Beaufort et al. Review of Revised Manuscript

While improvements were made in the revised version of Beaufort et al., the manuscript does not adequately describe the utility of the "thermal peak" metric (comment 1). The addition of a question and hypotheses were improvements but were tacked on the end of the introduction with little relevance to the presented problems and no rational as to how the authors came up with them (comment 2). Additionally, the manuscript lacks clarity, uses imprecise language, and has many technical errors that need to be addressed (comment 3).

Comment 1

(A) The authors are continuing to frame their metric as useful for understanding climate change, see lines 70-75:

"However, for management purposes in the context of climate change, annual stream temperature patterns may be unnecessary. Indeed, stream temperature metrics that focus on extreme periods (e.g. summer) are likely adequate ***to understand trends of increasing pressures on aquatic ecosystems*** [*emphasis added*].

To that end [*emphasis added*], our objectives are twofold: 1) to create of a harmonious stream temperature database for France, and 2) to develop empirical statistical models to predict a simple, ecologically relevant stream thermal metric that captures the magnitude of the stream temperature extremes at the regional scale."

As stated in my first review, the metric the authors are calculating in the study does not lead to better understanding of trends of increasing pressures on aquatic ecosystems (as promised in line 72). The "thermal peak" metric removes all temporal variation and thus any potential for understanding how summertime temperature extremes are trending due to climate change because the metric takes the mean across years. I would encourage the authors to really consider what this metric *can* and *cannot* tell us about stream temperature regimes. Again, I don't deny that the thermal peak metric is useful for understanding spatial variation in temperature extremes and this metric may be useful for management decisions. However, I disagree with the author's statement in their response to review that the spatial variation in stream temperature extremes are a result of increasing air temperature due to climate change. Imagine a scenario where the climate, and thus air temperature, has not changed, we would still expect there to be spatial variation in annual maximum temperatures across different streams and even in different locations, longitudinally, in the same stream (See Fullerton et al. 2015 for examples of spatial variation in stream channel temperature). (Referenced comment: "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*** [*emphasis added*]**.**")

(B) Lines 383-385: "…we employed a combined empirical approach that allowed us to identify, at the regional scale, ***a map of summer stream temperature maxima with important implications for aquatic species distributions under climate change*** [*emphasis added*]**.**"

What exactly are these important implications? A thorough explanation after this sentence or an example of how the authors imagine the map being used would provide much needed clarity on this point.

(C) Lines 393-394: "This map can presently be used to ***predict and manage future cold-water habitat streams*** [*emphasis added*], with potential for regular multi-annual updates."

I don't see how this map of summertime maximum stream temperature averaged over 2009-2017 can be used to predict where future cold-water habitat streams will be. I can make assumptions about what the authors are getting at here, but that is left unsaid. I am assuming the authors mean: This map can be used as an initial benchmark for understanding the summertime temperature extremes of streams across France and upon update with new data, we can determine where the thermal peaks are increasing the greatest, thus allowing managers to focus their efforts are the streams with greatest increase in thermal peak. This is possibly not what the authors meant here, but either way, the reader shouldn't have to guess as to what the authors are getting at. I do see the utility of the database, the presented maps, and their analysis of the different models, however the utility is not clearly presented in the manuscript.

Comment 2

In the Introduction, the authors lay out multiple problems: (1) lack of long-term stream temperature data, which is often spatiotemporally uncoordinated (see L 44-46), (2) stream temperature data are often supported or replaced by air temperature proxies (see L 46-47), and (3) other statistical gap-filling methods require a lot of data, which may be unnecessary for management purposes in the context of climate change (see L 66-67, 70-71). Later, the authors describe their question as "…what are the spatial patterns of stream temperature extremes in France and their drivers, and are these patterns consistent across modeling approaches?" (L 79-81). Previous to this question, there has been little mention of the drivers of stream temperature patterns nor a mention of the different modeling approaches in question. The Introduction would be greatly improved if the problems were more specific to their question. Additionally, the authors did not explain the rational for the listed hypotheses. Why would they expect the spatial patterns to be consistent across modeling approaches and the drivers predicted by the models to be different? The lack of rational for these hypotheses and the fact that their results exactly confirm this hypothesis leads me to think the authors retroactively came up with this hypothesis after they got the results they did. Although this may be an inaccurate assessment by me, the way the hypotheses are presented make it seem so.

The second hypothesis lists "stream size, air temperature, and groundwater contributions" as important to the thermal peak metric (L 82-83, 264-265), however, the manuscript uses drainage area "area" and mean monthly minimum specific discharge "q_{min}" as metrics for stream size and groundwater contributions, respectively. "Stream size" is a somewhat general description and can be described using many different metrics, thus use "drainage area" in hypotheses because that is the specific metric being used. In section 2.4.1, q_{min} is said to describe the low-flow regime of the site and later in the results this metric is used to describe groundwater contributions. The manuscript does not provide citations or rational as to how q_{min} is a metric of groundwater contributions in streams nor does the writing

describe that they are using q_{min} as a metric for groundwater contributions previous to results, thus the reader must make assumptions after the fact. Again, this could be solved by using "mean monthly min. discharge" in the hypotheses or a description of why the authors use q_{min} as a metric for groundwater contributions.

Comment 3

Additionally, there are numerous technical corrections, from erroneous inclusions/exclusions of words (e.g., L 73 "create ~~of~~ a harmonious", L 210 "we did [not] seek", and others), to mixed verb tense (e.g., L 104-115 use of "have" instead of "had", L 337-339, L374-376, etc.), spelling errors (e.g., L 340 "mulit-model", L 355 "wateshed", and others), and a general lack of precise language, e.g. (list not exhaustive):

- (L 8) "…data are decentralized…": what does "decentralized" mean in this context?
- (L 32-33) "… with different metrics to quantify the biological or ecological importance of each component": unclear what the authors are talking about here
- (L34) "…these metrics…": do you mean components here? The list in the previous sentence was "components" of thermal regimes
- (L 75) "… metric that captures the magnitude of the stream temperature extremes…": the phrase "stream temperature extremes" includes a maximum **and** a minimum, and the metric presented only addresses the annual stream temperature maximum.
- (L 94) "…harmonize all the disparate data sources.": what exactly is data "harmonization"?
- (L 98) "…stations without seasonal dynamics have been excluded…": What was the determining criteria for "no seasonal dynamics"?
- Table 2. Definition of "n" as count of days. These numbers seem like they are the numbers of sites used, not the number of days.
- Table 3. In "Hypothesized effect" of CI and HR. These are not hypothesized effects, just definition of terms.
- Figure 7. It is unclear why (a) and (b) are presented together here. The difference in models is shown in (a) and the difference in drainage area for one model is shown in (b). Why was the RF model chosen to display in (b) and not the other models?

As well as areas that could use an improvement to organization:

- (L 385) "We note here…": this sentence seems out of place here and is lacking a rational for the method that you *did* decide on. After reading this sentence I'm left wondering, well why didn't you just a simple mean of August data?
- (L 487) "Groundwater and shading proxies …": this sentence also seems out of place.