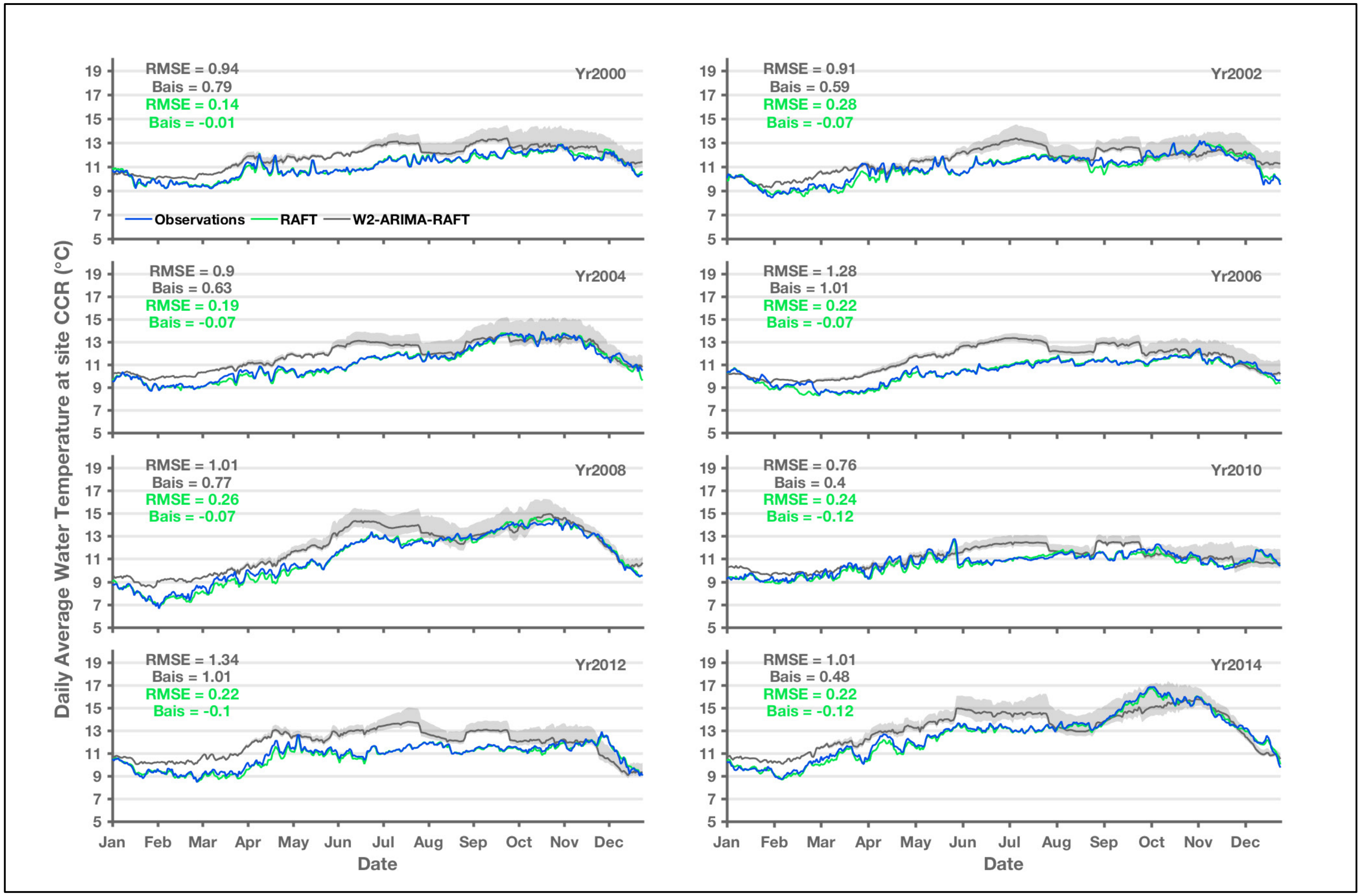

Figure 1-C: Time series of daily average observed river temperatures at gauging station CCR (blue line), along with RAFT predictions with known boundary conditions (green line), and RAFT predictions with boundary conditions supplied by the W2-AR models (grey line with 95th percentile confidence interval in shaded region) stratified by year from 2000-2015. RMSE and bias displayed for each stratum based on the 50th percentile estimates from 600 simulations.

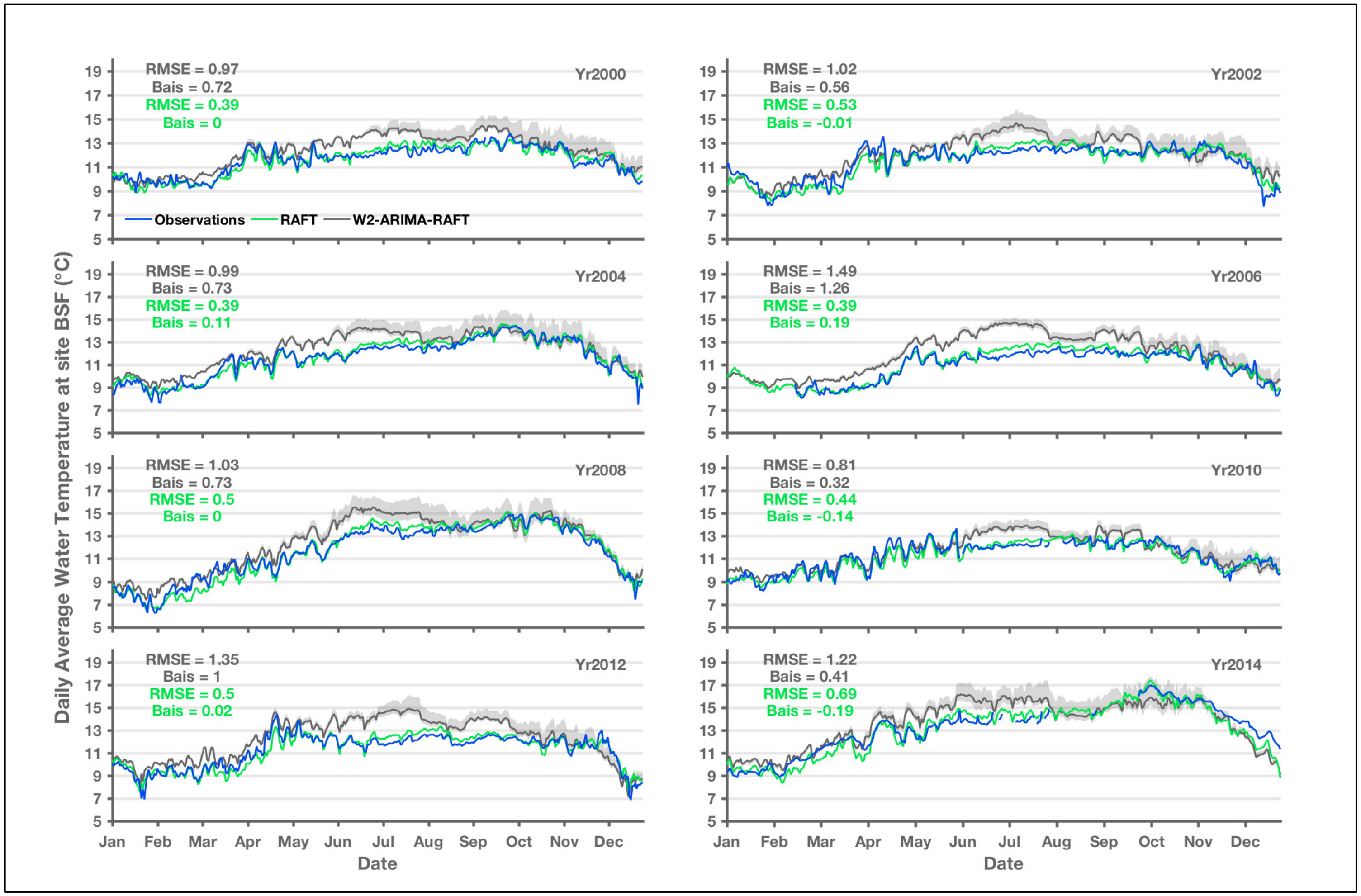

Figure 2-C: Time series of daily average observed river temperatures at gauging station BSF (blue line), along with RAFT predictions with known boundary conditions (green line), and RAFT predictions with boundary conditions supplied by the W2-AR models (grey line with 95th percentile confidence interval in shaded region) stratified by year from 2000-2015. RMSE and bias displayed for each stratum based on the 50th percentile estimates from 600 simulations.

Figure 3-C: Time series of daily average observed river temperatures at gauging station JLF (blue line), along with RAFT predictions with known boundary conditions (green line), and RAFT predictions with boundary conditions supplied by the W2-AR models (grey line with 95th percentile confidence interval in shaded region) stratified by year from 2000-2015. RMSE and bias displayed for each stratum based on the 50th percentile estimates from 600 simulations.

Figure 4-C: Time series of daily average observed river temperatures at gauging station BND (blue line), along with RAFT predictions with known boundary conditions (green line), and RAFT predictions with boundary conditions supplied by the W2-AR models (grey line with 95th percentile confidence interval in shaded region) stratified by year from 2000-2015. RMSE and bias displayed for each stratum based on the 50th percentile estimates from 600 simulations.

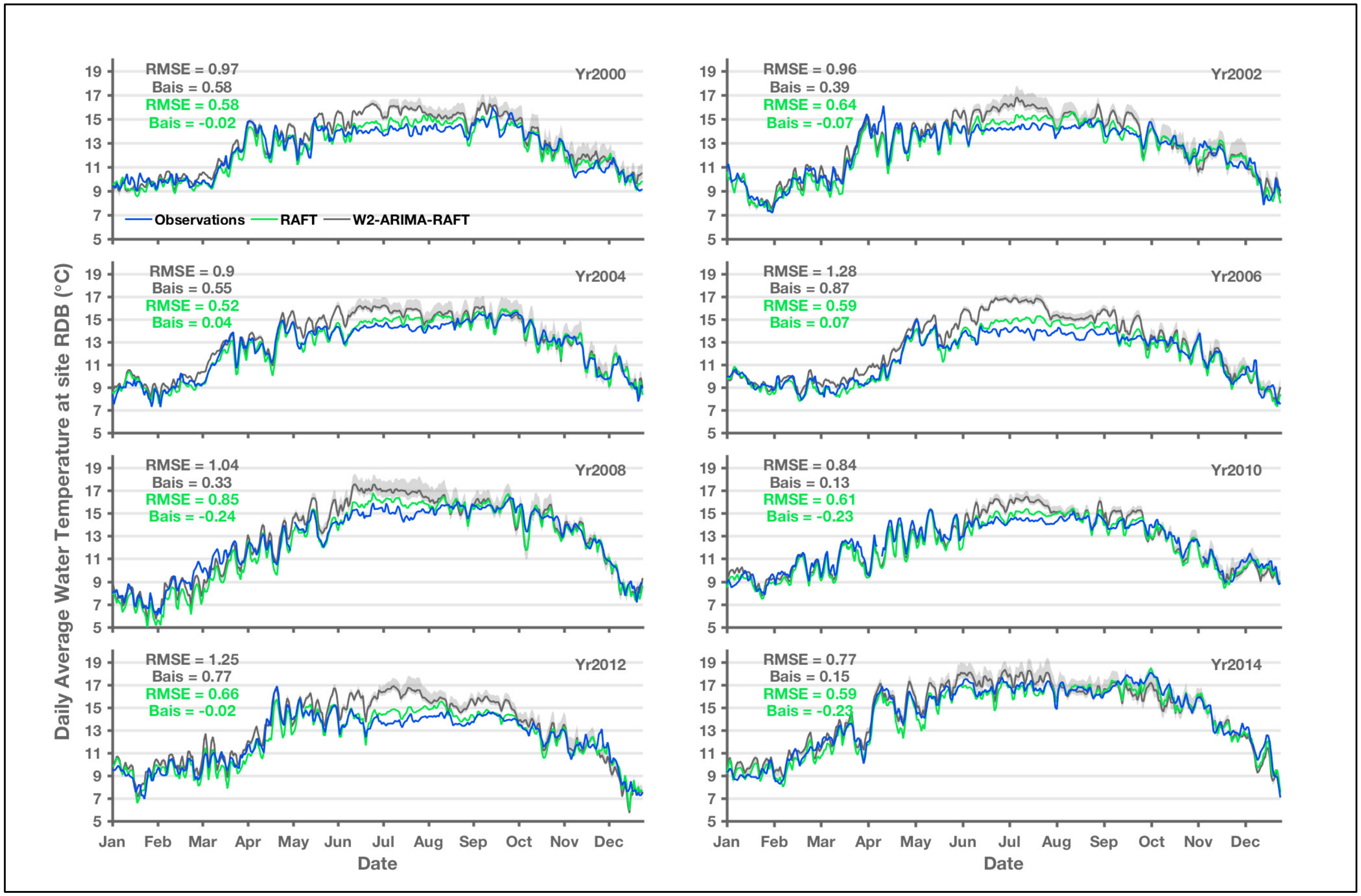

Figure 5-C: Time series of daily average observed river temperatures at gauging station RDB (blue line), along with RAFT predictions with known boundary conditions (green line), and RAFT predictions with boundary conditions supplied by the W2-AR models (grey line with 95th percentile confidence interval in shaded region) stratified by year from 2000-2015. RMSE and bias displayed for each stratum based on the 50th percentile estimates from 600 simulations.