Journal: HESS
Title: The thermal peak: A simple stream temperature metric at regional scale
Author(s): Aurélien Beaufort et al.
MS No.: hess-2021-218
MS Type: Research article
Author Response

Dear HESS editorial review board,

We thank the reviewers and editors at HESS for the opportunity to respond to comments and to revise our manuscript based on those comments. We believe the manuscript is greatly improved thanks to the careful attention paid by the two anonymous reviewers, and we present our updated manuscript below. First, we show our previously made point-by-point response to reviewers in blue, but now also include in red text the direct changes that were made to the manuscript. We hope that this will serve as a specific list of the changes made as requested by the editor.

Sincerely,

Jake Diamond and co-authors

Anonymous Referee #1:

Comments

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.

1. 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.
   We agree with the reviewer and changed the title to the following:
   Spatial extrapolation of stream thermal peaks using heterogeneous time series at national scale
2. 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.

We agree and will change the text as recommended. Changed on L7–10.

3. 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.

   We thank the reviewer for this important clarification and will rephrase the text as recommended and include the suggested literature. We will replace "However, these metrics can be accurately determined only if continuous time series of stream temperature are available (Jones and Schmidt, 2018)." with
   "However, many of these stream temperature metrics are strongly correlated (references), implying the utility of a single metric to understand stream temperature regimes."
   Replaced on L34–41.

4. 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.

We agree and will remove the referenced text: "The large spatial and temporal heterogeneity of the monitoring data precluded application of spatial autocorrelation methods, and we have therefore chosen to consider only non-spatial statistical models." Removed.

5. 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.

We agree with the Reviewer that the thermal peak calculation is somewhat involved, but we point out here that we did not know a priori that July and August would be the hottest months. The method used here is analogous to the one used for developing macroinvertebrate/fish typologies of France (Buisson and Grenouillet 2009), so we will include this rationale and relevant citations in the revised methods. To support our approach, we will also include a sensitivity analysis on the sites with annual data where we compare the T_p of the annual time series with the T_p limited to July and August. Finally, we will add text in Discussion to suggest simpler approaches. Changes made for clarity on L124–135. New sensitivity analysis is Fig A1.

6. 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.

We removed this reference to the PCA as it was unnecessary. Instead, we will include the following correlation matrix in Supplementary Material to illustrate the independence of variables used in this analysis. Please also see our related response to comment 7. Removed as requested and correlation matrix is now Fig A3.

7. 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.

We agree that this would greatly improve the clarity of the variable selection. Please see below for a revised table. Table added as Table 3.

8. 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?

   We will improve the clarity and detail on the techniques used in this section. For the multiple regression, we did not use any variable selection techniques. Our goal was not to have the best possible regression, but instead to use the already determined independent variables (see response to comment 6) to compare across modelling techniques (i.e., regression, ANN, random forest, multi-model). We will add text to make this more clear here and at the end of the Introduction. Changes made throughout L206–259 to improve clarity on methods.

9. 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.

   Indeed, we intended "predictions" instead of "parameters", which we will clarify. The reviewer is also correct with reference to the predictions being reach-specific; we will therefore include an "i" index in the equation. To clarify with regard to the mutli-model approach. the predictions made by the three models are included and each model prediction is weighted with a coefficient to match the observations as closely as possible. Hence, the coefficients are calculated only in relation to observations, so only where there are stations. Then, using this equation (7) with calculated coefficients, we extrapolate to simulate reach-T_p at the network scale (see Figure 5). We will be more clear on this in the Methods section. Clarified on L234–247.

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.

We agree that this analysis would be useful and will include a version of it in the Supplementary Material. We already have some prototypes of this analysis that are spatially explicit, which may be more informative than the suggested scatterplots, which we include below. We included the bivariate scatterplots and correlations as Fig 3d–g.

10. 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.
    We agree that this was not clear and will correct this in Methods text and the variable definition table. To clarify here, there is no regression. The air temperature predictions are simply SAFRAN reanalysis data. Clarified on L257–259.
11. 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 r2 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.
    We will simplify this terminology and clarify in the Methods to better present this information. Indeed, throughout this section we are referring to the same variable importance as described in part 2.4.2–2.4.4. These importance values are then summed to get cumulative importance; it was therefore necessary to standardize these importance terms. We do not check the explanatory power of the variables in the prediction itself, but we look at which variable each model used to obtain its prediction. We have chosen not to present the many other variables in this figure for both visual clarity, and because the

other variables' importance are negligible, as is evident by the high cumulative importance of the variables shown. We will, however, expand on the total explanatory power of each model and discuss minor relevance of the other variables in the Discussion. Clarified in Methods, and entire Results section 3.3 was revised with added Fig 5 to improve clarity of effects of the most important variables. We did not expand Figure 4 as we deemed it unnecessary to our overall story, especially in light of our new Question and Hypothesis on L79–84.

12. 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.

We agree and will reference this shortcoming in the Discussion while citing these relevant publications. Added on L390–394.

13. 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 r2 ~ 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.

We agree that SSNs are useful in these applications, and indeed, we conducted some benchmark tests on small region well covered by data (9000 km², 92 stations) for a robust estimation of parameters with the R package *SSN* (see figures below). SSN performed better than the other methods (decreased by 0.2°C for SSN model compared to random forest), which was encouraging. Additionally, by comparing the observed and estimated values, we can see that RF tends to underestimate the high values and to overestimate the low values. Still, the spatial patterns are very consistent among the two approaches, though there are important differences between the SSN and RF model estimates which can be +/- 2°C. The estimates of the SSN model are generally warmer than those of RF on the main major river axes and colder on the small tributaries. This is consistent also with observations. Unfortunately, due to the lack of an RHT with upstream-downstream information, we could not apply SSN at the scale of France.

So, while the presented models may not be optimal, we are confident the spatial patterns are correct. We will include a more detailed discussion of the pros/cons of the different models with the possibility of SSNs. Discussion of SSNs added on L426–440, with added Fig A5 and Table A1 to provide comparison.

14. 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.

We agree and will include a discussion of the pros/cons of the different models. In the Methods, we will remind that the multi-model approach is frequently used to account for uncertainties in studies of climate change impact and in hydrological forecasting systems. This approach was borrowed from the modellers community and carried out thereafter for predicting flow characteristics in ungauged basins (see Razavi and Coulibaly (2016), doi:10.1080/02626667.2016.1154558 for a recent application). More reliable predictions at ungauged locations are expected by combining single model estimations. In the Discussion, we will specify that whereas the multi-model has the best performance, it lacks the explanatory power and relative simplicity of the other approaches. Another benefit of the multi-model approach is that by leveraging multiple approaches, it can compensate for errors particular to individual models. This entire section (4.3, L442–464) was edited to describe differences in model structure and effects on important variables.

15. 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.

We agree that air temperature is not as common as it once was, but would argue that it is still in use because datasets like the one presented here are still relatively rare. However, we will reduce some of this stronger language throughout. We will further include some additional text to discuss how the dataset can be made available and used by ecologists in France and scientists more broadly. We note that part of the database (approximately 600 stations) is publicly available from Naïades, which we now include this in the Methods. The majority of other data is sparse, typically only with summer information. We have created a website to be able to share this data, and will include its information in our revisions (thermie_rivieres.fr). Edits made as requested in this section (4.4; L467–487) and information about how our dataset is useful was added on L491–493 with a link to the dataset on L374.

Anonymous Referee #2:

Comments

In Beaufort et al. the authors compiled stream temperature data for the entirety of France and Corsica, calculated an ecologically relevant summarizing metric ("the thermal peak"), and compared predictive models and an air proxy to determine the best model structure and predictor variables for extrapolating the thermal peak metric to all rivers in France/Corsica. Although the work done within is a valuable contribution, the manuscript would be improved by (1) altering the climate-change framing of the paper, (2) including a specific question or hypothesis, and (3) increasing specificity and clarity of the presentation. Regarding the first point, the paper frames the research as important for understanding stream thermal regimes under climate change. However, the authors calculate a metric which does not help us understand how climate change alters stream temperatures. This is a major issue. Additionally, there lacks a clear question or hypothesis in the introduction – another major issue.

We agree that the manuscript would be improved by reducing the focus on climate change in-of-itself, and honing the Introductory rationale towards the need to improved stream temperature datasets (our contribution) while acknowledging that climate change as the motivating driver. We also agree that ending the Introduction with a question and hypothesis will create a stronger rationale for our approach. Please see below for our more specific changes that will be made. We clarified our approach and intent, and added a specific question and hypotheses on L73–83.

1. Comments regarding emphasis on climate change:
   The introduction leads me to think that the authors would be explaining how stream temperature is changing in France/Corsica as a result of climate change.
   For example, see: Line 36-37: "However, the magnitude and direction of these expected changes will depend strongly on patterns of stream temperature change"
   Line 38: "Hence it is critical to describe and analyze the spatiotemporal variability of river thermal regimes."
   Line 71-74: "Indeed, stream temperature metrics that focus on extreme periods […] are likely adequate to understand trends of increasing pressures on aquatic ecosystems."
   And in the Methods: Line 102-103: "To address ecologically meaningful temperature metrics under climate change, we focused on the two hottest stream temperature months,

…" However, the metric the authors calculate does not describe temporal stream temperature changes. The described calculation of the thermal peak is an average of peak temperature across years for each logging station. The metric is thus only useful for understanding the spatial pattern of the thermal peak across France/Corsica for the years in this study. Although the compiled database itself may be a useful source for studies on climate change, the thermal peak metric calculated in this study does not describe changes in stream temperature through time.

Indeed, the authors even admit to this shortcoming in the discussion: Line 315-316: "The downside of the current approach is that it remains based on interannual metrics. Indeed, the non-concomitance of the time series does not allow us to compare extreme years (hot vs. cold)." More emphasis should be put on the thermal peak being a good comparison metric for rivers across space and may help managers understand which rivers currently exceed thermal tolerances of important biota.

These are all good points and we will alter the Introduction and Discussion accordingly. We will focus more on the main point of the paper throughout, which is the creation of a homogenous stream temperature database and the development of a simple metric to understand spatial variation in extremes, which of course, are a result of increasing air temperature due to climate change. Hence, we will remove/edit most of the referenced text here so that the objective is clear. We also will add text to lead to the eventual comparison of stream temperature records/extrapolation with air temperature, as our dataset offers an important improvement over often-used air-temperature proxies. We reduced the climate framing and removed text as requested, particularly in the Introduction by removing that language throughout, particularly in L25–42.

2. Comments regarding lack of question/hypothesis:

a. An objective is stated in the introduction (Line 73-74) but no specific question, hypothesis or prediction is explained. This left me wondering throughout the paper what the purpose of the research was. What questions motivated this research? What did you expect to learn from research at hand? What do you expect the thermal peak tell us about rivers in France/Corsica?

We agree that these additions will improve the rationale for this work, and we will include the following question and hypothesis in the Introduction. What are the spatial patterns of stream temperature extremes in France and their drivers, and are these patterns consistent across modeling approaches? We hypothesized that spatial patterns would be consistent, whereas the drivers would depend on the modeling approach used. We also hypothesized that stream size, air temperature, and groundwater contributions would emerge as important regardless of approach. Overall, we maintain our objective to create a harmonized database of daily stream temperature in France over the recent decade of warming as an important result of this work. We will demonstrate the importance of this result now by focusing the Introduction more on the current lack of such databases in the literature. We clarified our approach and intent, and added a specific question and hypotheses on L73–83.

To evaluate our hypothesis, we will include an additional result showing the direction and magnitude of variable effects on the thermal peak (i.e., not just the variable importance for each model) that will allow us to evaluate their influence. To illustrate an example result here, we will compare the spatial variability of the thermal peak in relation to air temperature and stream size (see below). Specific evaluation of this hypothesis is now evaluated in Figs 5 and 7.

b. Section 4.3 of the discussion aims to attach meaning to the most important explanatory variables across each model. There was no hypothesis about which explanatory variables were thought to be the most important predictors, therefore it seems the authors are fishing for an explanation without having a clear question, hypothesis or prediction about these explanatory variables. Since the author's purpose for making these models was to predict the thermal peak across all rivers of France/Corsica, it is inappropriate to interpret the explanatory variables in this way.

We agree that variable selection was not clear in original text. We will improve our rationale for the explanatory variables used by being more explicit in our questions and hypothesis (see our response above), and by including specific hypothesized effects for each variable in Table 2 (see below). Indeed, each variable has a precedent for controlling stream temperature in the literature and was evaluated for collinearity (see correlation matrix that will now be in the Supplementary Material) before application in the models. We also now include more text in the Methods providing rationale for our variable choices, where some of the variables have substantial literature support, some have more recent, but less tested support (i.e., effects of ponds and weirs), and some we chose to test in this paper (i.e., concavity index and hydrological regime). These clarifications in our rationale for variable selection allow us to more readily attach meaning to their relative importance and effect direction (on T_p) in the Discussion. Table 3 now includes hypothesized effects, correlation matrix was added as Fig A3, and rationale clarified on L182–184.

3. Comments regarding improvements on clarity and specificity:

   a. Line 88, sentence starting with "Hence, our main challenge was to pool...": This sentence was difficult to understand. Improve with a more thorough explanation, possibly breaking up this sentence into multiple.

   We agree this is confusing. We will clarify that the challenge was coalescing and harmonizing all the disparate data sources. Clarified on L86–104.

   b. Line 99, "data were … averaged into mean daily stream temperature data". Do you mean "daily mean" here? It is unclear if it is meant to be the mean daily temperature (averaged across years for a particular day) or the daily mean temperature (average temperature of a specific date).

   Indeed, we meant the latter: that the means of hourly data were taken to achieve daily mean temperatures. We will be more clear here. Changed on L101.

   c. Line 100, "1) hourly Tw anomalies". If Tw is defined as the daily mean temperature, how can you see hourly anomalies if there is one mean temperature per day? May need to define another variable to use here.

   Thank you for catching this. We will change this to "hourly stream temperature anomalies" to avoid conflating Tw with the hourly data. Changed throughout.

   d. Line 101, "Tair": Need to define this variable because it is the first time it is used.

   Fixed. Fixed on L103.

   e. Section 2.2, first paragraph. Why hold back on defining the "simple metric" within the first sentence? It is odd to hold back on defining the metric and then reveal the big mystery of what the metric is.

   We do not quite follow this comment, but we will add "…, the thermal peak." In the first sentence to clarify. Added on L124.

f. Section 2.2, second paragraph. $\overline{T_{w,30}}$ is defined as the "mean temperature of the 30 hottest consecutive days of each year". However, the data have already be constrained to July and August. How can you be sure that the 30 hottest consecutive days occur completely within the window between July 1 and August 31? In the last sentence you conclude that the hottest day of the year always occurred between July 28 and August 30. This implies that for the temperature time-series in which August 30th is the hottest day of the year, the 30 hottest consecutive days would include days in September. So, $\overline{T_{w,30}}$ should be re-defined in the paper as the mean temperature of the 30 hottest consecutive days in July and August. Be more specific about defining $\overline{T_{w,30}}$.

We now include in the Supplementary Material that our analysis shows that July and August are generally the hottest months of the year (see below for analysis on stations with annual data). Importantly, many of the stations, particularly those operated by fishing agencies, only have July and August data, which we will precise in the Methods. We agree that we should be more specific and clear in this section and will make the appropriate changes to the text noting that analyses are constrained to July and August. To support our approach, we will also include a sensitivity analysis on the sites with annual data where we compare the T_p of the annual time series with the T_p limited to July and August. Text clarified on L132–137 and Supplementary Fig A1 added.

g. Line 164, "SAFRAN data". Define what this data is.

We will include a definition in text that explains the following: (Système d'analyse fournissant des renseignements atmosphériques à la neige) is a mesoscale atmospheric analysis system for surface variables with reanalysis data at hourly time steps using ground data observations. Originally intended for mountainous areas, it was later extended to cover France. The detailed description can be found in Durand et al. (1993, 2009). Clarified on L174.

h. Line 198, "… all possible variables." Very broad. More specificity would be helpful, such as "all possible variables characterized in Table 2."

Agreed. Fixed.

i. Line 236-238, "This bias … in the regressions." These sentences are difficult to understand. Be more specific to improve clarity. The ending of one sentence is "…with only one year of observation.", and the next sentence begins, "In contrast, when there is only one year available…". These aren't contrasting clauses, thus making it difficult to understand what is being contrasted.

Agreed, we will improve clarity here. Clarified on L271–274.

j. Sentence starting on line 260, ending line 263, "The two most relevant…". Break into 2 sentences to improve clarity.

Agreed, we will improve clarity here. Clarified on L300–307.

k. Line 281: use of 14°C. Explain why 14°C important and why this value chosen as a cut-off point.

This was predominately chosen for ease of visualization. We will explain this in the figure and in text. Described on L362.

l. Figure 6C. Figure 6c shows probability distributions for extrapolated thermal peak. Firstly, the bins are not of equal size: the first bin (labelled <14) includes all temperatures between 6.3 (minimum temperature reported on line 277) and 14, a range of 7.7 degrees; the second through fifth bin have ranges of 3 degrees; and the final bin includes temperatures from 22 to 27, a range of 5 degrees. The 14- and 22-degree cutoff is not explained and thus seems arbitrary. Secondly, since temperature is continuous, it shouldn't be binned in this way, unless for some specific reason. Continuous distributions should be portrayed as "probability density functions". Each model would then have a continuous distribution, and each can be plotted on the same graph for comparison.

Originally, it was made this way to correspond to the cut-bins in Figure 5, which were made to best visualize the temperature distributions. We will made the requested edits to the figure. This figure is now Supplemental Fig A4 and is correctly binned.

m. Line 301, first sentence in discussion. Be more specific about the timeframe of the dataset: "the largest, regional summertime stream temperature datasets"

Agreed, we will include the timeframe of 2008–2018. Fixed on L372.

n. Line 326. Which "metrics based on observations", specifically, are you comparing to estimates of Tp?

We will more specifically refer to the fact the thermal peak calculated with observed data containing gaps $T_{p,obs}$ (i.e., without the climate correction) has larger biases than thermal peaks calculated with observations that are climate corrected $T_{p,fill}$. Clarified on L402–404.

o. Line 328. Further explain what is meant by "when applied at scale". How does that relate to figure 6B?

Good point, removed. Removed.

p. Line 352: "Higher minimum flows" would make more sense than "Larger minimum flows"

Agreed and changed. Changed on L455.

q. Line 354: Change "greater" to "more". "Great" implies a quality of shade and not a quantity. "More" implies a quantity.

Agreed and changed. Changed on L457.

4. Technical Corrections

a. Line 21: Use of "However" does not belong because this sentence doesn't relate to previous

b. Line 27: "…, and decomposition rates, and dictates animal … "

c. Line 33: "and levels" is repetitious with flows

d. Line 96: "… were previously…" change to "have been". "were previously" implies that a different study excluded these data. If that is the case, cite that study.

e. Line 155: "is not clear" should be "was not clear"

f. Line 221, equation 7: regression coefficient "a" is in the equation twice and "c" is missing.

g. Line 229: "cross-validation this 100 times"

h. Line 321: insert "between" between "correlation" and "Tair"

i. Line 325: remove "particularly"

j. Line 326: remove "a robust interannual metric". Only need to say "... sufficient to estimate Tp"

k. Line 329: change "to" to "of" in "climate corrections to temperature metrics"

l. Line 352: change "following" to "followed"

m. Line 373: change "representativeness" to "representation"

All suggested technical corrections will be made. All changes were made.