Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-171-AC1, 2016

Hydrology and Earth System Sciences Discussions
Open Access

Interactive comment on "Climate and basin drivers of seasonal river water temperature dynamics" *by* C. L. R. Laizé et al.

C. L. R. Laizé et al.

clai@ceh.ac.uk

Received and published: 9 December 2016

(1) P.2,l.26-27: irrelevant. Can be condensed.

Information requested by journal editor to acknowledge paper builds on first authors' PhD thesis. Sentence condensed as "This paper extends Laizé (2015)".

(2) P.3,l.19-20: possible, but please be more specific and add some reasoning.

The statement is backed up by a reference (Caissie, 2006).

(3) P.3,l.21, figure 1: perhaps add the symbols from equation 1 to highlight more which process is related to which heat flux.

The figure is being revised as suggested by the reviewer.

(4) P.3,l.31: Hrachowitz et al. (2010) would also fit in nicely here.

Citation added.

(5) P.4,l.4-6: I found this a bit exaggerated. There are in fact quite some studies that consider a range of catchment properties (e.g. Isaak and Hubert, 2001; Scott et al., 2002; Moore, 2006; Nelitz et al., 2007; Hrachowitz et al., 2010; Isaak et al., 2010). Please tone down and add at least these references.

The main point was in fact that they were very few studies in the UK (Hrachowitz et al. (2010) being one, and actually already cited in Table 2), and not that many, relatively speaking, internationally (suggested references are largely focusing on North America). Sentence (page 4 l 7-9) edited accordingly, with additional references (except for Scott et al. (2002) and Moore (2006), which we could not find based on the name and year only).

(6) P.4,l.25: table numbering is wrong. Table 2 is not referred to at all in the manuscript.

It seems there was a technical glitch when preparing the manuscript for upload. Indeed, current Table 2 was marked for deletion so that current Table 3 should have been Table 2, etc. We corrected the manuscript by deleting Table 2 and updating table numbers and references accordingly.

(7) P.4,l.29: "addresses" is unclear, maybe better to use "limits" or something similar

Text changed as suggested above.

(8) P.5,l.20: figure numbering wrong: figure 3 referred to before figure 2. Please also make this figure a bit more informative. Provide basin/river names and potentially include elevation information. Please also clarify why some observation sites are far from streams (e.g. in insets 2 and 3).

Current Fig 3 was meant to be Fig 2, and vice-versa, so was correctly referred to first. We swapped figures 2 and 3, and corrected numbering accordingly. All sites are on streams, but we only had access to a simplified river shapefile, which does not show smaller streams. A similar map was made with elevation as a background so could be provided as a replacement. Obtaining a more detailed river network to improve the inset is likely to take a significant amount of time.

(9) P.5,l.26: please provide more information on the actual data acquisition. Were the recorded values instantaneously measured temperatures or the averages over the logging intervals? How were the different sensors from the different studies placed and protected against radiative overheating? What about systematic uncertainties introduced by differential vegetation- and/or topographic shading at the different sites? Were the recorded data from the different studies pre-processed differently (e.g. filtering out overheating extremes)? What do different measurement precisions and accuracies of these different data sources imply for the analysis here? any systematic errors to be expected? And if not, why?

In Section 2.1, we cited the peer-reviewed papers related to the original datasets and covering the data acquisition. We also gave summary information. We feel that giving further details would require too much space. However, we clarified that fact that measurements are instantaneous whether they are manual or via a logger (l29, page 5). Regarding systematic differences between sites due to different recording processes, site characteristics, etc. (which are indeed to be expected), this was the main reason to use multi-level models. Multi-level models are models that take into account data structure; for example, if you had 2 sites, one shaded, one not, the regression slope and intercept for each site would be different to reflect that one site is, let's say, cooler on average and more responsive than the other.

(10) P.6,l.6-7: Please be a bit more specific. How was precipitation regionalized based on rain gauge data? Kriging? IDW? Thiessen? Other methods?

The text has been edited to clarify that precipitation data were derived from observed rain gauge data by using the natural neigbour interpolation method, which is a devel-

opment of the Thiessen approach.

(11) P.6,section 2.2: what about the uncertainties arising from the modelled climate data? How do they propagate through the temperature analysis here? Do they affect the overall interpretation?

The climate data are in fact deterministic (one set of climate data); some of the variables are in fact interpolated based on observations (eg precipitation), and we fitted one time series with other time series. In this sense, we did not analyse uncertainty as one may do with GCM outputs generating ensemble runs of several thousands. If one think in terms of how good CHESS data represent climate variables, we checked with our colleagues and they are of the opinion that the main weakness in the CHESS data was the downscaling of MORECS data from 40km to 1km, which may cause some variables to be overestimated in some parts of the UK; however, we had no sites located in those parts. Given the models performed reasonably well at predicting the observed water temperatures (conditional R-squared obetween 0.84-0.96), we consider that any uncertainty is acceptable and does not affect the overall analysis massively. In addition, with multi-level modelling, confidence intervals, although they can be calculated, are not considered as meaningful as for standard regression models.

(12) P.6,l.20-22: what is the reasoning behind investigating seasonal averages? Why only these? What is their ecological relevance? What about seasonal average daily (or 7-daily) maxima and minima? Would these not be more instructive? Just wondering.

Ecological relevance is with regards to phenology. A clarification on this was added in the introduction, page 3, l31. In addition, research fitted within a wider research on seasonal hydro-climatic patterns (eg Laize & Hannah, 2010). Minima and maxima would be of interest if investigating topics like lethal thresholds.

(13) P.7,l.4-5: where is this section? I cannot find it. This is relevant information and needs to be shown.

This information is actually in the Results section, not Methods (text was corrected). See comment 25 below; reviewer suggested this should move to the Methods section.

(14) P.7,l.5: what is meant by “permeability”? permeability of what? How was it determined? Catchment permeability in the sense of flashy impermeable catchments vs groundwater-fed catchments.

We added a clarification in the text. It is characterised by using catchment base flow index (BFI; described later in the text).

(15) P.7,l.6: not clear what is meant by this sentence

These basin properties are generally recognised in UK studies (those cited and many others) as modifiers of climate-hydrology associations. Sentence rephrased for more clarity.

(16) P.7,l.23ff: how was the spatial correlation structure between sites along the same rivers accounted for? What was the flow distance between the sites closest to each other?

It was taken into account by using multi-level modelling. As explained in the method section, the multi-level models were specified with 3 levels: data, data at a site, sites on a river. With the level representing rivers, the multi-level models were able to take into account the fact that two sites on the same river may have more similar records than two sites on different rivers due to their physical linking.

(17) P.7,l.24: it should at least be mentioned that linear models, in particular for the air-water temperature relationship, are oversimplifications and that for example logistic models can much better account for effects such as evaporative cooling (e.g. Mohseni et al., 1998).

We added this point in section 3 methods (and reference to Mohseni et al., 1998).

(18) P.7ff, sections 3.1, 3.2: I found this quite hard to follow. I would like to encourage

the authors to invest some more effort to describe this critical part of their analysis more clearly.

We appreciate the methods may be difficult to understand. We reviewed these sections attempting to clarify what we think may be the more confusing points (the reviewer’s comment is not specific in this regards).

(19) P.8,l.21-25: so how were the various combinations tested? Stepwise regression or best sub-set regression or some other method? What is AICc? How was it corrected for small-size datasets?

As described, all combinations were fitted (programmatically) and their AICs calculated. No stepwise or subsetting involved. The model with the lowest AICc was retained. Text was slightly edited to clarify the process. AICc is one standard option in R; the difference with AIC is a slightly modified formula putting more penalty on the number of parameters than with normal AIC. This is the suitable thing to do in this study (there are accepted rules about number of predictors v sample size). To the non specialist, this is actually a minor technical detail, which has been included for completeness. Detailing AICc any further would require to detail AIC, which is itself fully described in Akaike (1974); AIC and AICc are nowadays standard tools of the trade.

(20) P.10, section 4.1: not clear which explanatory variables were used in the individual models. All?

Yes, some predictors were used in all models (RI = 1), some predictors were used in some of the models constuting a set of best models (RI < 1).

(21) P.11,l.1: what does “adequately” mean? Please provide R2 and p-values. When using multi-level modelling (and moreover within the multi-model inference framework), R2 as commonly featured for regressions are actually not suitable, hence the current choice of showing observed vs predicted plots only. However, to address the reviewer’s point, there is an alternative R2 for multi-level models (“conditional R2”) by Nakagawa

and Schielzeth (2013), which we calculated and added to section 4.1 (reference to Nakagawa paper added too). Word "adequately" removed. Regarding p values, their conceptual equivalent within a MMI framework is the predictor relative importance RI used in the paper. A sentence in the method section 3.2 (l 25-26, page 9) and one in section 4.2.1 (l15-17, page 11) have been added to highlight this, and clarify that higher RI means more significant predictors.

(22) P.11,l.3ff, section 4.2: one thing that is completely missing here but that may be of considerable relevance is the potential collinearity (or correlation) between the predictor variables, which can potentially result in highly unstable and misleading model results. It will therefore be necessary to quantify the collinearity and evaluate to which degree it actually influences the results.

We were aware of possible collinearity issues and this was one of the reasons to use MMI. Collinearity gives inflated standard errors of parameter estimates. Approaches like MMI are fairly robust to even high levels of collinearity (see for example, Feckleton, 2011; Grueber et al, 2011); simply put, if there is high collinearity between two variables, then they don't appear in the same model, and don't force standard errors up. In addition, in this case, correlations between predictors were for most pairs below 0.5 (Pearson), and even for the more highly correlated pairs still within what is generally regarded as reasonable by statisticians. In the results/discussion, we aimed to take into account the implications of predictors co-varying when interpreting observed patterns.

(23) P.12,l.2: please clarify: are the percentage contributions in fact the proportions of the explained variance?

Sentence was slightly edited to clarify: for each record in the dataset, WT was predicted using the average model coefficients from Table 3, then the % contribution of each predictor to predited WT calculated (ie we made a time series of WT predictions and predictor % contributions).

(24) P.12,l.4, figure 5: please provide a unit for the y-scale in the figure. The unit of the

x-scale (%) seems to be wrong here. In addition, please be more specific: % of what?

The y-axis label ('%') was erroneously placed on the x-axis; figure amended. Captions explain what the % are for both sets of plots (we found that trying to abbreviate the % definition to fit it as an axis label did not really improve the figures).

(25) P.13,l.19ff: this needs to go into the methods section. Please also clarify why exactly these properties were chosen and provide a table with the relevant values.

The section on basin properties was originally in Methods but, because part of the analysis was used to confirm the selection of FEH properties (out of the available 19 properties) and because the Methods section was already quite long, we felt that it could be considered part of Results instead. We propose to move this back to Methods, where readers would more likely expect it. Regarding values, we assume the reviewer means the actual property values per site: we have originally included ranges of values only to save space (section 4.3) but we could include a table, perhaps as an appendix or as supplemental material. See also response to Reviewer #2 comments below.

(26) P.13,l.20: elevation not only related to wetness but clearly also to air temperature

We added this comment when elevation is introduced.

(27) P.13,l.26: area is proxy for discharge and thus for thermal capacity, but is also linked to elevation

We added this when the property is introduced. Note that this is also already mentioned in the Discussion.

(28) P.13,l.27: what is the reasoning behind using HOST/permeability? What is it expected to explain?

We were expecting groundwater-fed catchments to behave differently from impermeable ones (eg temperature regime influenced by groundwater inputs). We added this when the property is introduced. It is also covered in the Discussion.

(29) P.14,l.8: please also provide the individual p-values!

Please see response to similar comments re using MMI and selection with an information criterion (ie p values are not relevant). The models here were selected using MMI as per the main WT models. But, unlike for the main models, for which only the average model was featured, Table 6 lists all the models (and their R-squared and AIC) included in a MMI model set. This may give the impression the models were fitted using traditional approaches (eg removing predictors with high p values as not significant) although it was not the case.

(30) P.14,l.14-15: this is a sweeping generalization which needs to be toned down

Sentence revised.

(31) P.14,l.16: why should there be more small basins at higher elevations? Channel formation does not have anything to do with elevation, but rather with contributing area and local slope. There may be some correlation with elevation but it is not generally valid as posed here. what, however, is true is that, necessarily the opposite is true: there are more larger basins at lower elevations.

Sentence changed as suggested.

(32) P.16,l.5-6: this is possible, but not sufficiently substantiated by data here. I would argue that it is equally likely the indirect correlation is merely a model artifact without physical meaning (and potentially related to collinearity).

An early version of the manuscript actually made that very point. We re-instated it but kept the possible physical explanation as well.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-171, 2016.