Peer Review Report

Evaluating Water Temperature Modeling and Prediction in the Sacramento River Basin

Date

February 13, 2025

Originating Office

Bureau of Reclamation, California – Great Basin Region, Central Valley Operations Office

Reclamation Roles

Director or delegated manager: Levi Johnson, Office Manager, Central Valley Operations Office, Bureau of Reclamation

Peer Review Lead: Randi Field, Hydrologic Engineer, Central Valley Operations Office, Bureau of Reclamation

Peer Review Scope

Meteorologic dataset development and its application to water temperature modeling are the subjects of this peer review. Peer reviewers will be asked to provide responses relative to the following questions:

1. Is the design and methodology used to generate meteorological forecasts reasonable and appropriate for the intended application of seasonal water temperature management planning?

2. Are available data sufficient for forecast development?

3. Does the documentation include adequate information for replication and translation to other river basins?

The scope of the peer review did not include the selection of Water Temperature models used as the basis for this work, or the Climate Prediction Centers (CPC) seasonal three-month

outlooks used in the existing water temperature modelling platform. The water temperature models used in this work, HEC-5Q and CE-QUALW2 have been peer reviewed as part of their development. The CPC three-month outlooks themselves are also based on peer reviewed products.

Peer Reviewers

Peer reviewers were selected who have a broad range of familiarity with water temperature modeling and seasonal forecasting methodologies and meteorological products used in this work. The two selected reviewers were from the Bureau of Reclamation and are listed below:

Levi Brekke, Ph.D., P.E.

      Chief, Research and Development Office, Bureau of Reclamation, Denver, CO

Kristin Mikkleson, Ph.D., P.E.

      Hydrologic Civil Engineer, Bureau of Reclamation, Denver, CO

Summary of Reviewer Comments

Responses to the three specified questions in the peer review scope:

1. *Is the design and methodology used to generate meteorological forecasts reasonable and appropriate for the intended application of seasonal water temperature management planning?*

   Reviewer 1 Response: The design and methodology used to generate meteorological forecasts seems reasonable and appropriate for the intended application of seasonal water temperature management planning.

   Reviewer 2 Response: Yes, the historical methods used to generate seasonal water temperature forecasts that were analyzed in this study are appropriate for water management planning but only for the first month of the forecast. Beyond that initial month, forecasts appear to have no skill. This study did suggest additional methods or data that could be used to generate more skillful forecasts, although more investigation into these methods is warranted.

2. *Are available data sufficient for forecast development?*

   Reviewer 1 Response: Yes, albeit with potential limitations depending on how comments below are adjudicated with respect to section 3.1.3 (Assessment of forecast skill using CPC Hindcasts).

Reviewer 2 Response: Yes, the data used and developed as part of this project are adequate for forecast development. The study does a great job of outlining how successful the old forecasting methods are along with its limitations. The study then investigates a few plausible new models and develops a possible new historical climate dataset that could be used in forecasting; however, there are many more subsequent steps that need to be taken for Reclamation to actually adopt any of these new methods or data sources. It will be very important to someday determine how the uncertainty associated with climate inputs and forecasting methods translates to the uncertainty in forecasted water temperature which is much more possible now after the development of the python-based notebooks and the coupling of SUMMA to a water quality model.

3. *Does the documentation include adequate information for replication and translation to other river basins?*

   Reviewer 1 Response: Yes.

   Reviewer 2 Response: The document does contain adequate information for reproducibility and translation to other basins; however, it could be written a bit more clearly and I do have a few suggestions to help readers understand the study and its methods. I suggest adding an additional study domain section and conceptual figure detailing the methods, data, and tools used and produced in the study. The study domain section would enhance the readers understanding of what methods/data/models were applied where and could be linked to the conceptual image detailing the steps/methods taken and data used and/or developed to answer each study objective. Specifically, the report needs more detail describing how all the meteorological inputs for the water temperature model are developed given the variables in the GMET dataset are only air temperature and precipitation. In addition, I suggest defining all acronyms, abbreviations, and colors/lines on every figure. It can be a bit difficult to decipher what the figures are telling us without this information.

Additional feedback from the reviewers included minor editorial comments, re-organizing content for clarification, including a study domain sub-section within the methods section, enhanced details of the use of models and assessment metrics with supporting rationale, and a suggestion to include conceptual schematics. Recommendations were suggested to clearly distinguish content related to work scope and separating other topics from tool redevelopment. Further comments suggested including refinements to graphics/figures, enhancements for graphical comparison that demonstrate performance, and clarification of data interpretation. Several refinements were also presented to better frame how to approach the project's next steps.

All recommendations that were not deemed beyond the project scope (where noted in the table below) were accepted or addressed. A listing of review feedback and responses are presented in the table below.

Reviewer Name	Page(s) or other reference location	Priority	Comment	Agency Response
Reviewer 1	Executive Summary	Clerical	Summary is very detailed. Streamlining it might make it more effective. Suggestions for streamlining: e.g., Section "Research questions and methods" explains scope but is difficult to track (by comparison, Section 1.2 offers a nice logical outline and preview of work, and even that outline could be improved if it were binned by work phase as suggested in Section 4 Discussion). e.g., Section "Research questions and methods" would be clearer if questions or study objectives are listed near beginning of section. e.g., Section "Conclusions" could be simplified to clearly connect conclusions to work phase and questions/objectives, and it should clearly mirror or connect to Sections 4 and 5.	The Research Questions and methods clearly outline the different phases of the work and their tasks. In the Conclusions part, new references are added to note which tasks are being supported by the outcomes listed.
Reviewer 1	Executive Summary - Need for research	Clerical	2nd para, FITO – use throughout report or remove reference here in the Executive Summary; edit definition if kept (do you mean "Forecast Informed Temperature Operations"?)	This is a general overarching concept, worth mentioning here and also at the outset of the main body of the report. It's not a term that is needed

				throughout the report. Thanks for the correction on the meaning.
Reviewer 1	Executive Summary - Need for research	Clerical	2nd para, delete second to last sentence (seems redundant with sentence before it)	Done.

Reviewer 1	2.2	Clerical	Clarifying implication of met data set choice for stream temperature modeling	Good suggestion – this issue is now highlighted at the end of the first paragraph in section 3.1.
			o Gerber is a 6-hour station meteorology data set.	
			o GMET is a gridded daily meteorology dataset having variables that are total precipitation, mean temperature, and diurnal temperature range (DTR).	
			o Focusing on temperature range, comparing daily amplitude of 6-hour obs at Gerber to daily DTR at GMET grid cell overlying Gerber seems a little incongruent. This is partly about space (point Gerber, grid cell GMET). However, and maybe more significantly(?), this is about statistical metric: Gerber's daily amplitude of 6-hour values isn't necessarily the same as daily DTR from GMET grid cell overlying Gerber. I would expect GMET DTR to exceed Gerber 6-hr amplitude within a given day.	
			o If authors agree, then it might be good to point this difference out to readers and how that choosing between these air temperature data might affect stream temperature modeling results (assuming daily warming is sensitive to DTR vs. 6-hr amplitude forcing).	
Reviewer 1	2.3	Substantive	Describing skill metric options, choice and rationale	Additional text in this direction is added to the end of what is now Section 4.2.
			o Add para short previewing which skill metrics were considered, which were selected for evaluation, and why	

			(e.g., in the 2007 L3MTO work, Modified Heidke Skill Score was used to evaluate L3MTO credibility).	
Reviewer 1	3.1.1	Clerical	p. 13, second sentence starting with "In Figure 7…": What was the question that motivated preparing Figure 7, and what's the significance of "shifted distribution mean"	Figure 7 was a result of asking what a small sample distribution would look like after conditional resampling, and the shifted distribution mean shows an example of the impact on the central tendency of the resampled distribution. The plot also confirms that the new code is working as expected.
Reviewer 1	3.1.3	Clerical	Figures 11 and 12 and associated narrative are confusing. o Add text to identify source of "Tercile climate hindcasts" and relative to what climatology period. o Add text (caption or narrative) to describe what boxplots show – e.g., boxplots show distributions of a variable's daily values for the given year (?), hindcasts vs. observed at X location.	The source is described in Section 2.2.6. It's unclear what the question about climatology period means. The year range of the hindcast analysis is shown in the figures. The caption now has a description of the box and whisker symbol.

Reviewer 1	3.1.3 Figures 11 and 12	Clerical	o Clarify narrative to sharpen statements being made about the credibility of the seasonal climate (temperature) forecast and their applicability to condition met data inputs for stream temperature modeling, especially on whether credibility impressions now are the same as when CCALM was developed in 2007. The following elaborates on this comment:	We add text in Section 2.1 to comment on the likely changes in CCALM since it was introduced.
Reviewer 1	3.1.3 Figures 11 and 12	Clerical	Narrative says "To some extent, this is not surprising, given that the skill of seasonal-scale temperature forecasts (and associated tercile probabilities) is known to be relatively low (e.g., Baxter et al., 2016)…" It would be helpful to explain why this is the case.	Describing the myriad reasons why seasonal climate forecasts at regional scales have generally low skill, especially beyond a single season and/or that is unrelated to long term trends is beyond the scope of this report. We add some text including a paper reference to the topic as a starting point for the reader.
Reviewer 1	3.1.3 Figures 11 and 12	Substantive	When CCALM was developed, L3MTO skill was evaluated using MHSS, showing some skill which drove L3MTO application for condition met data inputs to stream temperature modeling. NOAA was also consulted to understand skill findings. NOAA suggested that when L3MTO forecasts had skill, it was only because the forecasts reflected long-term warming trend and not because there was any skill in predicting variation about that trend (latter seems consistent with Baxter et al. 2016?). Is this still true?	We did not assess this question in this project, but the PI believes this is still true. The seasonal forecasts from the models do capture the temperature trend, as shown in Lehner, F, AW Wood, D Llewellyn, DB Blatchford, AG Goodbody, and F Pappenberger, 2017. Mitigating the impacts of climate nonstationarity on seasonal streamflow predictability in the U.S. Southwest. Geoph. Res. Let., 44.

				https://doi.org/10.1002/2017GL076043 -- a paper written for an earlier S&T project.
Reviewer 1	3.1.3 Figures 11 and 12	Substanti ve	Narrative highlights that this analysis found skill at lead 0-month (i.e. Figure 11, applying Apr-Jun tercile forecast to condition met data in first month of that period) but not lead 1- or 2-month (i.e. Figures 12-13, applying Apr-Jun tercile forecast to condition met data in second and third months of that period). Do the L3MTO forecasts no longer follow warming trend? I.e. If L3MTO tercile forecasts are probability of tercile outcomes from a reference climatology period (recent 30 year period>?), and if there's been a warming trend through that climatology period, and if that warming trend is assume to persist during forecast development, then should the tercile probabilities be skewed warm and reflect that trend? I suppose the answer is yes, and they indeed do, but that variability about the trend is large and tercile outcomes error is still significant, resulting in low to no skill during the hindcasts period. If that's the case, then add narrative to explain that.	We did not assess the question of why the CCALM does not have skill in the hindcasts beyond the first months, and whether this is traceable to the L3MTO forecast skill or lack thereof. A discussion of the trend presence in observations and forecasts is now added before Figure 10a.

| Reviewer 1 | Figure 13 | Substantive | narrative is confusing:

o "the analysis showed that a moderate correlation (mid 0.40 decile) could be obtained" -- add text to indicate what is being correlated (e.g., is the April correlation of 0.44 reflecting correlation of hindcast and observed daily temperature values from all Aprils in the hindcast period (n = 30 days x 18 years) or is it correlation of hindcast and observed mean daily temperature value from all Aprils in the hindcast period (n = 1 daily mean x 18 years)? | The report now states that "the 20+ year hindcasts indicated that the approach brings a moderate correlation skill for monthly mean air temperature in the first lead month, but not at longer lead times." |

| Reviewer 1 | Figure 13 | Substantive | "roughly 80% of the variance in future stream temperature model temperature inputs is not explained even" -- Framing the evaluation in this manor may unnecessarily chill regard for the significance of leveraging seasonal temperature forecasts to condition met inputs for stream temperature modeling. Clarify text to sharpen intent and message.

§ Assuming the 80% of variance comment relates to the April correlation = 0.44 (or ~19.4% variance explained). If the correlation = 0.44 is based on correlating all daily hindcast vs observed daily values,

then this would seem to emphasize error associated with daily hindcast and observed sequences being out of sync through time.

§ Given that application of the seasonal climate forecasts and met inputs is to drive stream temperature modeling to inform ~monthly TCD operations decisions, the sequential alignment of hindcasts and observed daily met variability (which correlation measures) seems less important than the similarity of daily met distributions (which a test of distribution difference might measure). I'm thinking it might be more relevant to emphasize the latter. | These correlations are based on monthly means. The text is expanded to clarify this and to discuss the significance. |
|---|---|---|---|---|

Reviewer 1	Page 49 Bullet starting "another notable..."	Clerical	Regarding risk threshold of climate forecast use (e.g., "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."), consider removing content throughout or clarify reason for keeping it in and sharpen message	This is a key message as it makes clear that there are consequences for the temperature model inputs from making a conservative choice on the threshold. We retain it.
Reviewer 1	Page 49 Bullet starting "another notable..."	Clerical	The statement and paragraph are factually correct, but why are they being addressed in a section on climate forecast skill? § Rather, this statement and paragraph seem to be commentary on stream temperature planning risk policy, and values on reserving cold water resources for summer vs. autumn temperature support.	This is mentioned both at the place of the hindcast analysis and also in discussion section. These placements are appropriate.
Reviewer 1	Page 49	Clerical	This paragraph plust the first "other analysis" in Section 3.1.4 (plus Figure 14) offers a glimpse at how stream temperature modeling results are sensitive to p, but I'm not sure that's the subject of this report or emphasized within the scope shared in Section 1.2.	See comment above. One of the points/objectives of the project and report was to provide some analysis of the choices made in the approach (such as for p) and a tool for assessing such choices further in the future.

Reviewer 1	Page 49	Clerical	Given that tool redevelopment documented in this report does not depend on risk threshold choice in the application of this tool for stream temperature management planning, authors might consider removing this content. If this content is kept, authors should collect content on risk threshold selection into its own section and clearly indicate that this is commentary on risk management separate from documentation of tool redevelopment	Addressed. See comments on this topic above.
Reviewer 1	3.1.4	Clerical	Recommend removing Fig 15 and Table 1, and pare back discussion to only mention that the tool can ingest IRI tercile forecasts rather than those from CPC, and decision to do so would depend on understanding IRI forecast quality and usage logistics relative to that of CPC, which was beyond the scope of this analysis (if that's the case).	These are a demonstration of work done in support of one of the goals of the project, e.g. to enable assessment of alternative sources of climate information. They are appropriate to a project report and are retained.

Reviewer 1	3.1.4	Clerical	This section has three "other analysis" items: risk threshold selection, IRI forecasts, and new Python scripts for time series visualization. o On the first list item, see last comment on 3.1.3 suggesting either removal of this content, or moving it to a section that pools all content on risk threshold selection. o On the second list item, if Figure 15 and Table 1 are kept, then clarify to avoid confusion: § "In the example shown in Figure 15, the IRI forecasts (in blue) were generally warmer in the initial months, with higher exchange rate, than the CPC forecasts, which is in accordance with their tendencies over the first 4 months in Table 1" -- what is meant by higher exchange rate?	The phrase 'higher exchange rate' was removed.
Reviewer 1	3.1.4	Substantive	Table 1 compares CPC and IRI for one forecast issue (Mar 2003), which happens to show IRI was warmer than CPC in that instance, and for a location that's not identified. This seems to risk a reader jumping to an aggregate impression about IRI vs. CPC based on just one forecast issue comparison (Mar 2003).	Text is added to clarify the location. The example is provided to demonstrate the capability of tool for this type of analysis, demonstrating work done by NCAR supporting a project task. This is made clear.
Reviewer 1	3.1.4	Clerical	regarding discussion of python scripts for time series visualizations - this seems a bit extraneous but ok if the point is to highlight functionality of the redeveloped tool.	That is the point. These help document the work that was done.

Reviewer 1	3.1.5	Clerical	Overall, I'm wondering why this section is included in the report. It feels like it is more appropriately included in a user-guide for the tool and not this report which is about the tool's redevelopment. If kept, recommend moving to an appendix. Related thoughts:	This section helps document the work that was done. The project report summarizes work done for the project by NCAR. This material describes the tool that was built and explains the way the input choices are represented in the new tool design, providing more flexibility than was available in the spreadsheet.
Reviewer 1	3.1.6	Clerical	Start of section – It's kind of abrupt. Start with some Why. Before explaining that notebooks were development into a script version to provide input at multiple sites, offer a sentence or two on what was the problem or need of the tool described up to this point that was being addressed by developing this script multi-site version.	The section now starts with a sentence about this. It was covered previously in Section 2.1.
Reviewer 1	3.1.7	Clerical	Figures 17-19 – I question whether these are helpful. The narrative identifies that the scripted multi-site version has a main configuration file and a second input file; showing screenshots of each (Figs 17-18) are probably unnecessary. Further, they may be confusing because the screenshot details are not explained. If this was a user manual on how to understand the details of each, then screenshots are helpful, but that doesn't seem to be the case here. Similar comment wrt github repository and Figure 19.	A user manual is not scoped in this project. This section documents the work done by NCAR to develop the tool to fullfll project task requirements. It shows how it is configured so as to enable use by Reclamation.

Reviewer 1	3.1.8	Clerical	Github repository link did not work when I tried to access it	The repo links were checked by the author and they worked. The reviewer may not have had access since they are private during the course of the project.
Reviewer 1	3.1.9	Clerical	If section is kept in an appendix, then recommend keeping paragraph after Fig 19 caption that alludes to two-site example, maybe show some results from that example.	Showing more results from the example is beyond scope. This functionality was added and tested at the end of the project on request from Reclamation.
Reviewer 1	3.1.6	Clerical	Feels like this section could be moved to an appendix.	Comment addressed above. The section documents work done for one of the tasks in the project.

| Reviewer 1 | 3.2 | Clerical | Feels like 3.2-3.3 are disconnected from scope preview in Section 1.2.
o Section 1.2 says in third bullet: "In Task 4, we investigate 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."

o Section 3.2 says: "GMET daily temperature and precipitation datasets were created for the domain shown in Figure 2 with an initial ensemble size of 10 members… but ultimately the dataset was not exploited further in this project."

o Overall, the narratives in sections 3.2-3.3 indicate that GMET wasn't applied in tool redevelopment and that alternative climate forecast inputs weren't evaluated relative to CPC forecasts.

o Suggest modifying preview in Section 1.2 to focus on work completed rather than aspirations at the start. | GMET meteorological analyses were used in the validation of S2S alternative climate forecasts. Thus the Section 1.2 statement is correct. The specific GMET dataset (2 km) created for this project was not used in one way originally intended, i.e., creating alternative observational sources for use in the CCALM process. No action taken. |
| Reviewer 1 | 3.2 | Clerical | Figures 21-22 seem extraneous if discussion only says "Various exploratory visualizations of temperature and precipitation timeseries were created" and does not connect to the report's primary objective of tool redevelopment and validation | The visualizations convey information about the nature of the GMET dataset, the production of which was one of the tasks of the report. The figures help show that this work was undertaken. No action taken. |

Reviewer 1	3.2	Clerical	Discussion of the upward trend in Shasta region temperature seems to be more of a reflection of the station data driving GMET than of GMET; unsure it's relevant if focus in this section is on GMET.	Trends and variability in surface temperatures, reflected in station observations as well as derived gridded datasets such as one based on GMET, are relevant to the management challenge of maintaining cold water conditions in the rivers, which is the larger context of this report. No action taken.
Reviewer 1	3.3	Substantive	Since alt climate forecasts' skills were not compared to the skill of CPC tercile probability forecasts then maybe it's better to simply just describe alt forecasts available for future consideration. [Ideally, the project would have completed evaluation of alt climate forecasts and landed on results suggesting whether to continue using CPC forecasts vs. an alt source. But assuming time/resources prevent doing such evaluation at this time, then pare back narrative as commented above.]	The report discusses several optionals available for alternative sources of climate forecasts. We do not recommend any specifically but suggest that they warrant further study.
Reviewer 1	3.3	Substantive	2nd para re: NMME models "A broader investigation of this potential is warranted, as there are likely to be stronger options for longer lead air temperature prediction than either then NMME or the current CPC (and IRI) options." -- Saying "likely to be stronger options" is speculative unless you can substantiate with results.	This sentence has been truncated to simply 'A broader investigation of this potential appears warranted'.

Reviewer 1	3.3	Substantive	3rd para re: NCEP GEFS "At the shorter weeks 2-3 or 2-4 time scale, we found that the NCEP GEFS was generally superior in skill to other SubX models... skill likely surpassing that which can be gained from assigning the first seasonal CPC tercile prediction to the first month of the forecast." -- Similar comment as previous one – clarify or substantiate. Also, this is the first time GEFS is mentioned. Define acronym and describe it. Lastly, how does GEFS relate to SubX?	An example of the analysis results to substantiate this result is included in Figure 26. The entire climate analysis that was done is not recorded in this report as it was also supported by another project (it is part of ongoing S2S climate testbed work). We now note that GEFS is one of the SubX models. We define GEFS the first time it is introduced.
Reviewer 1	4	Clerical	I like how Section 4 discussion is outlined by work phases. Encourage authors to reorganize content in Sections 2 and 3 in similar fashion to improve readability – e.g., Let Section 2 be summary of Phase One methods and results (i.e. validation of migrating old tool and data to Python, modification of tool to support use of new stream temperature model, assessment of CPC L3MTO skill, and commentary on risk threshold effect – noting earlier comments to better isolate risk threshold selection content). Then let Section 3 summarize Phases Two and Three methods and results (GMET, other climate forecasts, SUMMA/RBM), given that Two and Three are common in that they demonstrate variations on methods and tools that may inform future directions for tool/application enhancement.	Done, Sections 2 & 3 were reorganized as suggested.

Reviewer 1	5	Clerical	Suggest relabeling this section as "Potential Future Enhancements" or "Future Directions" o This section presents several potential pursuits on advancing science applications that might enhance tool/application for Sacramento River FITO. o Branding these bullets as Recommendations might confuse some readers who think those recommendations rest on work completed in this effort rather than potential work set up this effort and/or hinted at in this report. For example:	Relabeled as 'Future Directions'
Reviewer 1	5	Substantive	"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." -- This conflicts with what's said in section 3.3 where the narrative asserted that alt climate forecasts are likely to be more skillful than the CPC products used in Phase One.	We clarify that we mean alternative climate forecasts 'not based on terciles'.
Reviewer 1	5	Substantive	"Instead, pursuing a more temporally-tailored strategy that merges 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 on? is recommended." -- The superior viability of this approach vs. that featured in the redeveloped tool seems	This sentences is reworded. Based on PI experience in other projects, we state that it is an option worth pursuing.

			speculative given that an evaluation wasn't done in this effort.	
Reviewer 1	5	Clerical	"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 European Center for Medium Range Weather Forecasting (ECWMF)." -- ECMWF products were not mentioned prior to this point in the report. It would be better to introduce ECMWF among the alt climate forecast options in Section 3.3 to set up the point made here.	A sentence is added in what is now Section 4.1. We don't use ECWMF in this work – it's just mentioned as a potential source.
Reviewer 2	General	Substantive	I suggest adding a study domain sub-section w/in the methods section. It would help the readers understand what data was generated for where and what models were applied where.	A study domain figure was inserted.
Reviewer 2	General	Substantive	This project is fairly complex and it is hard for the readers to understand what methods are used for what, why a GMET ensemble was generated, how SubX/NMME are used to generate forecast data, etc. I very much support creating a conceptual schematic that outlines the project methods and how they are related	We have revised the intro to make clear that all the elements of the project are not connected in a way that a schematic would capture. This should now be clear from the text.

Reviewer 2	General	Clerical	In all figures I suggest defining acronyms and abbreviations, along with line colors. If possible it would be better to have python-generated figures included as opposed to screen shots.	Creating publication quality figures is beyond scope for the project report.
Reviewer 2	General	Substanti ve	It is unclear how the entire meteorological variable suite needed to run a water temperature model is generated. How are solar radiation, wind, etc. generated from just precipitation and air temperature (which is what was generated with GMET)? I suggest increasing the clarity on this topic within the report.	The process to resample meteorological variables is described in Section 2.1.
Reviewer 2	General	Clerical	Define HUC	Done.
Reviewer 2	E.S, Need For research	Clerical	Rephrase "Technical demands have increased the challenge of satisfying improved forecasting of environmental conditions downstream of projects. "	Done.
Reviewer 2	E.S, Need For research	Clerical	Redundant statements at and preceding "Given recent extreme dry and warm year situations, there is also an urgent need for improving methodologies used in seasonal and real-time decision making. "	Revised to remove redundancy.
Reviewer 2	ES, Research Questions	Clerical	Define HEC-5Q	Done.

	and Methods			
Reviewer 2	ES, Research Questions and Methods	Substantive	Specify what rivers CVO forecasts water temperature on	The scope of the CVO forecasting for this project is outlined in Section 1.1. Broader CVO operations (eg all their river forecast operations in California) are not relevant to the project. The Water Temp. Mgmt. report is now referenced for further detail in Section 1.1.
Reviewer 2	ES, Research Questions and Methods	Clerical	define S2S	Done.
Reviewer 2	ES, Research Questions and Methods	Clerical	Last sentence not clear and is confusing.	Addressed in report edits.
Reviewer 2	ES, Research Questions	Substantive	Clarify difference between meteorological and climate datasets.	Addressed in report edits.

	and Methods			
Reviewer 2	ES, Research Questions and Methods	Clerical	Define NOAA SubX and NMME	Addressed in report edits.
Reviewer 2	Research Questions and Methods	Substantive	Clarify what the current models and data used by CVO are	Discussed in Section 1 and throughout the report. The exec. Summary does not contain this level of detail.
Reviewer 2	General	Clerical	It could be helpful if you have a section somewhere detailing your study domain and a bit of background about it.	Figure now included in Section 1.
Reviewer 2	Conclusions	Substantive	describe what a 'climate-conditioned deterministic analog' is	Discussed in Section 1 and throughout the report. The exec. Summary does not contain this level of detail.
Reviewer 2	2.1	Substantive	Suggest describing what a T value is and what statistic/test it is generated from	T-test is mentioned and a link to the Wikipedia student t-test is included.

Reviewer 2	2.4	Substantive	Describe SUMMA in more detail.	Section 5.1 describes SUMMA in detail.
Reviewer 2	2.4	Substantive	What does exploratory mean here? How were SUMMA and GMET used?	Exploratory is defined and SUMMA and GMET use is now explicitly explained in multiple locations.
Reviewer 2	3.1.1	Substantive	Define Small Sample within the context of this paragraph	Added parenthetical ("e.g., less than 50")
Reviewer 2	3.1.2,	Substantive	W2 model has been unmentioned until now - additionally clarify the data needs, and usage for this model	Additional details related to the W2 model are referenced via a link in the text (https://www.erdc.usace.army.mil/Media/Fact-Sheets/Fact-Sheet-Article-View/Article/554171/ce-qual-w2/)
Reviewer 2	3.1.4,	Substantive	Was W2 run for any of these forecasts? How does the uncertainty in input variables translate into the uncertainty in estimated water temperatures?	No, this comparison happened early in the project and was not revisited with W2. The question of how the input uncertainties translate into estimated water temperature is beyond the scope of the NCAR research effort, but was discussed qualitatively with project partners at various meetings during the project.

Reviewer 2	3.2	Substantive	Discuss uncertainty of GMET and purpose for generation.	This topic is now discussed in Section 3.1.
Reviewer 2	3.3	Substantive	Compare NMME models and skill discussed with current process.	A paragraph is added to section 4.2, before Figure 24. As recommended by this report (and discussed at multiple times during the project), a structured testbed effort toward determining and tracking the best possible climate forecasts with which to drive the seasonal temperature model input predictions would likely be valuable.
Reviewer 2	Figure 4	Clerical	What does 'wx' stand for? Suggest switching figure order given the order they are referenced in the text	The figure caption now indicates that "'Wx' is a common abbreviation for weather."
Reviewer 2	Figure 21, 22	Substantive	I suggest adding in analyses/data points to figure 21/22 showing how the local meteorology station compares with the generated GMET ensemble. Does the ensemble variability explain all of the observed data	This analysis could be done but the two are not expected to match exactly. The GMET analysis is for a 2-km grid cell mean at the location of Shasta Dam, and would almost certainly be biased relative to any point measurement at that location. The comparison would show this bias without the project having scope to delve into the details, raising multiple questions that this project was not designed to answer (but that could be addressed in another targeted project on generating and optimizing

				gridded spatially continuous datasets for providing inputs agency water models. Because this project included unplanned efforts related to developing inputs for two version of the water temperature models as well as multiple additional locations, detailed analyses of the GMET meteorology were not undertaken.