*Center for Independent Experts (CIE) Independent Peer Review of River Temperature Decision Support Tools*

*Michel Lapointe, PhD*

*December 2017*

**Executive Summary**

Overall, the model development team should be congratulated for assembling and testing a powerful decision support model and an accompanying user interface with good potential (more on minor web site weaknesses under ToR point 4, below). A broad range of software tools and data sources are meshed and incorporated in both final Reservoir and River models. Overall, the model specification and validation is well adapted to the nature of the flow and temperature, river morphology, land use and weather forecasting input data streams regionally available (some suggestions on possible additions to RAFT model specification are given under point 3 below).

Validation tests indicate generally quite good forecast and hindcast predictive performance. I would recommend that the research team present for technical users a slightly more detailed RAFT error analysis. In particular, it should distinguish between prediction errors for daily mean T and daily max T (especially focusing the analysis on seasons of particular interest) and it should further investigate possible sources of some apparent biases at some sites affecting higher T values (on this, see my further comments on RAFT validation, under point 3 "calibration and validation procedures").

**Background and description of reviewer role**

I am a fluvial geomorphologist with a background in physics. My CV is included. My main research expertise relevant here is in field and laboratory based studies of the determinants of salmonid habitat quality (for spawning, rearing and refuge habitats); this includes the geomorphic processes that control its spatial distribution and its sensitivity to various anthropogenic impacts such as forestry, road building, damming, climate warming. My main research has focused on Atlantic salmon and brook trout streams and rivers in eastern Canada, but I have also visited the Sacramento river and briefly consulted on the Trinity River project while on sabbatical in 2011.

More directly relevant, in 2000-2005 I have organized and led a multi researcher 5-year study on the effects of climate change on Atlantic salmon and brook trout stream thermal refugia in river systems in Quebec and New Brunswick. Over the period 2010-2015, I was responsible for leading the physical habitat components (thermal, chemical, hydraulic and sedimentary) in a cross-Canada 5 year project (NSERC Hydronet) on the impacts of hydro dams in Canada on fish habitat and populations.

I have studied all the supplied documents listed below.

M. Daniels, E. Danner. 2017. Technical Memorandum: Calibration and Validation of Water Temperature Models for the Shasta/Sacramento System.

Martin, B. T., A. Pike, S. N. John, N. Hamda, J. Roberts, S. T. Lindley, and E. M. Danner. 2017. Phenomenological vs. biophysical models of thermal stress in aquatic eggs. Ecology Letters. DOI: 10.1111/ele.12705

Pike, A., E. Danner, D. Boughton, F. Melton, R. Nemani, B. Rajagopalan, and S. Lindley. 2013. Forecasting river temperatures in real time using a stochastic dynamics approach. Water Resources Research 49(9):5168-5182. DOI: 10.1002/wrcr.20389

Danner, E. M., F. S. Melton, A. Pike, H. Hashimoto, A. Michaelis, B. Rajagopalan, J. Caldwell, L. DeWitt, S. Lindley, and R. R. Nemani. 2012. River Temperature Forecasting: a Coupled-Modeling Framework for Management of River Habitat. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5(6):1752-1760. DOI: 10.1109/JSTARS.2012.2229968

The Central Valley Temperature Mapping and Prediction (CVTEMP) website: http://oceanview.pfeg.noaa.gov/CVTEMP/

**My detailed comments and conclusions/recommendations with respect to the various Terms of Reference follow**

1. *Evaluation of the strength and weaknesses of the individual water temperature models as well as the process of linking the models, bringing attention to those weaknesses not adequately addressed in technical memorandum.*

Strength and weaknesses in sub-process inclusion and linkages.

W2 model. I am not an expert in physical limnology and lake stratification modeling. Moreover, thermal and flow sub process specifications details for the W2 model (created by other teams) were not given in listed readings nor reviewed here. However, given the input data sets used, this model seems reasonably well specified. Moreover, the calibration and validation show that it seems to perform adequately. A few more comments on W2 calibration/validation are given under point 3 below.

RAFT. One simplification in the RAFT model (which is duly acknowledged by the authors) is the assumption that heat exchange between groundwater and river water is purely by heat conduction through the sub-bed "conduction layer". A fuller specification would incorporate a heat advection term associated with groundwater seepage through the sub-bed into the water column. Along the Willamette River, Oregon (Lancaster and Haggerty, online report, reference below) advective heat fluxes from groundwater inflow through the bed were estimated as larger than conductive fluxes through the bed. At least in some seasons, this advection flux is probably significant along the Sacramento River and it might not be too difficult to incorporate, at least approximately, assimilating groundwater gradient and temperature data streams from apparently numerous available groundwater observation wells along the valley.

The authors note that RAFT simulations indicate that the conductive bed heat fluxes that are incorporated in RAFT are relatively weak. But as advective heat fluxes are not included in the heat budget, these RAFT simulations do not prove that these fluxes need also be weak. One can speculate, for example, that such extra heat exchange terms might become significant during any early summer periods of very low release discharges, in reaches where water tables are higher. However, modeling two way hyporheic water and advected heat exchanges is likely unnecessarily complex as it would involve much higher resolution (smaller than 2km) reach scale information on water surface profiles and bed morphology).

In conclusion, in certain periods and along certain reaches, advective heat fluxes from groundwater may not always be insignificant. To make the model more accurate in a broader range of conditions, this should be investigated, at first with back of the envelope computations. This would be particularly useful if the RAFT model were to be used for multi-year water release planning scenarios incorporating climate change, since USGS simulations seem to predict decades-scale changes in groundwater temperature and mass fluxes along the Central Valley.

Despite these caveats, and although sizable predictive discrepancies remain at certain places and times (see under point 3 below), the model without these bed heat advective components performed reasonably well overall, judging from the validation runs that are presented in readings.

*Investigation of the Temperature Impact of Hyporheic Flow: Using Groundwater and Heat Flow Modeling and GIS Analyses to Evaluate Temperature Mitigation Strategies on the Willamette River, Oregon*
*Final Report*
*Principle Investigators:*

*Stephen Lancaster (Asst. Prof., Geosciences, OSU)*
*Roy Haggerty (Assoc. Prof., Geosciences, OSU)*

2. *Evaluation of the methods used to incorporate uncertainty into predicting water temperature in Shasta Reservoir and Sacramento River down to Red Bluff, such as the use of variable meteorology and model parameters.*

I have no particular expertise here. The Kalman filtering approach used appears to be a coherent and sophisticated method to incorporate in the predictions estimates of the uncertainties in model parameters, in data inputs and in system state observations.

3. *Evaluation of the water temperature model calibration and validation procedure outlined in the technical memorandum and its ability to properly parameterize each water temperature model.*

The W2 and reservoir model validations are generally satisfactory in my view. It appears that observed (rather than predicted) Shasta release temperatures are often available. If I am not mistaken, these observed data (rather than W2 predictions) can then be used directly to predict (through an ARIMA model) Keswick release temperatures. Validations suggest that these Keswick release temperatures to the downstream river (as well as Shasta release predictions themselves) are often slightly overestimated. From the point of view of managing releases to avoid excessive temperatures downstream of Keswick, somewhat over-predicted T estimates at Keswick are arguably less dangerous than under-predicted ones.

RAFT (river model) validations. Although the cited "*Reclamation's 2008 OCAP Biological Assessment* (Chapter 2)" mentions mandated thresholds in daily mean T for egg survival, survival of juvenile Chinook is also mentioned as an objective. The latter can be sensitive to multi-hour mid-afternoon heat spells and this arguably should require the best possible model forecasts of the daily max, distinct from mean daily T thresholds.

In any case, one overall weakness of the submitted documents is that it is not always made clear in text and figure captions whether the RMSE and other prediction error stats reported are based on hourly T or mean daily T data. One may assume that these are errors in 15-minute values, but this should always be clarified in the error analysis sections. Since the prediction target that is cited in the Introduction concerns the daily mean T, a reader could infer that error stats refer to these data.

The challenges involved in predicting daily mean T and hourly T signals are not the same. For various technical reasons, one would generally assume that signal variability as well as prediction errors would be smaller in estimating daily mean T than in estimating daily max or daily min. Errors in predicting both daily max T as well as mean T should both be discussed. In particular, the submitted papers and internal reports (#) should present more detail on those specific biases that may be especially significant in management terms (such as error stats for daily maxima and mean T predictions specific to warmer days). They should also discuss possible sources for any observed biases of concern in model predictions.

Clearly, the overall prediction error stats reported from RAFT, generally under 0.5°C, are quite good, especially if these refer to hourly T values and not mean daily values. However, the task of reviewers is to raise questions that could lead to improvements. In this spirit, what follows may or may not be nit-picking. The Daniels et al. 2017 memorandum should be commended for presenting abundant error plots, stratified by site and month. However, the discussion of possible error sources could be deepened. The data plots provided suggest that, at certain places and times, observed versus predicted discrepancies can be substantial, reaching a few degrees C. Particularly notable (see Fig 22 in Daniels et al., 2017, Technical memorandum) was an apparent inability of the model in some validation runs **for June and July** to predict much higher temperatures than 17-18°C, while observed temps reached 19-20°C. The Fig 22 plots thus suggest a 1-3°C bias towards model under prediction of **max temperatures** these two months.

Daily max T overestimates are of course frequent and visible on many plotted time series. I emphasize any under-prediction biases here as these are more likely to yield to insufficient mitigation using cooler releases and, in turn, to more mortality. Oddly, this significant and systematic under-prediction bias for high temperatures on June and July panels of Fig 22 (and to some extent Fig 24) is not much commented on by the authors of that report, who mainly note overall model overestimates of T. The significance of this summary plot (Fig 22) is unclear, however, as similar plots by station in appendices B show that this underestimate bias mainly occurs at stations CCR and RDB (elsewhere in the documents submitted, there is a comment on particular model uncertainties at RDB diversion dam). The source of such local discrepancies should be investigated further and mitigated if possible.

Here again, it is not made clear in captions or text whether the data shown in Fig 22 is for hourly T or mean daily T values. Whichever it is may affect further possibilities for error reduction. Underestimates in peak daily T during low discharge periods may reflect an overestimation of the buffering effect of the heat sink produced by the relatively cooler river bed. Alternatively, they may reflect limits to prediction accuracy for sub daily cloudiness or wind episodes. For these reasons achieving errors under 1°C in peak daily T may well be much harder to achieve than similar uncertainties in mean daily T.

Overall, it would also be useful to have a sense of what proportion of RAFT prediction errors (ideally by months or location along the river) may be due to errors in weather forecasting versus errors in inflow input data or errors in heat exchange parameterizations. I have not found any substantial discussion of this topic in the submitted readings.

Assuming that available hindcasts yield reasonably good interpolations of observed historical weather series over the grid, would not a comparison of error stats in hindcast versus forecast model, segregated by month and reach, be informative in this respect? Alternatively, to what extent can T prediction errors be related to imprecise inflow volumes (i.e. imprecise input data) from the ungauged tributaries located below Keswick? If this is significant in certain periods and places, what is the potential to incorporate weather driven hydrological models over these ungauged watersheds?

*4 Evaluation of the implication of this work as decision support tools, bringing attention to the any potential for miss-use or miss-interpretation of this information to aid in fisheries and water management in California's Central Valley.*

Assuming uncertainties in daily mean or daily max T forecasts are always made clear to users, it is not apparent how this tool can be mis-used or mis-interpreted. The overall idea is easy to understand for an informed, technically minded public: releasing more cold water at Keswick can abate to some degree water temperatures during these waters transit time over a certain distance downstream, and the degree of abatement will be subject to how hot the weather will be during this few day transit period (thus the weather forecast).

If juvenile or pre-spawning adult carcasses have been found in the study reach after heat stress periods, the potential application of the tool not just to mitigate egg mortality but also juvenile or adult mortality should also be explored and discussed in the reports.

*5 Evaluation of the content made available in the CVTEMP website, bringing attention to content that was unclear and that could be improved.*

The CVTEMP website is an excellent initiative. The website is needed to allow the interested public to better understand the simulations. I do not have the local expertise to know if this website is <u>sufficient</u> given the needs of user/managers (a survey of relevant managers who actually investigate daily release scenarios and of deciders seems required with respect to this). The site will allow updated model outputs to be routinely consulted by busy water managers, who normally do not have the time to liaise and consult with the research team. An "interactive scenario tool" (such as the one illustrated on Fig 3c) of Danner et al., 2012) appears not to have been included on the site. I assume this is because providing in real time multiple scenarios with a wide range of user selected Keswick inputs (Q,T) might require too many computations on a daily basis . What is then the plan to provide this scenario tool (Fig 3c) to manager/users who are the ones most in need of the tool?

Under the river model page: incorporating redd and carcass locations is a very good idea. This data overlay could help generate hypotheses on mortality responses to thermal stress.

The site does not appear yet fully developed and the legends, captions and plots could be refined further. I note below a few other minor weaknesses, possibly easily repaired.

a) on the WATERSHED PAGE: the location of available gauge data is unclear <u>on that page</u>. Maybe refer readers of this page to previous page (ABOUT CVTEMP), for station locations? (i.e. someone consulting that page will wonder: the discharges given for Sacramento River refer to which site? I assume for the site below Keswick dam only?

b) on the RIVER MODEL/TEMP TIMESERIES page, the plots are often a bit hard to decipher. Overall, maybe too many lines are superposed within a small space and the nature of data plotted as the faint, dashed lines are not easily identified. Suggestions: rescale Y axis to expand plots (narrow the range of T shown)? OR introduce more buttons allowing user selected lines to appear in succession?
The ? button that gives a more detailed legend is indispensable: highlight/size up this button? Legends on this pop up remain a bit confusing: e.g. the hindcast box says "…if Keswick Q and T were known". What exactly does this mean? If these inputs were not observed or "known", what actually was used to drive hindcast simulations? The blue lines are referred to as "Forecast meteorology" lines (title given in blue boxes). This is a bit confusing since blue lines extend from hindcast to forecast domains.

6 *Provide a brief description on other aspects of the model not described above.*

The temperature tolerance model (Martin et al., 2017) seems well parameterized and provides a sound basis for resource conservation. Certainly, the biophysics involved in correcting lab based thresholds for actual flow conditions in redds seem sound. The data in their Fig 2 provides field based support to their model which accounts for the effect on oxygen uptake of both egg size and realistic water velocities in river substrate.

## Appendix 1: Bibliography of materials provided for review


M. Daniels, E. Danner. 2017. Technical Memorandum: Calibration and Validation of Water Temperature Models for the Shasta/Sacramento System.

Martin, B. T., A. Pike, S. N. John, N. Hamda, J. Roberts, S. T. Lindley, and E. M. Danner. 2017. Phenomenological vs. biophysical models of thermal stress in aquatic eggs. Ecology Letters. DOI: 10.1111/ele.12705

Pike, A., E. Danner, D. Boughton, F. Melton, R. Nemani, B. Rajagopalan, and S. Lindley. 2013. Forecasting river temperatures in real time using a stochastic dynamics approach. Water Resources Research 49(9):5168-5182. DOI: 10.1002/wrcr.20389

Danner, E. M., F. S. Melton, A. Pike, H. Hashimoto, A. Michaelis, B. Rajagopalan, J. Caldwell, L. DeWitt, S. Lindley, and R. R. Nemani. 2012. River Temperature Forecasting: a Coupled-Modeling Framework for Management of River Habitat. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5(6):1752-1760. DOI: 10.1109/JSTARS.2012.2229968

The Central Valley Temperature Mapping and Prediction (CVTEMP) website: http://oceanview.pfeg.noaa.gov/CVTEMP/

## Appendix 2: A copy of the CIE Statement of Work

# Statement of Work

**National Oceanic and Atmospheric Administration (NOAA)**
**National Marine Fisheries Service (NMFS)**
**Center for Independent Experts (CIE) Program**
**External Independent Peer Review**

*River Temperature Decision Support Tools*

**Background**

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf). Further information on the CIE program may be obtained from www.ciereviews.org.

**Scope**

The SWFSC Fisheries Ecology Division (FED) requests an independent review of the suite of temperature modeling tools they have developed for water and fisheries management in California's Central Valley. When Shasta Dam was built in the 1940s it blocked Sacramento River Winter-run Chinook (SRWRC) salmon from accessing the cold waters of their native spawning habitat. The quality (water flow and temperature) of their current habitat below the dam is now entirely controlled by releases from the dam, and because SRWRC are listed under the Endangered Species Act, dam operations must take into account the impacts on their spawning and rearing habitat. As a result, temperature compliance points have been established:

> From Reclamation's 2008 OCAP Biological Assessment, Chapter 2, pg. 2- 38 (http://www.usbr.gov/mp/cvo/ocap_page.html):
>
> "In 1990 and 1991, SWRCB issued Water Rights Orders 90-05 and 91-01 modifying Reclamation's water rights for the Sacramento River. The orders stated that Reclamation shall operate Keswick and Shasta Dams and the Spring Creek Power Plant to meet a daily average water temperature of 56°F as far downstream in the Sacramento River as practicable during periods when higher temperature would be harmful to fisheries. The optimal control point is the Red Bluff Pumping Plant. Under the orders, the water temperature compliance point may be modified when the objective cannot be met at Red Bluff Pumping Plant."
>
> Page 590 of the 2009 OCAP Biological Opinion starts off with RPA Action Requirements:
>
> http://www.westcoast.fisheries.noaa.gov/central_valley/water_operations/ocap.html

To aid in the water and fisheries management decisions, SWFSC has developed linked temperature models for the Shasta Reservoir and the Sacramento River to model how operations will impact water temperatures within the SRWRC spawning habitat. The SWFSC then developed a thermal tolerance model for SRWRC eggs (the most temperature sensitive life stage) and linked it to the temperature model. The combined suite of models allows for water and fisheries managers to evaluate how proposed seasonal water operations impact SRWRC eggs in a spatiotemporally explicit manner.

**Requirements**
NMFS requires three reviewers to conduct an impartial and independent peer review in accordance with the Statement of Work SoW, OMB Guidelines, and the Terms of Reference (ToR) below. The reviewers shall have working knowledge and recent experience in temperature modeling, with specific emphasis on water temperature modeling in both lentic and lotic fresh water systems (i.e. river and reservoirs), thermal performance modeling of ectothermic organisms with an emphasis on early life stage development in relation to temperature exposure, and experience linking physical and biological models. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

**Tasks for reviewers**
Each CIE reviewer shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Pre-review Background Documents: Review the following background materials and reports prior to the review. The contractor will provide these documents (via electronic mail or made available at an FTP site) to the CIE reviewers.

> M. Daniels, E. Danner. 2017. Technical Memorandum: Calibration and Validation of Water Temperature Models for the Shasta/Sacramento System.
>
> Martin, B. T., A. Pike, S. N. John, N. Hamda, J. Roberts, S. T. Lindley, and E. M. Danner. 2017. Phenomenological vs. biophysical models of thermal stress in aquatic eggs. Ecology Letters. DOI: 10.1111/ele.12705
>
> Pike, A., E. Danner, D. Boughton, F. Melton, R. Nemani, B. Rajagopalan, and S. Lindley. 2013. Forecasting river temperatures in real time using a stochastic dynamics approach. Water Resources Research 49(9):5168-5182. DOI: 10.1002/wrcr.20389
>
> Danner, E. M., F. S. Melton, A. Pike, H. Hashimoto, A. Michaelis, B. Rajagopalan, J. Caldwell, L. DeWitt, S. Lindley, and R. R. Nemani. 2012. River Temperature Forecasting: a Coupled-Modeling Framework for Management of River Habitat. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5(6):1752-1760. DOI: 10.1109/JSTARS.2012.2229968
>
> The Central Valley Temperature Mapping and Prediction (CVTEMP) website: http://oceanview.pfeg.noaa.gov/CVTEMP/

Desk Review: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the Contracting Officer's Representative (COR) and the CIE contractor.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

**Place of Performance**
Each CIE reviewer shall conduct an independent peer review as a desk review, therefore no travel is required.

**Period of Performance**
The period of performance shall be from the time of award through November 2017. Each reviewer's duties shall not exceed 10 days to complete all required tasks.

**Schedule of Milestones and Deliverables:** The contractor shall complete the tasks and deliverables in accordance with the following schedule.

| | |
|---|---|
| Within two weeks of award | Contractor selects and confirms reviewers |
| Within four weeks of award | Contractor provides the pre-review documents to the reviewers |
| October 2017 | Each reviewer conducts an independent peer review as a desk review |
| Within two weeks after review | Contractor receives draft reports |
| Within two weeks of receiving draft reports | Contractor submits final reports to the Government |

**Applicable Performance Standards**
The acceptance of the contract deliverables shall be based on three performance standards:
(1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each ToR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

**Travel**
Since this is a desk review travel is neither required nor authorized for this contract.

**Restricted or Limited Use of Data**
The contractors may be required to sign and adhere to a non-disclosure agreement.

**Annex 1: Peer Review Report Requirements**

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

3. The reviewer report shall include the following appendices:
    a. Appendix 1: Bibliography of materials provided for review
    b. Appendix 2: A copy of the CIE Statement of Work

**Annex 2: Terms of Reference for the Peer Review**

*River Temperature Decision Support Tools*

1. Evaluation of the strength and weaknesses of the individual water temperature models as well as the process of linking the models, bringing attention to those weaknesses not adequately addressed in technical memorandum.

2. Evaluation of the methods used to incorporate uncertainty into predicting water temperature in Shasta Reservoir and Sacramento River down to Red Bluff, such as the use of variable meteorology and model parameters.

3. Evaluation of the water temperature model calibration and validation procedure outlined in the technical memorandum and its ability to properly parameterize each water temperature model.

4. Evaluation of the implication of this work as decision support tools, bringing attention to the any potential for miss-use or miss-interpretation of this information to aid in fisheries and water management in California's Central Valley

5. Evaluation of the content made available in the CVTEMP website, bringing attention to content that was unclear and that could be improved.

6. Provide a brief description on other aspects of the model not described above.