Authors' response (Comment, Response, Change); page and line numbers refer to the marked-up revised manuscript.

Response to Reviewer #1

Comment #1: P.2,l.26-27: irrelevant. Can be condensed.

Response #1: Information requested by the journal editor to acknowledge that the paper builds on the first author's PhD thesis.

Change #1: Sentence condensed as "This paper extends Laizé (2015)".

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

Response #2: The statement is backed up by a reference (Caissie, 2006).

Change #2: None.

Comment #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.

Response #3: Figure revised as suggested.

Change #3: New figure inserted and caption edited accordingly.

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

Response #4: Agreed.

Change #4: Citation added (P3, L32 in revised manuscript with Track Changes).

Comment #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.

Response #5: 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).

Change #5: The sentence (P4, L5-9 in revised manuscript) was edited accordingly, with suggested additional references added, except for Scott et al. (2002) and Moore (2006), which we could not find based on the name and year only.

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

Response #6: It seems that there was a technical glitch when preparing the manuscript for uploading. Indeed, current Table 2 was marked for deletion so that current Table 3 should have been Table 2, etc.

Change #6: We corrected the manuscript by deleting Table 2 and updating table numbers and references accordingly.

Comment #7: P.4,l.29: “addresses” is unclear, maybe better to use “limits” or something similar.

Response #7: Agreed.

Change #7: Text changed as suggested above (P4, L32 in revised manuscript).

Comment #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).

Response #8: 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 did not show smaller streams. Smaller stream have been added, as well as elevation ranges as a background. Where available, river names were added.

Change #8: Updated Fig. 2, and corresponding caption.

Comment #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?

Response #9: 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. 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 two sites, one shaded, one not, the regression slope and intercept for each site would be different to reflect that one site is, for example, on average cooler, or one site is more responsive to direct sunlight than than the other.

Change #9: We added clarifications (measurements are instantaneous; P6, L3-4 in revised manuscript).

Comment #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?

Response #10: We clarified that precipitation data were derived from observed rain gauge data by using the natural neighbour interpolation method, which is a development of the Thiessen approach.

Change #10: The text has been edited accordingly (P6, L12-14 in revised manuscript).

Comment #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?

Response #11: The climate data are in fact deterministic (one set of climate data), some of the variables are 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 were 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 within 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.

Change #11: None.

Comment #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.

Response #12: Ecological relevance is with regards to phenology (clarification added in the Introduction). 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).

Change #12: Clarification regarding phenology added to Introduction (P3, L30-31).

Comment #13 P.7,l.4-5: where is this section? I cannot find it. This is relevant information and

needs to be shown.

Response #13: This information is actually in the Results section, but in response to Comment #25 below (and also to a similar comment from Reviewer #2), we followed the reviewer's suggestion that this should appear earlier in the manuscript, and moved it to the Data section.

Change #13: See Comment #25 below.

Comment #14: P.7,l.5: what is meant by “permeability”? permeability of what? How was it determined?

Response #14: We meant catchment permeability in the sense of flashy impermeable catchments vs groundwater-fed catchments. It is characterised by using the catchment base flow index (BFI; described later in the text).

Change #14: We clarified this point (P7, L15-16 in revised manuscript).

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

Response #15: These basin properties are generally recognised in UK studies (cited studies and many others) as modifiers of climate-hydrology associations.

Change #15: The sentence was reworded to improve clarity (P7, L16-17 in revised manuscript).

Comment #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?

Response #16: 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 one 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.

Change #16: None.

Comment #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).

Response #17: Agreed.

Change #17: We added this point to Section 3 Methods (P8, L17-20 in revised manuscript), including the reference to Mohseni et al. (1998).

Comment #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.

Response #18: We reviewed these sections and clarified the more confusing points (the reviewer's comment is not specific in this regards).

Change #18: Section 3.1 and 3.2 substantially edited and expanded (see specific edits as Track Changes, P8-11 in revised manuscript).

Comment #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?

Response #19: As described in the text, all combinations were fitted (programmatically) and their AICs calculated. No stepwise regression or sub-setting was done. The model with the lowest AICc was retained. We edited the text to clarify that 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). 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.

Change #19: Text edited to clarify selection process (P10, L7-15 in revised manuscript).

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

Response #20: All predictors were used. Predictors with RI equals to 1 were used in all models. Other predictors (with RI < 1) were used in some of the models constituting a set of best models.

Change #20: None.

Comment #21: P.11,l.1: what does "adequately" mean? Please provide R2 and p-values.

Response #21: When using multi-level modelling (and moreover within the multi-model inference (MMI) framework), R^2 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 R^2 for multi-level models ("conditional R^2") by Nakagawa and Schielzeth (2013), which we calculated and added to Section 4.1 (reference to the paper was added too). Word "adequately" was 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 Methods Section 3.2 and one sentence in Section 4.2.1 have been added to highlight this, and clarify that higher RI means more significant predictors.

Change #21: A paragraph presenting conditional R^2 (including reference) and providing values for present studies was added to Section 4.1 (P12, L27-29 and P13, L1-2, in revised manuscript). Sentence (P13, L5-9 in revised manuscript) was edited (in particular, we removed "adequately"; see Track Changes). One sentence added to Section 3.2 (P11, L21-25 in revised manuscript). One sentence added to Section 4.2.1 (P13, L12-14 in revised manuscript).

Comment #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.

Response #22: 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 do not appear in the same model, and do not force standard errors up. In addition, in this case, correlations between predictors were below 0.5 (Pearson) for most pairs, and the more highly correlated pairs were 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.

Change #22: None.

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

Response #23: For each record in the dataset, WT was predicted using the average model coefficients from Table 3, then the % contribution of each predictor to predicted WT was calculated (ie we made a time series of WT predictions and predictor % contributions).

Change #23: We edited the text to clarify the above (P14, L11-15 in revised manuscript).

Comment #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?

Response #24: The y-axis label ('%') was erroneously placed on the x-axis. 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).

Change #24: Figure amended (% relocated to y-axis).

Comment #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.

Response #25: The section on basin properties was originally in Data and 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 Data and Methods sections were already quite long, we felt that it could be considered part of Results instead. We appreciate that this can actually be confusing, and that readers would more likely expect this information earlier in the manuscript. We therefore opted to move a significant part of Section 4.3 to Data Section 2.4 Basin properties. Regarding a table of values, we assume the reviewer means the actual property values for each site; we included in the text of Section 4.3, the ranges of values for each property across the 35 sites. We believe this could be a good compromise in terms of information vs manuscript length, and would favour keeping the manuscript as it stands.

Change #25: Text from Section 4.3 moved to Section 2.4 (P7-8 in revised manuscript); see Track Changes for details.

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

Response #26: We added this comment when elevation is introduced.

Change #26: Comment added (P7, L31 in revised manuscript).

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

Response #27: We added this when the property is introduced. Note that this was already mentioned in the Discussion.

Change #27: Comment added (P8, L1-2).

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

Response #28: 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 was already covered in the Discussion.

Change #28: Comment added (P8, L4-6).

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

Response #29: 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. Please also see response #21 to similar comment re using MMI and selection with an information criterion (ie p values are not relevant).

Change #29: None.

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

Response #30: Agreed.

Change #30: Sentence revised (from P16, L24-25).

Comment #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.

Response #31: Agreed.

Change #31: Sentence revised (P16, L26-28).

Comment #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).

Response #32: An early version of the manuscript actually made that very point. We re-instated it but kept the possible physical explanation as well.

Change #32: Point above inserted in text (P18, L19-20 in revised manuscript).

Response to Reviewer #2

Comment #1: This article explores basin and climatic drivers of stream temperatures across the UK. While the authors do a nice job throughout stating what is novel with respect to the study, I have a hard time finding some of their results novel. They show that air temperature, and solar radiation, drive heat fluxes throughout the year. Their findings fall in line with 30+ years of stream temperature research. A potentially novel result is the inclusion of different climatic factors, and the modeling style that they use to include these factors. However, it is not clear what this information adds to predictive capacity for stream temperature across the UK. Does including these variables mean there is greater explanatory power? Tertiarily, they also relate models to basin properties. However, the basin properties that were included are not well described in the paper, and end up feeling tangential to the other results. I'm left wondering where the model(s) perform(s) well, and where they performs poorly, and how performance changes across different scales. Can this approach we used to improve modeling of stream temperatures? This is mentioned briefly at the end. As it stands, showing that models identify climatic variables as important seems to confirm what we already know. Showing, again, that basin properties influence these results is also potentially not new. I'm also left wondering about some of the implications of their data (in terms of temporal and spatial extent) for their conclusions. Overall, this is clearly a well-developed idea that will advance stream temperature research, but I am left feeling confused about broader implications, the sites in question, and whether this type of approach gets us any closer to improving our empirical modeling of stream temperatures.

Response #1: It is worth clarifying that we did not aim to produce a better predictive model of water temperature, but rather, the modelling exercise was a mean to gain better understanding of the large-scale spatial and temporal variability in climate–WT associations, and of the influence of basin properties on these associations. The modelling techniques (multi-level modelling and multi-model inference) are definitively novel in their application to water temperature. In particular, we could analyse data both from at site scale and at national scale at once with multi-level modelling. The sites covered a reasonably wide range of catchment types. The combined wider spatial patterns and site-specific responses related to basin properties help unravelling the relative influence of climate vs land surface control across scales.

Change #1: None.

Comment #2: Results, especially in tables and figures, are not presented in a way that enables easy interpretation by the reader. Table 6 means nothing to anyone but the authors. Table 5 – why is the FEH descriptor included, except for reference to Table 6? Why were the selection of descriptors used? Greater insight on which descriptors were included would be helpful. Section 4.3 for instance, refers to the abbreviations of FEH variables, but it would be much fewer words to just state the actual variables in text, and indicate FEH variables in parentheses

Response #2: We think this comment stems from the fact that we introduced the basin properties in the Results rather than in Data and Methods. As we stated in response to Reviewer #1 Comment #25, we moved the bulk of Section 4.3 back to Data, thus streamlining and clarifying which basin descriptors we used. Re Table 5, we included the FEH descriptor to highlight the fact that the basin property (eg elevation) is characterised by a specific descriptor (eg ALTBAR). Then, since these descriptors are indeed used for the results featured in Table 6, we thought it would help readers to make the connection. We believe we explain Table 6 clearly enough in the text. Regarding the latter point of FEH descriptors abbreviations, they are actually not abbreviations as such but short names (except for BFIHOST), so their explicit names is already given (eg ALTBAR = mean basin elevation above sea level). To clarify things, we swapped order of full name and FEH short name.

Change #2: Text from Section 4.3 moved to Section 2.4 (P7-8 in revised manuscript); see Track Changes for details; descriptors full names and short names swapped (P7-8).

Comment #3: The introductions to each section are not helpful, but I leave this up to the authors. I find that they detract from the reading of the manuscript.

Response #3: Experience with past papers showed that some readers benefit from these section introductions. We are inclined to keep them unless there is a need to reduce the length of the manuscript.

Change #3: None.

Comment #4: Sites with very different time scales of measurement where included. I get why this was done – there is not a lot of stream temperature data (a problem I am also having!). However, I'd like to know more about what is the effect? Were sites with 15-minute versus weekly and monthly data treated differently? With so many sites, it would be worth testing if 15-minute data were treated in the same way as weekly or monthly sites, what the effect on conclusions would be? If sites from weekly/monthly data were excluded, are conclusions different?

Response #4: The discrepancy in data time scales is handled by using multi-level modelling. So, for example, if sites based on 15-min data behaved differently from sites based on weekly data, the model would correct for that. However, it would not explicitly investigate what the effect of one over the other. We added a mention of this as possible future research in the Conclusions.

Change #4: Sentence added in Conclusions (P22, L13-16).

Comment #5: Unclear what kind of variability in terms of basin/river properties your paper explores – a figure to this effect would be a good contribution. For instance, where else would your results be comparable to? This would be helpful to know both in terms of stream temperature regime and basin properties.

Response #5: In Section 4.3 (moved to Section 2.4 in revised manuscript), we have included a paragraph giving the ranges of basin properties for the 35 sites. They do provide a fairly wide range of basin types in the UK. The original data sources include a mix of lowland permeable (eg from LOCAR, Tadnoll; the UK has most of its aquifers in lowland regions) and impermeable basins, upland impermeable (eg form Plynlimon), as well as small to medium basin sizes. Sites from the AWMN cover all types. The gap in coverage may be in terms of large basins (ie >1000 km2) but these are far less common in the UK than in other countries. Similarly to our response to Reviewer #1 Comment #25, we think that the manuscript as it stands provides a good compromise between information vs length.

Change #5: None.

Comment #6: Magnitude of fluxes depend not only on climatic variables, but also on water temperature. Is model able to include this interaction, as it is a key determinant of evaporation/condensation and convection/conduction?

Response #6: As it stands, the models are at their core linear regression with water temperature as the dependent variable. Feedbacks cannot be built-in. This would require a different type of method, possibly bespoke models handling iterative calculations. However, we explored this with the outputs from Fig 5 showing how the contributions of different predictors change with water temperature.

Change #6; None.

Comment #7: Need more information on descriptors. They're included haphazardly. Don't even know which predictors are included in the model.

Response #7: See responses to Reviewer #1 Comment #25, Reviewer #2 Comment #2 re Section 4.3 moved to Data. We believe the changes to the way basin properties are presented address this.

Change #7: Text from Section 4.3 moved to Section 2.4 (P7-8 in revised manuscript); see Track Changes for details.

Comment #8: Pg 3, Lines 10 – 20 – variables should have subscripts

Response #8: Agreed.

Change #8: Text amended as suggested (P3).

Comment #9: Pg 3, Line 28 – misplaced comma

Response #9: Agreed.

Change #9: Comma moved to its proper position.

Comment #10: Pg 4, line 6 – consider the role of basin properties with respect to what? There's several papers in the US that have investigated the role basin properties may play in determining the stream temperature regime – they do so from an empirical perspective

Response #10: We clarified we consider the role of basin properties with regards to stream temperature, and added the point about the empirical approach (also see response to see Reviewer #1 comment #5).

Change #10: Sentence (P4, L8-9 in revised manuscript) amended.

Comment #11: Pg 5, line 3 – it's not clear to me what you mean by 'not losing any information'

Response #11: This refers to the loss of information due to class-level averaging, a common step with classification-based analyses (already covered in a more detailed way on Page 4). We clarified this point.

Change #11: Clarification inserted (P5, L7 in revised manuscript).

Comment #12: Pg 6, section 2.2 – what impacts do you think using a 1km square meteorological dataset may have on your proposed conclusions? Are there any sites where microclimate could play a role?

Response #12: We agree there may be micro-climate effects (one co-author published several papers based on field site monitoring showing the impact of shading, etc.), but the focus of this study was the wider spatial patterns, which is quite novel, so we did not investigate this further. In addition, based on the information we had, we do not think there was any site where micro-climate effect was conclusively present. However, we mention this in the discussion (for example, highlighting how shaded river may behave differently from non-shaded). We added a mention of this point in the conclusion.

Change #12: Sentence added in Conclusions (P22, L29-33).

Comment #13; Section 3.2 was difficult to follow and written confusingly. Comments were included in parentheses and not explained fully. The importance of AIC weights was introduced, but there was little explanation of what this value tells the reader (does 'relative importance' mean a better model? More trustworthy model?)

Response #13: This particular section has been revised and streamlined to provide more clarity (see response to Reviewer #1 Comment #18).

Change #13: Section 3.1 and 3.2 substantially edited and expanded (see specific edits as Track Changes, P8-11 in revised manuscript).

Comment #14: Some missing words in section 3.3

Response #14: There was in fact an extra noun, which should have been deleted. Section has been checked and revised.

Change #14: Unnecessary word deleted (P11, L28).

Comment #15: Page 10 line 24: why was no predictor included for spring?

Response #15: As explained in that paragraph, for spring, the model getting the lowest AICc (ie the best model) was the model with random intercept only (ie the only difference between sites is with regards to the mean water temperature; all sites have the same response slope for all predictors). Therefore, no predictor was included as a random effect. We added a reminder that spring is a random intercept only ML model. We also expanded Section 3.1 to give more clarity regarding multi-level modelling, which in turns would help readers to understand what is meant in this paragraph.

Change #15: Reminder that spring is a "random intercept only" ML model (P12, L23). Section 3.1 expanded to give more clarity regarding random effects in multi-level models (P9, L22-28; see Track Changes for details).

Comment #16: Abbreviations make the results difficult to adjust – I know what short wave radiation is, but every time I see SWR, I get confused!

Response #16: We appreciate this problem, but we had to use abbreviations for the sake of conciseness (there are many references to the model predictors in the text), and following that, consistency required using these abbreviations all through out.

Change #16: None.

Comment #17: Pg 17 line 1: Most other studies only use AT because it so well predicts stream temperatures. While your models demonstrate association, how much better do they predict stream temperature than air temperature alone? Furthermore, you use gridded AT data, which is available everywhere. I find it much less likely that AT is unavailable at a site with a suite of other climatic variables.

Response #17: Our objective was not to build a better predictive model but to understand the various climate-stream temperature associations. As such, we do not claim that these particular models do necessarily better than AT-WT models (and did not investigate this), but that using other climate predictors in addition to AT could be informative in some cases. We added a few words to avoid misunderstanding on that point. Regarding the comment about AT being not available, it is true that if one uses gridded climate data, then AT is probably more likely to be available than some of the other variables: what we had in mind were field sites where air temperature measurements may be missing. However, this was a minor point however, so we deleted it to avoid distracting from the main message.

Change #17: Text inserted (P19, L12-13). Text deleted (P19, L18).

Comment #18: Pg 17, line 27 on – please rephrase out of list form

Response #18: Agreed.

Change #18: Text edited as requested (P20, L11-14).

Comment #19: Figure 4 should be improved – it is difficult to read axes. Model fits should be included.

Response #19: The size of the figure has been increased slightly to improve readability (figure has also been edited to amend one label, which was not displayed properly). Model fits (conditional R2) have been added to text. Given the conditional R2 are strictly speaking calculated for each model in a model set rather than for the average model (ie the average of all models in a model set), we think it is more accurate to leave the reader to gauge the average model fits visually than to add a mean R2 to the plots (see Reviewer #1 Comment #21).

Change #19: Figure amended. Paragraph giving conditional R2 inserted (P12, L27 to P13, L2, in revised manuscript).

Comment #20: Figure 5, please label the y-axes

Response #20: The y-axis label ('%') was erroneously placed on the x-axis.

Change #20: Figure amended.

Climate and basin drivers of seasonal river water temperature dynamics

C. L. R. Laizé[1, 2], C. Bruna Meredith[2,*], M. Dunbar[1,], D.M. Hannah[2]**

[1]{Centre for Ecology and Hydrology, UK}

[2]{School of Geography, Earth, & Environmental Sciences, University of Birmingham, UK}

[*]{now at: Scottish Environment Protection Agency, UK}

[**]{now at: Environmental Agency, UK}

Correspondence to: C. L. R. Laizé (clai@ceh.ac.uk)

Abstract

Stream water temperature is a key control of many river processes (e.g. ecology, biogeochemistry, hydraulics) and services (e.g. power plant cooling, recreational use). Consequently, the effect of climate change and variability on stream temperature is a major scientific and practical concern. This paper aimed (1) to improve the understanding of large-scale spatial and temporal variability in climate–water temperature associations, and (2) to assess explicitly the influence of basin properties as modifiers of these relationships. A dataset was assembled including six distinct modelled climatic variables (air temperature, downward shortwave and longwave radiation, wind speed, specific humidity, and precipitation) and observed stream temperatures for the period 1984–2007 at 35 sites located on 21 rivers within 16 basins (Great Britain geographical extent); the study focused on broad spatio-temporal patterns hence was based on three-month averaged data (i.e. seasonal). A wide range of basin properties was derived. Five models were fitted (all seasons, winter, spring, summer, and autumn). Both site and national spatial scales were investigated at once by using multi-level modelling with linear multiple regressions. Model selection used Multi-Model Inference, which provides more robust models, based on sets of good models, rather than a single best model. Broad climate-water temperature associations common to all sites were obtained from the analysis of the fixed coefficients, while site-specific responses, i.e. random coefficients, were assessed against basin properties with ANOVA. All six climate predictors investigated play a role as a control of water temperature. Air temperature and shortwave radiation are important for all models/ seasons, while the other predictors are important for some models/ seasons only. The form and strength of the climate-stream temperature association vary

depending on season and on water temperature. The dominating climate drivers and physical processes may change across seasons, and across the stream temperature range. The role of basin permeability, size, and elevation as modifiers of the climate-water temperature associations was confirmed; permeability has the primary influence, followed by size and elevation. Smaller, upland, and/or impermeable basins are the most influenced by atmospheric heat exchanges, while larger, lowland and permeable basins are least influenced. The study showed the importance of accounting properly for the spatial and temporal variability of climate-stream temperature associations and their modification by basin properties.

1 Introduction

River and stream water temperature (WT) is a key control of many river processes (e.g. ecology, biogeochemistry, hydraulics) and services (e.g. power plant cooling, recreational use); Webb et al. (2008). From the perspective of river ecology, WT's influence is both direct—e.g. organism growth rates (Imholt et al., 2013), predator-prey interactions (Boscarino et al., 2007), activity of poikilotherms, geographical distribution (Boisneau et al., 2008)—and indirect, e.g. water quality (chemical kinetics), nutrient consumption, food availability (Hannah and Garner, 2015).

Consequently, the effect of climate change and variability on stream temperature is a major scientific and practical concern (Garner et al., 2014). River thermal sensitivity to climate change and variability is controlled by complex drivers that need to be unravelled (a) to better understand patterns of spatio-temporal variability and (b) the relative importance of different controls to inform water and land management, especially climate change mitigation and adaptations strategies (Hannah and Garner, 2015). There is a growing body of river temperature research but there is still limited understanding of large-scale spatial and temporal variability in climate–WT associations, and of the influence of basin properties as modifiers of these relationships (Garner et al., 2014). This paper extends Laizé (2015)~~capitalises on the PhD research carried out by the first author (Laizé, 2015)~~.

River thermal regimes are complex because they involve many interacting drivers (Hannah et al., 2004, 2008). Caissie (2006) identified atmospheric conditions as the primary group of controls, with hydrology linked to basin physical properties (e.g. topography, geology) as secondary influencing factors.

The main climate variables (Fig. 1) which constitute an 'atmospheric conditions' group, can be identified by analysing the theoretical heat budget for a stream reach without tributary inflow, which may be expressed as (adapted from Hannah and Garner, 2015):

$$Q_n = Q^* + Q_h + Q_e + Q_{bhf} + Q_f + Q_a \qquad \text{Equation 1}$$

where Q_n is the total net heat exchange, Q^* the heat flux due to net radiation, Q_h the heat flux due to sensible transfer between air and water (sensible heat), Q_e the heat flux due to evaporation and condensation (latent heat), Q_{bhf} the heat flux to and from the river bed, Q_f the heat flux due to friction at the bed and banks, and Q_a the heat flux due to advective transfer by precipitation and groundwater.

The different components of Eq. (1) correspond to different processes, related to climatic and hydrological conditions. Q* corresponds to shortwave radiation (insolation from the sun) and longwave radiation (emitted towards the stream by clouds and overhanging surfaces such as vegetation, and reemitted back to space (lost) at water surface temperature). Q_h corresponds to convective energy exchanges between air and water (at the surface) causing heat loss or gain. Q_e represents heat loss by evaporation or gain by condensation. Q_{bhf} and Q_f do not relate directly to climate processes but rather local hydrological conditions. (Q_f can be assumed to be negligible in many systems; e.g. Hannah et al., 2008). Q_a corresponds to advective heat exchanges, e.g. inflow or outflow into the river reach, hyporheic exchange, groundwater. A direct, climatic component of Q_a is precipitation inputs, which is thought to have a limited contribution (Caissie, 2006).

These variables are not independent. Figure 1 features a schematic representation of the interactions between these variables. Downward short and long wave radiations increase WT but also air temperature, then there are exchanges between air and water, to influence sensible heating. Additionally, wind plays a significant role by increasing evaporative cooling and in modifying the air–water exchanges by increasing mixing (Hannah et al., 2008). The physical equations underpinning the role of wind can be found in Caissie et al. (2007).

A review of recent international water temperature research can be found in Hannah and Garner (2015). To date, most~~,~~ UK-focused studies (Table 1) tend to be either specific to a few monitoring sites, to have a limited geographical extent (i.e. focused with specific region of the country), and /or to consider few climate drivers. In addition, seasonality, which has huge ecological relevance with regards to phenology, ~~-~~is only explored formally in a small number of papers (e.g. Langan et al., 2001; Hrachowitz et al., 2010~~-~~). A major difficulty is to pair WT and climate monitoring sites, as monitoring is coordinated rarely, then to identify time series with long enough common periods of record. For example, Garner et al. (2014) undertook a England and Wales scale study and matched water temperature monitoring sites with climate and hydrological monitoring sites for 38 temperature sites out of ~ 3,000 sites in the Environment Agency's Freshwater Temperature Archive (Orr et al., 2014). Garner et al. (2014) is one of the ~~very ~~few studies ~~(~~internationally~~)~~ (eg Hrachowitz et al. (2010) in the UK; Isaak and Hubert (2001), Nelitz et al. (2007), or Isaak et al. (2010) in North America) to consider explicitly and empirically the role of a limited number of basin properties with regards to stream temperature.

In most of these studies, analyses are done on a site by site basis, which limits the extent to which broad patterns can be inferred (statistical results for a given site are only valid for that site); Caissie, 2006 emphasized this as a limitation when having to work across different spatial scales. In contrast, studies like Garner et al. (2014) group sites together using classification techniques to identify regional patterns. However, doing so causes a loss of information since data-points of all sites within a class are summarised and intra-class differences lost, and inferences at group level are not necessarily valid at site level. An alternative analytical/ statistical method, which can characterise broad patterns while preserving individual site information, should be investigated.

The following research gaps are identified (above): (a) climate–WT studies in the UK used a limited number of WT sites or climate explanatory variables (focus on air temperature links to WT) and /or are limited in geographical extent; (b) limited formal analysis of seasonality; (c) limited knowledge of role of basin properties as modifiers of climate–WT associations; and (d) need for alternative analysis method to optimise data utility.

Given this context, the aims of this study are (1) to improve the understanding of large-scale spatial and temporal variability in climate–WT associations, and (2) to assess explicitly the influence of basin properties as modifiers of these relationships. This paper resolves the issue of driving data availability by using a comprehensive and consistent set of modelled climate data (see Table 2~~3~~ below). With a period of records of 1984–2007 (24 years), for a total of 35 sites located on 21 rivers within 16 basins (providing a Great Britain wide geographical extent) six distinct modelled climatic variables were taken within 1 km of the sites. The study focuses on broad spatio-temporal patterns; hence it is based on three-month averaged data (i.e. seasonal). Such a temporal scale limits~~addresses~~ issues of temporal auto-correlation often found in water temperature time series (Caissie, 2006). The study also investigates a wider range of basin properties than previous studies.

Innovatively, this paper investigates both site and national spatial scales at once. Multi-level (ML) modelling with linear multiple regressions is applied as an alternative to site-specific or to classification-based analyses because it allows pooling of all site data together while taking into account data structure (i.e. observations at site, sites within same basin) as well as not losing any information due to class-level data averaging (Zuur et al., 2009). With this modelling technique, it is possible to investigate both study aims (i.e. the broad climate-WT associations common to all sites, and the site-specific responses which may be related to basin properties) within the same analysis framework. In addition, model selection used Multi-Model Inference (MMI), another state-of-the-art technique, which provides more robust models based on sets of good models rather than selecting a single best model (Grueber et al., 2011).

2 Data

With regards to research Aim 1 of this paper, observed river temperature data were assembled with a view to maximise spatial and temporal coverage as much as practically possible. To address the issue of mismatching monitoring networks, climate variables were obtained from a modelled dataset. The paired climate–WT dataset used in this paper has been published online via an open-access data repository (Laizé and Bruna Meredith, 2015). With regards to Aim 2, a comprehensive and consistent set of basin properties were derived for all study sites.

2.1 Water temperature data

WT data (unit: ºC) were collated from various research projects run by the UK's Centre for Ecology and Hydrology (CEH). The period of record, temporal resolution, and recording method of the individual datasets vary. These datasets totalled 41 sites, of which 35 were retained after quality-control (e.g. removal of duplicates; see Fig. 2~~3~~). As often the case, water temperature was not the main focus of these projects: fish for the River Frome (1 site, 1991-2009, 15-min logger; Welton et al., 1999), Great Ouse (1 site, 1989-1993, hourly logger), and Tadnoll (2 sites, 2005-2006, 15-min logger; Edwards et al., 2009) studies; impact of forestry on water quality for the Plynlimon catchment project (4 sites, 1984-2008, weekly manual recording; Neal et al., 2010); acidification monitoring for the UK Acid Water Monitoring Network (UKAWMN) project (10 sites, 1988-2008, monthly (not necessarily on same day) manual recoding; Evans et al., 2008); hydrological and biogeochemical processes for the LOwland CAtchment Research (LOCAR) project (17 sites, 2002-2011, 15-min logger; Wheater et al., 2006). Whether recording was done manually or with a logger, measures are instantaneous. Because these original projects were focused on natural rivers, the temperature data used herein may be considered as largely free from artificial influences (e.g. no industrial use for cooling or heated effluent discharges).

2.2 Climate data

The Climate Hydrology and Ecology research Support System (CHESS) dataset features six climate variables (Table 2~~3~~). CHESS is the forcing dataset for the Joint UK Land Environment Simulator model (JULES; Best et al., 2011). CHESS is a UK-wide 1-km grid dataset derived by downscaling the UK Meteorological Office Rainfall and Evaporation Calculation System (MORECS) 40-km grids (Hough and Jones, 1997), except for precipitation that were derived ~~is based~~ from ~~on~~ observed rain gauge data by using the natural neighbour interpolation method, which is a development of the Thiessen approach (Keller et al., 2006). For each 1-km cell, modelled daily time series of all variables are available for the period 1971–2007. The processes linked to AT, LWR, P, and SWR are given in the stream heat budget overview (see Introduction) and summarised in Table 2~~3~~. Specific humidity (SH) gives a measure of evaporation potential (i.e. the more humidity, the less evaporation due to reduced vapour pressure gradients; e.g. Hannah et al., 2008). Wind speed (WS) captures the various effects of wind in increasing evaporation (cooling) and convective air-water exchanges (cooling or warming) Each CHESS cell was matched to the study temperature site(s) it contained.

2.3 Seasonal time series

Firstly, sub-daily water temperature data were averaged at a daily time step (Frome, Great Ouse, Tadnoll, LOCAR) while spot measurements (Plynlimon, UKAWMN) were assumed representative of the day on which they were taken, although it is worth keeping in mind that they are only representative of daylight conditions. Secondly, daily water temperature data were matched by date to the daily climate data. Thirdly, seasonal averages were computed from these daily data for all variables. Seasons were defined as: December–February (winter), March–May (spring), June–August (summer), and September–November (autumn). For winter, these seasonal data for year *y* were based on data from December of year *y*-1 to

February of year *y* (e.g. for 1976, December 1975, January and February 1976). Lastly, five time series were derived from these data: one series per season at an annual time step (i.e. winter 2000, winter 2001, winter 2002, etc.), and one series with all seasons at a seasonal time step (i.e. autumn 2000, winter 2000, spring 2000, etc). These series and their related models are referred to as thereafter 'autumn', 'winter', 'spring', 'summer', and 'all seasons'.

2.4 Basin properties

Basin properties were derived from the UK Flood Estimation Handbook (FEH), the UK 'industry standard' for flood regionalisation studies, which includes 19 basin descriptors (Bayliss, 1999). A subset ~~selection~~ of descriptors was used. ~~, which are listed with detailed definitions in the Methods section.~~ First, the 19 catchment descriptors were derived for each site. Many basin properties co-vary, often substantially, and they are best interpreted as groups of properties ('meta-properties') rather than on their own. Descriptor specifications (Bayliss, 1999), pair plots, and correlation matrices were checked to identify likely groups of descriptors (for example, all FEH rainfall descriptors capturing basin wetness). Three groups were identifed, which~~These descriptors~~ relate to basin elevation, permeability (ie responsive impermeable v groundwater-fed basins), and size. These ~~,~~have been found ~~which are known~~ to modify climate-hydrology associations in UK basins ~~climatological links~~ (eg Bower et al., 2004; Laizé and Hannah, 2010; Garner et al., 2014). Then, a test run of the basin property analysis outlined in Section 3.3 (ANOVA) was performed in order to check that all FEH descriptors from a given group of properties had consistent associations (positive or negative) with each model predictor (considering basin properties significantly associated with site-specific coefficients only), while one FEH descriptor was retained to represent each meta-property.

The following meta-properties and their corresponding FEH descriptors were thus selected for the final analysis:

- Elevation/wetness ('elevation' hereafter): as noted in Laizé and Hannah (2010), basin elevation and wetness are very strongly correlated in the UK; the meta-property 'Elevation' is represented by the 'mean basin elevation above sea level' (m; FEH descriptor named 'ALTBAR'), and, for the winter model only, by the proportion of time basin soils are wet (%; FEH descriptor named 'PROPWET'), based on soil moisture time series classified as wet/dry days; highly correlated to rainfall); elevation is also related to air temperature;

- Size: basin area (km^2; ‘AREA‘) using its natural log; area is a proxy for discharge, thus for thermal capacity, and is also linked to elevation;
- Permeability: Base Flow Index from Hydrology of Soil Type (BFIHOST; dimensionless); ranging from 0 (less permeable basin) to 1 (more permeable); temperature regimes in groundwater-fed (permeable) basins are expected to be more influenced by groundwater inputs than in impermeable basins.

The 35 study sites are representative of a wide range of UK basin types in terms of the above properties: (1) upland/lowland (ALTBAR approximately within 20-700 m and PROPWET within 24-80%); (2) small and medium size (AREA ~0.5-415 km2); (3) impermeable/permeable (BFIHOST 0.24-0.92). In addition, the study sites feature combinations of all three meta-properties.

3 Methods

This section describes the analytical methods used. Firstly, as stated in the introduction, linear multiple regressions fitted with the Multi-level (ML) modelling technique was chosen as the core method because it allowed to analyse the multiple-site data in terms of both overall climate–WT associations (linked to research Aim 1) and site-specific responses (linked to research Aim 2; role of basins as modifiers of those associations). Although linear regressions are only approximating climate–WT associations (eg AT-WT associations are better described with logistic models; Mohseni et al., 1998), they were considered a sensible compromise. Secondly, with regards to overall climate-WT associations, ML model selection was done with Multi-Model Inference (MMI) to yield more robust models than with standard single model selection, especially given the number of climate predictors used. Lastly, any relation between site-specific climate-WT responses and basin properties were tested formally using an analysis of variance (ANOVA).

The study work flow is summarised in Fig. 3~~2~~: (a) WT observed data linked with (b) modelled climate variables, then (c) all converted to seasonal (three-month) average series used within (d) ML modelling / MMI framework producing (e) five output models (individual seasons and all seasons; Aim 1), and (f) sets of basin properties (Aim 2).

3.1 Multi-level modelling

To take into account the hierarchical nature of the water temperature dataset (e.g. data measured at the same site, sites located on the same river), ML modelling was used to build linear models with water temperature as the predicted variable, and the six climate variables as explanatory variables. When analysing multiple-site datasets, there are two common alternatives: (a) performing one regression for individual sites, or (b) one regression on all sites pooled together. On the one hand, site-specific regressions can make results highly uncertain (sites may have few data-points; fitting numerous regressions is more prone to identify spurious relationships, ie Type II errors). Thus, drawing out general patterns (e.g. variation between sites, effect of site characteristics) can be difficult. On the other hand, full pooling of sites ignores the clustering of samples within groups (eg measurements from a given site, or sites on the same river, may be more simialr), which may hide important differences between groups~~sites~~ and may cause problems with statistical inference (e.g. violation of the assumption of independence between samples, sites with large or small numbers of samples equally influencing the model outcome).

To overcome these issues, ML modelling can take into account the hierarchical structure in a dataset, ie the different 'levels' at which data can be grouped (eg data at sites, sites within basins, basins within countries), thus allowing~~s~~ for the pooling of data from multiple~~different sites while taking into account the hierarchical data structure~~. A ML model has two components, which correspond to generic patterns (i.e. similar to a regression on fully-pooled data) and to level-specific patterns. The generic patterns, which are described by the explanatory variables as in a standard regression, are called the 'fixed component' or 'fixed effects' of the model. The unexplained variation between levels (eg ~~i.e. site-specific~~ patterns specific to a site~~here~~) is termed the 'random component' or 'random effects'. The random component captures the fact that levels may respond differently to a given predictor. For example, stream temperature could be very responsive to climate at one site (high slope value) but unresponsive at another (low slope value). In some cases, levels may have the same response to predictors but may have differing averages, ie differing with regards to their intercepts (eg two sites with same temporal patterns but with one site systematically cooler than another due to local characteristics or recoding procedure); such ML models are commonly known as 'random intercept only'.

In our analyses, a three-level data structure was applied: individual observations (level 1) nested within monitoring sites (level 2) nested within river stretches (level 3). In addition, a time variable was included as a predictor to take into account any linear trend in the time series. To avoid instability issues when fitting models, the predictors were centred (i.e. predictor values minus their mean).

3.2 Model selection with multi-model inference

Following standard ML modelling practice (e.g. Zuur et al., 2009), the model selection was applied in two stages: (a) selection of the random component variables; (b) selection of the fixed component variables.

First, the random component selection was done as follows. W~~w~~ith all predictors included in the fixed component, all ~~models with the various~~ combinations of predictors in the random component were fitted. The models were then ranked using Akaike's Information Criterion (AIC; Akaike, 1974). AIC is used to ~~that~~ select~~s~~ models offering the best compromise between fit and predictor parsimony; a model with a lower AIC achieves a better rati of fit vs number of predictors. Note that ~~(~~a variation of AIC was used: AICc, which is AIC corrected for small-size datasets~~, AICc was used)~~. Selection was done for the four seasonal series as well as the 'all season' series. In each case, the single combination of predictors~~the random component~~ giving the lowest AICc was retained as the random component.

Secondly, with the random component selected, the fixed component model selection followed the MMI approach, which selects sets of 'good' models rather a single 'best' one. Using a traditional model selection technique, like stepwise regression, the single model with the best (i.e. ~~the~~ lowest) AICc would be selected. This presents two issues: (a) due to the algorithms underlying these types of selection techniques, some model formulations may end up not being tested thus causing a sub-optimal selection; (b) given models with similar AICc values have similarly good performance, it is not statistically correct to keep the lowest AICc model only as the best model and discard the others. MMI addresses these issues by selecting sets of good models. In practice, all possible combinations of predictors using from one to six of the climate variables~~predictors~~ described above were~~are~~ fitted. The resulting models were~~are~~ ranked based their AICc. All models within four points of the lowest AIC were~~are~~ selected (Zuur et al., 2009). Each set of models was then summarised as an 'average model' (predictor coefficients over all models in the set are averaged). ~~Grueber et al. (2011) cover the above points in details and give a very good example of such an application of MMI in a natural sciences context.~~

Akaike weights (Burnham and Anderson, 2002) were then calculated; these~~weights~~ are the re-scaled AICc scores of the models included in a MMI selection set. The weights, which add up to 1, ~~, which~~ give an indication of how ~~the relative~~ important relatively to each others are ~~ce of each~~ the models within a MMI set. For example, results showed that the 'all seasons' model is based on two models with Akaike weights 0.74 and 0.26: the former model has more influence on the resulting average model than the latter. ~~If only one model was tested, the weight would be one. Models with similar AICc scores have similar Akaike weights. Weights are used when reporting on MMI outputs. Then, following recommended statistical usage, all models within four points of the lowest AICc were selected (Zuur et al., 2009). Note that in some cases, there is only one model selected because its AICc is lower by more than four points from the next second model in line, and it would have the higher Akaike weight too.~~

The Akaike weights form the basis to calculate the Relative Importance (RI) of each predictor: RI is how one reports on the role of each explanatory variable in MMI.~~With MMI, the role of each explanatory variable is assessed using its relative importance (RI)~~. For a given predictor, RI is calculated as the sum of the Akaike~~ICc~~ weights (re-scaled AICc) ~~-~~of the models in which that predictor is included. RI ranges from 0 (variable never included) to 1 (included in all models). For example, results showed that the 'all seasons' model is based on two models with Akaike~~ICc~~ weights 0.74 and 0.26; the explanatory variable P is only included in the latter model, hence its RI is~~of~~ 0.26, while the other five predictors are in both models and have a RI of 1 (see Table 3~~4~~ below). With MMI, RI is analogous conceptually to predictor significance, assessed with p values,~~-~~in ~~a~~ standard regression~~.~~ Model. This is why p values are not calculated nor given in the Results section, but instead RI values for predictors are featured (a predictor with a higher RI is more significant). Grueber et al. (2011) cover the above points in details and give a very good example of such an application of MMI in a natural sciences context.

3.3 Analysis of basin property influence

For those explanatory variables that were included in the random effects (i.e. different sites can have different coefficients), any relation between site-specific coefficients and ~~site~~ basin properties was investigated by using maps and scatter plots of coefficients against basin properties, and by applying ANOVA to confirm observed patterns. For each coefficient and basin property, ANOVA is comparing formally (a) a model assuming there is no difference in coefficient between sites against (b) a model assuming the coefficient is function of the basin property. A basin property is considered having significant influence on the WT–climate variable relationship when the ANOVA p value is <0.05. To quantify the influence of these properties, either alone or combined, linear regressions of the site-specific coefficients against these properties were fitted.

4 Results

The result section has three parts:

- Selection and performance of the five models (all seasons, winter, spring, summer, autumn).
- Analysis of the fixed component of the five ML models to inform on climate-WT associations (research Aim 1); results are split in three sub-sections (relative importance of the predictors, form and strength of predictor-WT associations, relative contributions of predictors to modelled WT).
- Analysis of the random component of the five ML models to inform on site-specific climate-WT responses (for those predictors included as random effects), followed by ANOVA to assess the role of basins as modifiers of the climate-–WT associations. (research Aim 2).

4.1 Model selection and performance

As described above, selecting the five ML models was done in two stages. First, with all predictors included in the fixed component of the ML model, combinations of predictors as random effects were tested, and the combination yielding the lowest AICc was retained. As a result, the following variables were included as random effects (i.e. variables for which different sites have different coefficients): all seasons = AT and SWR; winter = SH; summer = P; autumn = SWR; no predictor was included for spring (random intercept only). Second, all combination of the predictors in the fixed components were tested with MMI. The number of models included in each final set as selected by MMI was: all seasons = 2; winter = 4; spring = 12; summer = 6; autumn = 14.

-With ML models, standard R^2 are not appropriate; conditional R^2 (Nakagawa and Schielzeth, 2013), which are analogue to standard R^2 but designed for ML models, were calculated. Conditional R^2 were: 0.96 for both all seasons models; 0.88 for all four winter models; within

0.88-0.89 (mean 0.88) for the 12 spring models (mean 0.88); within 0.84-0.85 (mean 0.84) for the six summer models; within 0.88-0.89 (mean 0.88) for the 14 autumn models.

With MMI, each set of models is summarised as an ‘average model’, for which a given variable coefficient is its average value over all models in the set. The average model coefficients are presented in Table 3~~4~~.~~ Thereafter, if unqualified, the term ‘model’ means the average model for a given set of selected models~~. All average models have good fits consistent with conditional R^2 given above, and ~~perform adequately (~~as evidenced by plots of modelled~~fitted~~ against observed water temperature data in Fig. 4~~)~~. Thereafter, if unqualified, the term ‘model’ means the average model for a given set of selected models

4.2 Relative influence of climate drivers

4.2.1 Relative importance of the predictors

As explained above, within the MMI framework, the significance of a predictor is captured with its relative importance RI in the selected model sets (RI = 0, predictor never retained; RI = 1, predictor retained in all models of set). Predictor RIs for all average models are given in Table 3~~4~~. First, there is no predictor with a zero RI for any average model. This means that all predictors are used in all or part of the sets of selected individual models. Predictors can be ordered by decreasing importance: AT (RI=1 for all models); SWR (RI=1 for four models, and 0.64 for the summer one); WS (RI=1 for two models, and 0.33-0.68 for others); SH (RI=1 for two models, 0.34-0.53 for others); P (RI=1 for one model, 0.15-0.41 for others); LWR (RI=1 for one model, 0.13-0.25 for others).

Second, each model has its own set of most important predictors (with RI > 0.50 as a threshold, i.e. predictor included in half of the selected individual models): all seasons, all predictors except P; winter, AT, SWR, WS, and SH; spring, AT, SWR, and WS; summer, all predictors; autumn, AT and SWR.

4.2.2 Form and strength of associations between climate predictors and water temperature

The section focuses on the fixed effect coefficients of the predictors (i.e. coefficients valid for all sites). Predictors AT, SWR and SH have positive coefficients for all models (i.e. increases of these predictors are associated with a consistent warming effect on water temperature).

Predictors LWR, WS, and P have positive or (mostly) negative coefficients (i.e. increases of these predictors are associated with warning or cooling, depending on season; Table 34).

The strength of the association varies with season. Comparing the absolute value of the seasonal coefficients for each variable (not between variables as they have different scales): AT, lowest in winter, highest in autumn; SWR, lowest in autumn, highest in winter; LWR, lowest in winter, highest in summer; WS, lowest in autumn, highest in summer; SH, lowest in autumn, highest in winter; P, lowest in summer, highest in autumn.

4.2.3 Relative predictor contributions

By definition, the predictors may have different units and orders of magnitude. Their coefficients cannot be compared directly to get an indication of their relative contribution to WT predictions. Instead, for each generic average model (see coefficients in Table 3), predicted WT values ~~predictions~~ were generated for the whole period of record, then ~~and~~ the percentage contributions of each predictor to these predicted WT ~~modelled~~ values were calculated (ie a time series of predicted WT and of percentage contributions for the six predictors). Boxplots of the percentage contributions for the six predictors and the five models are featured on the left-hand side of Fig. 5 (for readability, outliers are not displayed). The thick black central line corresponds to the median percentage contribution. The shorter the boxes and whisker extents are, the more constant are predictor contributions to modelled WT, with longer extents representing more variation. While, the boxplots inform about contribution differences between models, plotting predictor contributions against modelled WT (right-hand side of Fig. 5) shows that the contribution variability, for a given model, is in many cases related to WT rather than random (i.e. some predictors are more or less influential depending on thermal conditions).

AT is the main contributor except in winter (second to SH); its median contribution is around 12% for winter, and 30-35% for the other models. In all cases, AT contribution increases as WT increases (AT has more influence at warmer WT).

SWR influence is quite constant for all models (medians ranging from +4.5% to 7.5%; up to a maximum of +15.8% in winter) except autumn, for which it is very limited (median +0.13%). Within each model, SWR contribution is fairly stable across the WT range but showing slightly more variability for colder WT.

LWR is the second contributor for the 'all seasons' and the summer models. Its contribution is negative except for spring, but in all cases, the contribution decreases as WT increases (i.e. LWR has more influence on colder WT).

WS has a negative contribution for all models except autumn. WS is most influential for colder WT (e.g. down to a minimum of -13.70% for all seasons model, -11.74% for summer); its contribution decreases as WT becomes warmer (e.g. around -1% for most models). WS contributions are more variable for colder WT (ie more scatter right-hand side plots; Fig. 5) than for warmer WT. For autumn, WS has limited influence, with its contribution ranging from +0.17% to +0.90%.

SH contribution is highest in winter (main contributor with median +27.20%) and for 'all seasons', but otherwise limited for the other seasons (medians ranging +2.10% to +7.23%). SH contributions are independent from WT.

P has limited influence with its contributions ranging from -1.13% (minimum, spring) to +0.22% (maximum, winter). Its contributions show very little variability and no pattern in relation to WT.

4.3 Role of basin properties

The site-specific coefficients were initially mapped against elevation and permeability to explore basin modification of the WT–climate relationship, and any pattern linked to easting/northing. While there was no clear easting/northing pattern, the maps showed potential associations between coefficients and basin properties.

~~As explained above, a set of 19 catchment descriptors were derived for each site. Many basin properties co-vary, often substantially, and they are best interpreted as groups of properties ('meta-properties') rather than on their own. Descriptor specifications (Bayliss, 1999), pair plots, and correlation matrices were checked to identify likely groups of descriptors (for example, all FEH rainfall descriptors capturing basin wetness).~~ Then, ANOVA was run on those descriptors to identify the ones significantly associated with the model site-specific coefficients. ~~Finally, the descriptors for each meta-property were checked to confirm they have consistent associations (positive or negative) with each model predictor. Considering the basin properties significantly associated with the site-specific coefficients only, one descriptor was retained to represent each meta-property.~~

~~The following meta-properties and their corresponding FEH descriptors were thus selected:~~

- ~~Elevation/wetness (‘elevation’ hereafter): as noted in Laizé and Hannah (2010), basin elevation and wetness are very strongly correlated in the UK; the meta-property ‘Elevation’ is represented by the FEH descriptor ALTBAR (mean basin elevation above sea level; m) and, for the winter model only, by PROPWET (proportion of time basin soils are wet (%), based on soil moisture time series classified as wet/dry days; highly correlated to rainfall);~~
- ~~Size: AREA (basin area; km^2); using its natural log;~~
- ~~Permeability: BFIHOST (Base Flow Index from Hydrology of Soil Type (HOST); dimensionless); ranging from 0 (less permeable basin) to 1 (more permeable);~~

~~The 35 study sites are representative of a wide range of UK basin types in terms of the above properties: (1) upland/lowland (ALTBAR approximately within 20-700 m and PROPWET within 24-80%); (2) small and medium size (AREA ~0.5-415 km^2); (3) impermeable/permeable (BFIHOST 0.24-0.92). In addition, the study sites feature combinations of all three meta-properties.~~

Associations between meta-properties/descriptors and site-specific coefficients are showed in Table 4~~5~~. Note: no property was found to be associated with P coefficients in summer.

To quantify the influence of the properties, either alone, or combined, simple linear regressions of the site-specific coefficients were fitted and ranked with AICc following the MMI technique used above. Models are featured in Table 5~~6~~. The best models are the ones with the lowest AICc (displayed in bold characters); while all models featured are within four AICc points, hence are considered equally good (Zuur et al., 2009). Depending on the site-specific coefficient, the R^2 range from 0.125 (autumn SWR) to 0.411 (‘all seasons’ AT). In each case, a single regression (on BFIHOST or ALTBAR) is the best model AICc-wise, although most of the multiple regressions are within 4 AICc points so equally valid models. In the UK context, t~~T~~hese meta-properties are themselves not independent~~ in the UK~~: (i) high upland basins are more often ~~–~~impermeable~~ generally ~~because ~~(~~permeable geology predominantly occurs in the UK lowlands~~)~~; (ii) there are comparatively more larger~~small~~ basins at lower~~higher~~ elevations. Results in Table 5~~6~~ demonstrate this. For the ‘all seasons’ AT coefficient models, single regressions on BFIHOST, ln(AREA), and ALTBAR achieves a R^2 of 0.370, 0.284, and 0.127, respectively, but the multiple regressions with either two or all of them only achieve R^2 within 0.381–0.411. The comparatively small gain when adding several predictors is due to the three properties co-varying. Similar comments can be made on the other models.

5 Discussion

This section has two parts:

- Discussion of the ML modelling fixed components (national-scale patterns of climate-WT associations; research Aim 1); this includes outcomes of MMI, physical interpretation of the models, and dependence between climate-WT association and season/temperature.
- Discussion of the ML modelling random components (site-specific climate-WT responses to assess their modification by basin properties; research Aim 2); identified basin properties are first considered individually, then combined.

5.1 Influence of climate drivers

This section discusses results related to the fixed component of the ML models, which provide information on national-scale patterns (i.e. patterns valid for every sites used in the analysis). As explained above, these patterns would be analogue conceptually to those sought by using cluster analysis or fully-pooled regressions but without their shortcomings (e.g. loss of information, issues with dependent observations). The use of ML modelling adressed one of the limitation of empirical regression-based models, for which temperatures are predicted at specific sites only. Note: the four seasonal models are by definition related to the 'all seasons' model, since they are based on subsets of the same original dataset, so that seasonal patterns are not independent from the 'all seasons' patterns.

The six climate predictors investigated were identified as significant within the MMI framework (note: MMI applied to the selection of the fixed component part of the ML models only). Standard model selection techniques (e.g. stepwise) would have most likely excluded the predictors that are not retained in all models of the MMI selected model sets (i.e. predictors with lower RI values). In this regard, this study illustrated how MMI can be useful in picking the effect of secondary controls, otherwise masked by dominant primary drivers.

The models broadly make sense against known physical processes. In interpreting model results, it important to bear in mind that the aim of the study was to assess the relative empirical associations between WT and the set of climate drivers, therefore the models are not explicitly process-based. In addition, the climate variables are inter-related in some extent (e.g. P associated with more cloud cover, hence reduced SWR and greater SH), and the analysis is based on 3-month averaged data, which may cause some aspects of the physical processes to be lost by the averaging (e.g. distinction between variable like SWR, only contributing during daylight and others like LWR contributing continuously).

All models flag a close association between AT and WT. This finding is consistent with the literature: it is well documented that AT and WT are both influenced by similar climatic drivers (e.g. incoming radiation), and tend towards thermodynamic equilibrium (Caissie, 2006). Both variables consequently tend to co-vary positively, making AT a very useful predictor (as it has been widely demonstrated in the literature; e.g. Webb and Nobilis, 1997), although the association is partly causal only (Johnson, 2003). SWR (insolation from sun) is physically a positive input of energy; and it is appropriately captured in the models with positive coefficients. In this study, LWR is the downward component of longwave radiation (see Table 2~~3~~). From an energy budget perspective, LWR therefore corresponds to a positive flux toward the river water. Consequently, LWR contribution to WT should be positive. Results (Table 3~~4~~ and Fig. 5) show this is not necessarily the case. LWR corresponds to radiation diffused by clouds, so co-varies positively with cloud cover (in addition, a pairwise plot of the study dataset shows that within a given season LWR inversely co-varies with SWR). Therefore, the negative WT-LWR associations would either ~~most likely~~ be due to LWR acting as a proxy for processes driving colder water temperatures (e.g. cloud cover), or be a model artefact due to the LWR/SWR collinearity. SH represents the mass of water vapour in moist air. The rate of evaporation at the water surface is directly proportional to the SH gradient (the more humid the air, the lower the evaporation rate). All models give a positive association between SH and WT. As SH increases, the evaporation rate decreases, and consequently, cooling due energy loss as latent heat decreases as well. WS has a cooling effect by increasing evaporation at the water surface, which would be captured by a negative contribution to WT. In addition, WS plays a significant role in air–water energy exchanges by increasing mixing, which would manifest as increased cooling or warming depending on the AT-WT gradient. For all models but autumn, WS has an overall negative contribution (cooling). For the autumn model, the variable RI and its percentage contribution are both low, so the positive association has to be considered with caution. P have positive or negative coefficients depending on model. When rainfall occurs, its temperature may be higher or lower than that of the river depending on season. In addition, P can also act as a proxy for cloud cover, thus for reduced SWR and increased LWR. P has limited importance and percentage contribution in all the models, which is probably due to precipitations being event-based whereas other variables are continuous (e.g. AT).

The form and strength of the climate-WT association vary depending on season and on WT range, as showed by the variability in predictor coefficients and contributions. This most likely captures that the dominating climate drivers and physical processes (e.g. evaporation/condensation, radiative fluxes; see energy budget above) may change from one season to another, or within the same season, from colder to warmer weather conditions. As a consequence, the impact of short (e.g. seasonal climatic drought) and long term climate variability or change, and of mitigation schemes (e.g. increasing riparian tree shading) on stream temperature may not be uniform across time (e.g. higher long-term temperature increases in winter and spring; Langan et al., 2001).

Probably because AT performs very well as a predictor (e.g. Webb and Nobilis, 1997), m~~M~~ost empirical models have been based on single AT-WT regressions (Caissie, 2006) with very few using other climate predictors (e.g. AT and solar radiation; Jeppesen and Iversen, 1987). The present study demonstrated the potential of several other climate variables to contribute explanatory power (even if they are weaker predictors than AT), which can be beneficial when trying to tease out the relative influences of the various interconnected processes controlling water temperature regimes.~~, or when AT is not available at a site~~. Although this was not the primary objective of the study, the models could be used to generate seasonal water temperatures for the whole spatial and temporal extent of the CHESS datasets (whole country, 1971–2007 period of records), for example allowing to investigate broader geographical pattern, or the impact of extreme events like drought.

5.2 Role of basin properties

The analysis of the random component of the models (i.e. site-specific) identified permeability, elevation, and basin size as the main modifiers of the climate-WT response (note: unlike for the fixed component, the random predictors were selected using standard AIC, i.e. there is only one random component formulation for each of the five models). The use of ML modelling addressed the limitations of empirical regression-based models to work across different spatial scales (see above; Caissie, 2006). The basin properties are first reviewed individually, then together to assess how their respective influences may combine within a basin (i.e. are all influences cumulating, or one property dominating?)

For all models and for all predictors (all seasons AT, autumn SWR, winter SH), the more (less) permeable the basin, the lower (higher) the coefficients. Thus, water temperature in impermeable basins appears to be more sensitive to climate than in permeable basins. Indeed, in permeable basins, the temperature regime is comparatively more influenced by the groundwater input to the river; groundwater temperature tends to have more inertia and to have a damper effect on river WT (groundwater warmer than river in winter, cooler in summer) - see for example, Webb and Zhang (1999), Hannah et al. (2004), Caissie, 2006, Kelleher et al. (2012). This pattern is consistent with Garner et al. (2014), which used different temperature monitoring sites and basin properties to investigate air–water temperature associations only.

With regard to basin size, ~~results can be summarised as follows: (a)~~ with the 'all seasons' model, WT in smaller basins is more sensitive to AT but less sensitive to SWR than in larger basins. With the ~~; (b)~~ autumn model, WT in smaller basins is more sensitive to SWR.~~; (c)~~ With the winter model, WT in smaller basins is more sensitive to SH.

Although, there are seemingly contradictory patterns for SWR, this can be explained by the modelling. Where studies typically use only one variable to represent the whole climate (e.g. AT, Garner et al., 2014), several climate predictors are considered herein. As noted in the Introduction, AT and SWR co-vary in some extent. In the 'all seasons' model, AT and SWR were both selected to capture the between-site variability of the climate-WT response, while in the autumn model, only SWR was retained. As a consequence, in the autumn model, SWR represents climate control, most probably capturing part of the WT variability explained by AT when both variables are included as in the 'all seasons' model. Overall, WT is more sensitive to climate in smaller basins. Then, the inclusion of both AT and SWR in 'all seasons' allows to refine the assessment of river thermal sensitivity beyond climate as a whole, to different types of energy processes: smaller streams are more sensitive to air-water heat exchanges but less sensitive to radiative fluxes than larger streams. One can hypothesize that smaller streams have a lower volume of water to heat up than larger streams but also are likely to experience greater relative shading by riparian trees than wider rivers downstream.

This finding, at first, looks partly inconsistent with Garner et al. (2014), who concluded that larger basins were more sensitive to climate than smaller ones, because (i) headwater stream being located at the start of the network have less time than larger streams to reach equilibrium with AT further downstream, and (ii) headwater streams are more likely to be shaded (riparian woodlands, topography). However, Garner et al. (2014) was based on cluster analysis; small basins were included in one cluster only, which also included permeable basins. As a consequence, it is likely that permeability and size influences were in some extent confounded. In contrast, the sites used in this paper cover all combinations of size/permeability basin types. Secondly, as noted by Kelleher et al. (2012), within the small stream type, one needs to distinguish between shaded (i.e. due to with riparian woodland or topography) and exposed streams, with shaded streams behaving more like permeable streams. Only basin-wide land cover information was available for 29 out of 35 sites: 27 basins are under 20% woodland. While one cannot exclude woodland being concentrated on the riparian corridor of each site, it is sensible to assume the 35 sites have a mix of shaded and exposed streams. Although it would explain the pattern with 'all seasons' SWR (more shading, less incoming sun), the shaded headwater argument has to be considered inconclusive in relation to the wider climate controls.

With regard to basin elevation, results can be summarised as follows: (i) 'all seasons' model, WT in higher elevation basins is more sensitive to AT but less sensitive to SWR; (ii) winter model, WT in higher elevation basins is more sensitive to SH. These patterns can be explained partly by elevation, partly by the fact that permeability, size and elevation are not strictly independent in the UK. As noted above, elevation and rainfall co-vary greatly in the UK, so that upland basins are wetter than lowland basins, hence associated with greater precipitation (i.e. with more cloud cover and consequently, less influenced by SWR). In terms of basin types, the study sites have no upland permeable basins (the UK geology is such that this type hardly occurs in any case), plus high elevation basins tend to be smaller basins. The patterns observed with elevation, which are consistent with those for permeability and size, are most likely partly reflecting the upland basins are also largely impermeable and smaller.

Although each property has been statistically identified as having an influence, the latter point leads to investigating how these influences may combine. The regression models of site-specific coefficients against permeability, size, and elevation presented in Table 5~~6~~ provide some quantification of the influence of basin properties, both on their own, and combined. In each case, the best model uses a single basin property, although the retention of other properties in the MMI sets confirms the role of all three. In three cases out of four ('all seasons' AT, autumn SWR, winter SH), permeability (BFIHOST) is dominant. Therefore, the patterns described above would be primarily set by basin permeability, then by size and elevation. At one end of the spectrum, small, upland, and/or impermeable basins are the most exposed to atmospheric heat exchanges, at the other end, large, lowland, and permeable basins are the least exposed.

6 Conclusions

By focusing on a nation-wide set of water temperature sites and extensive climate dataset, this study addressed some of the limits of previous UK papers (limited number of WT sites, climate predictors, and /or geographical extent); it also investigated formally seasonal patterns, and, by using a wide range of basin descriptors, improved knowledge of the role of basin properties as modifiers of climate–WT associations.

With regards to the need to explore alternative modelling techniques to maximise data utility, ML modelling allowed to model climate-WT responses both at site and at national scales, thereby adressing the limitation of empirical regression-based models compared to deterministic models (Caissie, 2006). While the present ML models took into account discrepancies in temperature sampling (eg data from sites with 15-min recording may show different patterns from sites with weekly data), the effect of these discrepancies were not investigated explicitly, and would merit further research. In addition, the model selection based on the MMI approach permitted to investigate climate variables that would been most likely excluded by standard selection techniques, and identify their influence as secondary controls.

In relation to research Aim 1 (improved understanding of large-scale climate–WT associations), the modelling exercise showed that all of the six climate predictors investigated in this study play a role as a control of water temperature. AT and SWR are important for all models/ seasons, while LWR, SH, and WS are important for some models/ seasons only. The form and strength of the climate-stream temperature association vary depending on season and on water temperature. The dominating climate drivers and physical processes may change across seasons, and across the stream temperature range. The impact of climate variability or change, whether short or long term (e.g. seasonal supra-seasonal, or inter-annual climatic drought, long-term air temperature increaes), and the benefit of mitigation measures (e.g. increasing shading) on stream temperatures need to be assesed accordingly. While this study focused on wider spatial paterns, it is noteworthy that stream temperature could also be influenced by micro-climate effects (as far as metadata could be scrutinised, the study sites were free of such effects), future research could investigate how micro-climate and climate data spatial resolution may influence the models.

In relation to research Aim 2 (assessing influence of basin properties as modifiers of climate-WT associations), the study confirmed the role of basin permeability, size, and elevation as modifiers of the climate-WT associations. The primary modifier is basin permeability, then size and elevation. Smaller, upland, and/or impermeable basins are the ones most influenced by atmospheric heat exchanges, while the larger, lowland and permeable basins are least influenced (note: some basin types occur less frequently or hardly in the UK, e.g. upland permeable). This means that, in addition to seasons and temperature range, the impact of climate on stream temperatures and the benefits of mitigation schemes may vary with location. This study shows the importance of accounting properly for the spatial and temporal variability of climate-stream temperature associations and their modification by basin properties.

Data availability

The dataset used in this paper is available from the NERC EIDC open-access data repository (Laizé and Bruna Meredith, 2015).

Acknowledgements

The authors would like to thanks CEH colleagues for their help with (a) sourcing the water temperature datasets, Mike Bowes, Francois Edwards, Ned Hewitt, Mike Hutchins, and Gareth Old; and (b) retrieving the CHESS data, Eleanor Blyth, Douglas Clark, and Richard Ellis. The authors acknowledge financial support from the Natural Environment Research Council (NERC) through its National Capability funding to the Centre for Ecology and Hydrology, and its PhD funding to the first author. Some of the material in this paper was taken from the first author's PhD thesis.

References

Akaike, H.: A new look at the statistical model identification, IEEE Trans. Autom. Control, 19(6), 716-723, 1974.

Bayliss, A.: Flood Estimation Handbook, Volume 5: Catchment Descriptors, Institute of Hydrology, Wallingford, UK, 1999.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Menard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B. and Harding, R. J.: The Joint UK Land Environment

Simulator (JULES), model description - Part 1: Energy and water fluxes, Geosci. Model Dev., 4(3), 677-699, 2011.

Boisneau, C., Moatar, F., Bodin, M. and Boisneau, P.: Does global warming impact on migration patterns and recruitment of Allis shad (Alosa alosa L.) young of the year in the Loire River, France?, Hydrobiologia, 602, 179-186, 2008.

Boscarino, B. T., Rudstam, L. G., Mata, S., Gal, G., Johannsson, O. E. and Mills, E. L.: The effects of temperature and predator-prey interactions on the migration behavior and vertical distribution of Mysis relicta, Limnol. and Oceanogr., 52(4), 1599-1613, 2007.

Bower, D., Hannah, D. M. and McGregor, G. R.: Techniques for assessing the climatic sensitivity of river flow regimes, Hydrol. Process., 18(13), 2515-2543, 2004.

Broadmeadow, S. B., Jones, J. G., Langford, T. E. L., Shaw, P. J. and Nisbet, T. R.: The Influence of Riparian Shade on Lowland Stream Water Temperatures in Southern England and Their Viability for Brown Trout, River Res. Appl., 27(2), 2011.

Brown, L. E., Cooper, L., Holden, J. and Ramchunder, S. J.: A comparison of stream water temperature regimes from open and afforested moorland, Yorkshire Dales, northern England, Hydrol. Process., 24(22), 3206-3218, 2010.

Burnham, K. P. and Anderson, D. R.: Model selection and multimodel inference: a practical information-theoretic approach, Springer, 2002.

Caissie, D.: The thermal regime of rivers: a review, Freshwater Biol., 51(8), 1389-1406, 2006.

Caissie, D., Satish, M. G. and El-Jabi, N.: Predicting water temperatures using a deterministic model: Application on Miramichi River catchments (New Brunswick, Canada), J. Hydrol., 336(3-4), 303-315, 2007.

Crisp, D. T.: Water temperature of Plynlimon streams, Hydrol. Earth Syst. Sci., 1(3), 535-540, 1997.

Edwards, F. K., Lauridsen, R. B., Fernandes, W., Beaumont, W. R., Ibbotson, A. T., Scott, L., Davies, C. E. and Jones, J. I.: Re-introduction of Atlantic salmon, Salmo salar L., to the Tadnoll Brook, Dorset, Proceedings of the Dorset Natural History and Archaeological Society, 2009.

Evans, C. D., Monteith, D. T., Reynolds, B. and Clark, J. M.: Buffering of recovery from acidification by organic acids, Sci. Total Environ., 404(2-3), 316-325, 2008.

Evans, E. C., McGregor, G. R. and Petts, G. E.: River energy budgets with special reference to river bed processes, Hydrol. Process., 12(4), 575-595, 1998.

Garner, G., Hannah, D. M., Sadler, J. P. and Orr, H. G.: River temperature regimes of England and Wales: spatial patterns, inter-annual variability and climatic sensitivity, Hydrol. Process., 28(22), 5583-5598, 2014.

Grueber, C. E., Nakagawa, S., Laws, R. J. and Jamieson, I. G.: Multimodel inference in ecology and evolution: challenges and solutions, J. Evolution. Biol., 24(7), 1627-1627, 2011.

Hannah, D. and Garner, G.: River water temperature in the United Kingdom: changes over the 20th century and possible changes over the 21st century, Prog. Phys. Geogr., 2015.

Hannah, D. M., Malcolm, I. A., Soulsby, C. and Youngson, A. F.: Heat exchanges and temperatures within a salmon spawning stream in the cairngorms, Scotland: Seasonal and sub-seasonal dynamics, River Res. Appl., 20(6), 635-652, 2004.

Hannah, D. M., Malcolm, I. A., Soulsby, C. and Youngson, A. F.: A comparison of forest and moorland stream microclimate, heat exchanges and thermal dynamics, Hydrol. Process., 22(7), 919-940, 2008.

Hough, M. N. and Jones, R. J. A.: The United Kingdom Meteorological Office rainfall and evaporation calculation system: MORECS version 2.0-an overview, Hydrol. Earth Syst. Sci., 1(2), 227-239, 1997.

Hrachowitz, M., Soulsby, C., Imholt, C., Malcolm, I. A. and Tetzlaff, D.: Thermal regimes in a large upland salmon river: a simple model to identify the influence of landscape controls and climate change on maximum temperatures, Hydrol. Process., 24(23), 3374-3391, 2010.

Imholt, C., Soulsby, C., Malcolm, I. A., Hrachowitz, M., Gibbins, C. N., Langan, S. and Tetzlaff, D.: Influence of Scale on Thermal Characteristics in a Large Montane River Basin, River Res. Appl., 29(4), 403-419, 2013.

Jeppesen, E. and Iversen, T. M.: 2 Simple-Models for Estimating Daily Mean Water Temperatures and Diel Variations in a Danish Low Gradient Stream, Oikos, 49(2), 149-155, 1987.

Johnson, S. L.: Stream temperature: scaling of observations and issues for modelling, Hydrol. Process., 17(2), 497-499, 2003.

Kelleher, C., Wagener, T., Gooseff, M., McGlynn, B., McGuire, K. and Marshall, L.: Investigating controls on the thermal sensitivity of Pennsylvania streams, Hydrol. Process., 26(5), 771-785, 2012.

Keller, V., Young, A. R., Morris, D. G. and Davies, H.: Continuous Estimation of River Flows (CERF) Technical Report: task 1.1: Estimation of Precipitation Inputs. UK: 1-36, 2006.

Laizé, C. L. R.: Controls and modification of large-scale climate–hydrology–ecology associations, PhD thesis, University of Birmingham, 2015.

Laizé, C. L. R. and Bruna Meredith, C.: Water temperatures for the period 1984 to 2007 at 35 sites on 21 UK rivers, doi:10.5285/65400133-9cfb-4cf4-bffd-97f1c3752025, NERC-EIDC, 2015.

Laizé, C. L. R. and Hannah, D. M.: Modification of climate-river flow associations by basin properties, J. Hydrol., 389(1-2), 186-204, 2010.

Langan, S. J., Johnston, L., Donaghy, M. J., Youngson, A. F., Hay, D. W. and Soulsby, C.: Variation in river water temperatures in an upland stream over a 30-year period, Sci. Total Environ., 265(1-3), 195-207, 2001.

Malcolm, I. A., Hannah, D. M., Donaghy, M. J., Soulsby, C. and Youngson, A. F.: The influence of riparian woodland on the spatial and temporal variability of stream water temperatures in an upland salmon stream, Hydrol. Earth Syst. Sci., 8(3), 449-459, 2004.

Mohseni, O., Stefan, H.G., Erickson, T.R.: A non-linear regression model for weekly stream temperatures. Water Res. Res. 34(10), 2685–2693, 1998.

Nakagawa, S. and Schielzeth, H.: A general and simple method for obtaining *R^2* from generalized linear mixed-effects models, Methods in Ecology and Evolution, 4, 133-142, 2013.

Neal, C., Robinson, M., Reynolds, B., Neal, M., Rowland, P., Grant, S., Norris, D., Williams, B., Sleep, D. and Lawlor, A.: Hydrology and water quality of the headwaters of the River Severn: Stream acidity recovery and interactions with plantation forestry under an improving pollution climate, Sci. Total Environ., 408(21), 5035-5051, 2010.

Orr, H. G., Simpson, G. L., Clers, S., Watts, G., Hughes, M., Hannaford, J., Dunbar, M. J., Laizé, C. L., Wilby, R. L. and Battarbee, R. W.: Detecting changing river temperatures in England and Wales, Hydrol. Process., 2014.

Webb, B. W., Clack, P. D. and Walling, D. E.: Water-air temperature relationships in a Devon river system and the role of flow, Hydrol. Process., 17(15), 2003.

Webb, B. W., Hannah, D. M., Moore, R. D., Brown, L. E. and Nobilis, F.: Recent advances in stream and river temperature research, Hydrol. Process., 22(7), 2008.

Webb, B. W. and Nobilis, F.: Long-term perspective on the nature of the air-water temperature relationship: a case study, Hydrol. Process., 11(2), 1997.

Webb, B. W. and Zhang, Y.: Spatial and seasonal variability in the components of the river heat budget, Hydrol. Process., 11(1), 79-101, 1997.

Webb, B. W. and Zhang, Y.: Water temperatures and heat budgets in Dorset chalk water courses, Hydrol. Process., 13(3), 309-321, 1999.

Welton, J. S., Beaumont, W. R. C. and Ladle, M.: Timing of migration and changes in age structure of Atlantic salmon, Salmo salar L., in the River Frome, a Dorset chalk stream, over a 24-year period, Fisheries Manag. Ecol., 6(6), 437-458, 1999.

Wheater, H. S., Neal, C. and Peach, D.: Hydro-ecological functioning of the Pang and Lambourn catchments, UK: An introduction to the special issue, J. Hydrol., 330(1-2), 1-9, 2006.

Wilby, R. L., Johnson, M. F. and Toone, J. A.: Nocturnal river water temperatures: Spatial and temporal variations, Sci. Total Environ., 482, 157-173, 2014.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A. and Smith, G. M.: Mixed effects models and extensions in ecology with R, Springer, 2009.

Table 1. Climate–water temperature studies carried out in the UK.

Reference	Number of Sites	Number of Basins	Location	Number of Climatic Variables	Length of Study Period
Wilby *et al.* (2014)	36	2	central England	1	2 years
Garner *et al.* (2014)	38	38	England & Wales	1	18 years
Broadmeadow *et al.* (2011)	10	2	south England	3	3 years
Brown *et al.* (2010)	6	1	north England	2	2 years
Hrachowitz *et al.* (2010)	25	1	northeast Scotland	0	2 years
Hannah *et al.* (2008)	2	1	northeast Scotland	7*	2 years
Malcolm *et al.* (2004)	6	1	northeast Scotland	1	3 years
Hannah *et al.* (2004)	1	1	northeast Scotland	9*	6 months
Webb *et al.* (2003)	4	1	southwest England	1	5 years
Langan *et al.* (2001)	1	1	northeast Scotland	1	30 years
Webb and Zhang (1999)	2	2	South England	5	2 seasons
Evans *et al.* (1998)	1	1	west England	9*	17 days
Crisp (1997)	5	1	northwest Wales	1	3 years
Webb and Zhang (1997)	11	1	southwest England	4	2 seasons

* includes different measurements of related climatic variables

Table 2~~3~~. CHESS data.

Climate Variable	Abbreviation	Units	Process
Air temperature	AT	^{o}K	Convective energy exchanges at water surface; energy loss or gain
Long wave radiation	LWR	$W\ m^{-2}$	Downward energy bounced back by clouds; energy gain
Specific humidity	SH	$kg\ kg^{-1}$	Air moisture content; higher humidity reduces evaporation rate; energy loss (evaporation) or gain (condensation)
Precipitation	P	$kg\ m^{-2}d^{-1}$ ($mm\ d^{-1}$)	Advective exchanges; energy loss or gain
Short wave radiation	SWR	$W\ m^{-2}$	Downward direct energy (i.e. insolation); energy gain
Wind speed	WS	$m\ s^{-1}$	Increases evaporation (energy loss) and convective exchanges (air mixing; energy loss or gain)

Table 34. Generic response for the five average models.

	all seasons		winter		spring		summer		autumn	
	Coef.	*RI*	Coef.	*RI*	Coef.	*RI*	Coef.	*RI*	Coef.	*RI*
AT	0.5824	*1.00*	0.3955	*1.00*	0.6815	*1.00*	0.4969	*1.00*	0.6860	*1.00*
SWR	0.0055	*1.00*	0.0193	*1.00*	0.0073	*1.00*	0.0049	*0.64*	0.0003	*1.00*
LWR	-0.0149	*1.00*	0.0001	*0.13*	0.0020	*0.18*	-0.0126	*0.52*	-0.0013	*0.25*
WS	-0.1348	*1.00*	-0.0685	*0.68*	-0.0774	*0.63*	-0.3028	*1.00*	0.0181	*0.33*
SH	0.4664	*1.00*	0.6658	*1.00*	0.0772	*0.34*	0.1542	*0.53*	0.0507	*0.37*
P	0.0003	*0.26*	0.0007	*0.15*	-0.0041	*0.38*	-0.0004	*1.00*	-0.0045	*0.41*

Table 4~~5~~. Basin descriptors significantly related to site-specific model coefficients (ANOVA; $p \leq 0.05$).

Model	Predictor	Basin Meta-property	FEH Descriptor	Type of Association
all seasons	AT	Elevation	ALTBAR	Positive
		Permeability	BFIHOST	Negative
		Size	AREA*	Negative
all seasons	SWR	Elevation	ALTBAR	Negative
		Size	AREA	Positive
autumn	SWR	Permeability	BFIHOST	Negative
		Size	AREA*	Negative
winter	SH	Elevation	PROPWET	Positive
		Permeability	BFIHOST	Negative
		Size	AREA*	Negative

*tested on natural log

Table 5~~6~~. Linear regressions of site-specific coefficients as function of basin properties (models ordered by increasing AICc; best model in bold characters, all other models are within four AICc points of best model hence selected via MMI).

WT Model	Coefficient	Linear Regression	R^2	AICc
all seasons	AT	**BFIHOST**	**0.370**	**-31.3**
		BFIHOST+ALTBAR	0.403	-30.1
		BFIHOST+ln(AREA)	0.381	-29.3
		BFIHOST+ln(AREA)+ALTBAR	0.411	-28.3
all seasons	SWR	**ALTBAR**	**0.177**	**-277.5**
		ALTBAR+ln(AREA)	0.183	-275.2
		ln(AREA)	0.089	-274.0
autumn	SWR	**BFIHOST**	**0.125**	**-223.1**
		ln(AREA)	0.115	-222.6
		BFIHOST+ln(AREA)	0.136	-220.9
winter	SH	**BFIHOST**	**0.192**	**48.7**
		ln(AREA)	0.162	50.0
		BFIHOST+ln(AREA)	0.203	50.8
		BFIHOST+PROPWET	0.192	51.3
		PROPWET	0.123	51.6
		PROPWET+ln(AREA)	0.178	51.9

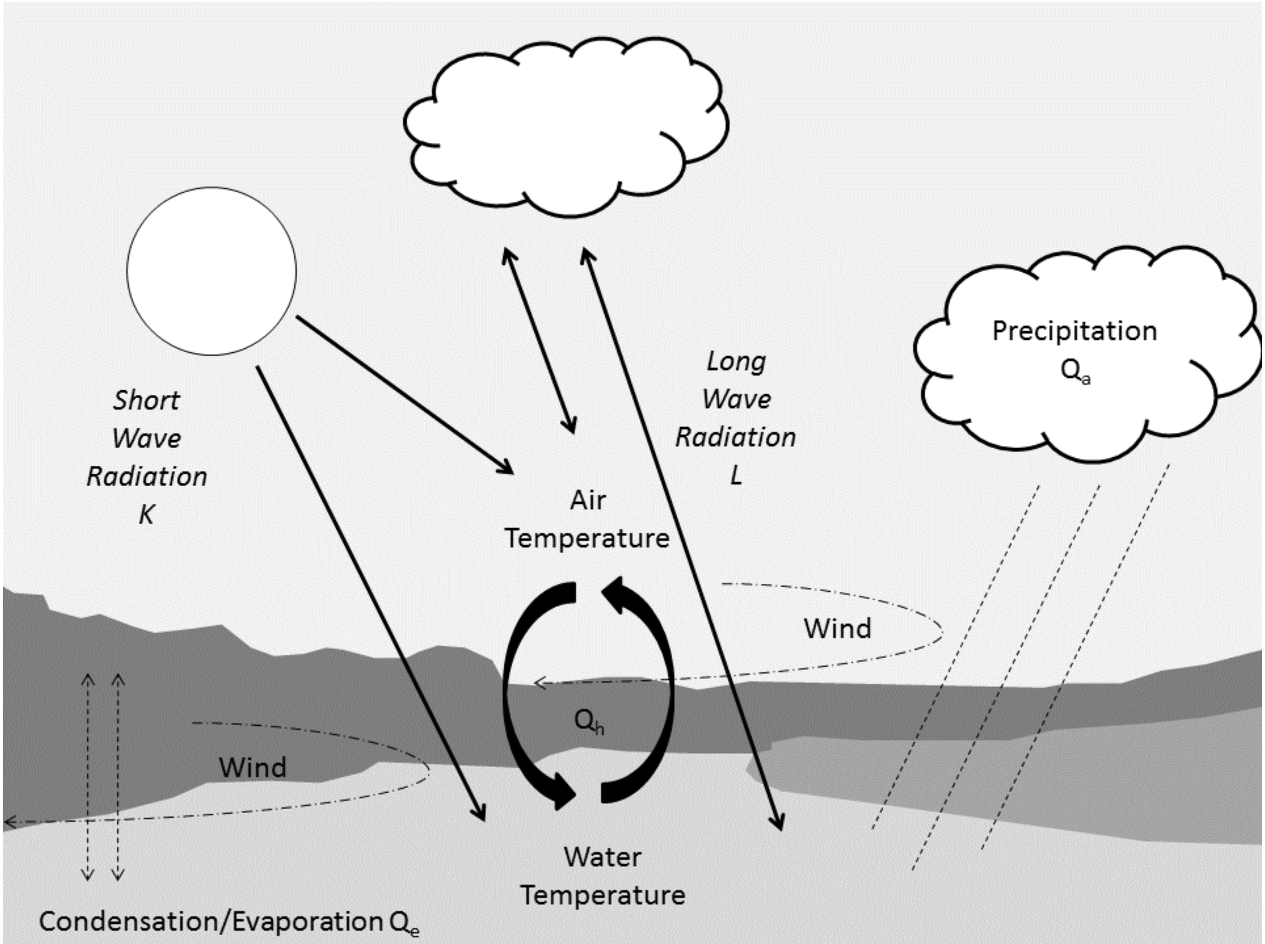

Figure 1. Multiple interdependent climate controls of water temperature; Q^* is the sum of K and L; Q_a also small advective fluxes due to inflow/outflow into river, hyporheic exchange, groundwater (not shown); [adapted from Caissie (2006) and Hannah et al. (2008)].

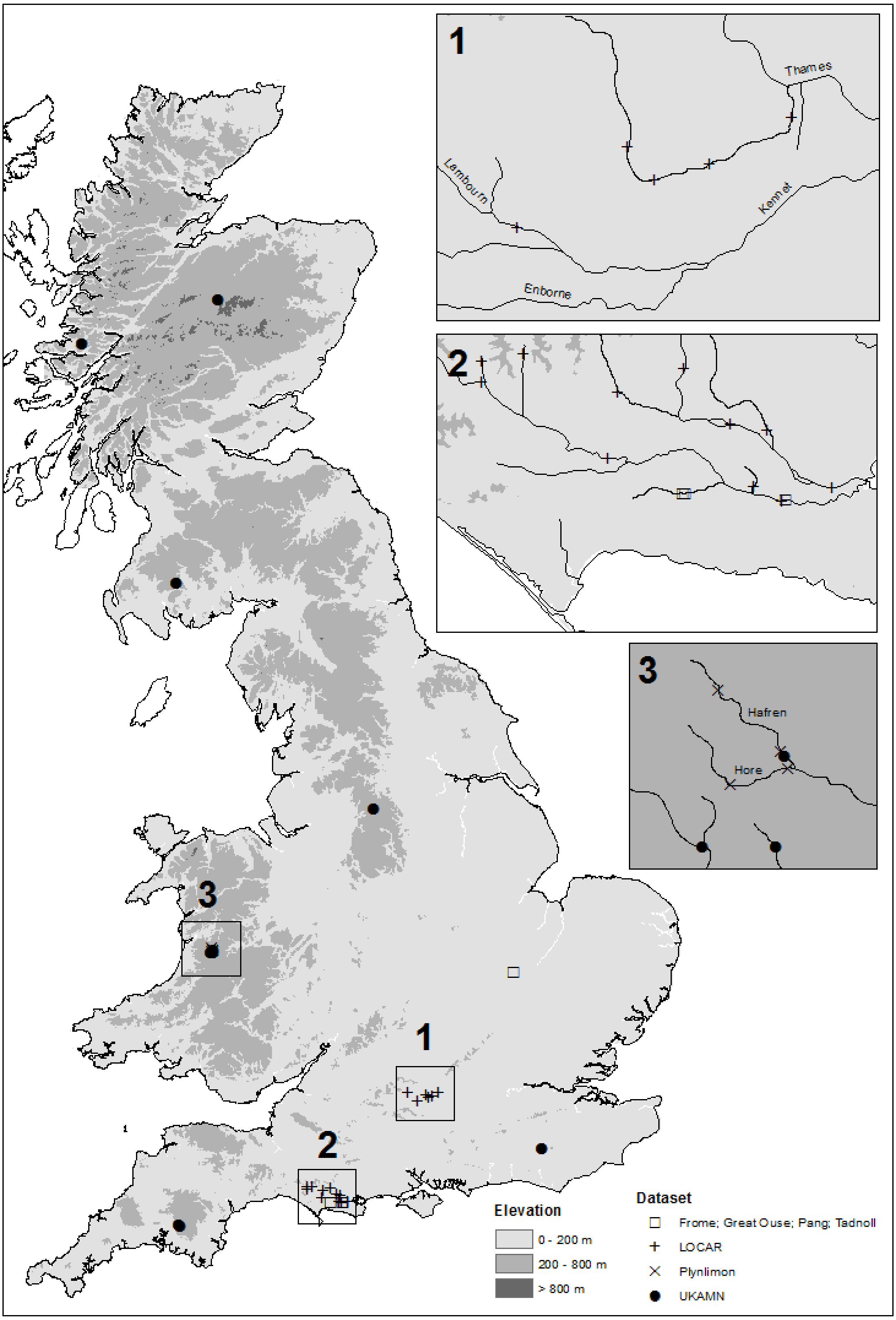

Figure 2~~3~~. Location map of the study sites.

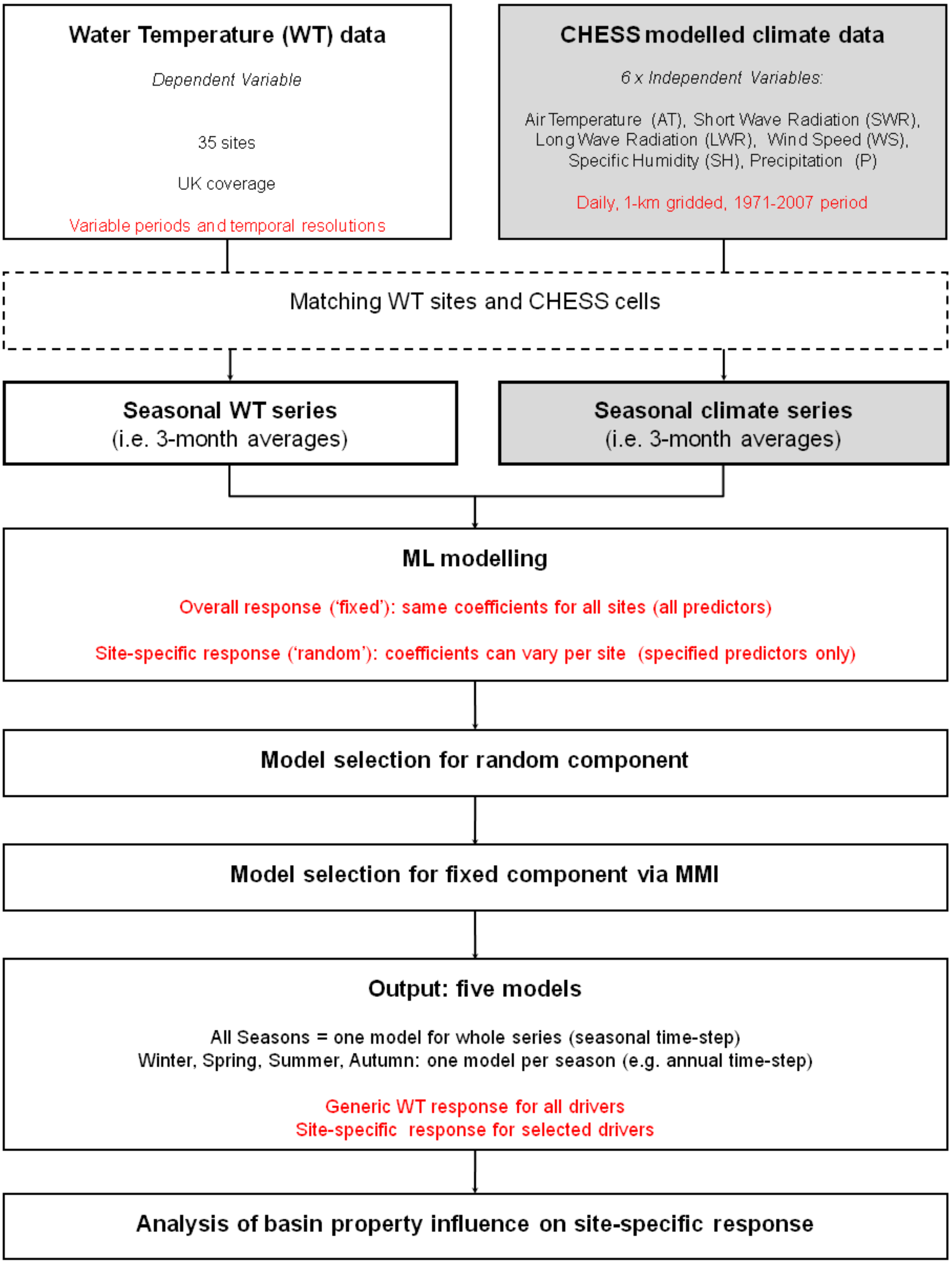

Figure 3~~2~~. Study flow chart.

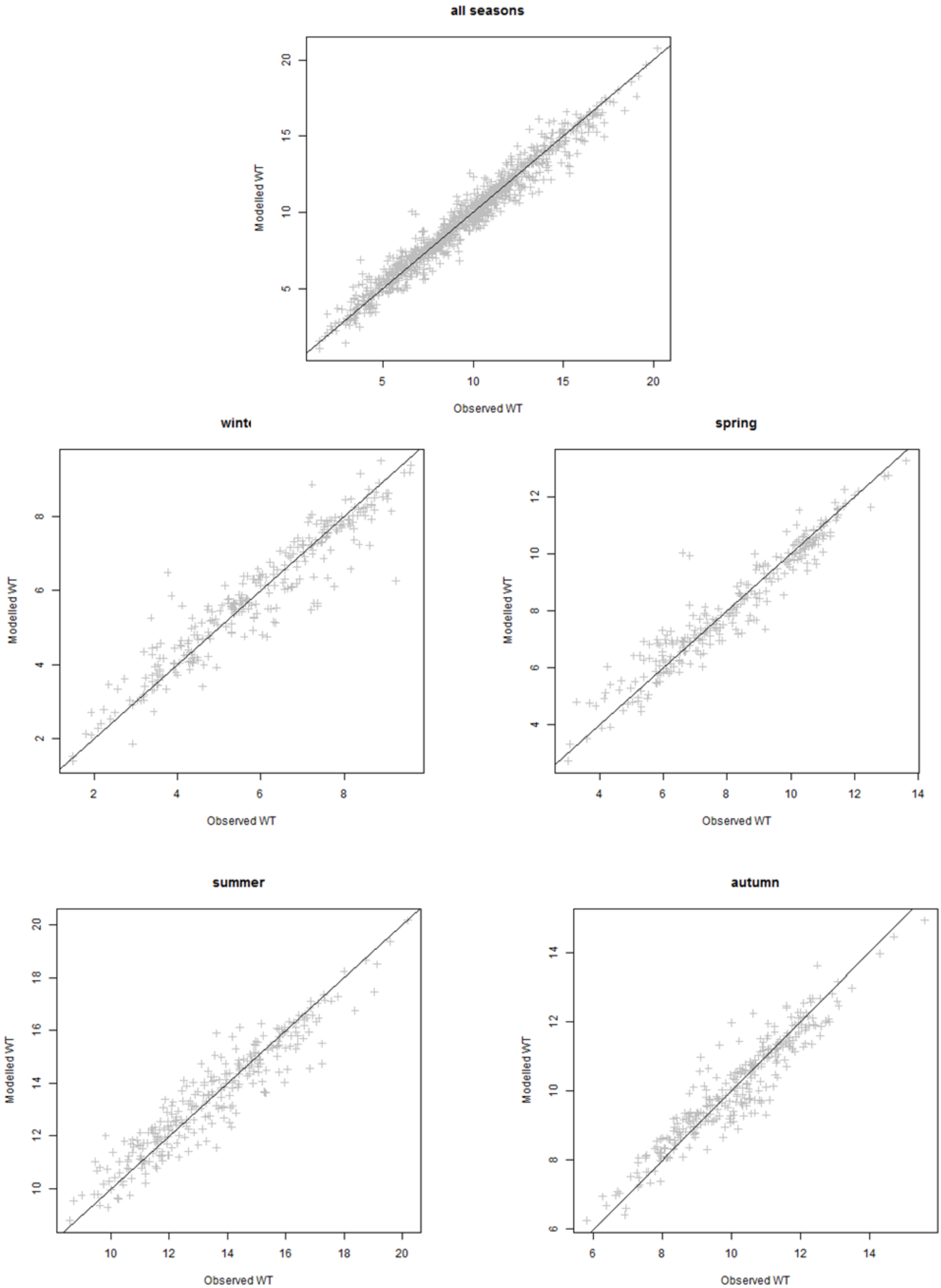

Figure 4. Plots of observed and modelled water temperature for the five models.

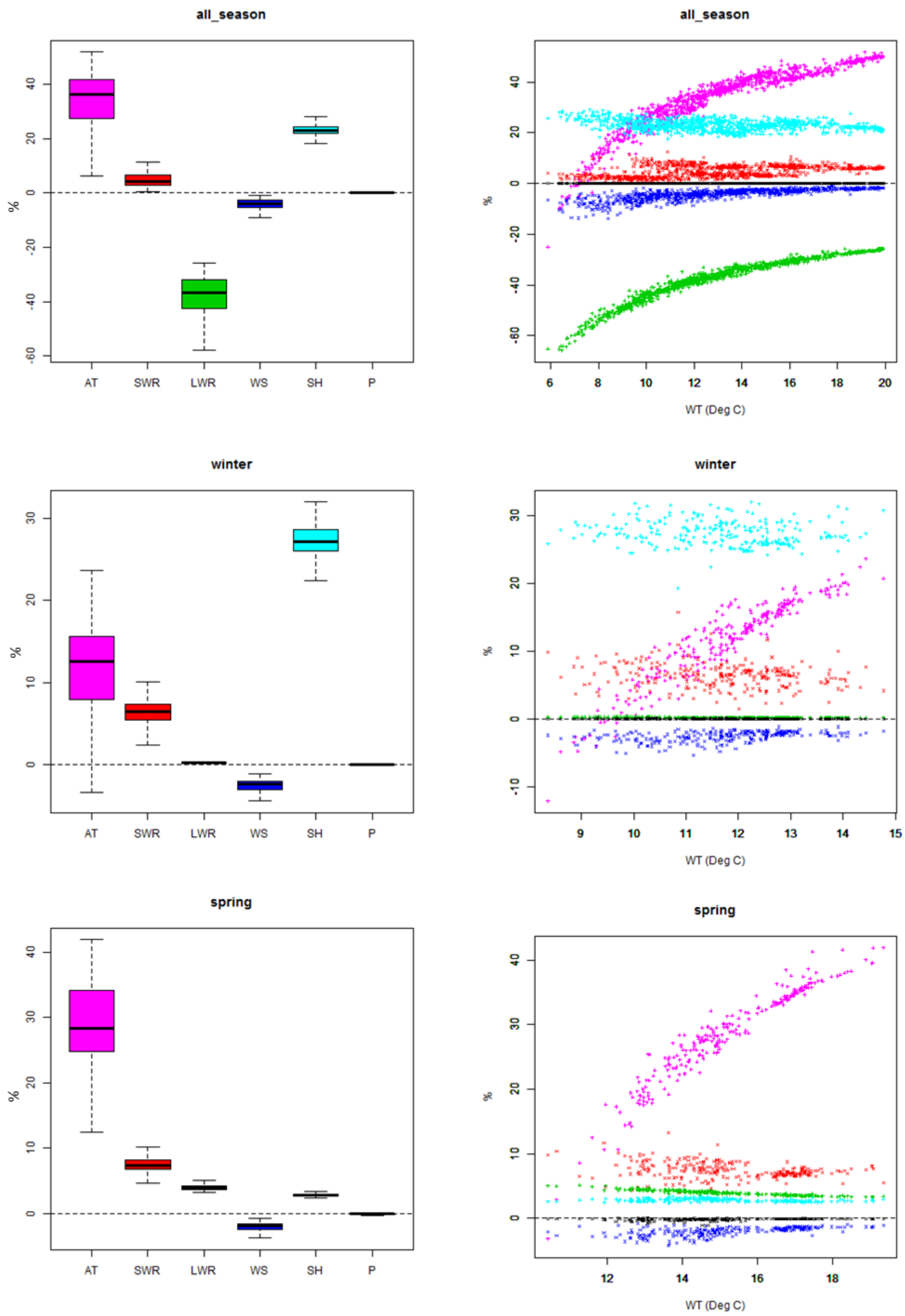

Figure 5a. Contributions of climate predictors to modelled WT (all seasons, winter, and spring): left-hand side, boxplots of percentage contributions of climate predictors to modelled WT values for all data-points (except outliers); right-hand side, scatter plots of percentage contributions of climate predictors to modelled WT values against modelled WT values for all data-points; colour-coding for all plots: magenta, AT; red, SWR; green, LWR; dark blue, WS; cyan, SH; black, P.

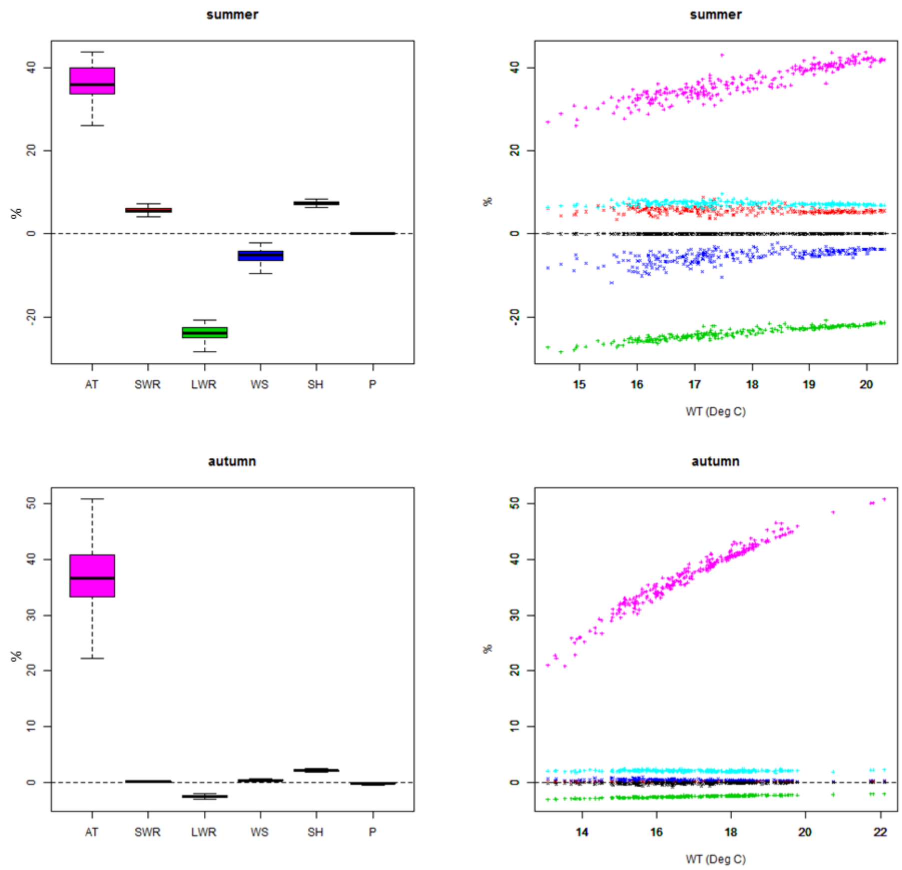

Figure 5b. Contributions of climate predictors to modelled WT (summer and autumn): left-hand side, boxplots of percentage contributions of climate predictors to modelled WT values for all data-points (except outliers); right-hand side, scatter plots of percentage contributions of climate predictors to modelled WT values against modelled WT values for all data-points; colour-coding for all plots: magenta, AT; red, SWR; green, LWR; dark blue, WS; cyan, SH; black, P.