

Comment on hess-2021-218

Anonymous Referee #1

Referee comment on "Spatial extrapolation of stream thermal peaks using heterogeneous time series at a national scale" by Aurélien Beaufort et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-218-RC1>, 2021

Beaufort et al. address an important topic in their manuscript, "The thermal peak: A simple stream temperature metrics at the regional scale"—namely, how does one develop accurate stream temperature information for an area the size of France that could be used in climate assessments or research on thermal ecology of lotic species? To accomplish this task, the author assemble a large national database of temperature measurements, summarize these records using a single metric called "The thermal peak", link this metric to site and watershed level descriptions derived from GIS sources, and then model the dataset using four different approaches to compare and contrast the outcomes. What the authors have undertaken is ambitious and to be commended, but I do have several reservations about the manuscript in its present form, as outlined below, that could be addressed to improve its overall quality.

- Consider a revision of the title so that it better represents the research question and issues at hand because it currently is focused a relatively minor methodological issue relating to how temperature records are summarized.

- Abstract. Add or revise the lead sentence to that it also frames the research question more broadly. For example, why do we care about or need stream temperature information? Climate change, water quality standards, thermal ecology could all be drawn on as motivating factors. I also disagree with the claim made in the lead sentence, that "spatiotemporally comprehensive stream temperature datasets are rare..." because the literature is full of stream temperature studies, and there are now many grassroots and state sanctioned monitoring programs. What's really the issue is that the data are scattered among many entities and rarely organized into a central

database. The fact that the authors have built such a large database for France during the course of this research shows that stream temperature data are common, and the database itself is a valuable contribution.

- Introduction, line 40. There is mention made here of thermal regimes and their components (frequency, magnitude, etc) and that continuous records, preferably of extended length are needed for accurate regime description. I disagree that this is the case as lengthy records are primarily useful for trend detection, as might be the case when describing the effects of climate change. More importantly, from the perspective of this manuscript is that many of the dozens of metrics that are often used to describe thermal regimes are strongly correlated. Thus, it is valid to focus on one (or a small set) summary metric, model it, and know that your representing a lot of the information about overall thermal regimes. This is the point you should make here, these three papers all provide good examples of the strong correlations among thermal metrics. Steel et al. 2016. Spatial and temporal variation of water temperature regimes on the Snoqualmie River network. *JAWRA Journal of the American Water Resources Association*, 52:769-787; Rivers-Moore et al. 2013. Towards Setting Environmental Water Temperature Guidelines: A South African Example. *Journal of Environmental Management* 128: 380–92; Isaak et al. 2020. Thermal regimes of perennial rivers and streams in the Western United States. *Journal of the American Water Resources Association*, 56:842-867.
- Methods line 90. The authors state that “the large spatial and temporal heterogeneity of the monitoring data precluded application of spatial autocorrelation models...” This isn’t an accurate statement, SSN models are perfectly suited to this type of temperature database, as two of the studies cited by the authors demonstrate (Detenbeck et al. 2017 and Isaak et al. 2017). Nonetheless, it's fine not to use SSN models and rely on other approaches so I would just delete this sentence.
- Line 118, Thermal peak metric. Derivation of this metric seems far more complicated than it needed to be, while also discarding valuable information about inter-annual variability by averaging over multiple years of observations at individual sites. Because the dates of the 30 warmest days will be different each year, it also adds some inconsistencies and creates complexities for processing the temperature records. The same information about the thermal regimes could have been obtained from a simple mean July or mean August temperature metric.
- Methods, line 170. Reference is made here to a principal components analysis but it’s unclear how this was employed or the effects it had on excluding variables from consideration.
- Methods, Table 2, explanatory variables. It’s not clear from the manuscript text how some of these variables would affect stream temperature. Please expand this table with another field labeled “Hypothesized effect” and briefly explain the rationale for considering each variable in the models, preferably with supporting citations.
- Methods, lines 197-215 describing the analysis techniques. Please provide more detail. The minimum requirement is providing enough information that a knowledgeable reader could replicate the analysis. In the case of the multiple regression, for example, how was model selection performed (e.g., AIC based, stepwise, best subsets, etc.). Was the potential for problematic multicollinearity assessed by removing highly correlated explanatory variables?
- Methods lines 218-222 describing the multi-model combination. It’s unclear how this was done exactly. The authors state, “estimates from each previously described model

were..." Usually parameters are estimated, so I think you really mean "temperature predictions from each previously described model were..." Moreover, those prediction combinations were presumably done for each reach within the network, so there should be an "i" subscript in the equation notation to denote this.

Also useful for comparing the models would be a multipanel figure containing a series of bivariate scatterplots showing the pairwise predictions from each combination of the models with the associated correlations shown. These correlations are quite high presumably, but one could also further explore the discrepancies between model predictions by analyzing the residual differences relative to the predictor variables.

- Results lines 244-245. It's unclear where the air temperature model predictions of stream temperature came from. Is the air temperature model a simple linear regression with air temperature the single predictor of stream temperature? If so, it should be mentioned and described in the preceding methods section with the other model types.
- Results, lines 259-265. Relevance of explanatory variables in the models. Inconsistent terminology in this section makes it difficult to understand how the explanatory variables are being assessed. Initial reference is given to "Explanatory power", later in the paragraph "cumulative importance" is referenced, and the accompanying Figure 4 refers to "relative importance." Are these all the same things and/or do they reference the r^2 statistic? Please clarify. Also, it would be useful to expand Figure 4 to see the effects of all the variables that were important contributors to each model, and to know what the total explanatory power was of each model.
- Discussion section, lines 315-316. Because of the way the thermal peak metric was calculated and model fits were conducted, by using temperature observations averaged across years, the ability to estimate inter-annual effects due to variability in air temperatures and discharge was lost. However, the stream temperature dataset certainly contains that information and it may be important to recognize and estimate in future model iterations because it can enable climate change forecasting. A technique for retaining both spatial and interannual temporal variation in model fits to similar stream temperature datasets was employed in both the Isaak et al. publications the authors cite and might be referenced in this section of the discussion.
- Discussion section lines 330-341 concerning spatial extrapolation by random forest models. It would be useful to expand this section and bring more balance to it with a discussion of the pros/cons of the various model types. For example, random forest models are easy to apply but are also generally known to overfit such that they can accurately predict a set of observations but may see performance declines when predictions are made at unsampled locations. They also have less robust means of model selection and significance testing than say multiple linear regressions. In all cases, the performance of the modeling techniques used here was less than that of SSNs applied to similar temperature datasets, which typically have $r^2 \sim 0.90$ and RMSE ~ 1.0 C but SSNs are labor intensive to apply in comparison to non-geospatial techniques and require specialized geospatial skills to fit.
- Discussion section lines 355-357 discussing differences among models in which explanatory variables are important. This to me, is one of the challenges and potential disadvantages to using a multi-model approach. It can result in a muddled inferential picture and therefore which variables might be important to emphasize to land managers or conservationists that are concerned about habitat restoration actions for stream temperature. For the multi-model approach to offer significant benefits, it seemingly should provide more robust and improved predictive performance, while caution is exercised regarding the interpretation of variables affecting the response

metric.

- Discussion section, lines 361-370 discussing the use of air temperature as a proxy for stream temperatures. While the use of air temperatures was common one or two decades ago, it's become much less common in recent years with the broad availability of stream temperature datasets and interpolated map scenarios like the author's have created here. Towards that end, it would be useful to discuss how your datasets will be made available to others so they can benefit from them. The large temperature observation dataset would be of great utility to researchers conducting thermal regime research, whereas the thermal peak scenarios could be used by aquatic ecologists in France developing species distribution models or assessing vulnerability to climate change.