BUREAU OF RECLAMATION

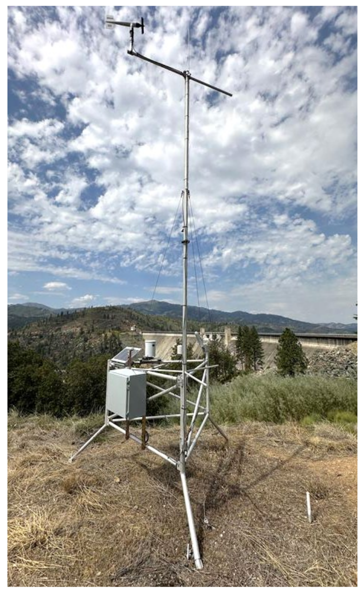
Final Report No. ST-2025-22050

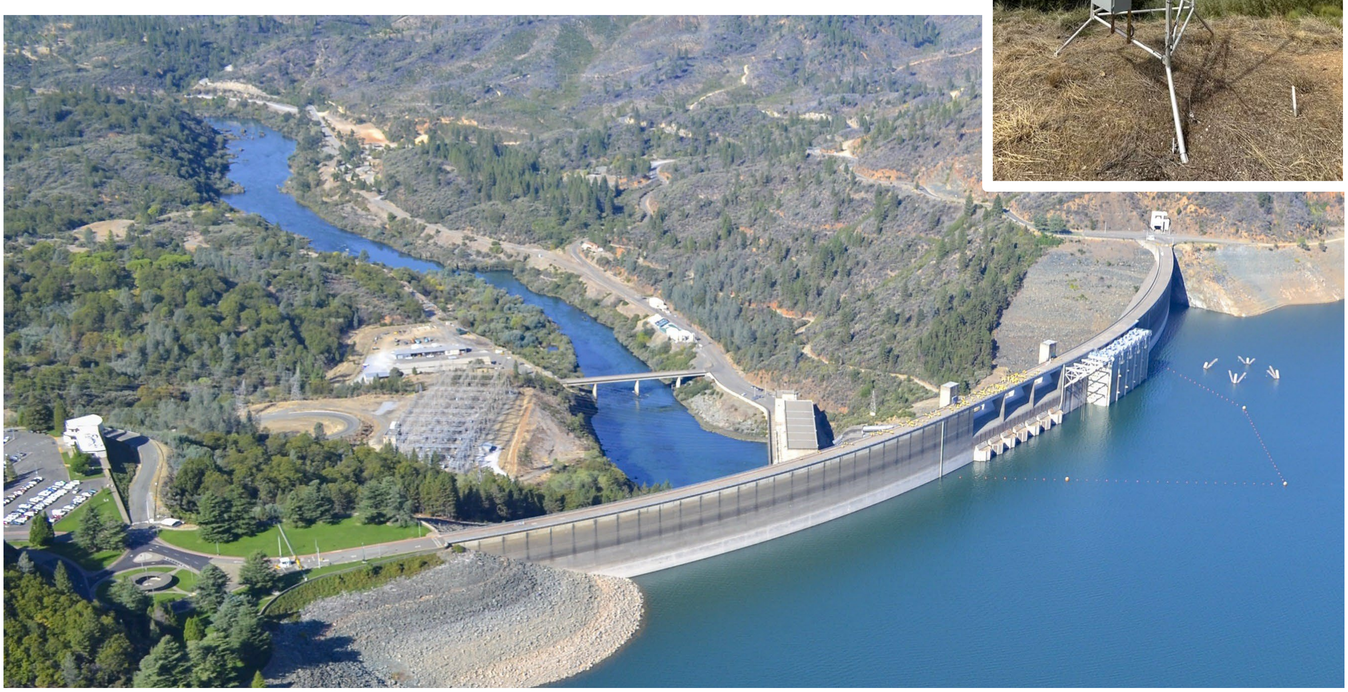
Evaluating Water Temperature Modeling and Prediction in the Sacramento River Basin

Research and Development Office

Science and Technology

Research Program





U.S. Department of the Interior

September 2025

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED *(From - To)*
09-29-2025	Research	January 2022 – September 2025

4. TITLE AND SUBTITLE
Evaluating Water Temperature Modeling and Prediction in the Sacramento River Basin

5a. CONTRACT NUMBER
24XR0680A1/RY15412022WP22050

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
1541 (S&T)

6. AUTHOR(S)
Andy Wood, Project Scientist IV
Randi Field, Hydrologic Engineer

5d. PROJECT ID NUMBER
Final Report ST-2025-22050

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NSF National Center for Atmospheric Research, CGD, Mesa Laboratory, 1850 Table Mesa Dr, Boulder, CO 80305

Bureau of Reclamation
Central Valley Operations Office
3310 El Camino Ave., Suite 300
Sacramento, CA 95821

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Science and Technology Program
Research and Development Office
Bureau of Reclamation
U.S. Department of the Interior
Denver Federal Center
PO Box 25007, Denver, CO 80225-0007

10. SPONSOR/MONITOR'S ACRONYM(S)
Reclamation

11. SPONSOR/MONITOR'S REPORT NUMBER(S) (if applicable)
Final Report ST-2025-22050

12. DISTRIBUTION/AVAILABILITY STATEMENT
Final Report may be downloaded from https://www.usbr.gov/research/projects/index.html

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This project focused on assessing and identifying avenues to improve Reclamation's current temperature modeling use of seasonal predictions of input meteorology, as well as on several related investigations. The effort developed a new overview for Reclamation and stakeholders of the methods and performance of the meteorologic forecasts currently being applied for water temperature simulation. To facilitate this effort, the current spreadsheet-based L3MTO method was duplicated in a set of Python scripts, which were transitioned to Reclamation for potential use in operations at Shasta Lake and other California facilities. These scripts make hindcast analyses feasible and provide flexibility in examining alternative inputs and ease in conducting supporting analyses. The scripts were used to test or demonstrate several variations on the approach, including using S2S forecasts from other sources, and to assess the skill of the approach and its likely upper limits of performance. In particular, the work showed that the temperature model input forecasts had mean monthly skill in the first lead month of the forecast and indicated potential for improvement through further development of the input climate forecasts as well as using alternatives to the climate-conditioned deterministic analog selection method. The project also implemented a California-wide daily gridded meteorological analysis using a tool called GMET, and a demonstration of a linked SUMMA-mizuRoute-RBM model for the Shasta and Trinity Lake drainage areas, demonstrating the feasibility of a distributed strategy for stream temperature simulation and potentially prediction.

15. SUBJECT TERMS
Stream temperature, seasonal climate forecast, hydrologic modeling, Sacramento River

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Randi Field
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U			916-979-2066

Mission Statements

The U.S. Department of the Interior protects and manages the Nation's natural resources and cultural heritage; provides scientific and other information about those resources; honors its trust responsibilities or special commitments to American Indians, Alaska Natives, Native Hawaiians, and affiliated Island Communities.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

Disclaimer

Information in this report may not be used for advertising or promotional purchases. The data and findings should not be construed as an endorsement of any product or firm by the Bureau of Reclamation, Department of Interior, or Federal Government. The products evaluated in the report were evaluated for purposes specific to the Bureau of Reclamation mission. Reclamation gives no warranties or guarantees, expressed or implied, for the products evaluated in this report, including merchantability or fitness for a particular purpose.

Acknowledgements

The Science and Technology Program, Bureau of Reclamation, sponsored this research. We also acknowledge high-performance computing support provided by NCAR's Computational and Information Systems Laboratory, sponsored by the US National Science Foundation.

Cover Image – Composite of two photographs showing an oblique aerial of Shasta Dam and the Sacramento River as it flows downstream of the dam, and a meteorological station located near the left (south) abutment of the dam. (Bureau of Reclamation)

Evaluating Water Temperature Modeling and Prediction in the Sacramento River Basin

Final Report No. ST-2025-22050

Prepared by:

NCAR Climate and Global Dynamics, Terrestrial Science Section, Boulder, CO
Andy Wood, Project Scientist IV

US Bureau of Reclamation, Central Valley Operations Office, Sacramento, CA
Randi Field, Hydrologic Engineer

Peer Review

Bureau of Reclamation
Research and Development Office
Science and Technology
Research Program

Final Report ST-2025-22050

Evaluating Water Temperature Modeling and Prediction in the Sacramento River Basin

RANDI FIELD Digitally signed by RANDI FIELD
Date: 2025.09.30 07:14:14 -07'00'

Prepared by: Randi Field, M.S.
Hydrologic Civil Engineer, Central Valley Operations Office, Bureau of Reclamation

KRISTIN MIKKELSON Digitally signed by KRISTIN MIKKELSON
Date: 2025.09.29 08:00:28 -06'00'

Peer review by: Kristin Mikkelson, Ph.D., P.E.
Hydrologic Civil Engineer, Applied Hydrology I, Bureau of Reclamation

This document has been reviewed under the Research and Development Office Discretionary peer review process, consistent with Reclamation Policy CMP P14. It does not represent and should not be construed to represent the Bureau of Reclamation's determination, concurrence, or policy.

Acronyms and Abbreviations

ACC	Anomaly Correlation Coefficient
CCALM	Climate Conditioned Analog Local Meteorology
CDFW	California Department of Fish and Wildlife
CNRFC	California Nevada River Forecast Center
CPC	Climate Prediction Center
CVO	Central Valley Operations
CVP	Central Valley Project
DTR	Diurnal Temperature Range
DWR	California Department of Water Resources
ECWMF	European Center for Medium Range Weather Forecasting
ENSO	El Nino Southern Oscillation
ESP	Ensemble Streamflow Prediction
FITO	Forecast Informed Temperature Operations
ft	foot/feet
GMET	Gridded Meteorological Ensemble Tool
HEC	Hydrologic Engineering Center
HSS	Heidke Skill Score
HUC4	Hydrologic Unit Code 4
L3MTO	Local Three-Month Temperature Outlook
MTC	Modeling Technical Committee
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NMFS	National Marine Fisheries Service
NMME	National Multi-Model
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
NWS	National Weather Service
OLS	Ordinary Least Squares
PDO	Pacific Decadal Oscillation
RBM	River Basin Model
Reclamation	Bureau of Reclamation
ROD	Record of Decision
S2S	Sub-seasonal to Seasonal
SOI	Southern Oscillation Index
SRTTG	Sacramento River Temperature Task Group
SubX	Sub-seasonal Experiment
SUMMA	Structure for Unifying Multiple Modeling Alternatives
TSC	Technical Service Center
USACE	United States Army Corps of Engineers
USGS	United States Geological Survey
W2	CE-QUAL-W2 Temperature Model
WTMP	Water Temperature Modeling Platform

Contents

Tables

Figures

Executive Summary

Need for research

California water management is going through unprecedented changes while simultaneously responding to more challenging operational conditions. In 2023, the Bureau of Reclamation (Reclamation) re-consulted on Central Valley Project Long-Term Operations which resulted in the issue of a Record of Decision (ROD) and Biological Opinions. These Biological Opinions serve as the basis of project operations and have increased Reclamation's responsibility for modeling and technical analyses to support decision making. Consequently, Reclamation expects to be required to deliver more technical analysis products and defend an increased number of analyses to external entities.

One area of improvement is connected to current practices and Forecast Informed Temperature Operations (FITO) and the need to develop an understanding of the meteorological variability and risk associated with the temperature management plans Reclamation's Central Valley Operations (CVO) office submits annually to regulatory agencies. These plans rely on long-range seasonal temperature forecasts, including meteorologic inputs, to provide inputs to water temperature models used to predict conditions downstream of Reclamation's reservoirs where fishery habitats are protected. Given recent extreme dry and warm year conditions, there is an urgent need for improving methodologies to generate these meteorologic data inputs used in seasonal and real-time decision making. This issue is particularly acute in the Sacramento River but is also a need that extends to other areas of the Central Valley Project (CVP).

Current water temperature modeling leverages meteorologic forcings and forecast input data from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC). The Local Three-Month Temperature Outlook product from the CPC indicates likelihood of above normal, below normal, and near normal air temperatures in future three-month time frames. This data feed is then translated to like historical months in the Shasta watershed, where level of risk is specified by the user. Reclamation is interested in understanding the current state of long-range forecast performance and assessing or improving forecast products and methodologies that support water temperature management.

Without addressing meteorologic forcings as described in this proposal, future legal temperature management challenges are expected, and partners will continue to lack a common understanding of temperature modeling inputs and uncertainty. For example, a practice of using conservative forecasts may result in risk bias. If not assessed, such bias has the potential to yield less efficient annual temperature management forecasts, less efficient use of cold-water pool resources, and increased forecasted temperature dependent mortality of fish. This is due to tradeoffs between early season use of cold water and its value when held in reserve to offset warm periods in the summer management season. There is a need for technical negotiations regarding temperature management to be reinforced by robust scientific support in the form of peer reviewed research, transparent operations, and journal article publication.

Research questions and methods

CVO currently uses HEC-5Q, a conceptual water quality model from the Corps of Engineers Hydrologic Engineering Center, to simulate water temperature in Shasta reservoir and downstream in the Sacramento River up to eight months in the future. To formulate inputs to HEC-5A, Reclamation uses the Local Three-Month Temperature Outlook (L3MTO) procedure developed in the late 2000s, which selects historical meteorological data sets similar (analogs) to the National Weather Service's (NWS) CPC three-month outlook terciles (above normal, normal, and below normal) projections, to use as inputs/forcings. The resulting downstream water temperature predictions are used by CVO to determine future facility/selective withdrawal operations, the ability to meet environmental requirements, and set real-time and seasonal management plans for submission to regulatory agencies and stakeholders.

CVO recently contracted work to improve Reclamation's current temperature model by developing a new framework and focusing on managed river reaches, which also supports analysis of the sensitivity of different parameters and inputs. The work of this project includes one effort that directly connects to the temperature modeling (analysis and upgrading of the meteorological forecasting tool), and several others that are related but not directly connected to the new temperature model/framework development effort – i.e., to investigate new meteorological forcings and uncertainties, new ways of incorporating Sub-seasonal to Seasonal (S2S) climate forecasts, and reservoir inflow temperature forecasts and uncertainty. These efforts collectively investigate alternative datasets (meteorological) and a different hydrology stream-temperature modeling approach.

The first phase of work evaluates and summarizes existing information, establishing a baseline based on the datasets and models used in current operations. This phase (Tasks 1 and 2) includes developing an overview for Reclamation and stakeholders of the skill, performance, and risk of the meteorologic forecasts currently being applied for water temperature simulations. An example of the unknown risk implications in the current approach is the use of compounded exceedance predictions – e.g., using a conservative exceedance value multiple months in a row – which is less likely to occur in reality than the prediction exceedance value used for each month. This practice may lead to releases that are not well balanced with risk/benefit for environmental and other operating objectives.

A second phase of the work (Tasks 3 and 4) investigates alternative methods for potential improvement of meteorological and climate inputs. Gridded Meteorological Ensemble Tool (GMET) surface meteorological analyses could be used to develop alternative and potentially improved inputs to the stream temperature modeling framework. Improved air temperature estimates could also be applied to reservoir and downstream temperatures, as well as to drive reservoir inflow and inflow temperature modeling. We also investigate the potential for climate forecast inputs to be derived from the NOAA SubX and National Multi-Model (NMME) climate prediction datasets.

The third phase of the work (Task 5) explores the potential for reservoir inflow and inflow temperature modeling and prediction using a National Center for Atmospheric Research (NCAR) model called the Structure for Unifying Multiple Modeling Alternatives (SUMMA) watershed model and mizuRoute channel routing model, the meteorological and climate outputs from phase 2, and the River Basin Model (RBM) for stream temperature modeling.

The final phase of the project targets multiple communication pathways to convey the results of the project. Communication is necessary for Reclamation to establish a common understanding of modeling information and uncertainty with partners. In addition, supporting future technical reviews with peer reviewed reports/journal articles will substantially enhance trust and credibility. To this end, the project team engaged in numerous interactions with agency partners in this arena to build trust, relationships, and collaboration agendas.

Conclusions

This project focused on assessing and identifying avenues to improve Reclamation's current temperature modeling use of seasonal predictions of input meteorology, as well as on several related investigations. Supporting Tasks 1 and 2, the effort developed a new overview for Reclamation and stakeholders of the methods and performance of the meteorologic forecasts currently being applied for water temperature simulation. To facilitate this effort, the spreadsheet-based L3MTO method was duplicated in a set of Python notebooks and analysis scripts, and later a command-line Python script, which are being transitioned to Reclamation for potential use in operations. These scripts made hindcast analyses feasible, reproducible, and provide flexibility in examining alternative inputs and ease in conducting supporting analyses. The scripts were used to test or demonstrate several variations on the approach, including using S2S forecasts (for Task 4) from other sources, and to assess the skill of the approach and its likely upper limits of performance. In particular, the work showed the temperature model input forecasts had mean monthly skill at significant levels only in the first month of the forecast, suggesting some room for improvement through further development of both the input climate forecasts, as well as potential replacement of the climate-conditioned deterministic analog selection method (discussed in detail in the report).

The work confirmed using the conservative p=0.25 risk threshold leads to temperature model input forecasts that are systematically biased high. This outcome is by design and means release planning will hedge toward anticipating higher temperatures than will occur on average. Whether this leads to inefficiencies in cold pool storage use depends on multiple factors, including the implications of temperature trends in the region with late summer temperatures particularly increasing.

Other tasks (3 and 5) in the project developed a multi-decadal (1970-2020) high resolution (2 km) ensemble surface meteorological analyses which was created for potential use in developing alternative, distributed inputs to the stream temperature modeling framework. A process-based SUMMA-mizuRoute-RBM model implementation for the drainage areas of Shasta and Trinity reservoirs, including additional downstream drainage area for each, was also implemented. This first linkage of the SUMMA-mizuRoute capability with a temperature model demonstrated the potential for distributed water temperature estimates along an intermediate channel resolution, but the models were not calibrated and validated in this study.

The main outcomes of this project were a detailed quantification of the performance of the water temperature forecasting resources used by CVO and the creation of alternative datasets that could offer improvements over those used in current practice. This information will illuminate the risk targets and tools being used for Shasta downstream temperature management, which are not currently well understood. The documentation will also be useful as a reference point in interactions with other stakeholders and joint management groups. The efforts of this proposal focus mostly on Shasta Dam related temperature modeling, but the resulting refactored meteorological prediction tool is designed to be applicable to other Reclamation reservoirs for which downstream stream temperatures influence release decisions (e.g., selective withdrawal, downstream stream temperatures, and habitat protection measures). CVP facilities at Folsom Dam on the American River and New Melones Dam on the Stanislaus River are examples of additional locations where fishery protection is of interest. The approach is also presented as an example of a design that can be used to foster ongoing development and evaluation, so that the current method can be viewed as a starting point versus a fixed capability to be used for years to come.

1.0 Introduction

1.1. Project background

Management of water resources in California requires modeling and prediction of water resources systems and associated environmental conditions. External partner agencies have increasingly expressed interest in the details of Reclamation's modeling in recent Biological Opinions and other documentation. Improving Reclamation's ability to communicate data, assumptions, data uncertainty and risk in warm, dry conditions through modeling has become increasingly important given growing water demands and recent temperature variability in California. It is imperative that Reclamation provide clear and detailed descriptions and assessments of its modeling capabilities and demonstrate strategic efforts to improve internal modeling processes.

CVO is a partner in the joint management of the Sacramento - San Joaquin River systems, which includes Shasta Lake and Dam. Reclamation operates Shasta Dam for multiple objectives, including regulating water temperature 60 river miles downstream of Keswick Dam. Keswick Dam is the regulating reservoir that impounds water downstream of Shasta Dam. The Temperature Control Device (TCD) structure at Shasta Dam enables control over the depths from which water releases are taken, as deeper waters are typically colder than surface waters. Selective withdrawal release decisions are guided in part by downstream water temperature forecasts (up to eight months lead time) with the goal of meeting required temperature targets downstream for fishery purposes. These FITO consist of reservoir and stream temperature model simulations driven by meteorological sequences drawn from historical observations and conditioned on seasonal climate forecasts from the NOAA CPC. A general map of the temperature management areas is shown in **Figure 1**, highlighting the Sacramento River reach and other areas where CVO provides seasonal FITO information. Greater detail about Reclamation water temperature management is provided in the "Water Temperature Management in Reservoir-River Systems through Selective Withdrawal" Report (Reclamation, 2017). A broad overview of the FITO workflow is displayed in **Figure 2** where key model inputs, such as future forecasted conditions, and results are made available to various temperature model configurations via automated pre and post-processing routines.

The accuracy and reliability of forecasts are critical to calculate water temperature exceedance risk and mitigate mortality of at-risk fish populations downstream of Shasta Dam. Forecasted conditions biased towards lower temperatures can underestimate releases needed to compensate for warmer than expected conditions and risk mismanagement of cold-water pool resources, incurring temperature/fishery impacts. On the other hand, forecasts that overpredict higher temperatures could mean cold-water pool resources would be over-exploited early in the forecast period, risking the limiting of release options for avoiding potentially greater fish egg mortality later in the hotter periods of summer. There is an urgent need for greater clarity and community understanding of the quality and uncertainty of such projections, and the implications of specific choices made in creating the forecasts. Given the importance of stream temperature management in California, passe as well as new steps to begin to improve the models, methods and datasets that Reclamation apply to this management challenge.

Figure 1.—A general study domain figure showing the location of Lake Shasta, the temperature management reach, and the Fairfield location used in the current L3MTO.

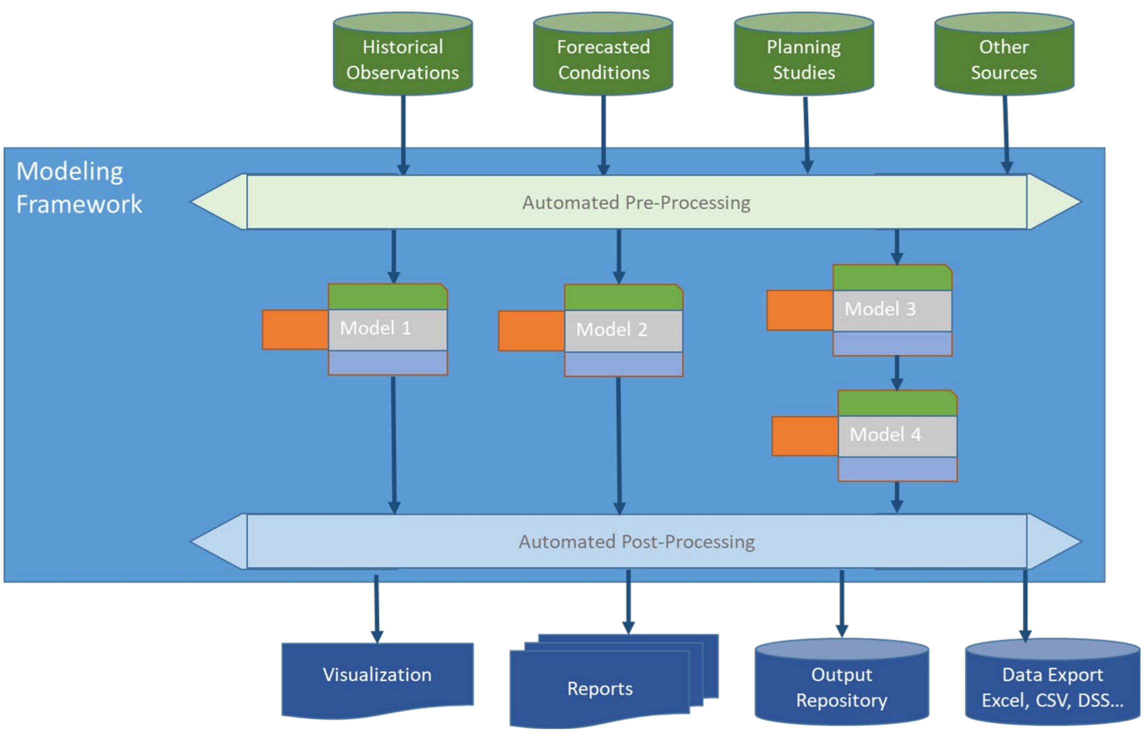


Figure 2.—Overview of a conceptual FITO workflow demonstrating the critical inputs, data adjustments made via automated pre-processing, the flexibility to use multiple water temperature prediction model configurations, and automated post-processing capabilities to improve efficiency and usefulness of water temperature prediction results.

1.2. Study objectives and approach summary

The project consisted of a three-year effort by the Reclamation California - Great Basin Region, CVO, and the NCAR to assess the current meteorological and river flow and temperature datasets and methods used to set operating criteria at Shasta Dam, and to begin investigating avenues to improve them.

The project sought to tackle this goal though assessing the utility of existing meteorological forcing datasets as inputs to water temperature modeling, investigating the potential benefits of using S2S climate predictions for water temperature prediction, and performing exploratory work on Shasta Lake inflow temperature modeling and forecasting. Several tasks were outlined to conduct the work.

- **Tasks 1 and 2** focused on gathering, discussing, and evaluating datasets used in current stream temperature simulation and forecasting.
- **Task 3** involved the application of the GMET (Bunn et al., 2022) to provide an alternative and potentially improved dataset for deriving inputs for the temperature model. The long record of the GMET dataset (from 1970 to present) was intended to shed light on trends and variability in the climate inputs driving the water temperature modeling, provide a forcing dataset for calibrating the hydrology and stream temperature model, and potentially be used as distributed inputs for stream temperature forecast downscaling.
- In **Task 4**, we investigated whether climate forecast inputs from new sources may offer better inputs to the current analog generation approach based on CPC tercile probability forecasts, using the GMET meteorological analyses as a target climatology for validation. The potential sources of new inputs were the NMME (Kirtman et al, 2014) and the NOAA SubX subseasonal forecast dataset (Pegion et al, 2019).
- In **Task 5**, we began implementing the SUMMA and mizuRoute hydrology and channel routing models for major inflow locations to Shasta (such as the Pit, the McCloud, and Sacramento Rivers), and linked the model output to the RBM (Yearsley et al, 2009) model for simulating stream temperature, although this effort did not proceed to a full model calibration.
- **Tasks 0 and 6** focused on project management, documentation of the study, and closeout activities. To this end, the project included periodic but extensive overall interactions and presentations with other groups working on aspects of the stream temperature modeling and forecasting, and multi-agency partners and consultants. The project also formed a steering committee of members drawn from multiple agencies, and conducted semi-annual committee updated review meetings throughout the project.

1.3. Study team and partners

The research team included Randi Field, Mechele Pacheco, Donna Garcia, and Kevin Thielen (and in the first year, Michael Wright) from Reclamation and Andy Wood, Yifan Cheng, and David Yates from NCAR. Yifan departed midway through the project after leaving NCAR for an academic faculty position. The project also coordinated regularly with other groups in California, including the Central Valley Project operations and water temperature modeling development project teams, the Modeling Technical Committee (MTC) and the Sacramento River Temperature Task Group (SRTTG), and members of the Water Temperature Modeling Platform (WTMP) project team. Lastly, the project convened a steering committee, which reviewed project progress several times per year and included members from the California Department of Water Resources (DWR), NOAA California Nevada River Forecast Center (CNRFC), NOAA [CPC and National Marine Fisheries Service (NMFS)], the California Department of Fish and Wildlife (CDFW), and Reclamation.

2.0 Tasks 1 and 2 Approach and Results

2.1. Approach: Evaluating datasets and methods used in current stream temperature simulation and forecasting

Tasks 1 and 2 focused on gathering, discussing, and evaluating datasets and methods used in current stream temperature simulation and forecasting. At the start of the project, the seasonal stream temperature forecasting model (HEC5Q) received input generated by a climate forecast conditioned analog resampling method implemented in an Excel spreadsheet. The Excel spreadsheet is titled the "*Sacramento-Valley L3MTO-based Like Year Selection and Meteorological Sequencing Tool for supporting MP-CVOO Stream Temperature Planning*", and references its two major components, a particular climate forecast product (the L3MTO) and the conditional weather generation approach (the Like Year Selection and Meteorological Sequencing). Because this is a long and specific approach title, we refer to the general analog-based approach here as "climate conditioned analog local meteorology" (CCALM). The CCALM method, developed in the late 2000s by Dr. Levi Brekke of Reclamation (Brekke et al, 2008), used NOAA's L3MTO climate predictions, developed by the NOAA CPC around that time.

Most if not all seasonal forecasting approaches that provide high-frequency, local-scale meteorological input sequences for analysis or modeling have two major components. One is a source of large-scale climate prediction information, and the second is a conditional weather generator, or a means of generating weather-scale sequences that are informed by the climate predictions. In this context, CCALM uses the CPC L3MTO tercile forecasts as the first component (climate prediction), and the conditional analog sequence selection as the second component, weather generation. Each of these components could be done differently, but this approach is not uncommon in operational applications.

The L3MTO tercile-based temperature prediction is a forecast of the probability that the mean temperature at a single meteorological station location for a future season will fall into either above, below, or normal categories (defined by the terciles of the historical observations). The L3MTO forecast product is statistically downscaled to local weather stations from the CPC's official Three-Month Temperature Outlooks, and was created to provide more accurate local representation of the spatially coarse official forecasts. According to Timofeyevna et al. (2024), the L3MTO has evolved from its original form: i.e., in 2021, "*L3MTO was phased out of operations, and users were referred to interactive map displays of CPC 8–14-day, 1-month, and 3-month outlooks, first developed by NWS WFO Pendleton, Oregon. Comparison between these two products is not direct because the interactive display provides spatially interpolated values rather than L3MTO's downscaled information. The interactive map provides users with a high-resolution climate and a point-and-click interface to obtain a pie chart of forecast probabilities for below-, normal-, and above-tercile outcomes that resembled L3MTO output.*" The current Reclamation practice uses this modified form of the L3MTO. In practice, the exact method behind the L3MTO calculation likely does not make a large difference in the quality of the product. For variables such as temperature, departures from normal are regionally widespread (i.e., have a long spatial correlation length), and this is equally if not even more true for forecasted tercile probabilities of occurrence; hence the difference between estimating such terciles directly at a station or extracting them from continuous national field would be difficult to detect. Moreover, it is reasonable to assume over the two decades since the L3MTO was introduced, the quality of the underlying seasonal outlooks has improved, or at least has not degraded, which warranted NOAA superseding the older method.

The L3MTO spreadsheet contains all the datasets required to make a multi-season meteorological input prediction (for the HEC-5Q model) after the L3MTO seasonal tercile predictions are obtained. One copy of spreadsheet is created for each initialization of the forecast, which have a monthly operational update frequency beginning in March of each year. The spreadsheet performs a number of operations, as depicted in the graphic of **Figure 3**. The major calculations include:

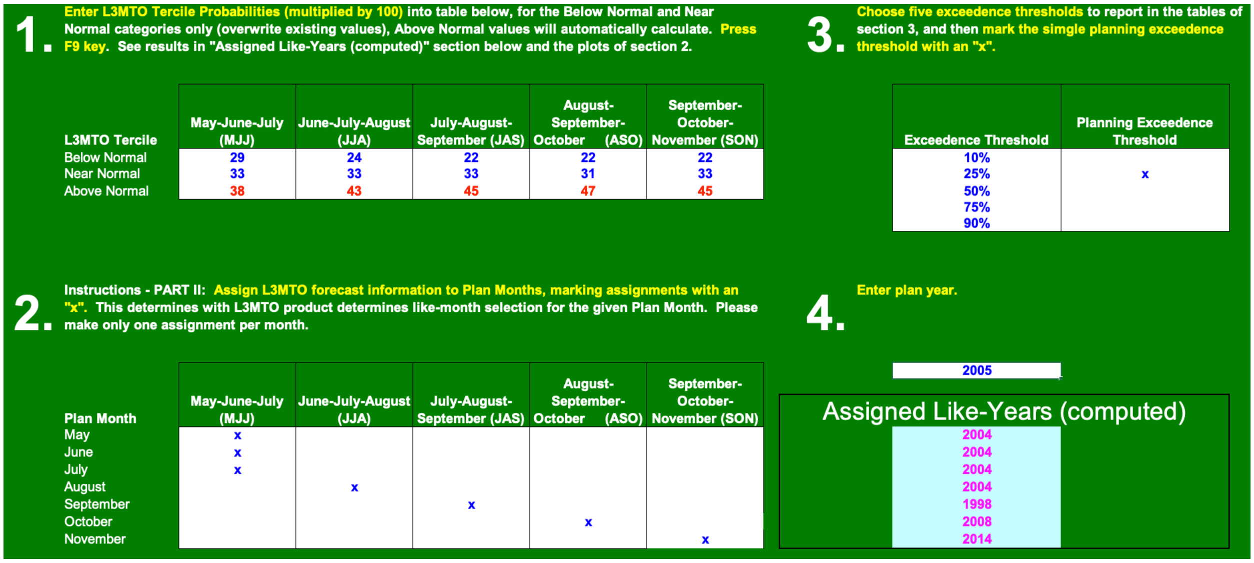
1. Obtaining L3MTO forecasts of tercile probabilities for seasonal temperature (e.g., May-June-July, MJJ) for meteorological station near Shasta Lake. Based on a correlation analysis to find the most skillful local stations, the L3MTO station location selected was Fairfield, CA.
2. The tercile probabilities are used in a weighted resampling (N>1000) of the local station records to create a conditional temperature distribution for each seasonal forecast period based on a larger sample. Note, the seasonal forecast periods overlap (e.g., MJJ, JJA).
3. From the larger sample distribution, a quantile of the forecasted seasonal temperatures is chosen, depending on a specified risk tolerance (exceedance probability). For example, a p=0.25 risk tolerance setting indicates that a higher than median forecast quantile is chosen, reflecting a choice to hedge the forecast toward higher temperatures.
4. The resulting forecasted seasonal mean temperature value (e.g., MJJ) is used to select the year from the historical record at Fairfield with the closest matching seasonal temperature.
5. The selected "analog" year for each season is assigned to one or more months in the forecast horizon (e.g., the analog year for MJJ may be assigned to April, May, June and/or July).
6. For each month, the assigned analog year is used to extract the observed meteorological forcing record used as input to the stream temperature (and reservoir) model. In this case, the forecast record comes from a meteorological station at Gerber Reservoir, one of the closest stations to Shasta Dam, and has a timestep of 6 hours.
7. These monthly time slices of meteorology are concatenated to create the forecast forcings used to run the stream temperature model.

The spreadsheet tab for manually entering tercile forecast probabilities, assigning risk thresholds and season-to-month choices is shown in **Figure 4**. This "analog re-sampling" practice is used elsewhere in Reclamation where high frequency (e.g., daily or sub-daily) inputs to models are required and there is a desire to impose a climate information signal. Wood et al. (2021)

summarizes an application for constructing daily inflow sequences for a reservoir management model, one of several such project applications. It is similar to the NWS's Ensemble Streamflow Prediction (ESP) method, and to the climate informed version originally pioneered for Great Lakes water level forecasting (Croley, 1996).

To facilitate analysis, experimentation, and enable greater flexibility in the application of the L3MTO-based method, this project ported the spreadsheet calculations into a Python notebook, which is an interactive coding approach that streamlines code testing and development. The developmental notebooks were later converted to a Python command line script more suitable for operational use. The notebook was used to assess various questions such as whether the use of CPC-based future weather inputs to the temperature model adds measurable skill in their prediction, and for which lead times the strategy is effective. This work is described in Section 3.1.

Figure 3.—A screenshot of the L3MTO-based seasonal temperature model input forecasting spreadsheet.


1. Enter L3MTO Tercile Probabilities (multiplied by 100) into table below, for the Below Normal and Near Normal categories only (overwrite existing values), Above Normal values will automatically calculate. Press F9 key. See results in "Assigned Like-Years (computed)" section below and the plots of section 2.

L3MTO Tercile	May-June-July (MJJ)	June-July-August (JJA)	July-August-September (JAS)	August-September-October (ASO)	September-October-November (SON)
Below Normal	29	24	22	22	22
Near Normal	33	33	33	31	33
Above Normal	38	43	45	47	45

3. Choose five exceedence thresholds to report in the tables of section 3, and then mark the simgle planning exceedence threshold with an "x".

Exceedence Threshold	Planning Exceedence Threshold
10%	
25%	x
50%	
75%	
90%	

2. Instructions - PART II: Assign L3MTO forecast information to Plan Months, marking assignments with an "x". This determines with L3MTO product determines like-month selection for the given Plan Month. Please make only one assignment per month.

Plan Month	May-June-July (MJJ)	June-July-August (JJA)	July-August-September (JAS)	August-September-October (ASO)	September-October-November (SON)
May	x				
June	x				
July	x				
August		x			
September			x		
October				x	
November					x

4. Enter plan year.

2005

Assigned Like-Years (computed)
2004
2004
2004
2004
1998
2008
2014

Figure 4.—A screenshot of a control tab in the L3MTO-based seasonal temperature model input forecasting spreadsheet.

2.2. Results: Evaluating datasets and methods used in current stream temperature simulation and forecasting

2.2.1. Conversion of Excel Spreadsheets to Python Notebooks and Scripts

As discussed in Section 2.1, the operational spreadsheets were analyzed and converted into a Python format. A small collection of the spreadsheets (one per forecast date or "planning month", from years including 2017 and 2021) was shared with NCAR, which first reproduced the operations in a notebook form (facilitated by NCAR jupyterhub development environments). All data tables stored in the spreadsheet were converted into CSV format text files so the Python version could read these from disk, based on file paths specified in a configuration section, rather than having the datasets themselves hardwired into the method – that is, each Excel spreadsheet used for a monthly Planning update contained multiple tabs storing the local observational datasets (and various transformations of them, such as seasonal averages) used in the method. This arrangement means substituting other datasets would be a manual task. The notebook also allows for identification and potential generation of multiple analog selections, which could produce an ensemble of temperature model input predictions.

The notebook (named **apply_l3mto.hec5q.ipynb**) enabled additional visualizations beyond those present in the original spreadsheet, which helps to validate output. An exact bit-for-bit and value-for-value comparison was not possible due differences in the implementation of functionality (such as random sampling) by Excel and Python. Instead, the sampling operations (values sampled according to specific temperatures and years) and sequencing of selected analog flows were checked during notebook development to ensure consistency and validity of the code functionality. An example of the new approach configuration section of the first notebook is shown in **Figure 5**.


1. Settings for L3MTO-based process

```
[2]: fcstYear        = 2003
     fcstMon         = 4
     fcstLen         = 8
     seas2monNdx     = [1, 1, 2, 3, 4, 5, 6, 7]                              # index of forecast seasons that are assigned to each plan month (lead)
                                                                             # i.e., if fcstMon=4, the first element = 1 will assign AMJ to lead 1 (Apr)
     planningThresh  = [0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]      # percentile of the forecast distribution used to select closest year
     climoYears      = [1971,2000]                                           # climatology years for calculating tercile thresholds
     sampleYears     = [1961,2005]                                           # sampling period years from historical record
     nSample         = 1000                                                  # size of sample to use in estimating quantiles
     nAnalog         = 5

     tercFcstFile    = '../cpc_fcst/sflist.Apr2003.cpc_terc_fcst.74516.csv'   # CSV file containing tercile forecasts
     outMetFile      = 'Gerber_met.cpc_terc.thr-0.25.apr2003.alt-seq.csv'     # output file
```

Figure 5.—Screen shot illustration of L3MTO forecast python script settings that were ported from the original Excel spreadsheet version of the approach.

During development, examples of the notebook's method of validation are given in **Figures 5-7**. In **Figure 6**, histograms of historical climatology (climo) and conditionally-resampled historical seasonal mean temperature values are compared (taken from the Fairfield local observation site that is currently used in the CCALM approach for Shasta Lake), showing the shifted distribution mean. The Fairfield location was selected during the original method development based on having the highest skill of the original L3MTO locations in predicting Shasta Reservoir area seasonal temperatures. **Figure 7** (left chart) shows how the 1000-member sample (red) extracts values from the historical climatology, leading to the distribution functions in the right subfigure, which also shows the distribution shift resulting from conditional resampling, and the application of a risk threshold of exceedance probability p=0.25 applied to select the analog target temperature for this particular season. Such visualizations help establish that the notebook code is producing the intended calculations and also illustrate some of the particular behavior of the method. For example, though conditional resampling of climatologies is common, it is not typically applied to small-sample (e.g., less than 50) climatological datasets (e.g., historical means for a given season at a given location), thus the resulting distributions can be strongly influenced by individual values.

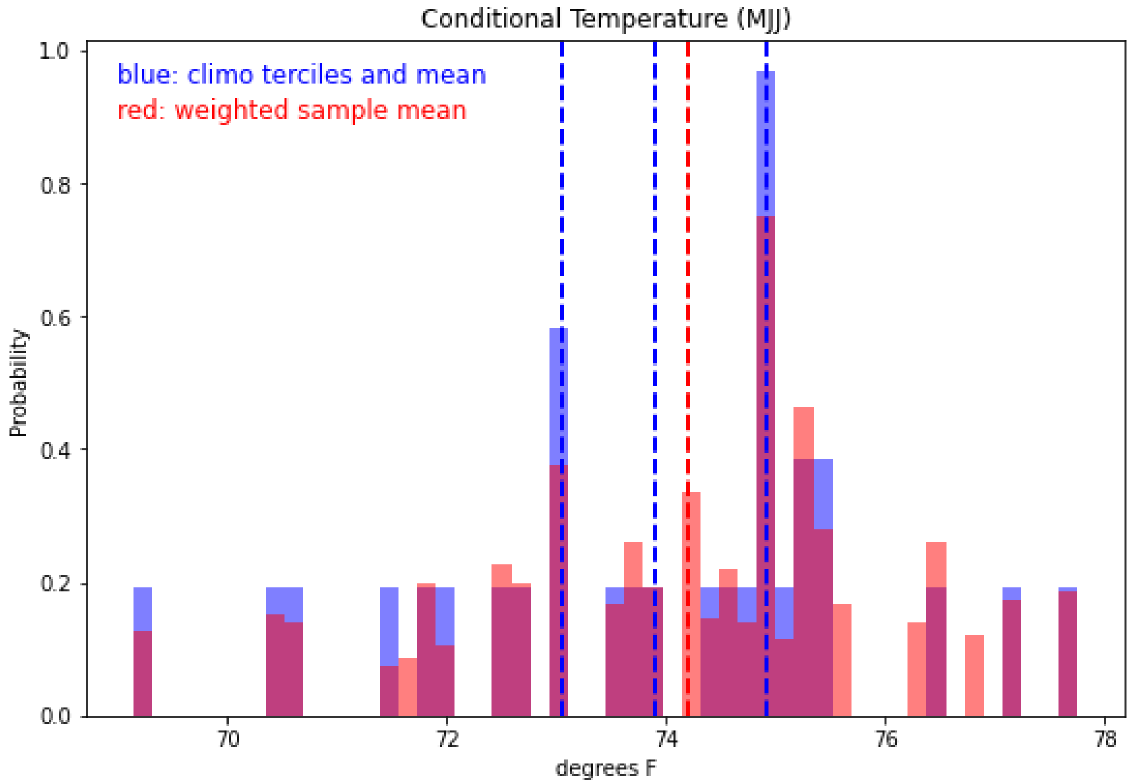


Figure 6.—Visualization of the Python script histograms of a seasonal temperature forecast, comparing historical (climo) values (blue) with the conditional sample values (red) during May, June, and July (MJJ) from 1980 – 2020.

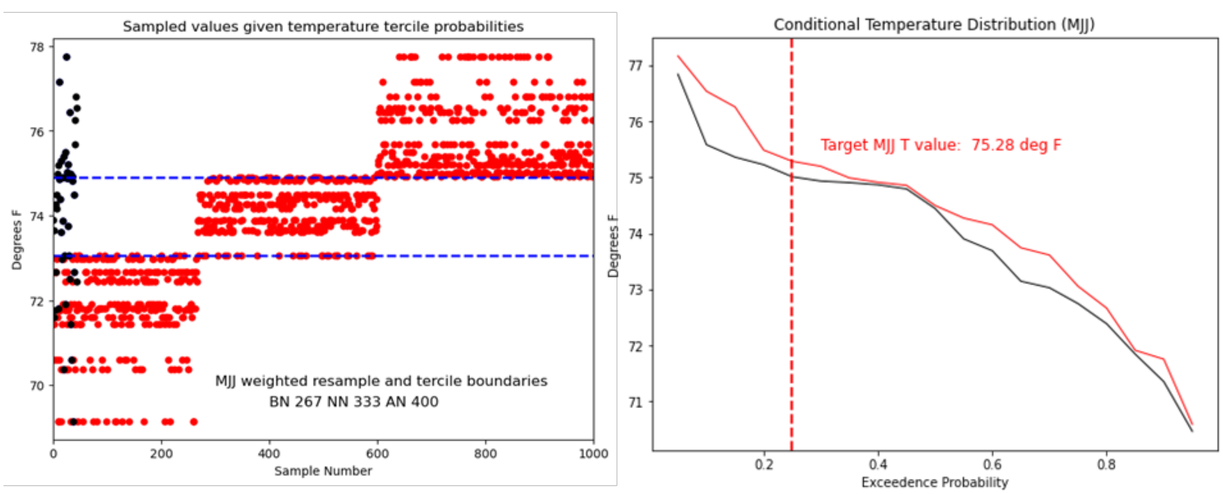


Figure 7.—Visualization of the Python-based conditional sampling of a seasonal temperature forecast, showing (left) each of 1000 selected sample values (red) from the historical observations (black), and historical tercile boundaries shown in blue; and (right) as a conditional distribution function (red) relative to the observed distribution (black), with a target value selected by a threshold of p=0.25 exceedance probability.

The notebook also produces a visualization (plot) of the sequenced meteorological output (i.e., the input to the temperature model). **Figure 8** shows an example of this output for the HEC5Q model, which helps confirm that it is a plausible sequence for the forecasted time period.

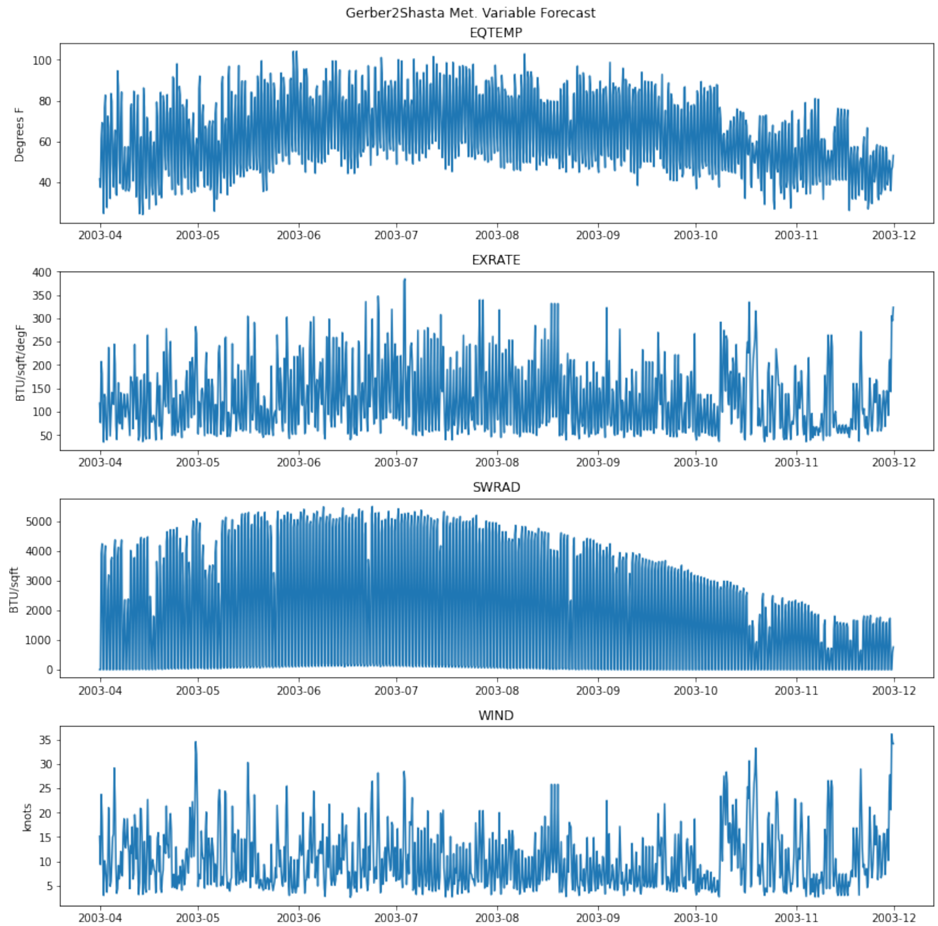


Figure 8.—Plot of the seasonal forecast of meteorological inputs to the HEC5Q model (example from the initial Python script). The variables (top to bottom) are air temperature, exchange rate, downward solar radiation, and wind speed. The values are resampled from local in situ meteorological station data records with the exception of the exchange rate, which is derived by Reclamation.

2.2.2. Conversion of Python L3MTO approach to support CE-Qual-W2 modeling

After the initial work to emulate the spreadsheet L3MTO approach in Python and create a flexible tool for analysis, the larger water temperature modeling development focus shifted toward supporting CE-Qual-W2 water quality models (herein also referred to as W2 for short); https://www.erdc.usace.army.mil/Media/Fact-Sheets/Fact-Sheet-Article-View/Article/554171/ce-qual-w2/ and required a conversion of the notebook and subsequent scripting and analysis, leading to an updated version (called **apply_l3mto.conv_w2.ipynb**). This script was more streamlined and configured to work with the larger (and hourly) input variable list required by the W2 model relative to HEC5Q, using observed values from a model local input file for the Redding, CA CIMIS meteorological station (Station ID KRDD). Other aspects of the L3MTO procedure remained the same, and some of the analytical plots included in the notebook during development were removed. An example forecast from the new notebook for the W2 model inputs is shown in **Figure 9**. Note that the last field (surface pressure) is a constant in the input files and is a non-sensitive parameter to the water temperature model.

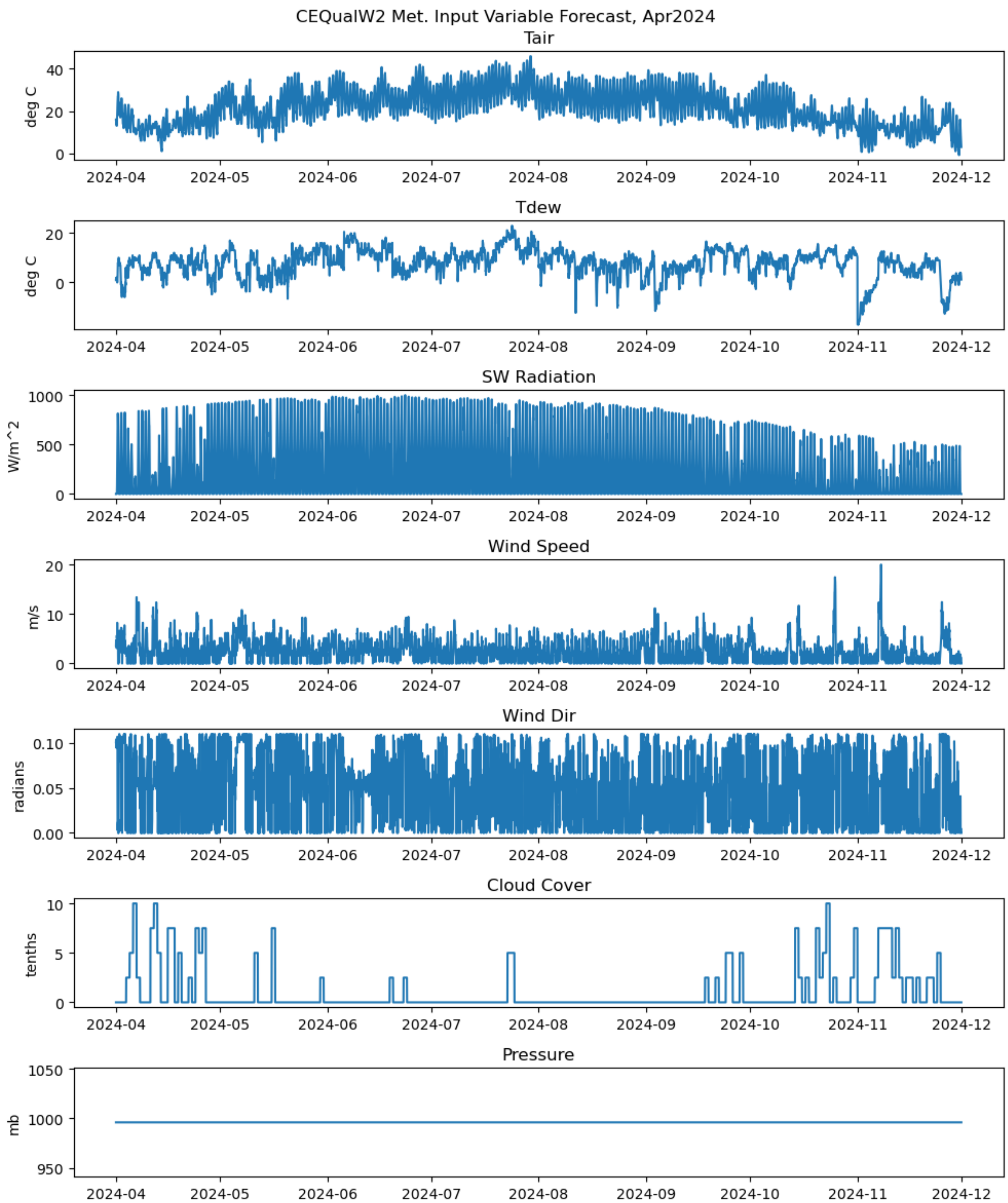


Figure 9.—Plot of an example forecast produced by the Python L3MTO forecast script for CEQualW2.

2.2.3. Assessment of forecast skill using CPC Hindcasts

As the Python-based notebooks were being developed, a number of exploratory analyses were conducted related to the L3MTO-process. Tercile climate hindcasts from the year 2000 to the present (i.e., 2024) were downloaded and processed for use by the new L3MTO scripts, focusing on several common textbook statistics used as skill metrics (monthly mean correlation, mean absolute error and bias). Using the default method choices from the current L3MTO spreadsheets (e.g., season to month assignments and risk thresholds), the 20+ year hindcasts indicated the approach brings a moderate correlation skill for monthly mean air temperature in the first lead month, but not at longer lead times. An April planning month hindcast analysis for the first month lead time (April) is shown in **Figure 10**.

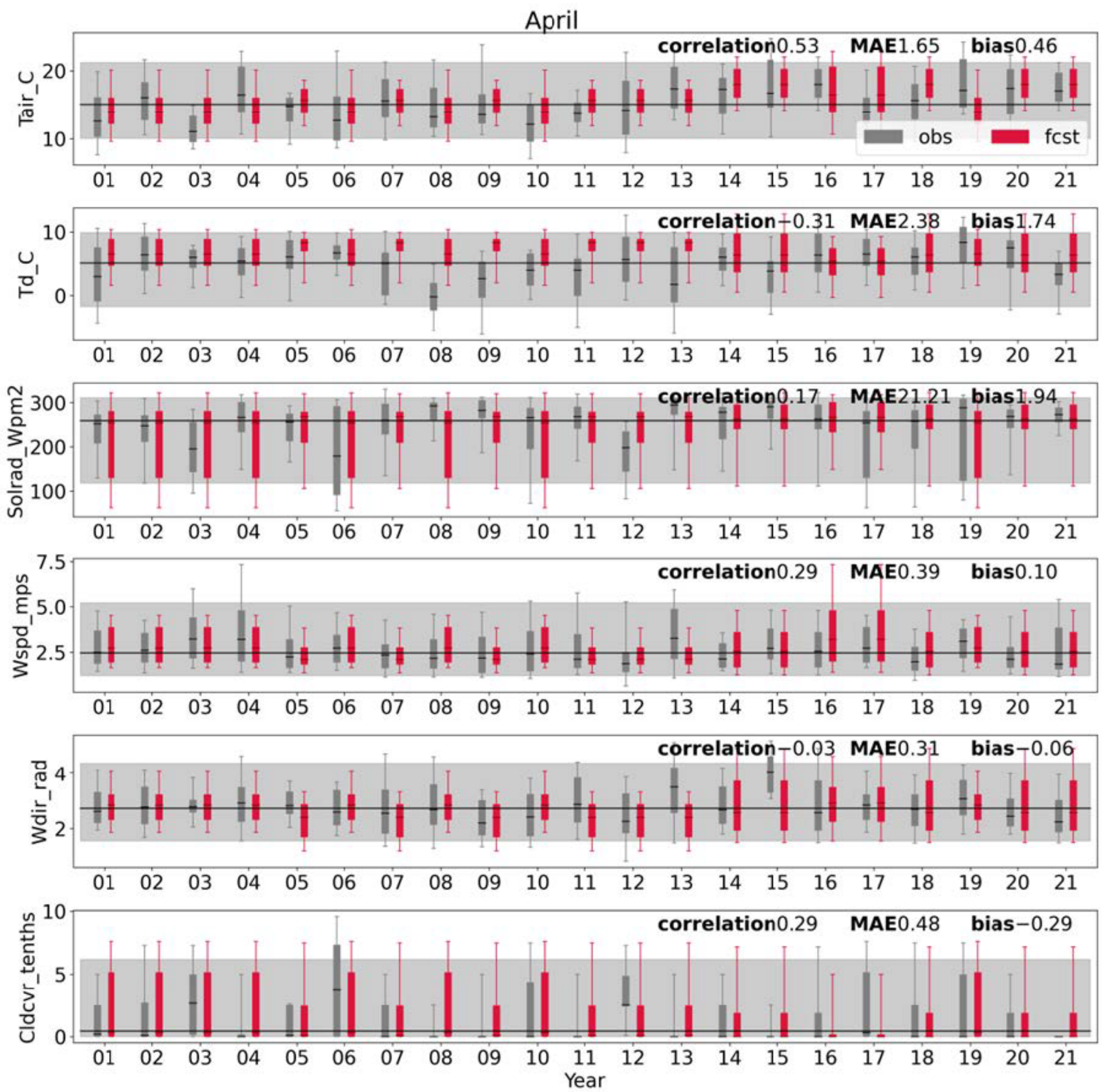


Figure 10.—April planning month hindcasts from 2001-21 for the CCALM process, for the April forecast month. Each standard box and whisker symbol shows the distribution of daily values in each month.

In general, for samples of this size range, a correlation below an absolute value of 0.3 indicates a non-significant relationship. Here we expect to see a significant correlation and low error and bias in the temperature-related variables, given the conditioning climate inputs for the forecasts is temperature. Correlation ranges from -1.0 for a perfect inverse prediction to 1.0 for a perfect prediction. The units of error and bias are the same as the units of each variable. The equations for correlation, MAE and Bias are as follows.

Correlation (Pearson's r) = $$\frac{\sum_{i=1}^{n}(f_i-\bar{f})(o_i-\bar{o})}{\sqrt{\sum_{i=1}^{n}(f_i-\bar{f})^2}\sqrt{\sum_{i=1}^{n}(o_i-\bar{o})^2}}$$

Mean Absolute Error (MAE) = $$\frac{1}{n}\sum_{i=1}^{n}|f_i - o_i|$$

Bias = $$\frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)$$

Where:

- f_i and o_i are individual forecast and observed values
- $\bar{f}$and $\bar{o}$ are the means of forecasts and observations, respectively
- n is the number of data points

By the 2nd and 3rd lead months in the April planning forecasts (May and June, shown in **Figures 11-12**), the forecast correlation skill (shown in the top right of each subfigure) drops to statistically insignificant levels. This is even true for air temperature, which had the highest correlation between hindcasts and forecasts during the month of April. To some extent, this is not surprising, given that the skill of seasonal-scale air temperature forecasts (and associated tercile probabilities) is known to be relatively low (e.g., Baxter et al., 2016). A discussion of the myriad reasons for the low skill of seasonal climate forecasts at regional scales is beyond the scope of the report, but the published literature of the last 30 years offers perspective on this question, with one seminar review paper being Goddard et al. (2001). In general, the skill of temperature forecasts (those from CPC included) are known to benefit from a general warming trend in many regions, which is captured in the forecast system. Limits in skill result from unpredictable variations around this trend. Indeed, the temperature variables in Figures 10-13 show visual signs of a trend in both observations and forecasts, with 4-5 warmer than average years clustered at the end of the record, though the 21-year period is too short to estimate a trend with statistical confidence.

The approach for translating tercile forecasts to a single sequence for use in modeling also adds uncertainty (randomness, noise) to the process. A recommended practice in seasonal forecasting is to use ensemble predictions, which provide many future sequences to characterize the uncertainty, but in many water applications, it is not yet technically feasible to run more than one or just a few sequences through the subsequent modeling and analysis tools, and the decision processes are not designed to accommodate ensemble predictions. This is an area for future development.

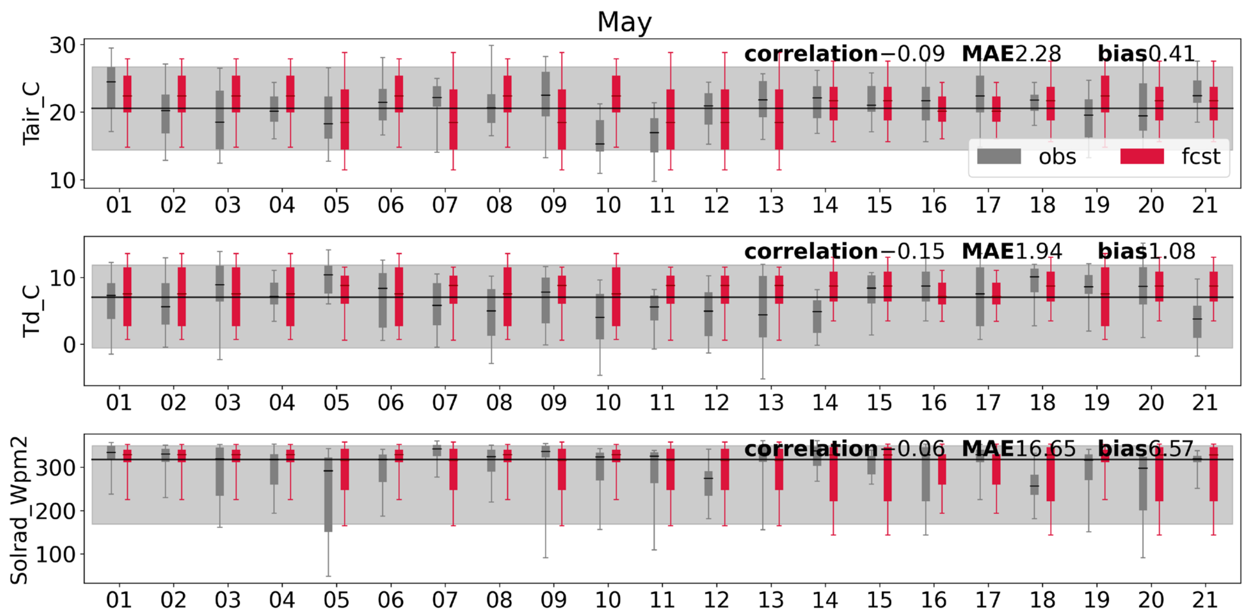


Figure 11.—April planning month hindcasts from 2001-2021 for the L3MTO process, for the May forecast months (lead month 1), for 3 W2 variables associated with temperature. Each standard box and whisker symbol shows the distribution of daily values in each month.

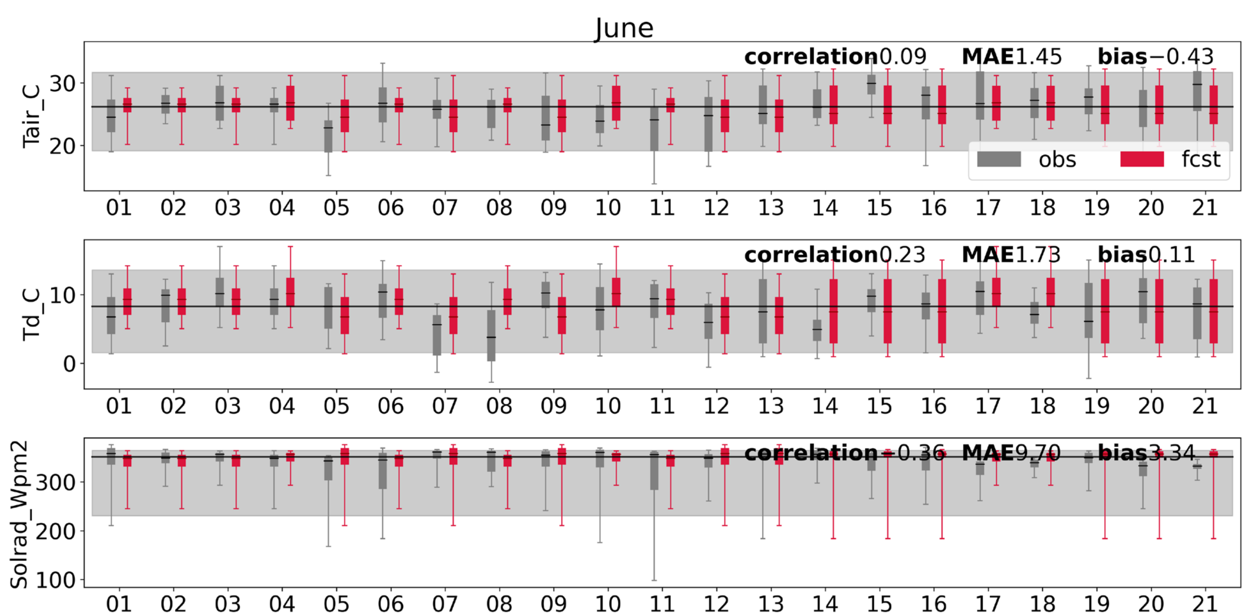


Figure 12.—April planning month hindcasts from 2001-2021 for the L3MTO process, for the June forecast months (lead month 2), for 3 W2 variables associated with temperature. Each standard box and whisker symbol shows the distribution of daily values in each month.

To bracket the level of uncertainty added by the L3MTO conditional resampling process, we created a set of "perfect tercile hindcasts", in which the actual observed tercile in each year was assigned a 100% probability of occurring. We ran these forecasts through the L3MTO (Python) notebooks for the same hindcast period to quantify the upper limit of skill that could be achieved. Narrowing the focus to the temperature forecast input variable (which is most directly connected to the climate temperature forecast), the analysis showed that a moderate correlation (mid 0.40 decile) could be obtained through each lead month of the forecast (the first 3 months are shown in **Figure 13**). This result means that roughly 80% of the variance in future stream temperature model temperature inputs is not explained even when a perfect tercile climate forecast is used.

Another notable indication from the analysis is that when using the conservative p=0.25 risk threshold, the perfect forecasts do produce temperature model input forecasts that appear to be systematically biased high (between 0.37 and 1.65 degrees Celsius in this example, as reflected in the inset statistics). For context, the calibration target of most water temperature models is less than 1 degree Celsius, so a bias of 1.65 degrees Celsius is not unreasonable. This outcome is not surprising (it is by design) and means that the release planning will hedge toward anticipating higher temperatures than will occur on average.

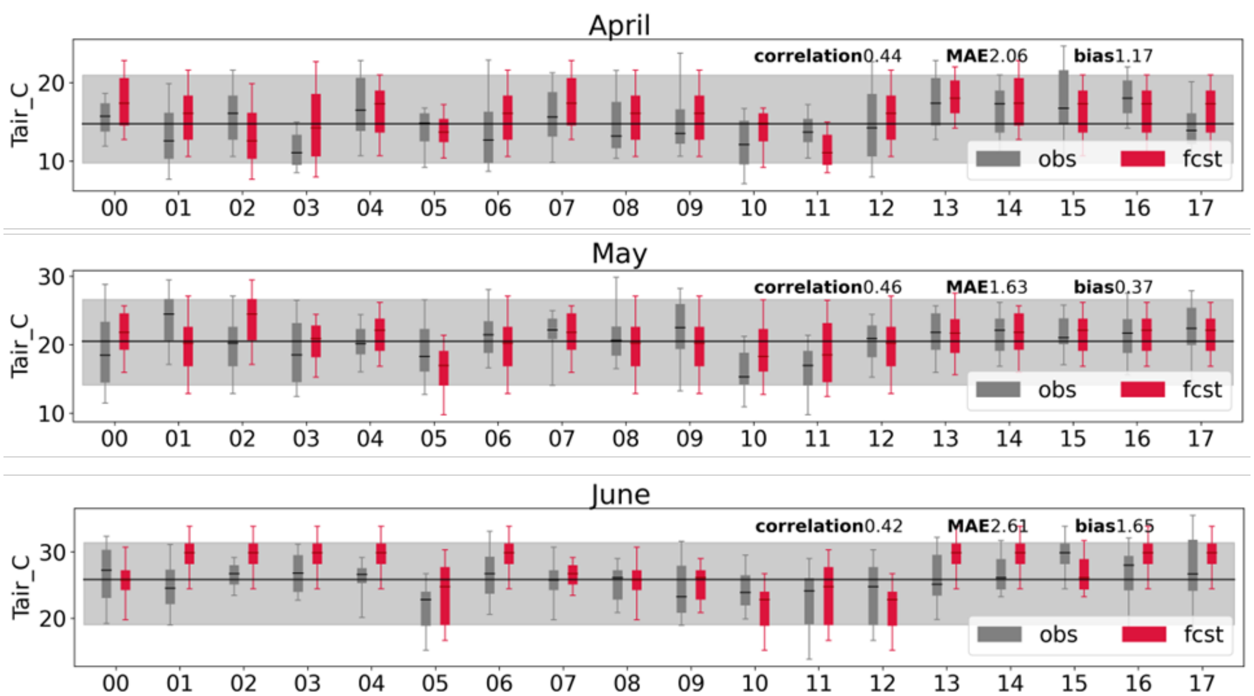


Figure 13.—April planning month hindcasts from 2001-2021 for the L3MTO process, but using "perfect forecasts" for the May and June forecast months (lead months 1 and 2) and for the air temperature variable.

These analyses were undertaken midway through the project, and motivated investing project time into assessing the climate forecast skill of other sources of operational predictions, as summarized in Section 3.3.

2.2.4. Demonstration of International Research Institute based forecasts and other analyses

One of the motivations for moving from a manual Excel based framework for forecasting to a Python scriptable and automatable format was to facilitate experimentation with different choices in applying the L3MTO approach. Although the project did not end up taking a deep dive into this analysis, a few examples of the types of investigation that are now facilitated were explored. One was visualizing the impact that the choice of risk threshold would have on the forecast (undertaken with the HEC5Q version of the approach). **Figure 14** shows a comparison of the forecast outcome from using a p=0.10 (red) versus p=0.90 (blue) exceedance threshold, indicating a slight tendency for the 0.10 choice to lead to higher temperatures and solar radiation.

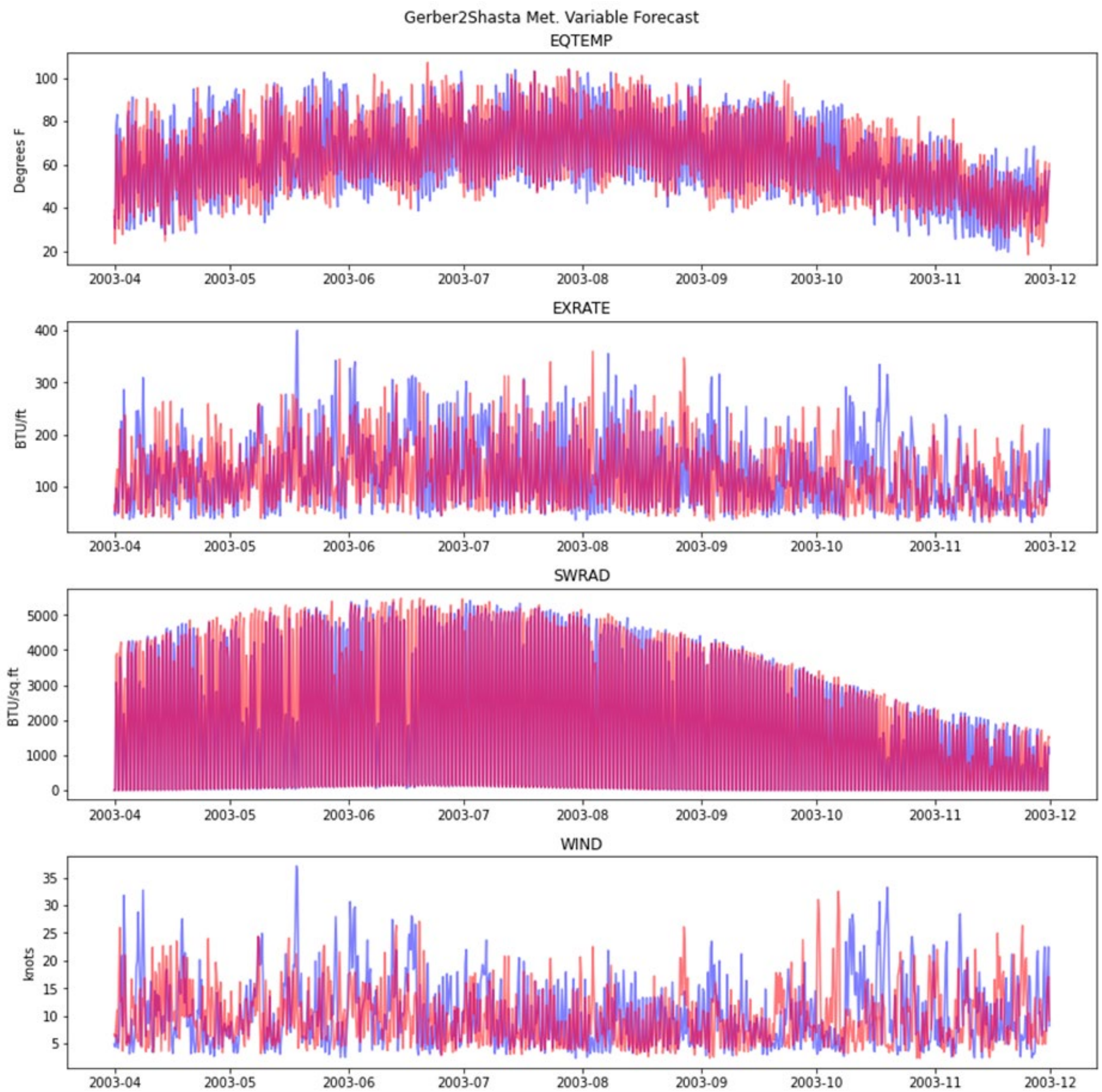


Figure 14.—Comparison of HEC5Q input meteorology forecasts using two different temperature threshold conditions: p=0.10 (red) versus p=0.90 (blue).

Another example of an analysis that was easily performed by the new tool was swapping CPC tercile forecasts for similar seasonal forecasts from the International Research Institute (IRI) https://iri.columbia.edu/our-expertise/climate/forecasts/seasonal-climate-forecasts/, which has long been affiliated with NOAA but makes forecasts with a more global scope. In the example shown in **Figure 15**, the IRI forecasts (in blue) selected for northern California region were generally warmer in the initial months than the CPC forecasts, which is in accordance with their tendencies over the first 4 months in Table 1, which shows Above Normal (AN), Near Normal (NN) and Below Normal (BN) probabilities over the first 4 seasons, after which the forecasts were near climatology for both centers. Clearly the nature of the input climate forecast does translate into the meteorological temperature model inputs to some degree, thus if IRI forecasts were found to be more skillful and available at required operational latencies, they would warrant attention. We note that this example is for demonstration purposes only.

Table 1.—Example comparison of CPC and IRI forecasts for March 2003 seasonal tercile predictions

Lead Time	CPC BN	CPC NN	CPC AN	IRI BN	IRI NN	IRI AN
Season 1	27.77	33.33	38.89	22.0	33.0	45.0
Season 2	26.74	33.33	39.92	24.5	33.0	42.5
Season 3	28.35	33.33	38.31	24.5	33.0	42.5
Season 4	30.00	33.33	36.66	24.5	33.0	42.5

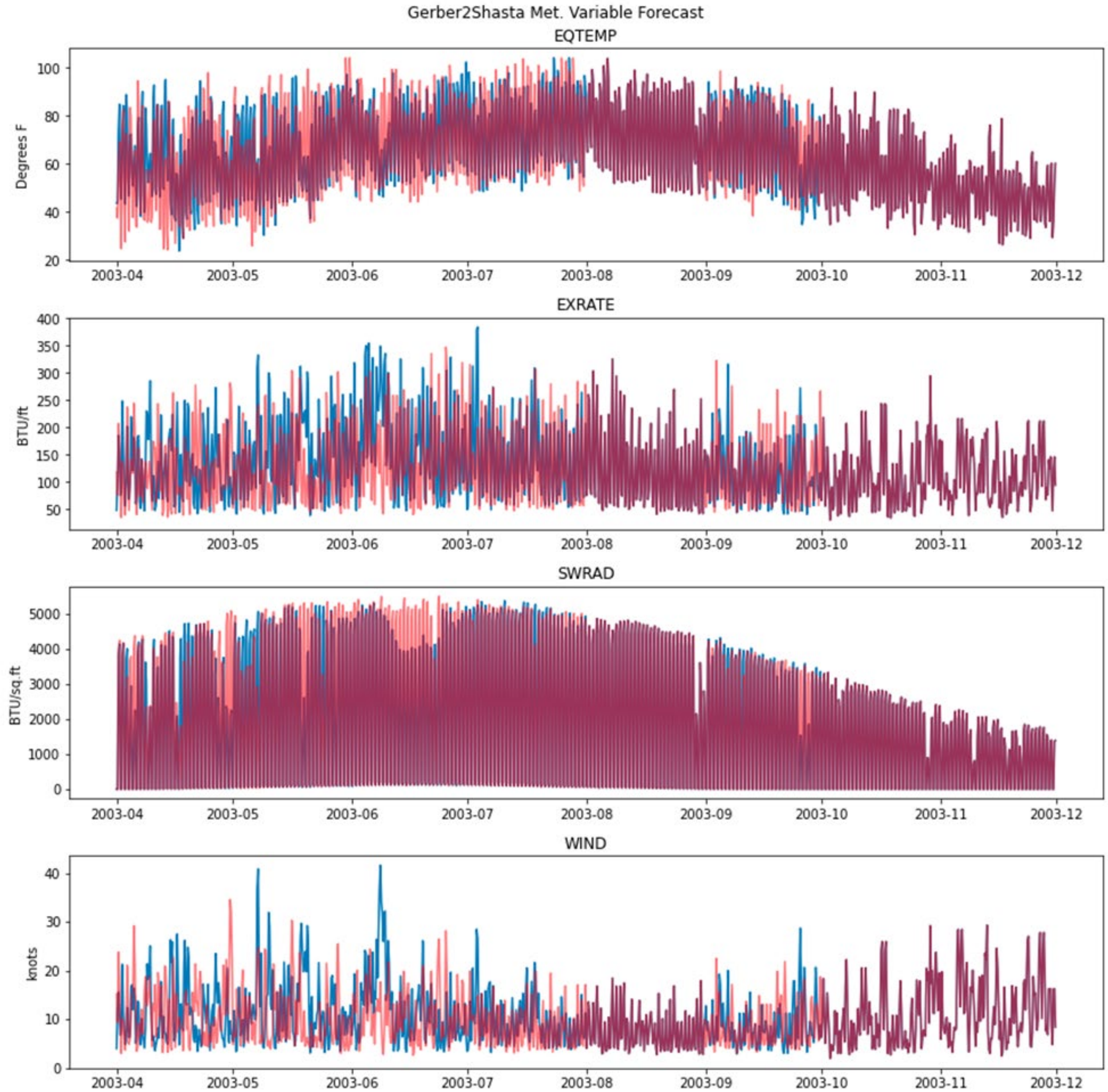


Figure 15.—Comparison between forecasts produced using IRI tercile seasonal temperature predictions (blue) versus CPC predictions (red).

A number of analysis scripts were written to visually compare the observed and forecasted meteorological timeseries produced by the new Python-based L3MTO script, which is helpful to augment the statistical analyses illustrated in **Figures 10-13.** An example output of the plotting script is shown in **Figure 16**, which includes all non-constant W2 model input variables, for an April L3MTO forecast initialization.

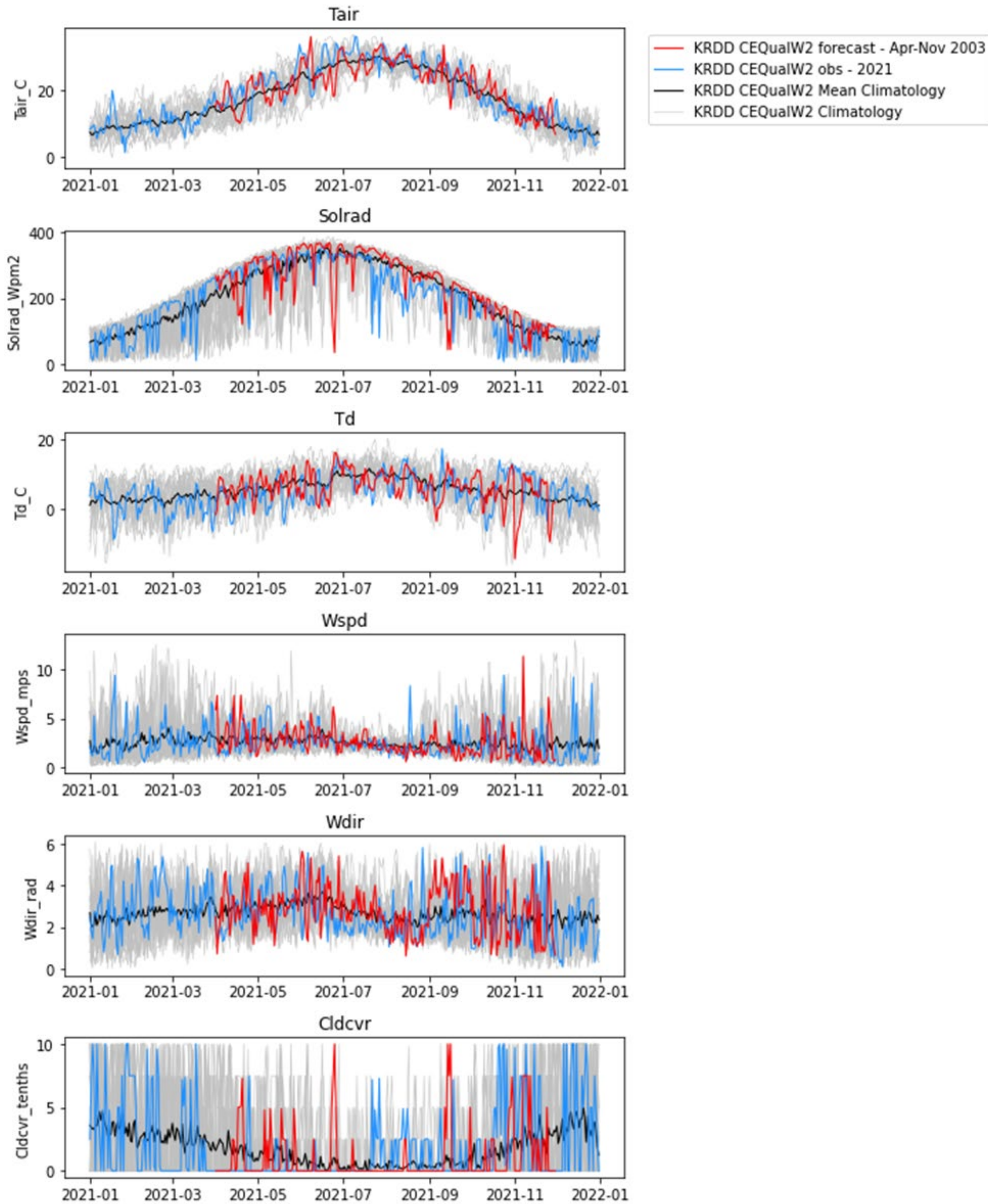


Figure 16.—Visualization of a Python L3MTO script generated forecast versus observed meteorology.

2.2.5. Development of command-line script multi-site version and datasets

The Python (currently version 3.10.13) notebooks were converted into a script version that was tested at the command line to provide input at multiple sites. The main script (named **apply_l3mto.cequalw2.v2.py**) takes two input files in a machine and human-readable text format called TOML (https://toml.io/en/). The first is the main configuration file (**Figure 17**), which currently contains a range of settings that were previously in the notebook, as well as input and output file paths and filename templates. The templates are patterns that are filled in as the script loops over multiple sites, which are specified in a second input file (**Figure 18**), the name of which is indicated in the configuration file as the "siteInfoFile". This file provides information about the locations that the python script will loop over in the form of arrays.

The current python (version 3 or higher) environment to run this script includes the following module dependencies: **datetime, sys, calendar, numpy, pandas, matplotlib, toml, and hecdss**. The latter is used in the notebook version and will be incorporated into the script after further refinement and specification.

The command line script has been tested in a unix environment, where it is run as follows:

> python apply_l3mto.cequalw2.v2.py

The script currently outputs meteorological forecasts for the sites listed in the siteInfoFile, in CSV format, although DSS formatting is also being added to the script. In addition, it outputs a daily timestep version of the forecast and the retrospective climatology, and timeseries plots of both daily and model timestep (i.e., hourly) variables.

The directory structure for the script, inputs, and outputs is user-specified (in the configuration file). Currently, the directory structure used for development is as follows:

- SacMet/ — top level, main Github repository directory
    - scripts/ — command line script and configuration (toml) files
    - notebooks/ — development notebooks and miscellaneous analysis
    - inputs/
        - clim_fcst/ — climate forecast input files
        - local_met/ — local meteorology files
    - output/
        - met_fcst/ — meteorological forecasts
        - ancillary/ — supporting output data files and plots

All scripts and small data files are being maintained in a repository located at https://data.usbr.gov/catalog/8140. This repository (a screen shot is shown in **Figure 19**) is currently private for team members, but can be made public at a later date and also forked to a new hosting domain.

```
#### Configuration settings for running the L3MTO process to generate seasonal temperature forecast inputs to the CEQUALW2 model
#    that is currently part of the WTMP system in Calfornia.  The L3MTO process was developed by Levi Brekke in 2009.
#    Version: 2.0
#    Author:  A Wood, NCAR, Oct 2024

# ---- input settings
fcstYear       = 2024                                          # year of the forecast
fcstMon        = 4                                             # month of the forecast start
fcstLen        = 8                                             # length in months of the forecast

seas2monNdx    = [1, 1, 2, 3, 4, 5, 6, 7]                      # index of forecast seasons that are assigned to
                                                               #   each plan month (lead)
                                                               #   i.e., if fcstMon=4, the 1st element = 1 will
                                                               #   assign AMJ to lead 1 (Apr)
planningThresh = [0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25]  # percentile of the forecast distribution used to
                                                               #   select closest analog year

climoYears     = [1981,2017]                                   # climatology years for calculating tercile thresholds
sampleYears    = [2000,2017]                                   # sampling period yrs from hist. record -> used to
                                                               #   select analog local sequences
#sampleYears    = [1961,2005]                                  # sampling period years from historical record

nSample        = 1000                                          # N sample used in estimating conditional forecast distribution
nAnalog        = 5                                             # number of closest analogs years to output

fixRandomNumbers = true                                        # 'true': generate the same random starting seed
                                                               #   repeatedly so that the results of repeated runs
                                                               #   do not change (useful in testing/debugging)

# ---- forecast site information (arrays of variables describing different script output sites)
siteInfoFile         = './site_metadata.l3mto_cequalw2.toml'

# ---- input local target obs Tair file for estimating conditional seasonal analogs (matches site of CPC downscaled input)
inputLocalMetDir      = '../inputs/local_met/'                 # location of local temp model inputs used in resampling
inputClimFcstDir      = '../inputs/clim_fcst/cpc/'             # location of tercile based climate forecast files

# ---- local obs ceQualW2 input file that are resampled to make the forecast
localRetroMetFileTPL  = 'cequalw2.station_obs.{localMetStnID}.csv'

# ---- output file information
outputModelMetDir         = '../output/met_fcst/'
outputInfoDir             = '../output/ancillary/'
tercFcstFileTPL           = 'cpc_tercile_fcst.{year}{mo}.{clim_loc_id}.csv'       # template of CSV filename for tercile forecasts
outputMetFileTPL          = 'met.w2.cpc_terc.{year}{mo}.{localMetStnID}.csv'      # output file forecast file in timesteps of the
                                                                                  #   temp. model inputs
outputMetFileDlyTPL       = 'met.w2.cpc_terc.{year}{mo}.{localMetStnID}.dly.csv'  # output forecast file (daily)
outputRetroMetFileDlyTPL = 'retro_met.w2.{localMetStnID}.csv'                    # output daily aggregation of historical model met.

# ---- END ----
```

Figure 17.—L3MTO-process Python script run-time configuration file.

```
#### Local water temperature model site metadata used in the L3MTO process to generate seasonal temperature forecast
#    inputs to the CEQUALW2 model that is currently part of the WTMP system in Calfornia.
#    The L3MTO process was developed by Levi Brekke in 2009.
#    Version: 2.0
#    Author:  A Wood, NCAR, Oct 2024

# ---- forecast site information (arrays of variables describing different script output sites)
#      to add more locations, extend each array with appropriate information.
siteLabel       = ['shasta', 'trinity']     # user specified label for site
climFcstId      = [ '74516',    '74516']     # ID of site used in filenames for climate forecasts
localMetStnID   = [   'KRDD',     'TCAC1']     # label for site used in local WTMP model input files

# ---- local obs monthly temperature files matching the location of the tercile climate forecast
localMonTobsFile = ['1a_mon_hist_T_shasta_dam.csv', '1a_mon_hist_T_shasta_dam.csv']
```

Figure 18.—Site information metadata file.

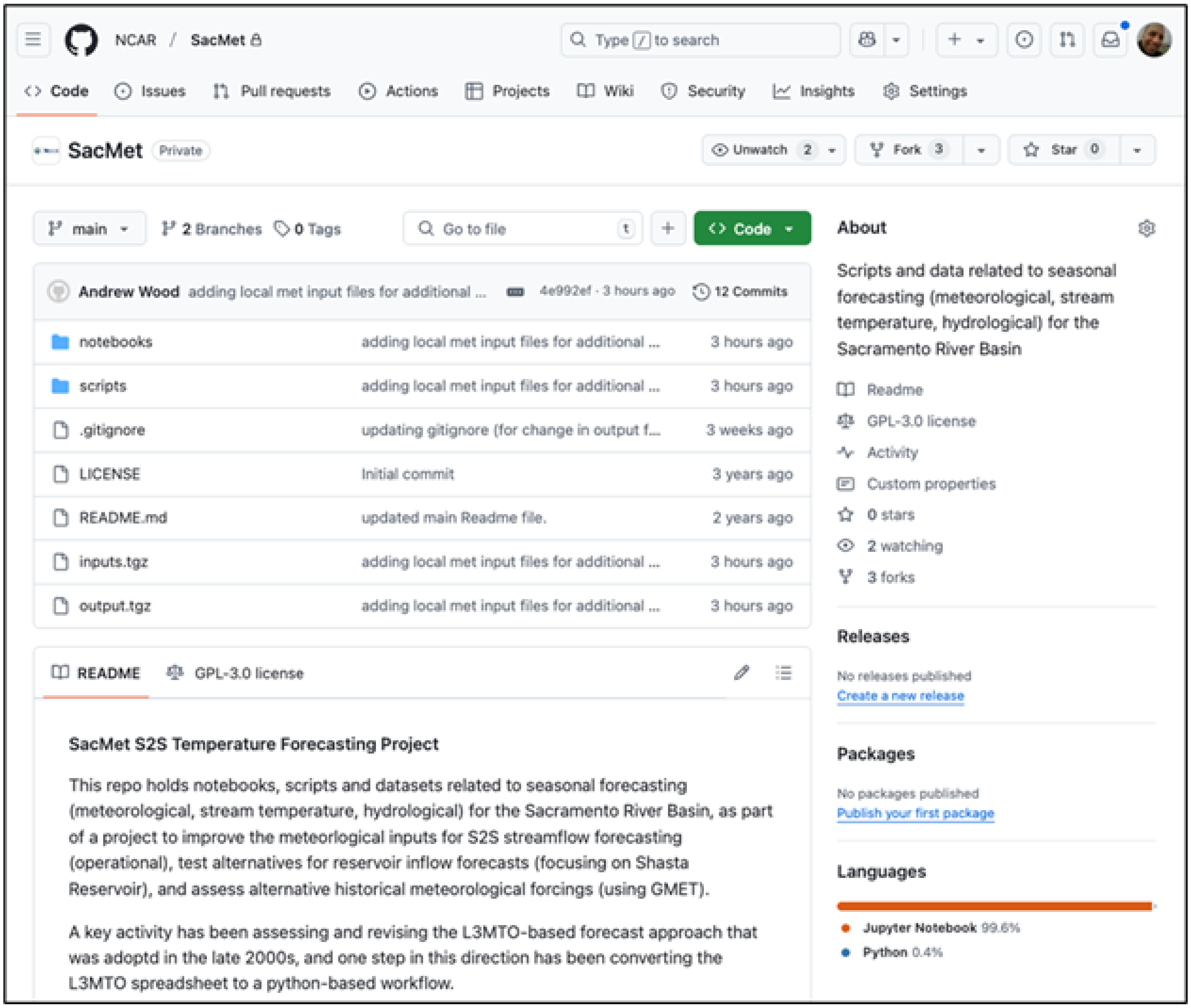


Figure 19.—Screenshot of SacMet project Github repository containing scripts and small datasets.

A two-site example of this script usage was created and run, but input local meteorology files were created for the following sites, based on data files provided to NCAR (by Mike Deas): KRDD/Reading, Trinity/TCAC1, Folsom, American, Lewiston, New Melones/Green Spring, Stanislaus/GreenSpring. Some of the input data files appeared to be the same for multiple sites. A UNIX shell script for reformatting (named **format_ceqw2_input.csh**) was written to combine and reformat the data files (one per variable) into inputs for the Python L3MTO forecast script. The observational records for the sites included in this work spanned 1/1/2000 to 12/31/2021.

To add new sites to the approach, similar temperature model input datasets must be added to the existing collection. In addition, analyses of whether different climate forecast locations are more skillful for each of the sites than the current default location (from the pre-existing HEC5Q L3MTO spreadsheet process) would need to be conducted. The site arrays in the metadata input file would need to be extended.

2.2.6. Accessing CPC Tercile Forecasts and Hindcasts

The weblink to CPC tercile forecasts can be downloaded from: https://ftp.cpc.ncep.noaa.gov/cpcfcsts/archives/long_lead/data/. A command line script **SacMet/inputs/clim_fcst/cpc/get_cpc_terc_forecasts.cs**h in the repository first retrieves a file for all the station locations and lead times, containing both precipitation and temperature tercile forecasts. The file can return one or many forecast files at once, depending on user settings within. Table 2 shows the file format and example data.

Table 2.—Tercile Forecast Data

STN	YYMMDD/HHMM	TBLW	TNRM	TABV	TCAT	PBLW	PNRM	PABV	PCAT
69002	240619/0100	33.33	33.33	33.33	4.00	33.33	33.33	33.33	4.00
69007	240619/0100	33.33	33.33	33.33	4.00	33.33	33.33	33.33	4.00

Footnotes:
STN = station ID
TBLW, TNRM, and TABV values are percentages of forecasted temperatures being below, normal (average), or above historical temperatures
PBLW, PNRM, and PABV values are percentages of forecasted temperatures being below, normal (average), or above historical precipitation
TCAT and PCAT values are indices with 4 being equally likely, 1 being more likely to be below normal (average), and 3 being more likely to be above normal.

For this work, the station ID matching Fairfield, CA was used:

74516 TRAVIS AFB CA US 3826 -12193 19

The shell script loops over specified station IDs and extracts the temperature tercile forecasts into a CSV text file with the following name/format, which has one row per lead time in the CPC forecast, each of which has 3 tercile probability forecasts (for the below normal (BN), normal (NN), and above normal (AN) categories) – Filename: **cpc_tercile_fcst.202406.74516.csv**:

The file has the following format:

BN,NN,AN
33.33,33.33,33.33
33.33,33.33,33.33
29.84,33.33,36.83

Other alternatives for retrieving and formatting the CPC tercile forecasts are encouraged, as this approach has not been refined beyond the needs of this research. Also, forecast stations for use with locations other than Shasta (and possibly the nearby Trinity) reservoirs have not been identified. A list of other available stations is included in the repo (named **stns_ll90.txt**).

3.0 Task 3 Approach and Results

3.1. Approach: GMET application to the upper Sacramento River basin

Supporting Task 3, the project applied a meteorological forcing generation approach called the Gridded Meteorological Ensemble Tool (GMET) to provide an alternative and potentially improved dataset for deriving inputs for the stream temperature model. The multi-decadal record of the GMET dataset could be used to analyze trends and variability in the climate inputs driving the water temperature modeling, to provide a forcing dataset for calibrating the hydrology and stream temperature modeling, and also provide spatially-distributed inputs for stream temperature forecast downscaling, if needed, after further temporal disaggregation. GMET has typically been run at a daily timestep to produce daily precipitation, mean temperature and diurnal temperature range (DTR). Because the input meteorology of the two water temperature models that were used in this study have a sub-daily (HEC5Q is 6 hourly and CE-Qual-W2 is hourly) timestep, GMET outputs could not be used directly, and care would be required in the temporal disaggregation to ensure that realistic sub-daily temperature cycle characteristics are preserved.

GMET methodology is based on multiple linear and logistic regression using static geophysical attributes to predict precipitation and temperature across a latitude/longitude grid (Newman, 2015; Bunn et al., 2022), with prior applications ranging from 1/8th degree to 0.025 degree resolution. Regression errors are used to condition spatially correlated Gaussian random fields for ensemble generation. The spatial regression approach for interpolating situ meteorological observations uses spatially distributed information as predictor fields in an ordinary least squares (OLS) linear regression to explain the spatial distribution of point in situ observations. In this project's application of GMET, the spatial predictors are static geophysical attributes (north-south and east-west slopes, elevation, latitude, and longitude). The regression was applied to predict daily precipitation, mean temperature, and DTR for each target grid cell, based on the current observed values of those variables within a sample from the 30 nearest meteorological stations and given their relationship to the local terrain features at the station locations. This strategy generates dynamic (time-varying) uncertainty estimates that are driven by daily observed meteorological conditions.

GMET was applied for the period 1970 to present at a 2-km grid resolution, yielding daily precipitation and temperature minima and maxima. Here, a small ensemble of 10 forcing members was generated to validate the usability of the GMET dataset. The configuration used 1,550 meteorological stations that had been quality controlled and gap filled, and the result grid domain and station locations are shown in **Figure 20**. The station data preparation approach was used previously in other applications (e.g., Mendoza et al, 2017), but is not well described in those. Relatively straightforward checks for plausible data range, temperature observation ordering, repeated values are applied. Station record gap filling relies on quantile mapping of nearby best-correlated stations through an iterative updating process (Newman et al., 2015).

Before use in modeling applications, GMET output meteorological fields typically must be post-processed to further transform them to the exact spatial configurations (which could be points or polygons) and temporal resolutions required by the model of interest. When disaggregating to sub-daily time resolutions and expanding the variable suite to include meteorological parameters used by models such as CEQualW2 or SUMMA, a common approach used in recent GMET applications is to apply a software tool called MetSim (Bennett et al, 2020), which applies the algorithms of MT-CLIM (Running et al., 1987). Pending a clear subsequent step to use the GMET based forcings in the water temperature modeling, MetSim was not run for this project. In the later hydrological modeling, an existing GMET forcing that did use MetSIM was applied.

Note, during this project, but in a separate effort (funded by NOAA, led by PI Wood), GMET was upgraded and converted to a python-based application called the Geospatial Probabilistic Estimation Package (GPEP; Tang et al., 2024; https://github.com/NCAR/GPEP), although only GMET was used in this work. GPEP has greatly expanded functionality relative to GMET, particularly related to the availability of machine learning methods for spatial modeling of meteorological fields.

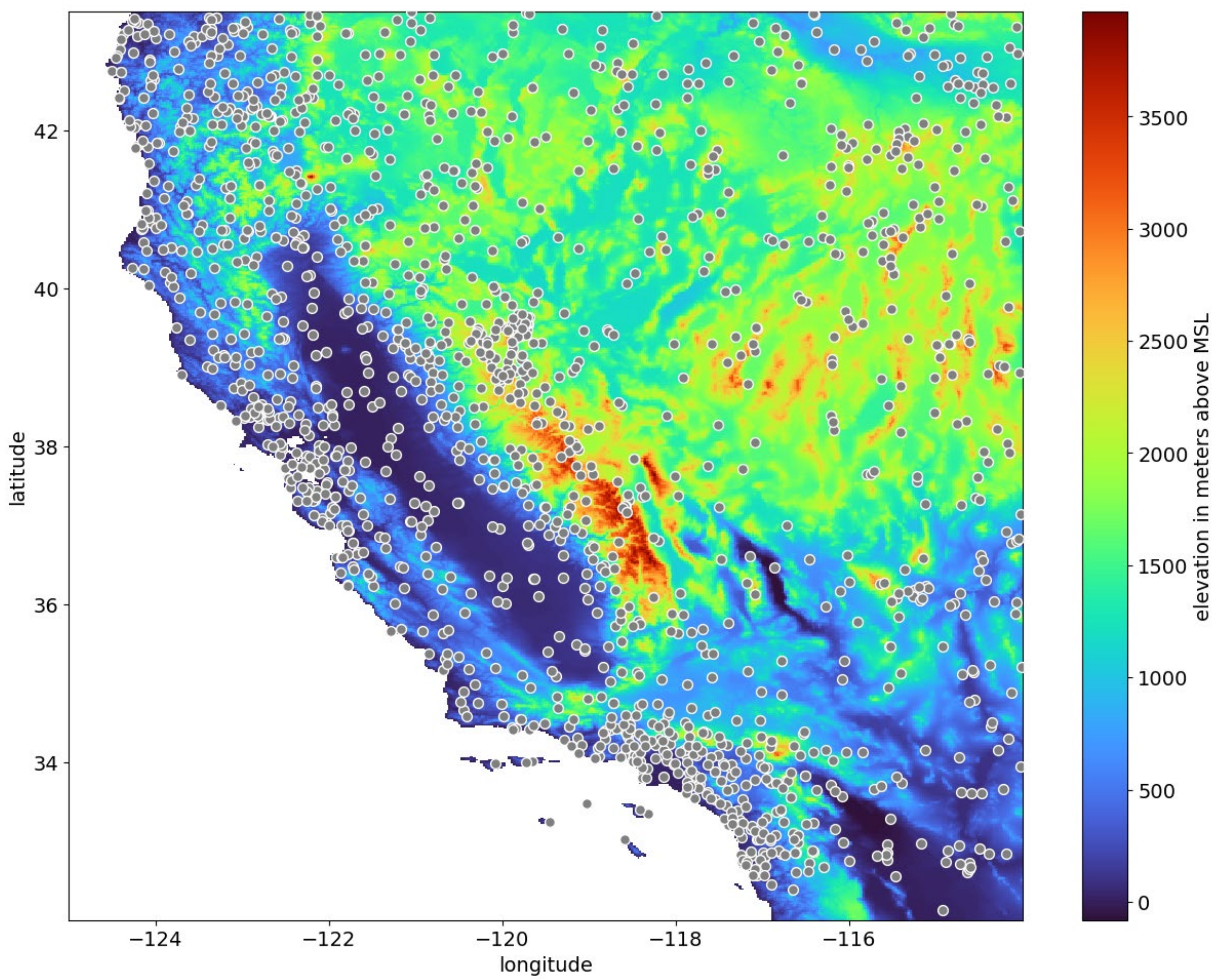


Figure 20.—The geographic extent and elevation range of the 2-km resolution GMET surface forcing dataset for California and adjacent drainages, also showing meteorological station locations.

3.2. GMET application to the upper Sacramento River basin

GMET daily temperature and precipitation datasets were created for the domain shown in Figure 2 with an initial ensemble size of 10 members from 1970 to 2022. Creating ensembles after the GMET regression fields have been estimated is a fast computational procedure (taking only minutes), so the generation of a larger ensemble was postponed pending an applications need in the project. The resulting dataset, including the gridded regression fields and the ensemble members, is approximately 500 Gb in size (uncompressed NetCDF format). **Figure 21** shows a snapshot of a gridded temperature field at 2-km resolution. Various exploratory visualizations of temperature and precipitation timeseries were created to illustrate the potential for trend and variability analysis (**Figures 22-23**), but ultimately the dataset was not explored further in this project.

The apparent upward trend in temperature in the Shasta region is, however, notable – especially in summer, auguring increasing need to safeguard cold water resources to offset temperature spikes late in the season. In any case, the forcing dataset, which has a finer spatial resolution than most other comparable datasets, is available for future research. That said, a newer dataset that is similar to this one, and may be of higher quality – CONUS-wide and including more meteorological stations – at 2 km resolution is being created using GPEP, under NOAA and USACE funded projects at NCAR. A distinctive difference between the GMET/GPEP dataset and other common surface forcing datasets is their ability to quantify uncertainty (time and space varying), which datasets from PRISM or Daymet or NLDAS2 do not provide. The means to maintain and improve them is also available to Reclamation as a result of their project sponsorship. In general, GMET meteorological ensemble uncertainty reflects the strength of the connection between the terrain-related predictors and the observations of a weather field on each day. Often temperature is strongly associated with elevation, due to atmospheric lapse rates, while the precipitation interaction with terrain features is weaker and dependent on multiple drivers such as wind speed and direction, as well as orography. Precipitation uncertainty tends to be relatively larger than temperature uncertainty, but this varies by meteorological regime.

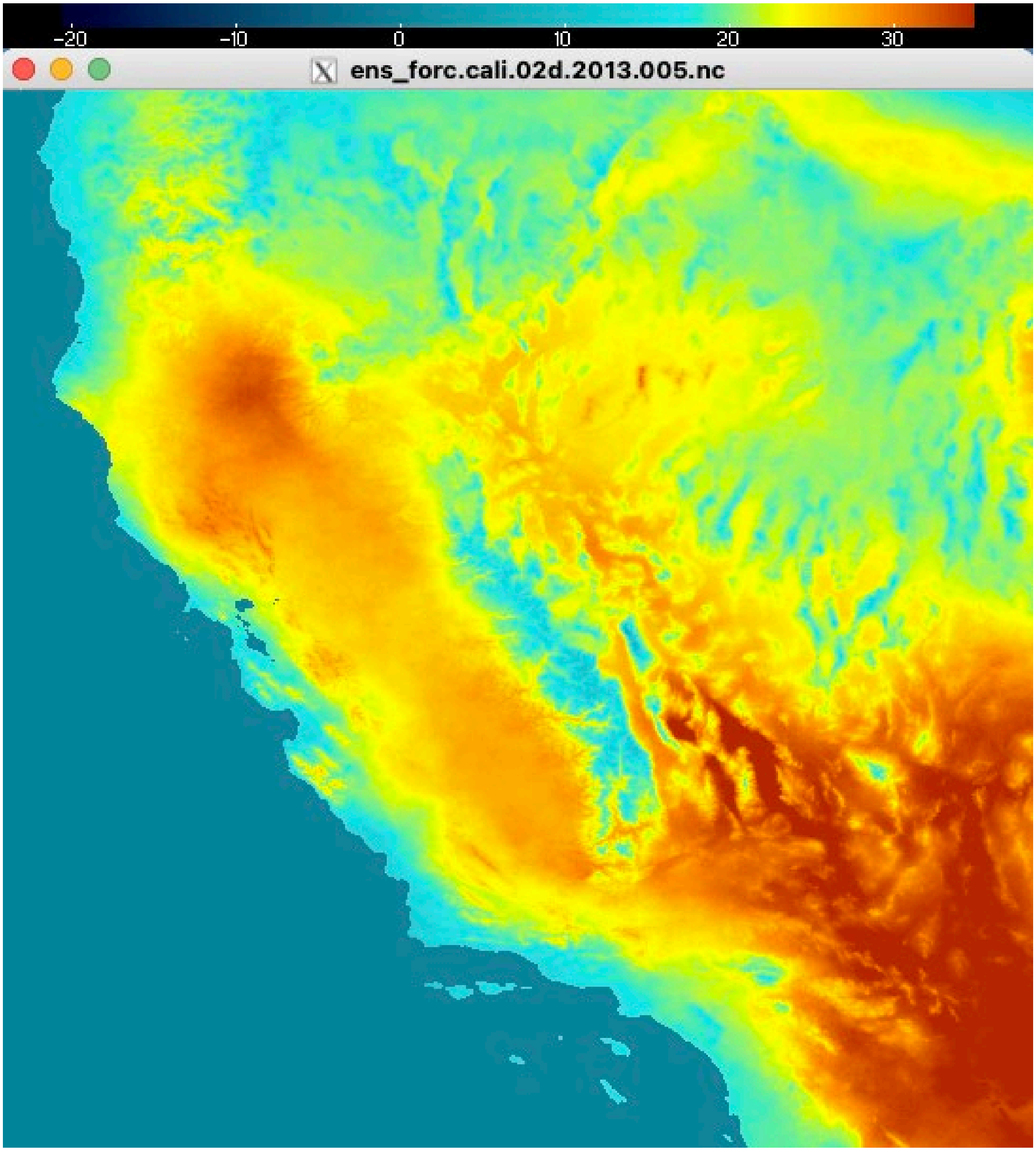


Figure 21.—A ncview software screenshot example of a daily temperature analysis from the 2-km resolution GMET-based surface forcing dataset for California and adjacent river basins areas, showing mean temperature (Celsius) on June 6, 2013.

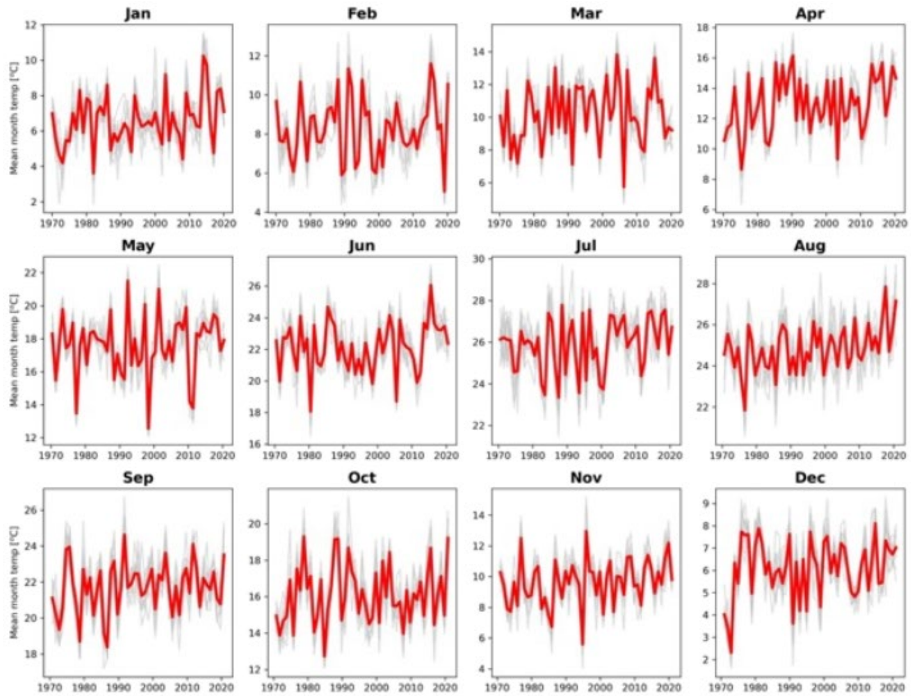


Figure 22.—An analysis of trends and variability from a GMET 10-member ensemble temperature analysis for the Shasta Dam location. The red line indicates the mean of the ensemble members, while the grey lines show their individual variations, illustrating that they could provide uncertainty in a full-scale ensemble implementation (i.e., 30-100 ensemble members).

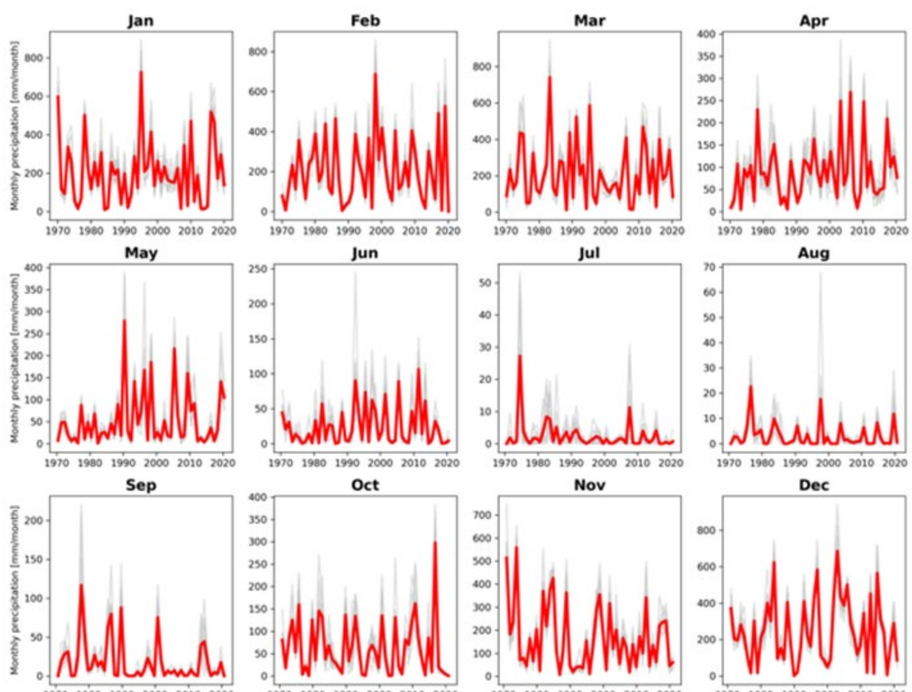


Figure 23.—An analysis of trends and variability from a GMET 10-member ensemble precipitation analysis for the Shasta Dam location.

4.0 Task 4 Approach and Results

4.1. Approach: Alternative S2S climate forecast processing and evaluation

For **Task 4**, we investigated whether climate inputs from a different source may offer better inputs to the current analog generation approach, and use the GMET meteorological analyses as a target climatology for validation. The current CCALM process uses official climate tercile probability forecasts of local temperature (the L3MTO) from NOAA CPC, which have a long operational history (back to the early 1990s), to conditionally resample historical local observation-based model input sequences to provide a seasons-ahead forecast. Other approaches for climate prediction on S2S time scales exist and have comparable or potentially greater skill for the coarse-scale climate variables used to condition the second component of the approach, which is the conditional local weather generation (model input sequences). One is to use climate indices – e.g., the El Nino Southern Oscillation (ENSO) index, the Southern Oscillation Index (SOI), the Pacific Decadal Oscillation (PDO) index, and others to indicate near future climate tendencies, given their known teleconnections to North American continental climate variability. A second strategy is to use dynamical climate model seasonal forecasts, which have advanced in skill over the last 3 decades, albeit slowly, and small operational ensemble collections of these are now commonly run in various climate forecast centers. One is the NMME (Kirtman et al., 2024), which includes 7 global climate models, forecasting for up to 9 months lead time (and updating once per month). Another is the Sub-seasonal Experiment (SubX) sub-seasonal ensemble, which predicts climate out to 45 days (Pegion et al, 2019), is led by NOAA and includes a similarly sized suite of global weather and climate models. Highly regarded S2S climate forecast ensembles are now freely available in near real-time from the European Center for Medium Range Weather Forecasting (ECWMF), but we do not assess these here.

In this project, we explored whether skillful climate predictions could be found from either or both of the SubX and NMME models for the region encompassing the Sacramento River basin. The project helped to advance a national-scale S2S climate testbed effort that was also being used to seek similar findings for the Pacific Northwest [sponsored by the United States Army Corps of Engineers (USACE)]. The testbed built from resources and a strategy initially developed years ago in a NOAA-funded project (led by A. Wood at NCAR and Sarah Baker of Reclamation; Baker et al., 2020), that established a Hydrologic Unit Code 4 (HUC4) level analysis of S2S climate predictands from SubX and an earlier version of the NMME. The layout of the HUC4 regions is shown in **Figure 24**. Skill evaluation for the ensemble dynamical climate forecast systems was conducted using multi-decadal hindcasts from each system, focusing on identifying the skill of different combinations of climate forecasts for aggregated periods that are well-suited to operational workflows.

An example of such a tailored arrangement was explored here (and is now being adopted in other NCAR projects) is illustrated in **Figure 25**, in which the first forecast week is handled using a weather forecast (preferably an ensemble) with explicit hourly to daily sequencing of meteorological events, after which the remainder of the first month is handled as an aggregated climate forecast period, which increases the signal-to-noise ratio of the prediction. Subsequent months or seasons are similarly time-aggregated, working with climate predictand durations (e.g., 1 month, 1 season) that are available from a wide range of operational products or approaches. The shared multi-project work under this task focused on the entire US in building the data processing workflows and analysis scripts required, but specific analysis was conducted and presented for the upper Sacramento River basin.

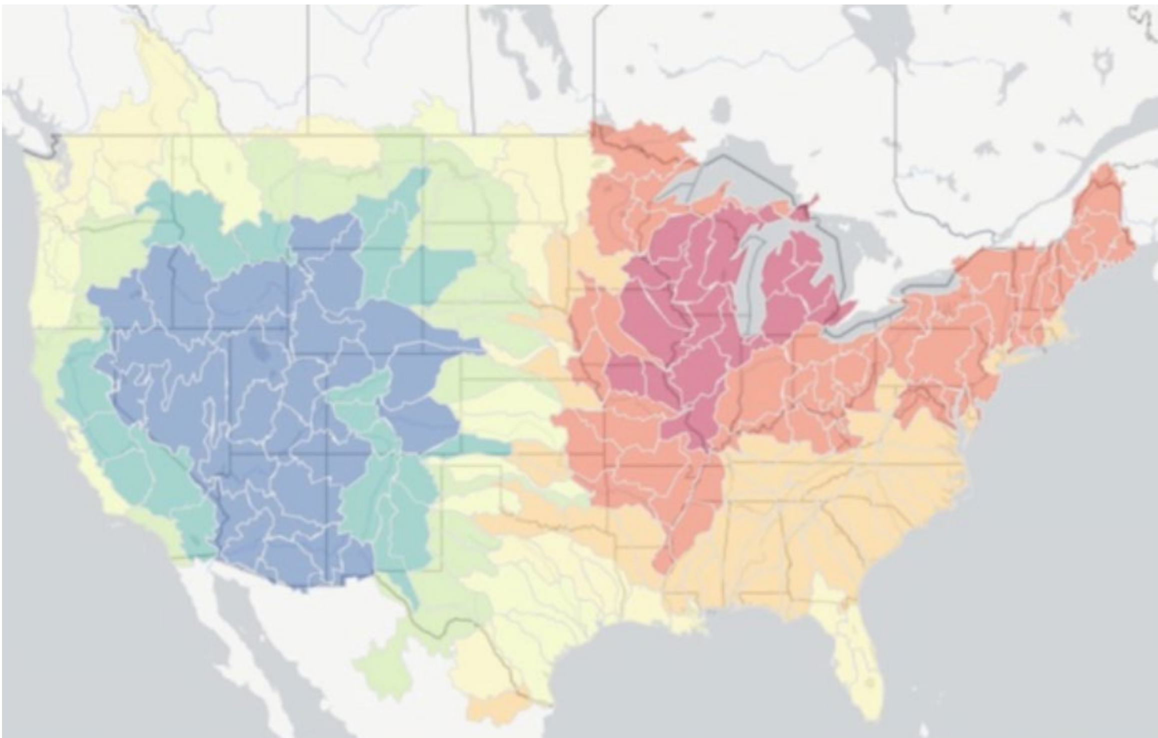


Figure 24.—S2S Climate Forecast Testbed regions (USGS HUC4) used for analysis of potential climate prediction sources across multiple projects at NCAR, as described initially in Baker et al. (2020).

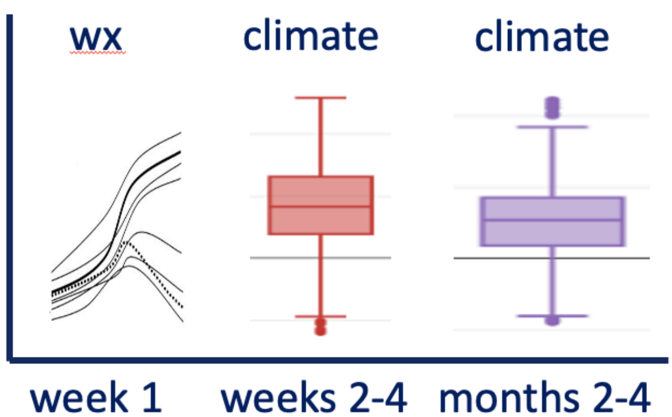


Figure 25.—Illustration of the conceptual sequencing of weather and climate approaches for the first four months of a prediction period, for use in assigning available methods and datasets in an operational workflow. “Wx” is a common abbreviation for weather.

4.2. Results: Alternative S2S climate forecast processing and evaluation

This project was one of several that contributed to a multi-project effort toward identifying and benchmarking potential sources of improved S2S-scale climate forecasts at watershed relevant scales. An NCAR "S2S Climate/Water Testbed" effort pre-existed this project, having started under NOAA funding that supported initial work by PI Wood and Reclamation's Sarah Baker, and is continuing as enabled by the joint interest in better S2S climate forecasts as they arise in various ongoing projects (e.g., from National Aeronautics and Space Administration (NASA), USACE and Reclamation). As discussed in Section 2.3, the effort aims to identify the best currently available climate (precipitation and temperature forecast) in various categories: sub-seasonal, focusing mainly on weeks 2-4, and seasonal for lead times after the first forecast month. These time frames align with different categories of intrinsic climate predictability as well as different climate forecast system products.

We evaluated sub-seasonal climate forecasts from different SubX models (for weeks 2-4) and NMME models (for longer lead seasons) at the HUC4 scale, both CONUS-wide and for unit 1802 (Sacramento River Basin), using both long term statistical analyses (e.g., **Figure 26**) and visualizations (such as the monthly NMME verification shown in **Figure 27**). We found that several NMME models and model combinations, with and without post-processing, could provide weakly skillful predictions (based on a correlation skill score). A broader investigation of this potential appears warranted, although a comprehensive comparison of these results with the CPC Official Seasonal Outlooks for Temperature was not undertaken in this study.

For example, the best NMME model results for the region in **Figure 26**, at lead time 2 (i.e., 1 month before the start of the 3-month season), have an Anomaly Correlation Coefficient (ACC) of nearly 0.4, which corresponds to a skill score (ACC2) of 16% (variance explained above climatology). Current published assessments of CPC skill are limited and also use different metrics (which prevents direct comparison), but several that could be found which focus on a shorter lead time (0.5 months, which is generally more skillful) and suggest skill scores ranging from 5% (as in Figure 2 of https://www.climate.gov/news-features/blogs/enso/seasonal-verification-part-3-verify-vengeancethis-time-its-probabilistic) to between 10% and a maximum of 30%, depending on the month at: https://www.cpc.ncep.noaa.gov/products/verification/summary/index.php?page=map.

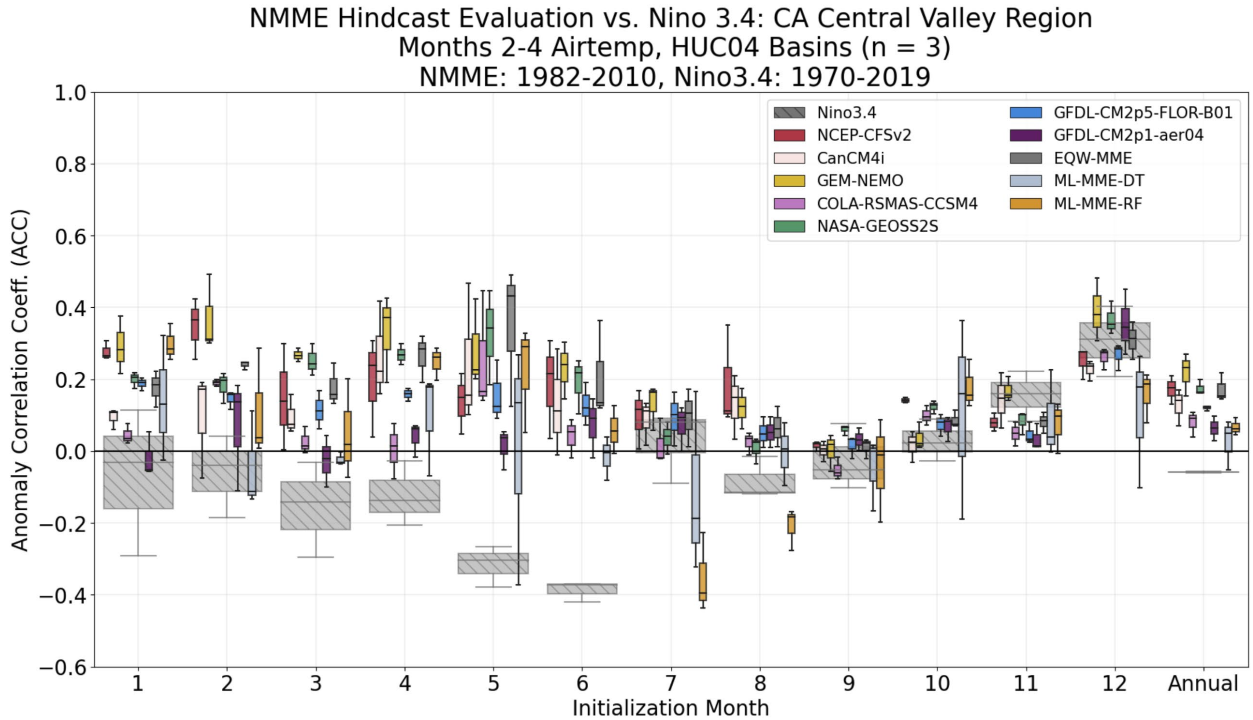


Figure 26.—Example of an S2S Testbed hindcast skill evaluation benchmarking months 2-4 (season 1) air temperature skill from the NMME climate forecast models, the Nino 3.4 teleconnection index, and various combinations, including an equal weight mean of the NMME and two machine learning post-processing variations (random forests and decision trees, RF and DT). The focus region is three HUC4 basins in Northern CA (HUC4 IDs 1801-1803).

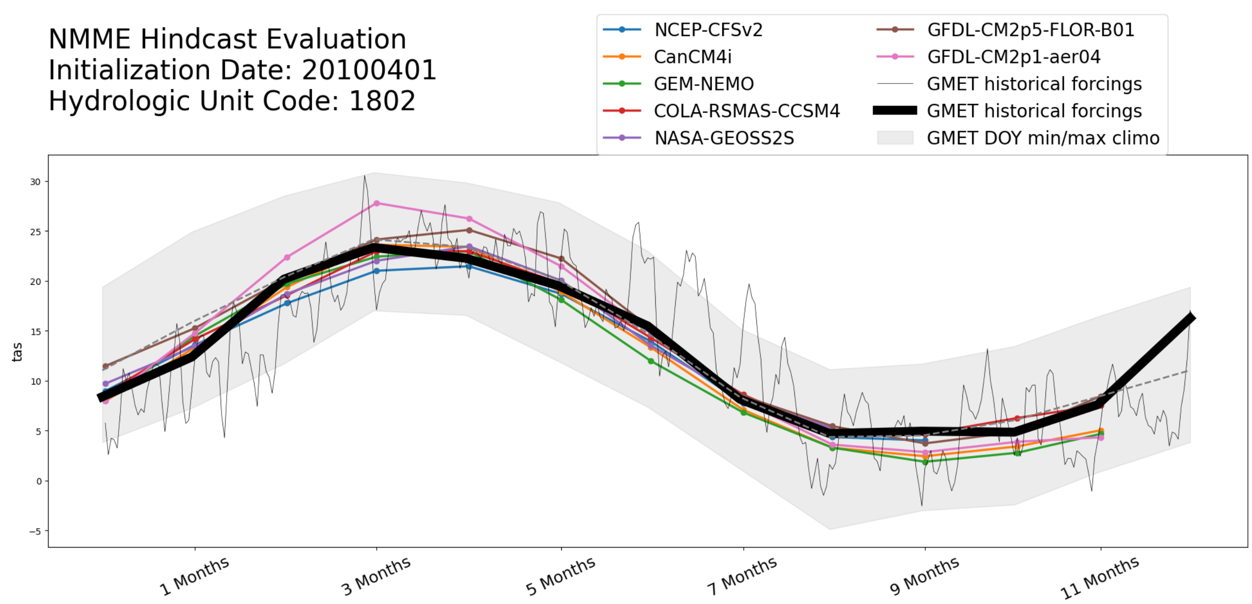


Figure 27.—Example of an analysis of a NMME forecast for monthly temperature versus a historical observation and climatology from GMET, for the HUC-4 of the Sacramento R. basin. The thick black line is the mean of a forcing ensemble.

Looking at the shorter weeks 2-3 or 2-4 time scale, we found that the NCEP GEFS (Global Ensemble Forecast System, one of the SubX models) was generally superior in skill to other SubX models, and would provide a justifiable solution for assigning climate variability in the first month of the forecast, with skill likely surpassing that which can be gained from assigning the first seasonal CPC tercile prediction to the first month of the forecast. An example of the analysis conducted for this effort is shown in **Figure 28**, which contrasts monthly predictions for sub-seasonal air temperature (tas) predictands from different SubX models in the Sacramento River basin HUC4. The ACC is a common statistic used in verifying climate forecasts, measures the correspondence between the anomalies (differences from "normal") of the forecasts versus those in the observations. Another common skill score for categorical predictions (such as tercile-based forecasts) is the Heidke Skill Score (HSS), which measures skill of a forecast by comparing the hit rate of the forecast (the tercile with the largest probability being the one in which the observation verified) versus the hit rate of an equal probability forecast (for below, normal and above average). The HSS was used in the development of CCALM. In assessing the alterative climate forecasts, we used ACC because we are interested in surveying their skill more generally, without assuming that we would use them only in tercile format. Other metrics (especially those measuring both median and spread of the ensemble) would also be appropriate. Both HSS and ACC can be easily found in common statistical textbooks or free online resources.

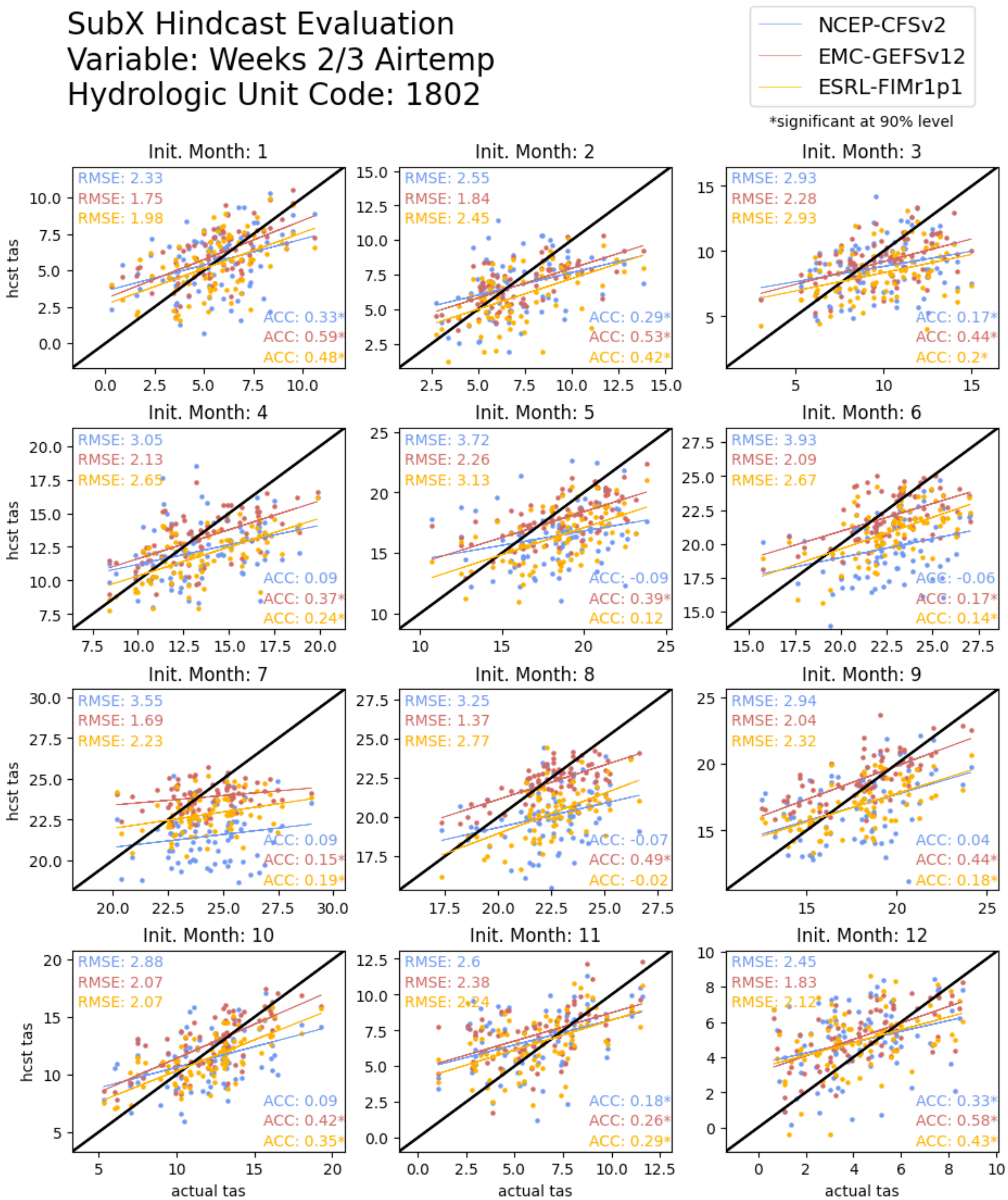


Figure 28.—Skill analyses of three high-ranking SubX models for weeks 2-3 subseasonal air temperature forecasts (tas) for the HUC4 region closest to Shasta Dam.

5.0 Task 5 Approach and Results

5.1. Approach: Stream temperature inflow modeling and prediction

Process-based hydrological models are widely implemented tools used to simulate water, energy, and momentum fluxes within or between different hydrologic domains and subsystems. Over last few decades, process-based hydrological models of various concepts and structures have been modified and tested to obtain hydrologic predictions across a range of spatial (gridded, regional, continental, global etc.) scales to handle different scientific and engineering problems. Yet the collection of available models, which differ in arbitrary specifics, do not provide for systematic and controlled assessment and research into the relationship between model performance and specific modeling decisions, including model structure, parameters and parameterizations. To address this challenge, Clark et al. (2015a, 2015b) created a flexible modeling framework, namely the SUMMA, as a platform for testing and benchmarking different modeling approaches and parameterizations, different process representations across spatial scales, and different representations of spatial variability and hydrological connectivity. SUMMA is written in FORTRAN and is publicly available from an online code repository at: https://github.com/CH-Earth/summa, with online documentation at: https://summa.readthedocs.io/en/latest/SUMMA_documentation/. SUMMA's original development was supported through a NOAA grant to NCAR, but Reclamation and other agencies have since contributed substantial support in multiple projects, including using SUMMA in various river basin locations (such as the Rio Grande, watersheds of the California Sierra Nevada, the Bighorn River basin). Greater detail is provided in the references above.

SUMMA simulates hydrologic fluxes including surface runoff and subsurface discharge. The mizuRoute multi-method channel routing model (Mizukami et al, 2016) is implemented in this project to route SUMMA hydrologic total runoff (surface and subsurface) through the basin's stream channel network and calculate streamflow. The network resolves the stream reaches and key flow locations, and the network may be at a finer or coarser spatial scale than the hydrologic simulation network. MizuRoute development has also been supported by Reclamation project funding over the past decade.

For stream temperature simulation, the River Basin Model (RBM; Yearsley et al., 2009) was applied through a one-way coupling to calculate stream temperature for the stream segments simulated by mizuRoute. RBM is a particle-tracking stream temperature model that solves the time-dependent one-dimensional heat advection equation using a mixed Eulerian-Lagrangian approach. Water temperature is calculated for a specific stream segment based on the upstream water temperature, inflow into the stream segment, the dominant heat exchange at the air-water surface, and inflow and temperature of water advected from tributaries. RBM thus requires simulated reach-level flow data generated by mizuRoute, as well as meteorological forcings, which are taken from the SUMMA forcings and GMET. It also requires assumptions about runoff temperature, which, as in prior applications of RBM, is handled by optimizing regressions related air temperature to runoff temperature and accounting for the presence of snowmelt. Given the importance to CVO activities of other objectives of this project, the inflow and inflow temperature modeling component of this project was scoped as being exploratory (a stretch goal), setting the stage for more detailed analysis or continuation if warranted.

5.2. Results: Stream temperature inflow modeling and prediction

Detailed descriptions of the baseline SUMMA configuration used in this work are provided in a prior Reclamation project report (Wood et al., 2021). In brief, SUMMA was implemented at a spatial scale of lumped United States Geological Survey (USGS) HUC12 catchments and run on a 3-hourly timestep with GMET and MetSIM based input meteorological forcings, remapped to the HUC12 modeling spatial units. The model network is defined by the reach-based global MERIT-Hydro Flowlines network (Yamazaki et al, 2019), after extracting stream channel segments local to URG basin, and adding necessary routing parameters. A unit hydrograph (UH) routing method (termed "impulse response function" or IRF in mizuRoute) was applied. For **Task 5**, we implemented SUMMA and mizuRoute models for major inflow locations to Shasta (such as the Pit, the McCloud, and Sacramento Rivers), and later extended the model domain to include the Trinity Reservoir drainage and downstream areas to control points at Trinity R. nr Burnt Ranch CA (USGS 11527000) and the Sacramento R at Red Bluff CA (USGS 11378500).

The SUMMA and mizuRoute modeling domains are shown in **Figure 29**, which delineates the HUC12 catchments and MERIT stream segments as well as river gaging locations. NCAR consulted CNRFC and reviewed their upper Sacramento R modeling boundaries to determine consistent inclusion of catchments, as some of the upper parts of the basin are endorheic, and do not connect (perhaps in all but the most extreme wet years), while appearing topologically connected based on river routing networks.

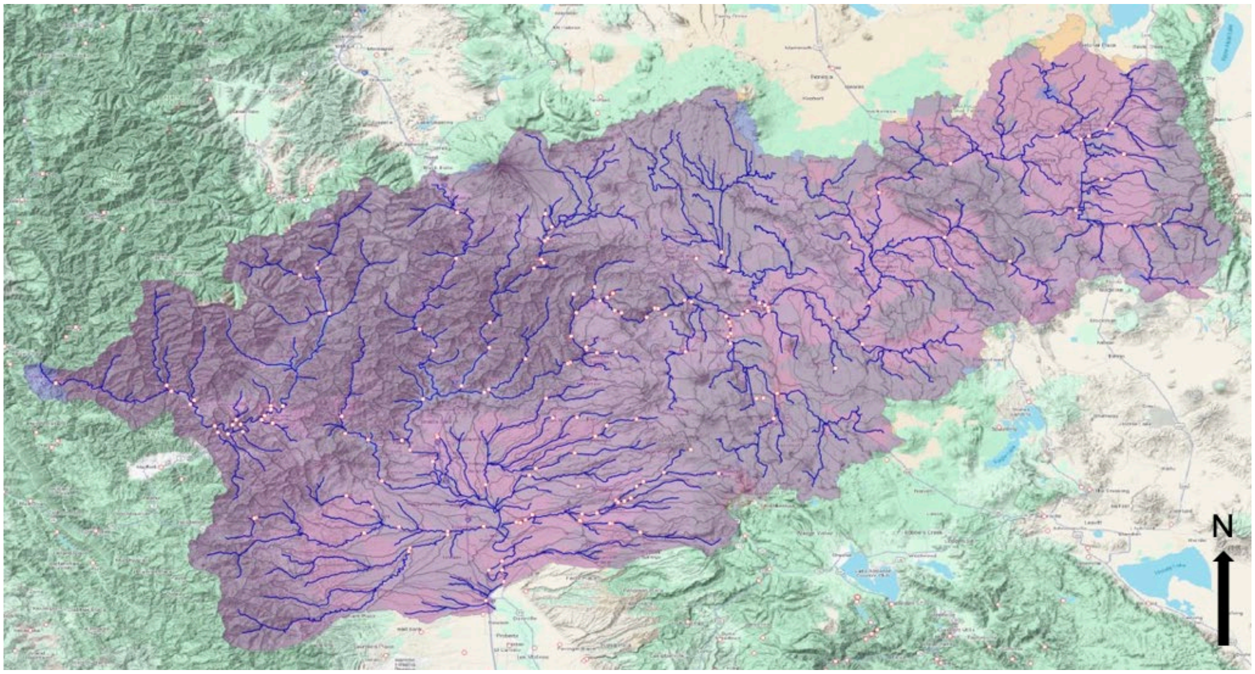


Figure 29.—Simulation domain for the SUMMA HUC12 and mizuRoute MERIT-Hydro routing network for the Shasta and Trinity reservoir drainages, including additional downstream areas. Blue lines indicate stream segments and light pink dots indicate USGS gaged locations, and shaded areas indicate the watersheds included in the model.

We linked the SUMMA and mizuRoute models to a new implementation of the RBM stream temperature model and ran several proof-of-concept simulations. This project is the first to couple RBM with SUMMA-mizuRoute for stream temperature simulation. An illustration of the resulting stream temperatures (before the SUMMA modeling was extended to the Trinity-related domain) averaged over a simulation year is shown in **Figure 30**, showing the gradual rise in temperature from basin headwaters to reservoir inflow locations. This modeling effort was initially scoped as a stretch goal of the project, and the project did not proceed to a full model calibration or evaluation. The models that were created remain as potential resources to be advanced and applied in future projects.

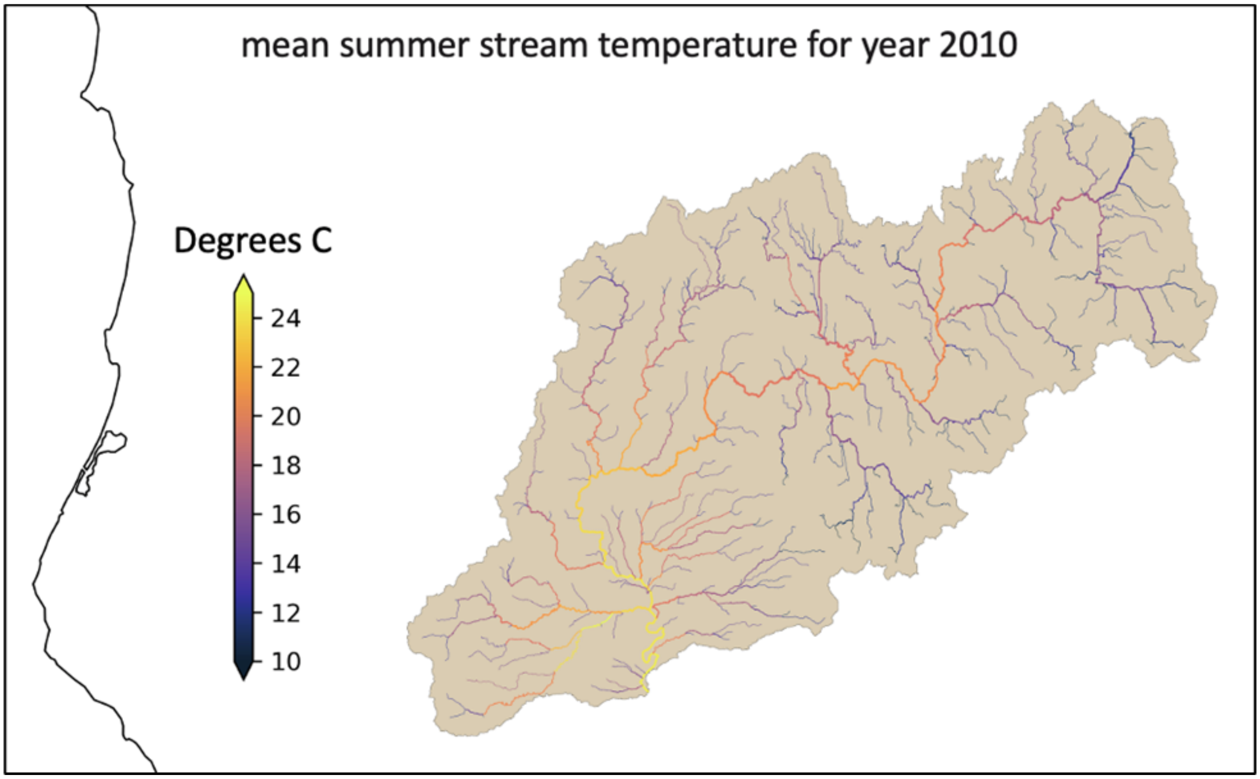


Figure 30.—Prototype RBM-simulated mean stream temperature for the SUMMA-mizuRoute model in the Lake Shasta drainage area. The north-south direction is aligned with the temperature scale bar.

6.0 Task 6 Approach and Results

6.1. Outreach and coordination

The project includes significant activities supporting both Project Management (Task 0) and Documentation and Closeout Process (Task 6). Regular monthly and/or biweekly team meetings have been conducted throughout the project. In addition, there were over a dozen meetings with various partnering efforts and teams including Reclamation's water temperature modeling development team, and with the project steering committee which included Reclamation, DWR, CNRFC, CPC, NMFS, DWR, and CDFW. During these meetings, the NCAR team reported on progress, presented overviews and results from the project, and external partners gave feedback on the progress and methods used in this project. In addition, the project team gave a seminar for the Reclamation Hydrology & Hydraulics community of practice webinar series, and the NCAR PI (Wood) supported other Reclamation efforts, such as attending and presenting on seasonal forecasting at a Klamath River multi-day water supply workshop, and on streamflow hindcasting at a Reclamation Reservoir Operations workshop. The project team also visited Sacramento, Shasta and Folsom dams, and met in person with CNRFC personnel to discuss the project. Lastly, the NCAR PI met with a group from DWR and Scripps to present and discuss collaboration related to a community seasonal climate forecast testbed.

7.0 Discussion

This project focused on assessing and identifying avenues to improve Reclamation's current use of input meteorology in seasonal predictions of stream temperatures, as well as several related investigations. A four-pronged approach was followed to address the project objectives. These major efforts are recapped below, with key project outcomes or findings and recommendations for future work summarized after each.

The first phase of work evaluated and summarized existing information about the current practice of preparing seasonal outlooks for inputs to water temperature modeling, quantifying a baseline based on the datasets and models used in current operations. This phase (Tasks 1 and 2) included developing an overview for Reclamation and stakeholders of the method and performance of the meteorologic forecasts currently being applied for water temperature simulation.

- To facilitate this effort, the spreadsheet-based L3MTO method was duplicated in a set of Python notebooks and analysis scripts, and later a command-line Python script, all of which are maintained in an online Github repository shared with Reclamation team members and an associated consultant. These scripts were needed to make hindcast analyses feasible and provide flexibility in examining alternative inputs and ease in supporting analysis.
- The method was updated from the HEC-5Q model support in place at the start of the project, to support a newer CE-QUAL-W2 model, and is now being transitioned to Reclamation for potential use in operations. The Python-based tool can facilitate operational flexibility (such as running ensembles from different analogs, input climate forecasts or method parameter choices). Such variations were tested, including using tercile forecasts from the IRI center, although their retrieval was less scriptable (a script was written to download forecast map images and extract the pixel color identifying the tercile probability in the Shasta region, but this is not an ideal solution; others will exist using the IRI data library).
- Assessment of 20 years of hindcasts initialized in Spring indicated that the approach provides significant but moderate skill (correlation of monthly means of 0.4-0.5) in the first month of the forecast period, but not thereafter. An evaluation of perfect tercile climate forecasts showed that the method itself (climate-conditioned analog selection) cannot provide higher skill due to uncertainties arising from factors such as the small historical sample size for analog selection coupled with the coarse conditioning provided by tercile-resolution climate forecasts.
- Another notable indication from the analysis is that when using the conservative p=0.25 risk threshold, the perfect forecasts produce temperature model input forecasts that are systematically biased high. This outcome is by design, and means that release planning will hedge toward anticipating higher temperatures than will occur on average. Whether this leads to inefficiencies in cold pool storage use depends on multiple factors, including the implications of temperature trends in the region with late summer temperatures particularly increasing.

A second phase of the work (Tasks 3 and 4) investigated alternative methods to generate the climate inputs needed for seasonal forecasting. A multi-decadal (1970-2020) GMET high resolution (2 km) ensemble surface meteorological analyses was created for potential use in developing alternative, distributed inputs to the stream temperature modeling framework used in seasonal forecasting. The project also investigated the potential for more skillful climate forecast inputs to be obtained from the NOAA SubX and NMME climate prediction datasets, and discussed the feasibility of obtaining short-range (7 or 10 day) weather forecasts from CNRFC.

- The developed GMET dataset can be acquired at NCAR and limited analysis was conducted to query trends in temperature and precipitation forcings, but it was not used in further analyses in this project. It remains as a potential resource should the temperature modeling (which uses point observations from a limited number of stations across northern California) seek a more distributed approach to providing model inputs. Other projects at NCAR are now upgrading the 2-km dataset, however, and that work is the recommended option if such work is of interest.
- The project envisioned a strategy for temperature forecasting inputs to the model based on the use of operational weather predictions (ideally from CNRFC) for the first week, followed by a sub-seasonal climate forecast for weeks 2-4, followed by seasonal climate forecasts for months 2-4 and thereafter. The weather forecast time scale was not analyzed for this project, but hindcast-based analysis of the major dynamical climate forecast systems at the time scales, and empirical (e.g., climate index-based) climate forecasts at the seasonal scale, suggest that slightly better sources of climate prediction information could be used in this application. In this region, the GEFS model predictions for weeks 2-4 temperature and NMME-based predictions for months 2-4 temperature show potential. Longer lead times were not assessed, as expectations are lower for skillful climate forecasting beyond the first 1-2 seasons, and a default to climatology for those may be appropriate.

The third phase of the work (Task 5) explored the potential for reservoir inflow and inflow temperature modeling and prediction, using the SUMMA watershed model and mizuRoute channel routing model coupled to the RBM stream temperature model.

- A SUMMA-mizuRoute-RBM model implementation was created and run for the drainage areas of Shasta and Trinity reservoirs, including additional downstream drainage area for each. This first linkage of the SUMMA-mizuRoute capability with a temperature model demonstrated the potential for distributed water temperature estimates along an intermediate channel resolution, but the models were not calibrated and validated in this study.

Throughout all phases of the project, multiple communication pathways were used to discuss methodologies and convey the results of the project.

- The project team engaged in numerous interactions with agency partners knowledgeable of seasonal forecasting efforts, presented project work at multiple multi-partner meetings and at semi-annual intervals to a multi-agency advisory board.
- A peer reviewed reports/journal article based on material in this report and more extended analysis is planned.

8.0 Future Directions

- A new approach is recommended, given that the L3MTO-based analog method has a relatively low upper bound of potential skill, and alternative tercile seasonal forecasts (such as from IRI) are not likely to have a large impact on improving the skill of the temperature modeling inputs. An option worth investigating further (based on the PI's experience in other projects) is a more temporally tailored strategy, not represented as terciles, that merges ensemble weather forecasts over the first 7-10 days, followed by using sub-seasonal climate forecasts from GEFS for the remainder of month 1, followed by either an empirical (index-based) or NMME forecast from month two onward. Continued benchmarking of different sources of climate predictions in these lead times is recommended, including ensembles that are now freely available in near real-time from the ECWMF. We recommend that interactions with climate and weather forecasting centers continue so that new developments are incorporated into the water temperature model meteorological input generation process, even while it continues to the use the analog-based approach, CCALM, which is not bound to use only L3MTO-based climate forecasts.
- While maintaining the tercile-conditioned deterministic analog approach at present (using improved conditioning factors if found), the use of other temperature input sequence approaches should continue to develop. For instance, an alternative small ensemble of meteorological inputs can be created that sequences weather forecasts at the temporal resolution of the water temperature model with climate-adjusted resampled historical sequences (similar to an NWS Ensemble Streamflow Prediction, or ESP, forecast) at longer timescales. If an ensemble cannot be run through the water temperature models, an alternative would be to run low, median, and high traces to bracket the uncertainty in future conditions. This approach would not use the current Python L3TMO script, however.
- Current water temperature modeling (and input forecasting) is linked to in situ hourly meteorological observations from a small number of stations. The gridded forcing dataset methods offer an alternative, which could be used in the generation of reach-based distributed inputs to the water temperature models, possibly enhancing their representation of gradients in temperature throughout the river system. If the temperature modeling inputs expand from using only in situ observations to being able to use

estimated products (e.g., gridded, ensemble, synthetic datasets), other potential modifications are also possible that might enhance the skill of a prediction. For instance, the recent tendency toward summer wildfires that create widespread smoke and reduced radiation at the surface may exert a cooling effect relative to clear sky conditions; such phenomena could be factored into predicted meteorological sequences.

- The transition of the existing L3MTO spreadsheet tool to a Python based command-line tool with configurable inputs (e.g., it can forecast for multiple sites at once, and, with a suitable workflow wrapper, over multiple dates or with multiple assumptions) was motivated to enable hindcasting and testing of alternatives over many initializations of the method. We recommend that this concept of tool design – i.e., to support both development and operations – be adopted so that the current process can continue to evolve as new methodological opportunities arise.

9.0 Data Availability

All project scripts and small data files are being maintained in a repository located at https://data.usbr.gov/catalog/8140.

Core software used in the project, such as models, are contained in online repositories, including:

- https://github.com/NCAR/summa (public)
- https://github.com/NCAR/mizuRoute (public)
- https://github.com/NCAR/GMET (public)
- https://github.com/NCAR/GPEP (public)
- https://github.com/UW-Hydro/RBM (public)

10.0 References

Baker, S.A., B. Rajagopalan, and A.W. Wood. 2020. “Application of Postprocessing to Watershed-Scale Subseasonal Climate Forecasts over the Contiguous United States.” J. Hydrometeorology, 21(5), 971-987. https://doi.org/10.1175/JHM-D-19-0155.1

Baxter, S. 2016. “Why Aren’t We Beating a Categorically Warm Forecast? Science and Technology Infusion Climate Bulletin, 41st NOAA Annual Climate Diagnostics and Prediction Workshop.” Orono, ME, 3-6 October 2016. https://www.weather.gov/media/sti/climate/STIP/41CDPW/41cdpw-SBaxter2.pdf

Brekke, L.D., J.J. Barsugli, G.J. McCabe, and R. Yaworsky. 2008. “Using NOAA CPC Local Three-Month Temperature Outlooks to support fisheries management.” Poster presentation, Amer. Met. Soc. 22nd Conference in Hydrology, New Orleans. https://ams.confex.com/ams/88Annual/techprogram/paper_134398.htm

Bunn, P.T.W., A.W. Wood, A.J. Newman, H. Chang, C.L. Castro, M.P. Clark, and J.R. Arnold. 2022. “Improving station-based ensemble surface meteorological analyses using numerical weather prediction: A case study of the Oroville Dam crisis precipitation event.” J. Hydromet. 23(7), 1155-1169. https://journals.ametsoc.org/view/journals/hydr/23/7/JHM-D-21-0193.1.pdf

Clark, M.P., B. Nijssen, J.D. Lundquist, D. Kavetski, D.E. Rupp, R.A. Woods, J.E. Freer, E.D. Gutmann, A.W. Wood, L.D. Brekke, J.R. Arnold, D.J. Gochis, and R.M. Rasmussen. 2015a. “A unified approach for process-based hydrologic modeling: 1. Modeling concept.” Water Resour. Res. 51, 2498–2514. https://doi.org/10.1002/2015WR017198

Clark, M.P., B. Nijssen, J.D. Lundquist, D. Kavetski, D.E. Rupp, R.A. Woods, J.E. Freer, E.D. Gutmann, A.W. Wood, D.J. Gochis, R.M. Rasmussen, D.G. Tarboton, V. Mahat, G.N. Flerchinger, and D.G. Marks. 2015b. “A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies.” Water Resour. Res. 51, 2515–2542. https://doi.org/10.1002/2015WR017200

Croley, T.E. 1996. “Mixing Probabilistic Meteorology Outlooks in Operational Hydrology.” J. Hydrologic Engineering, Volume 2, Issue 4. https://doi.org/10.1061/(ASCE)1084-0699(1997)2:4(161)

Goddard, L., S.J. Mason, S.E. Zebiak, C.F. Ropelewski, R. Basher, M.A. Cane. 2001. “Current Approaches to Seasonal-to-Interannual Climate Predictions.” International Journal of Climatology, 21. 1111 – 1152, 10.1002/joc.636.

Kirtman, B.P., D. Min, J.M. Infanti, J.L. Kinter, D.A. Paolino, Q. Zhang, H. van den Dool, and et al. 2014. “The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction.” Bulletin of the American Meteorological Society 95 (4): 585–601. https://journals.ametsoc.org/view/journals/bams/95/4/bams-d-12-00050.1.xml

Mendoza, PA, A.W. Wood, E.A. Clark, E. Rothwell, M.P. Clark, B. Nijssen, L.D. Brekke, and J.R. Arnold. 2017. "An intercomparison of approaches for improving predictability in operational seasonal streamflow forecasting, Hydrol." Earth Syst. Sci., 21, 3915–3935, 2017.

Mizukami, N., M.P. Clark, K. Sampson, B. Nijssen, Y. Mao, H. McMillan, and many authors. 2016. "mizuRoute version 1: a river network routing tool for a continental domain water resources applications." *Geoscientific Model Development*, *9*(6), 2223–2238. https://doi.org/10.5194/gmd-9-2223-2016

Pegion, K. and Coauthors. 2019. "The Subseasonal Experiment (SubX): A Multimodel Subseasonal Prediction Experiment." Bull. Amer. Meteor. Soc., 100, 2043–2060. https://doi.org/10.1175/BAMS-D-18-0270.1.

Reclamation. 2017. "Water Temperature Management in Reservoir-River Systems through Selective Withdrawal." Reference Technical Memorandum for Central Valley Project Operation, California. September. https://www.usbr.gov/mp/bdo/docs/cvp-wtm-selective-withdrawal-2017-09.pdf

Running, S. W.; R. Nemani, and R.D. Hungerford. 1987. "Extrapolation ofmeteorological data in mountain terrain, and its use for simulating forest evapotranspiration and photosynthesis." Canadian Journal of Forest Research, 17: 4 72-483.

Tang, G., A.W. Wood, A.J. Newman, M.P. Clark, and S.M. Papalexiou. 2024. "GPEP v1.0: a Geospatial Probabilistic Estimation Package to support Earth Science applications." Geosci. Mod. Dev. 17, 1153–1173. https://doi.org/10.5194/gmd-17-1153-2024

Timofeyeva-Livezey, M. and Coauthors. 2024. "NWS Regional and Local Climate Services: Past 20 Years, Present, and Future." Bull. Amer. Meteor. Soc., 105, E832–E847, https://doi.org/10.1175/BAMS-D-22-0284.1

Wood, A.W., J. Sturtevant, J.L. Barrett, and D. Llewewellyn. 2021. "Improving the robustness of southwestern US water supply forecasting." Report to the Science and Technology Program, Research and Development Office, US Bureau of Reclamation, Denver, USA. https://www.usbr.gov/research/projects/download_product.cfm?id=3029

Yamazaki, D., D. Ikeshima, J. Sosa, J. Emanuel, Y. Lin, R. Tawatari, T. Sayama, and S. Kanae. 2019. "MERIT Hydro: A high-resolution global hydrography map based on the MERIT DEM." Water Resources Research, 55(6), 5053–5073. https://doi.org/10.1029/2019WR024873

Yearsley J. 2009. "A semi-Lagrangian water temperature model for advection-dominated river systems." Water Res. Rsrch., 45, W12405. https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2008WR007629