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

Responses to Referees

Anonymous Referee #1:

We thank Referee #1 for their detailed review of our manuscript. We have broken out your individual comments, which are numbered, and responded to each accordingly in blue. We hope that our comments address and clarify any issues or concerns that they may have.

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 are considering alternative titles including the following:
   Spatial reconstruction of simple stream temperature metric at regional scale: the thermal peak
   Spatial extrapolation of stream temperature thermal peaks using heterogeneous time series
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. For example: "Spatial reconstruction of stream temperature is relevant to water quality and fisheries management, yet large, regional scale datasets are rare because data are decentralized and not coherent scattered among many entities."

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.”

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.

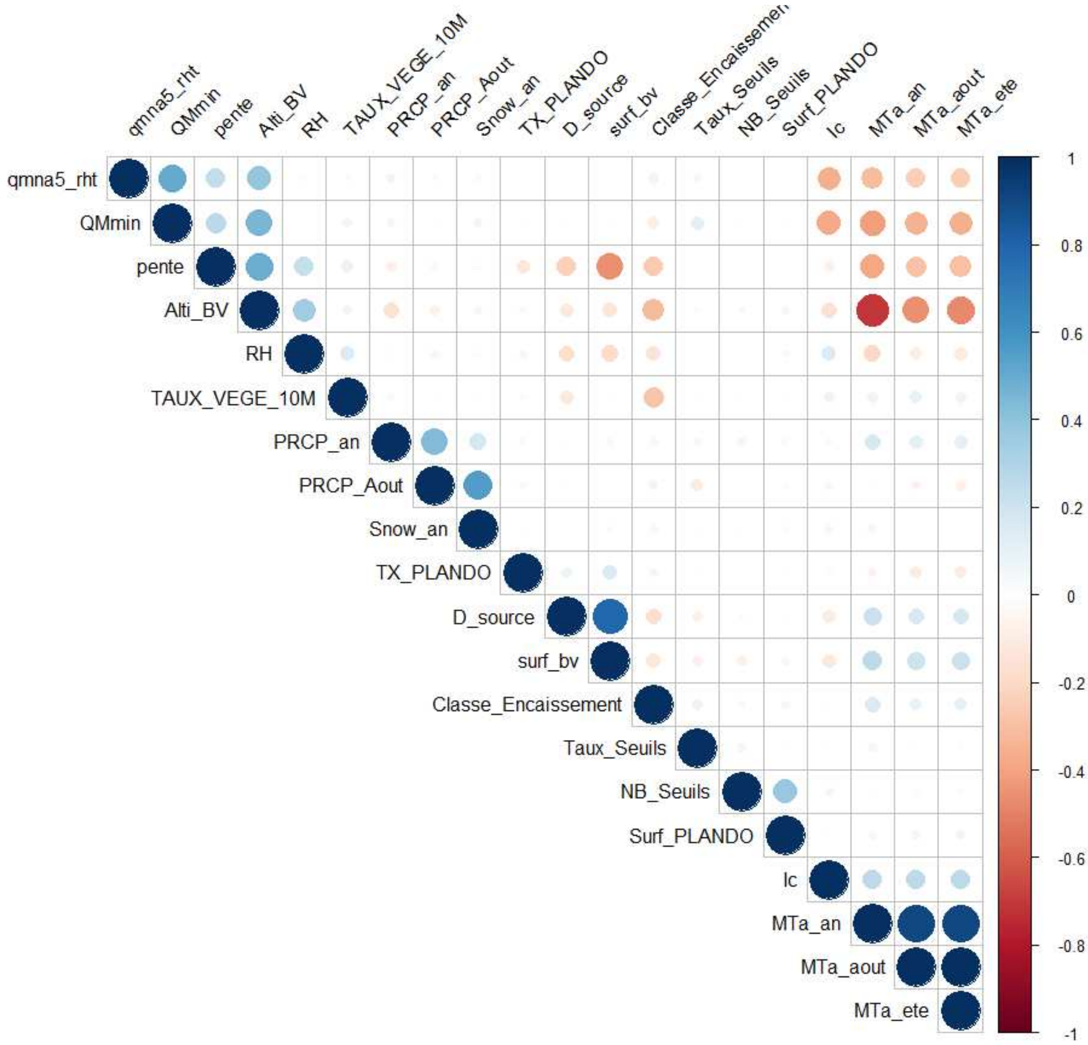
   We agree and will remove the referenced text: “The large spatial and temporal heterogeneity of the monitoring data precluded application of spatial autocorrelation methods, and we have therefore chosen to consider only non-spatial statistical models.”

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.

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.



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 2. List of explanatory variables used in models.

Category	Variable	Notation	Source	Possible effect
Climate	Mean annual precipitation (2009 –2017) [mm]	P_{annual}	SAFRAN	Contrast between climatic regimes (Moore et al, 2013)
	Mean summer precipitation, July–August (2009 –2017) [mm]	P_{summer}	SAFRAN	Influence of heat budget Caissie, 2006
	Mean annual snow accumulation (2009 –2017) [mm]	S_{annual}	SAFRAN	Heat budget, meltwater influence, Caissie, 2006, Webb et al, 2008
	Mean summer air temperature, July–August (2009 –2017) [°C]	T_{summer}	SAFRAN	Positive effect related to heat budget, relative Moore et al, 2013 (index of the thermal summer climate)
Hydrology	Mean annual specific discharges [L s^{-1} km^{-2}]	Q_{mean}	RHT	Thermal capacity influence, Caissie, 2006
	Mean monthly minimum specific discharge with a return period of 5 years* [L s^{-1} km^{-2}]	q_{min}	RHT	Proxy of base flow index, Chang and Psaris, 2013
	Concavity index† [-]	CI	RHT	Proxy of water storage in the catchment (snow or groundwater), tested in this paper
	Hydrological regime‡ [-]	HR	RHT	Contrast between hydrological regimes, tested in this paper
Catchment characteristics	Mean catchment elevation [m]	elev	RHT	Negative effect given the relation with air temperature (Isaak and Hubert, 2001)
	Drainage area [km²]	area	RHT	Proxy for width-depth ratio of streams (Hrachowitz et al, 2010), and thermal capacity (Imholt et al, 2013)
	Mean streams slope over the catchment [m km^{-1}]	slope	RHT	Affect river hydraulics and thus thermal advection and exposure time to incoming radiation (Daigle et al, 2010)
	Riparian vegetation cover ratio in 10 meters buffer (%)**	veg	SYRAH	Negative effect, as decrease exposure to diurnal radiation (Moore et al, 2005)
	Linear weir density upstream of stations (# km^{-1})**	weirs	SYRAH	Potentially heating effect (Chandesris et al, 2019)
	Areal weir density upstream of stations (# km^{-2})**	weir area	SYRAH	Potentially heating effect (Chandesris et al, 2019)
	Pond cover ratio upstream of stations (%)**	ponds	SYRAH	Potentially heating effect when ponds and shallow reservoirs release warm water from overflow (Seyedhasemi et al, 2021)
	Stream incision class **	SI	SYRAH	Potential proxy of exchange of water with hyporheic environment, Webb et al, 2008

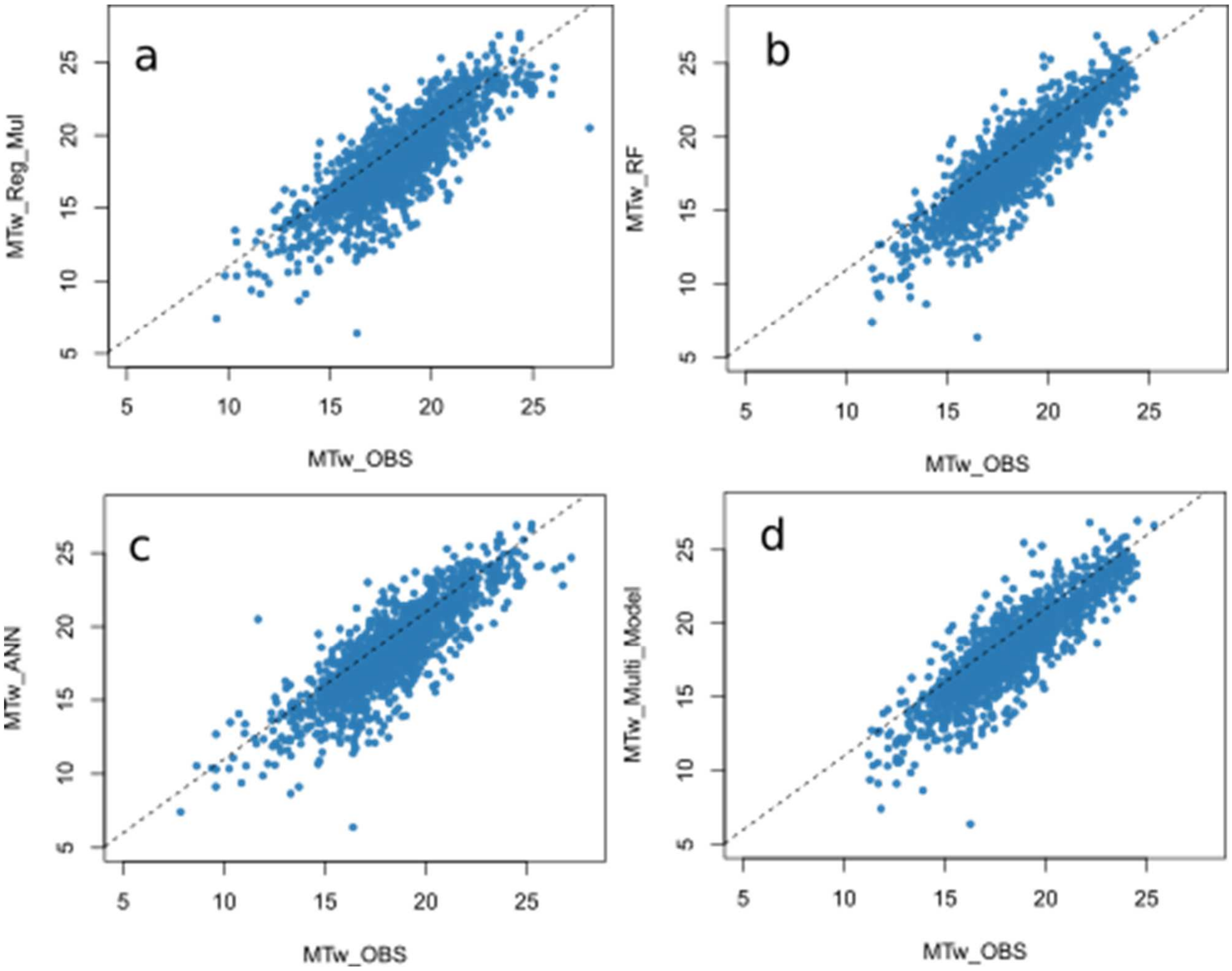
8. Methods, lines 197-215 describing the analysis techniques. Please provide more detail. The minimum requirement is providing enough information that a knowledgeable reader could replicate the analysis. In the case of the multiple regression, for example, how was model selection performed (e.g., AIC 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.

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.

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.



(a) Multiple regression model vs. observations ; (b) random forest model vs. observations ; (c) ANN model versus observations and (d) Multi-model combination versus observations

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.

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.

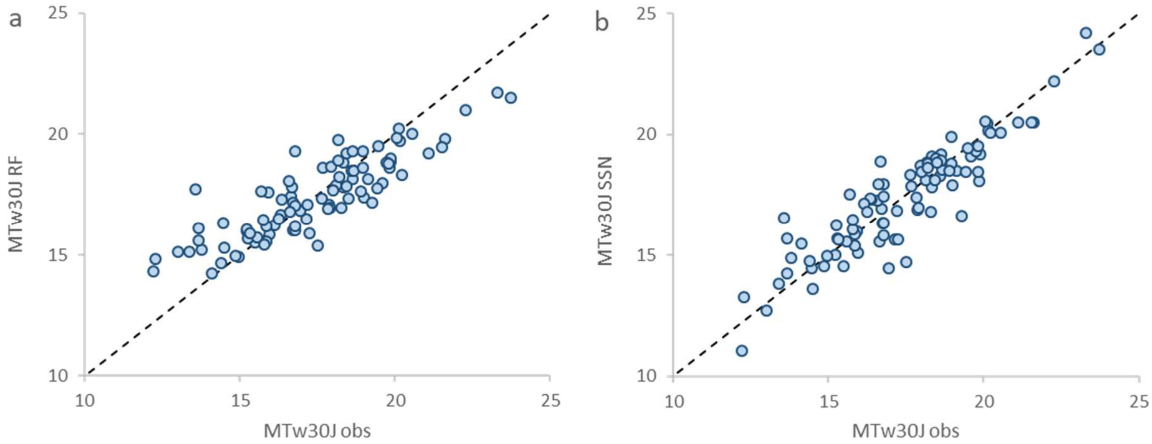
12. Discussion section, lines 315-316. Because of the way the thermal peak metric was calculated and model fits were conducted, by using temperature observations averaged across years, the ability to estimate inter-annual effects due to variability in air temperatures and discharge was lost. However, 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.

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.



Model	RMSE	NSE	Biais	BiaisAbs
RF	1.24	0.44	0.01	0.95
SSN	0.99	0.81	0.05	0.76

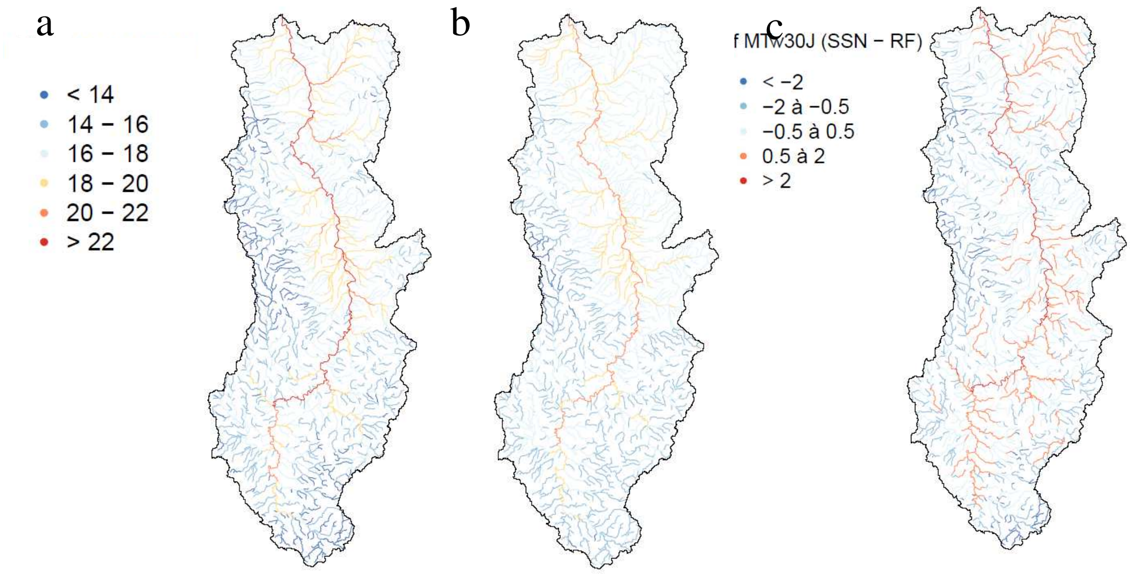


Figure comparing thermal peak estimates from (a) SSN, (b) RF, and (c) the difference between RF and SSN.

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.

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).