Author Response on "Can river temperature models be transferred between catchments?" by Faye L. Jackson et al.

We are pleased that the anonymous reviewers considered that the paper was "technically sound", would be "of interest to the readership of HESS" and that the questions are of "present interest to the community". We also note that none of the reviewers raised any issues in relation to the analysis that was presented or the methods, we have therefore focussed our revisions on the introduction and discussion, with some additional illustration of the covariate data. We thank the reviewers for their useful and thoughtful comments and for identifying additional literature. We have attempted to incorporate these comments into a revised manuscript, while also recognising the challenges associated with integrating four different perspectives and the need to maintain a clear and concise manuscript. Below we provide specific responses to the reviewer's comments and identify where we have made changes to the manuscript.

We consider that we have adequately addressed the issues that have been raised and hope that the paper is now suitable for publication in HESS.

Referee #1

Overall, the paper is technically sound.

- We are pleased to see the reviewer agrees that the paper is technically sound

A short review of the literature on the use of GAM in water temperature modelling would be a welcome addition.

- We are unsure whether the reviewer would like to see a discussion of the use of GAMs more generally in river temperature models, or the use of GAMs in combination with river network smoothers to account for spatial covariance. If the latter, then we are only aware of one previous paper doing this (Jackson et al., 2017). More broadly we are only aware of three other modelling papers that use GAMs to model river temperature and these are quite different from the current application. Li et al. (2014) use GAM's to model temporal variability in Tw as a function of Ta allowing for a temporally variable slope and intercept that varied smoothly over time; Orr et a. (2015) used GAMs to model temporal trends in Tw from a diverse set of logged and spot sampled temperature data; Wehrly et al. investigated the relative ability of GAMs, linear mixed models and kriging to predict spatial variability in water temperature but did not incorporate a combination of landscape covariates and consideration of spatial covariance in a single model (i.e. kriging model did not contain covariates and GAM / LMM did not consider network related covariance).
- Given the very few examples of GAMs being used in river temperature modelling, difficulties in incorporating such a review in the current paper structure and the fact that these issues are discussed in Jackson et al. (2017) we have not added further discussion of this topic to the current paper and hope that the referencing of Jackson et al. (2017) will suffice.

In my opinion, the main weakness of the manuscript is in the discussion. Two main points need to be further discussed:

1. The challenge of inter basin transferability using air temperature needs to be further addressed and potential next steps identified. For instance, the readers may ask the following question: Is it the seasonal component or the residual component of air temperature that make it so difficult to transfer? Could an alternative model be envisaged in which the parameters of the air temperature seasonal harmonic be estimated/transferred?

- In this paper we focus explicitly on the issue of spatial variability in Tw during summer (as do many other studies). In this context, the seasonal variability is something to be examined at another time, although recent studies have shown that seasonal variability in local Tw~Ta relationships can indeed be modelled (Li et al., 2014). In this study we tried to understand and predict the spatial variability in Tw_{max}, the value of Tw at a single point in time (hottest 7 day period) from covariates that include the corresponding value of Ta, namely Ta_{max}. For Ta_{max} to be an accurate (precise and unbiased) predictor of the spatial variability in Tw_{max} requires the Tw_{max}~Ta_{max} relationships to be broadly consistent between catchments. For example, if the within-catchment relationships between Tw_{max} and Ta_{max} are linear, the slopes and intercepts must be similar across catchments to use a Tw_{max}~Ta_{max} relationship developed in one catchment to predict Tw_{max} from Ta_{max} in a different catchment. Our study finds that the relationships are not consistent across catchments and that models using Ta_{max} developed in one catchment might not transfer well to another.
- It is well documented that (temporal) Tw~Ta relationships can vary substantially between sites and catchments due to the effects of other controls including discharge, groundwater – surface water interactions, hydrogeology and landuse (e.g. Tague et al., 2007). If such controls differ between catchments (for example if one catchment has a higher groundwater component than another) then this would likely lead to spatial relationships between Tw_{max} and Ta_{max} that differ between catchments (particularly since differences are accentuated at high temperatures).
- Because Ta and Tw respond to similar physical drivers (Johnson, 2003), Ta is a good predictor of the temporal variability in Tw once that relationship has been established at a site. However, in terms of processes, Ta is not the dominant control on Tw (e.g. Hannah et al., 2008) and there are therefore major challenges in predicting how this relationship should vary spatially. It is this understanding that is required to make Ta a good predictor of the spatial variability in Tw. The finding that Tw~Ta relationships vary widely between catchments (both temporally and spatially as we demonstrate for the Tw_{max}~Ta_{max} relationship) with varying characteristics is not new; it is simply that we have shown this to be a problem for predicting spatial variability in Tw from Ta when models are transferred.
- One approach that could be considered for future work would be to model the Tw~Ta relationship and then allow this relationship to vary spatially depending on other covariates that are known to affect this relationship. For example, in the context of the current paper, the intercept and slope of a linear Tw_{max}~Ta_{max} relationship could be related to catchment-scale landscape or hydrological covariates. We have added some discussion of this to the paper. However, such an approach would require more than the four catchments considered here. It is also important to remember that, in many cases, managers do not have access to data from a wide range of catchments

and are forced to make management decisions about one catchment based on models developed in another. In this context, our paper suggests that models using landscape covariates would be more robust than those also using some Ta metric.

- Given that this issue was raised by several reviewers we have revised the paper to further emphasise why we would not expect Ta alone to necessarily be a good predictor of the spatial variability in Tw at a fixed point in time, when models are transferred between catchments or regions.

2. The problem of the impossible air-water temperature relationship at Bladnoch needs to be further explained. This is very unusual. I suspect that it is caused in part by station locations on this basin and by the fact that the samples used in the model only include air temperatures ranging between 18.5 and 20.5 deg C (figure 4)?

- We agree that the physically implausible relationship could be caused by the limited range of Ta_{max} which was much smaller in this catchment (due to its physiographic characteristics) than the others, coupled with the way the landscape and hydrological controls vary across the catchment and in particular act at the station locations. We have added text to reflect this. We note that, whilst there is sufficient evidence of nonlinearity in the data to force a smooth effect of Ta_{max} in the single catchment model, there is not sufficient evidence to do so in the multi-catchment model (when the evidence of non-linearity in the Bladnoch relationship is swamped by the lack of evidence of non-linearity in the other catchments). The slope of the Bladnoch Tw_{max}~Ta_{max} relationship in the multi-catchment model is negative (but not significantly so) (Figure 8c). However, our main point, which remains valid, is that care is required when transferring Tw~Ta relationships among catchments (or regions).

Referee #2

In this paper, the authors explore the transferability of statistical models to predict a metric of maximum summer stream temperature. They use data from four catchments in Scotland collected during one summer season. Consistent relations with landscape variables were found; however, the relation between stream temperature and air temperature was inconsistent among catchments, and was even physically implausible in one. The authors conclude that, overall, the ability to transfer statistical models among catchments is limited without further research to gain a better understanding of inter-catchment differences.

Considering the high level of concern about rising stream temperature and the increasing number of papers focused on modeling stream temperature over the last decade or so, the topic is timely and would be of interest to the readership of HESS.

- We are pleased that the reviewers considers that this paper will be of value to the readers of HESS and believe that we can adequately address the reviewers concerns below

I have a number of concerns about this work in its present form.

1. It is difficult to judge the novelty and significance of this work because the authors have not effectively placed it into the context of previous research on the topic. It is unclear what specific knowledge gaps are being addressed, or what new knowledge has been generated. Although the authors do cite a number of relevant, related studies (e.g., Hrachowitz et al., 2010; Chang and Psaris, 2013), they do not adequately address how their results are similar to or differ from those in previous studies. In addition, a number of papers not cited have addressed landscape-scale modelling of stream temperature, including Isaak and Hubert (2001), Scott et al. (2002), Tague et al. (2007), Wehrly et al. (2009) and Moore et al. (2013). Some of the previous papers have focused on extensive regions and thus have implicitly demonstrated that models based on landscape variables can be applied consistently across multiple catchments.

- We agree that we did not cite all of the papers that model Tw as a function of landscape covariates and/or Ta; rather we cited a selection of papers that illustrate the various approaches and statistical methods that have been explored. We have added additional references where this broadens the discussion. However, none of the studies raised by the reviewer explicitly demonstrate the consequences of transferring models developed in one catchment to another catchment and none of the suggested studies consider network structure in the underlying models, an extremely important part of the transferability story. We have thus further clarified this uniqueness of our study in the introduction.
- Isaak and Hubert (2001) developed a spatial model for a single river (catchment), but did not assess the ability of this model to predict within or between (transfer) river catchments or regions. Similarly Scott (2002) investigated the effects of landscape covariates on river temperature (and water quality) in a single large river catchment (series of sub-catchments) but again did not assess between-catchment transferability of the resulting models.
- Tague et al. (2007) investigated Tw~Ta relationships across sites in Western Oregon and observed substantial differences in Tw~Ta relationships, much of which could be explained by large scale differences in hydro-geology. This paper adds to the existing references already in our paper that suggest Tw~Ta relationships can be highly variable between sites and catchments and provides further support for our suggestion that Ta models may not transfer well from one catchment to another. Indeed the abstract for Tague et al. (2007) states "In this study we show that, in regions where groundwater inputs are key controls and the degree of groundwater input varies in space, air temperature alone is unlikely to explain within-landscape stream temperature patterns". We are more than happy to cite this paper, providing further support for our findings. However, we note that again this paper does not explicitly demonstrate the consequences of transferring Tw~Ta relationships between catchments as we do in our paper.
- Moore et al. (2013) and Wehrly et al. (2009) both fit large scale Tw models using landscape covariates and air temperature as predictors, the former focussed on maximum mean weekly temperatures and the latter on mean July temperatures. Wehrly et al. investigate a range of statistical models, including linear mixed models that account for (Euclidean) spatial structure, and find that across the large spatial scales investigated, the models that considered

spatial structure performed best (but not by a large margin). Both of these papers examined the performance of predictions (within the dataset) using random selections of sites. This provides a measure of the performance of the models to **interpolate** across the model space. Wehrly et al. include terms to account for spatial structure and this allows Tw predictions to vary spatially (independent of the covariates); however it is unclear how well the model would perform when predicting Tw in new regions (e.g. into adjacent states). Moore et al. acknowledge that there could be some negative spatial bias in their model in the north east, indicating that the covariates may not be completely transferable or that there is additional spatial variability that should be accounted for. Indeed Figure 7 in Moore suggests that the Tw~Ta relationship is not constant across sites and, consequently, using a 'global' Tw~Ta relationship (i.e. fitting an average relationship across all sites) could introduce biases into predictions depending on the spatial distribution of the within-site response gradients. To determine the ability of the Moore model to extrapolate rather interpolate would have required the model to be fitted regionally and then the predictions tested in other regions (for example, using cross-validation with one sub-region excluded at a time). The reviewer notes that "some of the previous papers [presumably the Moore and Wehrly papers] have focused on extensive regions and thus have implicitly demonstrated that models based on landscape variables can be applied consistently across multiple catchments". This is true up to a point. However, because the models are fitted at such a large scale, it does not follow that the models will be good predictors of local (e.g. within-catchment) spatial variability in Tw. For managers operating at a more local (catchment) scale, areas e.g. at risk due to sustained high temperatures might not be well identified because the local Tw~Ta relationship is different to the global Tw~Ta relationship.

- Our paper investigated the consequences of **extrapolation** rather than **interpolation**. We investigated the consequences of transferring model predictions between catchments (i.e. to areas outside of the model domain). This is often required when widespread data are not available, for example when financial and logistical considerations have focussed data collection around a few critical rivers. The consequences of **extrapolation** have not been investigated previously as far as we are aware. In particular, we show that for our catchments, models that only use landscape covariates provide more reliable predictions that those using a Ta metric. Further, because there is also increasing interest in the use of spatial statistical river network models (and recognition of the need to use these models where data collection is focussed on particular rivers) we think it important to quantify the relative accuracy of predictions for new sites within a river catchment for which there are already data, with those for catchments for which there are no data (and only landscape covariates / Ta can provide predictions). This has not been explored in previous papers. Finally, given the widespread use of Ta in large scale models we find it useful to highlight that Tw~Ta relationships are not universally applicable and that this can result in catchment or regional biases. Although Tague et al. suggest this could be an issue, we explicitly demonstrate some of the problems.
- We strongly disagree that the knowledge gaps being addressed are unclear, and this issue was not raised by the other 3 reviewers. Nevertheless, we have added material to the introduction to make the issues clearer and further clarify the value of the current study.

What was the sampling design? Were sites selected randomly within some predefined strata (e.g., based on catchment area)? This point is important, because a carefully designed sampling scheme can minimize issues with multi-collinearity and enhance model identifiability.

- The sampling design for the sites is described in detail by Jackson et al. (2016) which is cited in the paper. The design aimed to address the issues raised by the reviewer. In short, we tried to cover the environmental range of all the predictor covariates across space (in each of the catchments). It is a strategically designed network with careful quality control. Unfortunately it is not possible to entirely exclude collinearity given the spatial structure of the rivers involved and their physiographic / hydro-climatological context, but we did exclude strongly correlated covariates from the analysis (see methods section)

The authors note that the relation with air temperature is inconsistent. In discussing this point, they draw upon the results of studies of the temporal relation between stream and air temperature. However, it is not valid to draw inferences about spatial patterns from temporal relations. See Luce et al. (2014) for a discussion of stream thermal sensitivity. The authors should focus on relations between stream and air temperature in a spatial context. The cited paper by Fellman et al. (2014) did try to include air temperature as a spatial covariate but did not find a significant relation. However, their sample size was only 9 and thus their analysis had limited power. A number of studies have found significant relations between stream and air temperature in a spatial context (e.g., Tague et al., 2007; Wehrly et al., 2009; Moore et al., 2013).

- We agree that, ideally, we should focus on relations between stream and air temperature in a spatial context. However, whilst other studies have found significant spatial relationships between stream and air temperature (as we did in the Tweed and Dee and, admittedly implausibly, in the Bladnoch), as far as we are aware ours is the first study to investigate differences in spatial relationships between catchments or regions. To explain why such differences might exist, we feel that it is entirely valid to consider the widely reported variation in temporal Tw~Ta relationships both between sites and catchments. Clearly, the more temporal Tw~Ta relationships vary between sites, the harder it will be to establish significant spatial Tw~Ta relationships, and the less precise predictions based on such spatial relationships will be. Equally, just because there is variation in the temporal relationships between sites, it doesn't necessarily mean that a spatial relationship developed in one catchment will differ from that in another. However, if the temporal relationships vary between catchments, for example because of differing ground water influences, then there is no reason to expect consistent spatial relationships across catchments, particularly when considering maximum summer temperatures, when differences between catchments will be accentuated. One of the papers suggested by the reviewer also uses these arguments. Tague et al. illustrate that the temporal Tw~Ta response varies depending on geological setting and state that "In this study we show that, in regions where groundwater inputs are key controls and the degree of groundwater input varies in space, air temperature alone is unlikely to explain

within-landscape stream temperature patterns". We have revised our discussion so that we are clearer about the possible effects of between-site and between-catchment variation in the temporal Tw~Ta relationship.

- We have included the additional Luce et al. (2014) reference which also shows considerable variability in Tw~Ta relationships (or climate sensitivity).
- We acknowledge that other papers have found a relationship between Tw and Ta in large scale spatial models. We agree that average relationships can be observed over large spatial scales. Indeed, we could have forced our multi-catchment model to have a common Tw_{max}~Ta_{max} relationship across catchments. However, instead we demonstrate that the Tw_{max}~Ta_{max} relationships are not consistent across catchments, and consequently that predictions from a model which assumes a common response would be biased. We have added discussion of this topic to the paper.

The authors should consider more thoroughly the reasons for the "physically implausible" relation between stream and air temperature for one catchment. Presumably it reflects a confounding effect of some variable not included as a covariate. For example, Hrachowitz et al. (2010, p. 3383) found that stream temperature tended to increase with elevation, which they attributed to the fact that upper elevations were not forested. Apart from this within-catchment scale, was the among-catchment variation in stream temperature consistent with the spatial pattern of air temperature? Perhaps air temperature can be effectively used at some spatial scales and not others? This could be an interesting point to address with reference to the broader literature.

- We believe that this response is most likely due to systematic spatial variability in covariates that affect the gradient of the temporal Tw~Ta relationship between sites (e.g. geology, hydrogeology, landuse), coupled with a limited range of observed values of Ta_{max} (see response to referee 1). If the gradient of the temporal Tw~Ta relationship varies among sites in a spatially systematic way, then the spatial Tw_{max}~Ta_{max} response could be more complex as seen here. We agree that this could be because of some covariate that are not included in the model, or because of some interaction between the covariates that are included. However the exact cause of the pattern in the Bladnoch is supposition. We have added discussion of this issue in our revision.
- We are a little unclear about precisely what is being asked. However, the estimated catchment effects in the multi-catchment LS model are not obviously related to the average Ta_{max} in each catchment (although with only four catchments, there is little power to pick up such a relationship).

The study is fundamentally constrained by the limited sampling in both time and space. Although the authors acknowledge some implications of the small sample size, including the inability to include interaction terms, they do not fully address how the small sample size has constrained their ability to draw inferences. Two key points follow.

a. The authors do not provide sample sizes, but inspection of Figure 1 suggests about 20 to 30 per catchment. These are not large sample sizes, especially for the application of multiple regression. One guideline is that

roughly 10 samples are required to support each predictor variable. Hence, the authors are fundamentally unable to incorporate potentially important predictors or interactions among predictors. Studies with greater sample sizes have been able to incorporate more predictors, leading to broader insights into landscape-level controls on stream temperature (e.g., Isaak and Hubert, 2001; Scott et al., 2002; Wehrly et al., 2009).

- The sample sizes (which are clearly stated in line 15, page 3 of the original manuscript) are 59 Dee, 34 Tweed, 25 Spey, 19 Bladnoch. We accept that the sample sizes are modest (especially in the case of the smaller Bladnoch catchment), but we consider them adequate considering the strategic nature of the network which covers the environmental range of the covariates and avoids site redundancy. Further, the sample sizes were sufficient to fit 'full' models for each catchment which included all the main explanatory variables. We agree that there is a danger of over-fitting when there are many covariates and small sample sizes; hence our use of AICc / BIC to penalise more complex models. We were not seeking to develop complex models, but rather simpler models that would be more likely to be transferable (e.g. Millidine et al., 2016). Although it was not the primary focus of the study, we note that our multi-catchment models, which were fitted to 137 sites, contained a limited number of covariates. We therefore consider this to be undue criticism.

b. The study only covers one season, in which temperatures were low and substantial rain fell. It is therefore unclear whether the results are specific to this one period. Perhaps in a warmer, drier summer, there would be greater spatial variability and perhaps different predictor variables would dominate.

- It is true that the dataset is constrained to one year. We believe that greater spatial variability could occur in a hotter, drier year, but we also believe that the ranking of inter-site differences in temperature would remain similar (we have seen this in our longer term data). We therefore believe that in other years the LS models would be capable of predicting hotter and cooler areas of the catchment, but not of adequately characterising between site variation or absolute temperatures. For those catchments with a feasible LS_Ta model, it is possible that better predictions of absolute temperatures could be obtained, but inter-site variability could well be greater. This would need further investigation. We expect that the same covariates would remain important even in other years, but of course cannot be certain; this would again require investigation at a future date. We have added discussion of the topic to the paper.

The authors mention the effect of continentality on stream temperature, but the causal mechanism is unclear. I could imagine that the effect arises through the effect of continentality on air temperature, yet this seems inconsistent with the findings related to air temperature. Alternatively, could it reflect variations in precipitation and thus streamflow?

- Yes, the effects of continentality could relate to both Ta and rainfall (with greater rainfall over the mountainous areas). Continentality might have a greater effect in the larger east coast catchments and indeed there was a positive relationship between Tw_{max} and Ta_{max} in the Tweed and Dee catchments.

In the conclusion, the authors suggest that further work should investigate the modelling of among-catchment variability. It might be useful for the authors to take a first run at this by examining whether the among-catchment variability is correlated with some catchment-scale measure of air temperature (or some other relevant variable).

- This is an interesting proposition for future work when we are able to bring in more catchments (and indeed this work is underway) but seems of limited value here where we have considered just four catchments.

In Tables 2 and 4, the authors should include the standard error of estimate or the root-mean-square error from validation. It would be interesting to see a comparison of the precision of their models with that found in previous studies using similar temperature metrics.

- The RMSE, standard deviation and bias of the models in Table 2 are presented in Table 3. E.g. for the Dee you would look at the Dee donor catchment, applied to the Dee.
- We do not include performance metrics for the models in Table 4 because they are not transferable (they include catchment effects and interactions that cannot be transferred more widely). We only report these models to illustrate that the responses differ between catchments.
- We note that there are relatively few studies available that model T_{max} (many models focus on mean temperatures that appear to be more spatially predictable), not all of these provide comparable performance metrics, none of these assess transferability (extrapolation) or include a RNS. However, we have added some text to the discussion to try and address this issue, focussing on within catchment performance of the single catchment models which is most closely analogous to previous studies.

The authors should provide some more information about the RNS model, which is not as commonly applied as network models based on spatial covariance functions. Are there limitations related to sample size? For example, for network models based on spatial covariance functions, a general guideline is that one needs at least 50 samples.

- A detailed description of the RNS is provided in O'Donnell et al., (2014), with application to temperature models described by Jackson et al., (2017). We do not want to repeat these papers here. The limitations on sample size come from the combination of the RNS and the covariates and in the case of this study required a minimum of 19 loggers (text has been added to this effect). Previous exploratory work suggested that the RNS required up to 6 d.f. for models of Tw_{max} in the river Spey (Jackson et al., 2017). In this paper we therefore allowed up to 7 degrees of

freedom; the Bladnoch did not contain a RNS, and the Tweed, Spey and Dee used 3.91, 4.2 and 6.8 d.f. respectively in the LS models.

Referee #3

General comments: This paper presents empirical basin-specific water temperature models and tests for their transferability to other basins. The authors provide a discussion about the water temperature response to individual landscape variables and to air temperature, and about the importance of the other basin specifics that cannot be explained by the extracted landscape variables. These are questions of present interest to the community. The chosen case study/ data set are interesting, as they provide a large enough water temperature range in each basin, and one basin that shows significantly different from the others.

- We are glad that the reviewer considers the study to be of present interest to the community.

However, the number of sites in each basin is limited vs number of covariates considered and effectively used in the models.

- This issue was also raised by reviewer 2 and we consider our previous response to be appropriate here as well: We accept that the sample sizes are modest (especially in the case of the smaller Bladnoch catchment), but we consider them adequate considering the strategic nature of the network which covers the environmental range of the covariates. Further, the sample sizes were sufficient to fit 'full' models for each catchment which included all the main explanatory variables. We agree that there is a danger of over-fitting when there are many covariates and small sample sizes; hence our use of AICc / BIC to penalise more complex models. We were not seeking to develop complex models, but rather simpler models that would be more likely to be transferable (e.g. Millidine et al., 2016). Although it was not the primary focus of the study, we note that our multi-catchment models, which were fitted to 137 sites, also contained a limited number of covariates.

The text is concise and precise, with appropriate level of detail. Tables and Figures are useful.

- We are pleased the reviewer agrees that the text is concise, precise and appropriate

Specific comments : One may guess that the Bladnoch basin shows significant physiographical differences, in addition to geographical location, as compared to the other 3 basins, but it would be interesting to see it on comparative graphs – maybe just box plots comparing the landscape variates for each basins, as in Figure 2 for the water temp? Intercorrelation issues are mentioned, but should be further discussed. In particular: could it explain the unphysical relationship between water and air temperatures in the Bladnoch basin?

- These issues were also raised by other reviewers. As such we have added a scatter plot matrix to the methods section to illustrate the distribution of the catchment covariates and their correlations. As stated in the methods, Strongly correlated covariate combinations were excluded from the analysis.
- We have added additional discussion of the Bladnoch Tw~Ta relationship in response to this and other review comments.

P. 2, l. 24: I agree with the choice of focusing on only one temperature metric in this paper, but the statement "The principles explored in this paper are likely to be similar across water temperature metrics" seems a bit forward. Variability metrics and others such as number of degree-days could show less transferable than average and extreme values.

- We are simply suggesting that the general principles and issues are likely to be relevant to other temperature metrics. We have adjusted the wording accordingly.

P. 4, l. 13: "A 'full' model was first fitted which included all the available covariates": all available covariates, or only those that were not rejected due to strong (>0.8) intercorrelation?

- Only those not rejected due to high correlation with other variables. Wording updated.

Technical corrections : Just a suggestion: rename the predicted variable throughout the paper, to reflect the way it is computed and the fact that it is not the maximum August temperature (e.g. Twmax –> Tw7d max).

- Because we define the term early on we consider that the current abbreviation is adequate.

Anonymous Referee #4

The paper by Jackson and colleagues assesses the performance of statistical temperature models fit to data from one basin when applied to another basin. Successfully doing so should ultimately depend on the degree of similarity between two basins and the adequacy of the temperature sample and model calibration within the first basin. In basins where the covariance structure among predictor variables used in the model is similar, one would expect a high degree of transferability because temperature measurements are being taken and applied along similar predictor variable ranges and combinations of ranges. In basins where the covariance structure is appreciably different, one would expect low transferability. It would be interesting, therefore, for the authors to provide tables for the four catchments summarizing the pairwise correlation coefficients among all the predictor variables at the observation sites (below the diagonal) and those same variables representing reaches in the full network (above the diagonal). The authors could compare those correlation matrices among the four basins to determine which were most (dis)similar and the representativeness of their temperature samples, which could provide insights to their results in Table 3.

- We are not sure that we completely follow the reviewer's logic here. The covariates were chosen to represent physical processes. We cannot see why their inter-correlations would affect transferability unless the reviewer is suggesting that catchments with similar covariate covariance structures will be more physically similar and thus be more liable to have similar river temperatures? Nevertheless, we have added a scatter plot matrix to the methods section along with information on the correlation between covariates.

Bladnoch, for example, is the main outlier and one might expect its covariance structure to differ most from the other three basins.

- As discussed above, the main issue with the Bladnoch is with the Tw~Ta relationship. We consider that this is likely due to the limited Ta range and / or systematic effects of other covariates on the Tw~Ta relationship. We have added discussion of these issue to the manuscript following comments by other reviewers (see response above)

Also useful would be tables that summarized descriptive statistics for the temperature observations and network reach characteristics in the four basins. Oftentimes, simply knowing the mean, median, standard deviations, minima, and maxima is very informative for understanding datasets and the performance of statistical models.

- We have added a scatter plot matrix with distributional summaries to the methods section.

Thinking about this question more broadly, one might also ask whether there's ever utility in transferring a model developed and calibrated for one area to another area? The author's models developed using data pooled across basins and fit with a RNS seem to provide the best of most worlds – the largest sample size to provide robust parameter estimates, highest predictive performance, ability to account for local variation with the RNS, and ability to incorporate (or test for) basin level effects. Where those effects aren't significant the data can be pooled and separate categories retained in the model structure only for those basins that are appreciably different. That said, the transferability question presents a challenge if the issue involves application to a basin that is entirely lacking in temperature data.

- We have further clarified this issue in our introduction. Scotland has 16006 catchments running to the sea. Clearly monitoring all these catchments would be impossible. At present, in the UK and many other countries there are simply inadequate resource to monitor everywhere. It is therefore essential to be able to predict to areas with no data and to know the possible consequences of doing so. This is the focus of this paper and it stems for a genuine research and management need that has not been addressed by the current literature.

For that topic, the authors cite work by Millar as an example of providing informed priors that would be worth developing and exploring in more detail. There's also a large literature on the ungauged basin topic that could be drawn on here (for a

recent review, see Hrachowitz et al. 2013. Hydrological Sciences Journal 58:1198-1255). However, it's also the case that reliable miniature temperature sensors are inexpensive and easily deployed, thereby making new monitoring efforts straightforward and decreasing the probability that ungauged basins will remain so indefinitely.

- We don't agree with this comment. See issues above in relation to the number of catchments in Scotland. Furthermore, although the loggers are getting cheaper it is challenging and expensive to deploy and download the loggers (staff, equipment, travel, accommodation), calibrate the loggers and maintain appropriate databases. We mention these issues in our introduction. Because of these limitations many areas in Europe are still without an adequate quality controlled temperature monitoring network and extrapolation or informed interpolation to catchments without data will be required for management purposes.

Remote sensing imagery from the MODIS satellite has also been used recently to accurately predict stream temperatures (McNyset et al. 2015. Water 7:6827–6846), which in some regards renders the ungauged basin topic moot given the global availability of this imagery source. And as the authors demonstrate here and in their previous work (Jackson et al. 2016), it's possible to derive many useful model predictors from geospatial datasets that are available for most basins.

- Just because it is possible to obtain predictors of Tw across many catchments does not mean that you can accurately predict Tw, as shown by this paper. Although Ta data are available for any catchment, the relationships between Tw and Ta varied substantially. It is likely that similar issues exist with remotely sensed data although we have not seen this explored in detail.

So I wonder if a more useful tact for thinking about 'transferability' wouldn't be attempting to establish a scalable modelling framework that can be consistently applied (i.e., transferred) to other basins as new data become available? The focus then largely becomes one of efficient sampling design, which the authors have experience with (Jackson et al. 2016) and that others have developed network design theory for (Som et al. 2014. Environmetrics 25:306- 323). In this context, the RNS models, like the similar SSN models (Ver Hoef et al. 2006), provide another advantage in their ability to borrow strength from other areas when fit to across-basin datasets, thereby minimizing the number of new temperature samples required in previously ungauged basins.

- We do not agree. Of course it is always sensible to incorporate new data as it becomes available and better predictions will be available where there are data (as shown in this paper). But it is simply not going to be possible to monitor everywhere, in all years and the covariance components of temperature models cannot be transferred between catchments. It is therefore important to understand all the issues that will be encountered when making

predictions to catchments without data. This is exactly what the current study does and we think that the results are extremely valuable for scientists that are facing similar challenges.

Incorporating a temporal dimension to the purely spatial models considered here by the authors would also be a way to reduce costs associated with new temperature monitoring in a scalable framework.

- We agree that spatio-temporal models would be a potentially useful approach for future research and have added discussion of this topic to the paper.

Rotating panel designs could be used wherein most sites are monitored for short periods (e.g., 1-2 years) to capture spatial variation in a dataset while a small number of sites are monitored indefinitely to capture temporal variation. The sites representing spatial variation would not all have to be monitored in the same years because the temporal sites could be used for standardization. Isaak et al. (2010) provide an example of models fit to similar space-time temperature datasets that were gradually accumulated over the span of fourteen years.

- If the sample design used in this paper was only intended to provide a representative sample of the rivers (some sort of spatially balanced design like GRTS or similar) then we could see that this might be possible. However, the sample design that we employed carefully and systematically covered the environmental range of the covariates present in the catchments. As such it is not sensible to randomly drop sites as this affects the data range that is covered. Furthermore we consider it is important to gather information on inter-annual variability in temperature responses across the complete range of covariates. We therefore consider it important to retain the structure of the current network. We agree that, in the future, if it were possible to develop a spatio-temporal model, it could be useful to supplement the current design with some additional sites to increase spatial coverage on a rolling programme. However, there would also be a number of technical and methodological challenges to be overcome. For example, it is unclear whether the RNS varies between years depending on prevailing hydroclimatic conditions. Short term deployments would not be able to address such issues and further methodological developments may be required to allow the RNS to interact with other drivers e.g. Ta or discharge. We have added some discussion of this topic to the manuscript.

References:

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, doi:10.1002/hyp, 2008.

Isaak, D. and Hubert, W.: A hypothesis about factors that affect maximum summer stream temperature across montane landscapes, J. Am. Water Resour. Assoc., 37(2), 351–366, 2001.

Jackson, F. L., Malcolm, I. A. and Hannah, D. M.: A novel approach for designing large-scale river temperature monitoring networks, Hydrol. Res., 47(3), 569–590, doi:10.2166/nh.2015.106, 2016.

Jackson, F. L., Hannah, D. M., Fryer, R. J., Millar, C. P. and Malcolm, I. A.: Development of spatial regression models for predicting summer river temperatures from landscape characteristics: implications for land and fisheries management, Hydrol. Process., 31, 1225-1238, doi:10.1002/hyp.11087, 2017.

Johnson, S. L.: Stream temperature: scaling of observations and issues for modelling, Hydrol. Process., 17(2), 497–499, doi:10.1002/hyp.5091, 2003.

Li, H., Deng, X., Kim, D. and Smith, E. P.: Modelling Maximum Daily Temperature Using a Varying Coefficient Regression Model, Water Resour. Res., 2014.

Luce, C., Staab, B., Kramer, M., Wenger, S., Isaak, D. and McConnell, C.: Sensitivity of Summer Stream Temperatures to Climate Variability in the Pacific Northwest, Water Resour. Res., 50(4), 3428–3443, doi:10.1002/2013/WR014329, 2014.

Millidine, K. J., Malcolm, I. A. and Fryer, R. J.: Assessing the transferability of hydraulic habitat models for juvenile Atlantic salmon, Ecol. Indic., 69, 434–445, doi:10.1016/j.ecolind.2016.05.012, 2016.

Moore, R. D., Nelitz, M. and Parkinson, E.: Empirical modelling of maximum weekly average stream temperature in British Columbia, Canada, to support assessment of fish habitat suitability, Can. Water Resour. J., 38(2), 135–147, doi:10.1080/07011784.2013.794992, 2013.

O'Donnell, D., Rushworth, A., Bowman, A. W., Scott, M. E. and Hallard, M.: Flexible regression models over river networks, J. R. Stat. Soc. Ser. C (Applied Stat., 63(1), 47–63, doi:10.1111/rssc.12024, 2014.

Scott, M. C., Helfman, G. S., McTammany, M. E., Benfield, E. F. and Bolstad, P. V: Multiscale influences on physical and chemical stream conditions across blue ridge landscapes, J. Am. Water Resour. Assoc., 38(5), 1379–1392, doi:10.1111/j.1752-1688.2002.tb04353.x, 2002.

Tague, C., Farrell, M., Grant, G., Lewis, S. and Rey, S.: Hydrogeologic controls on summer stream temperatures in the McKenzie River basin, Oregon, Hydrol. Process., 21, 3288–3300, doi:10.1002/hyp, 2007.

Wehrly, K. E., Brenden, T. O. and Wang, L.: A Comparison of Statistical Approaches for Predicting Stream Temperatures Across Heterogeneous Landscapes, JAWRA J. Am. Water Resour. Assoc., 45(4), 986–997, doi:10.1111/j.1752-1688.2009.00341.x, 2009.

Can spatial statistical river temperature models be transferred between catchments?

Faye.L.Jackson[12], Robert.J.Fryer[3], David.M.Hannah[1], Iain.A.Malcolm[2]

[1] School of Geography, Earth and Environmental Science, University of Birmingham, Birmingham, B15 2TT, UK
[2] Marine Scotland Science, Scottish Government, Freshwater Laboratory, Faskally, Pitlochry, PH16 5LB, UK
[3] Marine Scotland, Marine Laboratory, 375 Victoria Road, Aberdeen, AB11 9DB, UK

Correspondence to: F. L. Jackson, F.Jackson@marlab.ac.uk

Abstract. There has been increasing use of spatial statistical models to understand and predict river temperature (Tw) from landscape covariates ~~has increased rapidly~~. However, it is not financially or logistically feasible to monitor all rivers and the transferability of such models has not been explored. This paper uses Tw data from four river catchments collected in August 2015 to assess how well spatial regression models predict the maximum 7 day rolling mean of daily maximum Tw (Tw_{max}) within and between catchments. Models were fitted for each catchment separately using (1) landscape covariates only (LS models) and (2) landscape covariates and an air temperature (Ta) metric (LS_Ta models). All the LS models included upstream catchment area and three included a river network smoother (RNS) that accounted for unexplained spatial structure. The LS models transferred reasonably to other catchments, at least when predicting relative levels of Tw_{max}. However, the predictions were biased when mean Tw_{max} differed between catchments. The RNS was needed to characterise and predict finer scale spatially correlated variation. Because the RNS was unique to each catchment and thus non-transferable, predictions were better within catchments than between catchments. A single model fitted to all catchments found no interactions between the landscape covariates and catchment, suggesting that the landscape relationships were transferable. The LS_Ta models transferred less well, with particularly poor performance when the relationship with the Ta metric was physically implausible or required extrapolation outside the range of the data. A single model fitted to all catchments found catchment-specific relationships between Tw_{max} and the Ta metric, indicating that the Ta metric was not transferable. These findings improve our understanding of the transferability of spatial statistical river temperature models and provide a foundation for developing new approaches for predicting Tw at unmonitored locations across multiple catchments and larger spatial scales.

Key words: transferability, water temperature, landscape controls, spatial statistical models, spatial variability, air temperature, Scotland

1 Introduction

River temperature (Tw) is a key control on the health of aquatic systems (Webb et al., 2008) and is particularly important for the growth, survival and demographic characteristics of cold water adapted species such as salmonids (Elliott and Elliott, 2010; Gurney et al., 2008; Jonsson and Jonsson, 2009; McCullough et al., 2001). Rising Tw will influence fish populations by altering the thermal suitability of rivers (Comte et al., 2013; Isaak et al., 2010, 2012). Thus models that can; 1) identify areas most affected by thermal extremes, 2) improve understanding of spatio-temporal variability of thermal regimes, 3) predict the potential effects of climate change and 4) illustrate opportunities for thermal moderation, such as riparian tree planting (Hannah et al., 2008; Hrachowitz et al., 2010), are important for fisheries management. Large-scale models are required to provide information at the spatial scales appropriate to management decisions i.e. catchment (Chang and Psaris, 2013; Hrachowitz et al., 2010; Imholt et al., 2011, 2013; Jackson et al., ~~2016a~~2017; Steel et al., 2016), regional (Hill et al., 2013; Isaak et al., 2012; Ruesch et al., 2012) and national scales.

Although process based models provide important mechanistic understanding at small spatial scales, their intensive data requirements prohibit their use at larger scales (Jackson et al., 2016~~b~~). In contrast, empirical models of Tw rely on Tw observations and explanatory covariates (e.g. altitude or air temperature) which can often be derived remotely at relatively low cost. The development of affordable, reliable, accurate Tw dataloggers has led to a rapid increase in Tw monitoring (Sowder and Steel, 2012), to the point that staff time, data storage and quality control are often now the greatest limitations on data collection (Jackson et al., 2016~~b~~). At the same time, there have been substantial developments in spatial statistical modelling approaches (Ver Hoef et al., 2006, 2014; Ver Hoef and Peterson, 2010; Isaak et al., 2014; Jackson et al., ~~2016a~~2017; O'Donnell et al., 2014; Peterson et al., 2013; Rushworth et al., 2015), monitoring network design (Dobbie et al., 2008; Jackson et al., 2016~~b~~; Som et al., 2014), spatial datasets (e.g. shapefiles incorporating covariates such as in "The National Stream Internet Project" (Isaak et al., 2011) or gridded air temperature datasets (Perry and Hollis, 2005a, b)) and spatial analysis tools (Isaak et al., 2011, 2014; Peterson et al., 2013; Peterson and Ver Hoef, 2014).

While continuous river temperature data are routinely collected in some areas, resulting in large regional temperature datasets and associated models (e.g. Wehrly et al. 2009; Moore et al. 2013), this is far from universal. ~~Despite these developments,~~ In many cases financial and logistical considerations limit data collection making it~~it is still~~ impractical to monitor all rivers. For example, in Scotland there are 16006 river catchments (unique rivers running to the sea), including 629 catchments >10km^2, but there was no systematic nationwide quality controlled river temperature data collection until 2015 (Jackson et al. 2016). ~~and, f~~For management purposes, it is therefore often necessary to predict Tw at unmonitored locations, both within (Hrachowitz et al., 2010; Jackson et al., ~~2016b~~2017; Peterson and Urquhart, 2006) and between catchments (Isaak et al., 2014). In recent years, it has become increasingly common to develop and apply spatial statistical river network models that incorporate network covariance structure to predict spatial variability in river temperature (e.g. Isaak et al., 2014; Jackson et al., 2017). It is widely acknowledged that these models can dramatically improve predictions of river temperature where sufficient observational data exist, but the covariance component of the predictions cannot typically

be transferred between catchments. Despite these ~~importance of this~~ important issues, there has not yet been an assessment of the transferability of spatial statistical Tw models between catchments; i.e. the ability of a model developed in one catchment to predict Tw in another. This paper investigates the ability of spatial statistical Tw models to predict Tw at unmonitored locations within and between catchments.

The principles explored in this paper are likely to be relevant to other ~~be similar across~~ water temperature metrics so, for brevity, this study focuses on maximum summer temperature, a metric which is prevalent in the recent literature, reflecting its importance for the survival of cold water adapted fish (Chang and Psaris, 2013; Hrachowitz et al., 2010; Jackson et al., ~~2016a~~2017; Malcolm et al., 2008; Marine and Cech, 2004).

Models are fitted using two sets of covariates. The first set contains landscape covariates which can be generated from readily available spatial datasets and have been the focus of many previous studies of spatial variability in river temperature (e.g. Hrachowitz et al., 2010). Due to increasing interest in the use of air temperature (Ta) to predict spatial variability in water temperature (e.g. Jonkers & Sharkey, 2016), the second set contains a metric of air temperature in addition to landscape covariates.

The paper addresses the following objectives:

1. Develop statistical models for predicting maximum summer water temperature from landscape covariates in four separate river catchments.
2. Determine whether models containing an air temperature metric explain more of the variation in maximum summer water temperature than those only containing landscape covariates.
3. Assess the transferability of models containing only landscape covariates or both landscape and air temperature covariates between catchments
4. Produce single models of maximum summer water temperature for all four catchments using both sets of covariates and consider their potential for transferability at larger (e.g. national) scales.

2 Methodology

2.1 Water temperature data and metric

Tw data were obtained from monitoring sites in four catchments; the Bladnoch in Western Scotland and the Dee (Aberdeenshire), Spey and Tweed in Eastern Scotland (Fig.1). These catchments are Special Areas of Conservation for Atlantic salmon and form part of the Scotland River Temperature Monitoring Network (SRTMN) (Jackson et al., 2016~~b~~). Details of the network, including design and quality control procedures, are given in Jackson et al. (2016~~b~~). The catchments all contain an adequate numbers of Tw dataloggers to develop Tw models on a single catchment basis with 59, 34, 25 and 19 sites in the Dee, Tweed, Spey and Bladnoch, respectively. The choice of catchments ensured a broad geographic coverage across Scotland with a wide environmental range of landscape covariates (Jackson et al., 2016~~b~~).

Data were collected at 15 minute intervals throughout August 2015. The maximum temperature was calculated for each day and used to produce a 7 day rolling mean of maximum temperatures. The metric of maximum temperatures used in this study (Tw_{max}) was the maximum value of this 7 day rolling mean. This metric was preferred to a single observation of Tw as it characterises the occurrence of sustained high temperatures which are thought to be most ecologically damaging.

2.2 Model covariates ~~and river network basis functions~~

Detailed discussion of the landscape covariates and their calculation can be found in (Jackson et al., 2016~~a~~). In brief, the covariates were: elevation (Elevation), upstream catchment area (UCA), percentage riparian woodland (%RW), hillshading / channel illumination (HS), channel width (Width), channel gradient (Gradient), channel orientation (Orientation), distance to coast (DC) and distance to the sea along the river (RDS). Table 1 summarises how the covariates were calculated. Before model fitting, Gradient, UCA and Width were log transformed to reduce skewness and HS was centred by subtracting the median value from all observations.

An air temperature metric (Ta_{max}) was calculated for each site from the gridded UKCP09 Ta dataset (available from the UK MET Office). See Perry and Hollis (2005a, 2005b) for details of this dataset. Analogous to the calculation of Tw_{max}, Ta_{max} was given by the maximum of the 7 day rolling mean of daily maximum air temperatures in August 2015.

Figure 2 illustrates the distribution and correlation among covariates included in the single or multi-catchment models (excluding strongly correlated (> 0.8) covariate pairs, see below for details) for each of the four catchments and for the global (four catchment) dataset.

2.3 Modelling

Ten models of Tw_{max} were developed: two models for each of the four river catchments using either 1) landscape covariates only (LS models) or 2) landscape covariates and Ta_{max} (LS_Ta models) and two models for all four catchments combined, again using either 1) landscape covariates only (multi-catchment LS model) or 2) landscape covariates and Ta_{max} (multi-catchment LS_Ta model). The modelling process differs slightly between the single and multi-catchment models and these are described in turn. All analysis was done in R version 3.2.3 (R Core Team, 2015).

2.3.1 Single catchment models

The set of covariates was first reduced to avoid problems of collinearity. If two covariates were strongly correlated (Pearson correlation coefficient >0.8) in any one catchment, one of the covariates was dropped from the set available for modelling for all catchments. This ensured all the LS models were based on a common set of covariates (UCA, %RW, HS, Orientation, DC) as were the LS_Ta models (Ta_{max}, UCA, %RW, HS, Orientation, DC).

The relationship between Tw_{max} and the covariates was explored using generalised additive models (GAMs) with Gaussian errors and an identity link (Wood, 2001). A 'full' model was first fitted which included all the available covariates from the reduced dataset and a river network smoother (RNS) (see below):

$$Tw_{max} \sim s(covariate_1) + \ldots + s(covariate_n) + RNS$$

Here, n is the number of covariates (n = 5, 6 for LS, LS_Ta models respectively) and $s(covariate_i)$ denotes that covariate i was fitted as a smoother. The amount of smoothing was estimated from the data (Wood, 2001), with each smoother constrained to have at most 2 degrees of freedom (df) based on the expected simplicity of Tw_{max} responses to the covariates. The RNS is included to account for spatial structure in the data that cannot be explained by the covariates. The RNS is a modified version of that developed by O'Donnell et al. (2014), with the amount of smoothness at a confluence controlled by the proportional influence of upstream tributaries weighted by Strahler river order (Strahler, 1957) and fitted using a set of 'reduced rank' basis functions. See Jackson et al. (~~2106a~~2017) for full details. The RNS was allowed up to 7 df based on knowledge of RNS complexity for the Spey (Jackson et al., ~~2016a~~2017). To ensure the RNS did not account for variability that would otherwise be explained by covariates, RNS basis functions were excluded if they were strongly correlated (>0.8) with any of the covariates. Thus, base 1 was removed from the Spey and Dee RNSs due to correlations with DC. In the LS_Ta models, base 2 was also removed from the Spey RNS due to correlation with Ta_{max}. The model was fitted by maximum likelihood using the "mgcv" package (Wood, 2016) in R.

All possible subsets of the full model were then fitted. The final model was that with the lowest Bayesian Information Criterion (BIC) or Akaike Information Criterion corrected for small sample size (AICc) that contained no terms significant at the 5% level. The choice of Information Criterion was based on the desire to penalise more complex models that would be unlikely to transfer well (Millidine et al., 2016). Thus, BIC was used for the Dee and Tweed where there were more sites and AICc was used for the Bladnoch and Spey where there were fewer sites. Terms in the final model with 1 df were replaced by linear terms.

In common with similar modelling studies (Hrachowitz et al., 2010; Imholt et al., 2011; Jackson et al., ~~2016a~~2017; Ruesch et al., 2012), no interactions were considered between covariates due to data constraints.

2.3.2 Multi-catchment models

Covariates were excluded if they were strongly correlated (>0.8) across the entire multi-catchment dataset. The reduced set of covariates was Elevation, UCA, %RW, HS, Gradient and Orientation for the LS model, and Ta_{max}, UCA, %RW, HS, Gradient and Orientation for the LS_Ta model. The RNS basis functions were the same as those included in the single catchment models.

A 'starting' model was fitted of the form:

$$Tw_{max} \sim Catchment + s(covariate_1) + \ldots + s(covariate_n) + RNS{:}Catchment$$

where Catchment is a categorical variable allowing a different mean level for each catchment and RNS:Catchment denotes a separate RNS for each catchment. The covariate smoothers were given a maximum of 2 df and the RNS a maximum of 7 df for each catchment. The model was then refined in a backwards and forwards stepwise procedure which considered a) replacing smooth covariate effects by linear terms and then dropping them altogether; b) dropping the RNS by Catchment term altogether; c) adding interactions between the covariates (either linear or smoothed) and Catchment. An interaction between a covariate and Catchment would indicate inter-catchment differences in the relationship between Tw_{max} and the covariate, suggesting that the model might not transfer well to new catchments. Interactions between the covariates were not considered. Model selection was based on BIC. Finally, any non-significant terms ($p > 0.05$) in the final model were removed.

2.3.4 Model performance and transferability of single-catchment models

The ability of single-catchment models to predict Tw_{max} within the catchment they were developed (the donor catchment) was assessed using Leave-One-Out-Cross-Validation. Each site was removed in turn, the final model was refitted, and then Tw_{max} was predicted at the missing site using a) using all model terms (i.e. the covariates and the RNS if present) and b) only covariates (i.e. excluding the columns in the model matrix relating to the RNS). The prediction using all model terms should outperform that using only covariates because it incorporates the extra information about spatial structure that is captured by the RNS. However, a RNS from one catchment cannot be used to predict in another because the river networks will differ. The prediction using only covariates therefore provides a benchmark for assessing the transferability of models between catchments, since it measures how well a model will transfer to a catchment that is identical in all but its river network.

Transferability to another catchment (the target catchment) was assessed by using the model from the donor catchment to predict Tw_{max} at the monitoring sites in the target catchment. As RNSs cannot be transferred, only covariates were used in the predictions (i.e. the columns in the model matrix due to the RNS were ignored).

Three performance metrics were calculated: Root Mean Square Error (RMSE) (Eq.1), which measures overall performance (accuracy), Standard Deviation (SD) (Eq.2), which measures how well a model can predict within-catchment spatial variability (precision), and Bias (Eq.3).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{s=1}^{n}(\hat{x}_s - x_s)^2} \quad (1)$$

$$\mathrm{SD} = \sqrt{\frac{1}{n}\sum_{s=1}^{n}((\hat{x}_s - \bar{\hat{x}}) - (x_s - \bar{x}))^2} \quad (2)$$

$$\mathrm{Bias} = \bar{\hat{x}} - \bar{x} \quad (3)$$

where x_s and $\hat{x}_s$ are the observed and predicted Tw_{max} at site s, $\bar{x}$ and $\bar{\hat{x}}$ are the mean observed and predicted Tw_{max} in the catchment, and $_n$ is the number of sites in the catchment. Standard deviation was used rather than variance, so that all three metrics are on the same scale and can be compared. Model performance was also illustrated by plotting observed Tw_{max} against predicted values and comparing this to a 1:1 line. Points close to the 1:1 line indicate precise unbiased predictions, points consistently displaced above or below the line indicate biased predictions, and high scatter about the line indicates imprecise predictions. The consequences of predicting outside of the environmental range of a given model was shown by coding sites as "in" or "out" of range.

3 Results

Across Scotland, August 2015 was wetter than the 1981-2010 mean (MET Office, 2016) and this was reflected in relatively low Tw. Rainfall in Eastern Scotland (which covers the Spey, Dee and Tweed) was 107% of the 1981-2010 mean, whereas rainfall in Western Scotland (which covers the Bladnoch) was only 98% of the 1981-2010 mean (MET Office, 2016). Maximum air temperature was the same in Eastern Scotland as the 1981-2010 mean maximum and 0.2°C cooler in Western Scotland over the same period (MET Office, 2016).

Figure 1 shows the spatial variability in Tw_{max} across the four catchments and Figure 2 summarises the distribution of Tw_{max} by catchment (bottom left diagonal panel). Median Tw_{max} in the Dee (15.1°C), Tweed (15.6°C) and Spey (15.6°C) were broadly similar, but median Tw_{max} in the Bladnoch (16.4°C) was ca. 1°C higher (Fig. 2). The range of Tw_{max} was 5.7, 5.9, 6.0 and 5.5°C in the Bladnoch, Dee, Spey and Tweed, respectively (Fig. 2).

3.1 Single catchment models

All four LS models were simple (Table 2), explained much of the variance in Tw_{max} (76.6-85.6%) and contained similar positive relationships between Tw_{max} and UCA (Fig. 3). This relationship was near linear until ca. 100km^2 and then levelled off in the Bladnoch (Fig. 3d), smooth, but near-linear in the Spey and the Tweed (Fig. 3b, c) and linear in the Dee (Fig. 3a). The magnitude of the effect was similar across catchments at ca. 4°C. Three models contained a RNS, which explained much of the variance; 61.7, 13.9 and 63.7% in the Dee, Tweed and Spey respectively (Table 2). The Tweed model also had a negative linear effect of %RW.

The LS_Ta models always had a better BIC / AICc than the corresponding LS models, but were typically more complex, always required more df, and only explained a greater % variance in the Bladnoch and the Tweed (Table 2). For the Tweed, the LS_Ta model used only covariates, whereas the LS model required a RNS to account for unexplained spatial structure. For the Bladnoch, the LS_Ta model included UCA and Ta_{max}, whereas the LS model only included UCA.

In common with the LS models, UCA was in all the LS_Ta models (Table 2) and the direction, shape and magnitude of the effects were consistent with the LS models (Fig. 4, top row). Ta_{max} was in all the LS_Ta models except the Spey (Table 2).

There was a positive linear relationship between Tw_{max} and Ta_{max} in the Dee and Tweed (Figure 4e, f) and a U shaped response in the Bladnoch which is physically implausible, increasingly so when extended beyond the range of Ta_{max} observed in the Bladnoch (Fig. 4g). Orientation had a small positive effect on Tw_{max} in both the Dee and Tweed (Fig. 4h, i) with higher temperatures for a N-S orientation than an E-W orientation. There was also a negative linear effect of %RW and a positive smoothed effect of HS in the Tweed, and a positive linear effect of DC in the Spey (Fig. 4j, k, l, respectively).

3.2 Transferability of single catchment models

The transferability of the LS and LS_Ta models is summarised by their RMSE, bias and standard deviation in Table 3 and illustrated in Figs. 5 and 6 respectively. All the models performed well within catchments (i.e. in the catchments where they were developed) when all model terms (i.e. both covariates and the RNS) were used in the predictions, with a bias of < 0.1°C in absolute value and a RMSE of < 1°C. The LS_Ta models always had a lower RMSE than the LS_models. As expected, within-catchment predictions were poorer when only the covariates were used (excluding RNS), with a median RMSE of 1.2°C and a maximum RMSE of 1.8°C.

The rest of this section focusses on the predictions, both within and between catchments, using only the covariates. For the catchments in Eastern Scotland (Dee, Tweed and Spey), the RMSE, bias and standard deviation of any model was broadly similar whether it was used to predict for the donor catchment or to the other two target catchments. The RMSE of the LS models tended to be lower than that of the LS_Ta models (median 1.3 and 1.7°C respectively). The LS and LS_Ta models both had median absolute biases of 0.3°C and median standard deviations of 1.1 and 1.4°C respectively. RMSE is a combination of bias and standard deviation, so the RMSE of both sets of models was generally dominated by the standard deviation.

Predictions involving the Bladnoch, either as donor or target catchment, tended to be poor. The Bladnoch is in Western Scotland and was warmer than the other catchments (Fig. 2). The Bladnoch models always over-predicted Tw_{max} in the other catchments and the Dee, Tweed and Spey models all under-predicted Tw_{max} in the Bladnoch (Fig. 5, 6). This often led to substantial bias and hence RMSE. The Bladnoch LS_Ta model had the largest biases, which were also due to the implausible relationship with Ta_{max} (Figure 4g). The Dee, Tweed and Spey had reasonable standard deviations when transferred to the Bladnoch (median 1.0 and 1.1°C for the LS and LS_Ta models respectively) which suggests that, despite having poor RMSE, the models still could be used to predict areas of relatively high or low Tw_{max} within the Bladnoch (rather than absolute values of Tw_{max}). The same is true of the Bladnoch LS models when transferred to the Dee, Tweed and Spey (median standard deviation 1.3°C). However, the Bladnoch LS_Ta model had a high standard deviation (median 3.3°C) when transferred to the Dee, Tweed and Spey, again due to the implausible relationship with Ta_{max}.

3.3 Multi-catchment models

The multi-catchment LS model included Catchment, UCA, %RW, Elevation and a RNS for each catchment (Table 4). By fitting a single model to all four catchments it was possible to assess whether covariate effects were consistent across

catchments and thus transferable to new catchments or regions. None of the covariates interacted with catchment. The Catchment effect indicates inter-catchment differences in mean Tw_{max} having accounted for the landscape covariates; in particular, higher Tw_{max} in the Bladnoch (Figure 7d). In common with the single catchment LS models, there was a positive smooth relationship between Tw_{max} and UCA with an effect size of ca. 3°C (Figure 7a). There was also a negative linear relationship between Tw_{max} and both %RW and Elevation, with effect sizes of ca. 1°C and 2°C respectively. The model explained 84.4% of the variance, comparable to the single catchment LS models. The RNSs explain less of the variance than in the single catchment models (Tables 3, 4).

The multi-catchment LS_Ta model explained 83.2% of the variance and contained Catchment, UCA, %RW, Ta_{max} and a RNS for each catchment (Table 4, Figure 8). None of the landscape covariates interacted with catchment. However, the Ta_{max} relationship did interact with catchment, (Fig. 8a-d), with positive relationships in the Dee and Tweed and negative (albeit non-significant) relationships in the Spey and Bladnoch. This suggests that relationships with Ta_{max} are non-transferable and Ta_{max} would not be a good predictor of Tw_{max} in new catchments.

4.0 Discussion

Even with the introduction of relatively cheap and accurate dataloggers it is not financially or logistically possible to monitor everywhere. Consequently, there is a need to develop models to understand and predict river temperatures at large spatial scales to inform evidence based management of rivers and fisheries even where extensive local temperature data collection does not exist. Spatial statistical models offer great promise in this respect. However, to date, the transferability of these models has not been considered. This study fitted separate models of Tw_{max} to data from four catchments and transferred these models between catchments. Models containing only landscape covariates typically contained similar covariates and covariate responses, and performed better than models containing Ta_{max} when transferred between catchments. A physically implausible model transferred particularly poorly. The covariates alone often explained much less of the spatial temperature variability than when a RNS was added, but provided the only means of predicting temperature in new catchments with no or limited data (a minimum of 19 loggers was required to fit the full models including covariates and RNS). A single ~~M~~model~~s~~ fitted to all four catchments ~~combined~~ suggested common responses to landscape covariates, but inter-catchment differences in mean temperature and in the relationships between Tw_{max} and Ta_{max}. These findings are discussed in more detail below.

4.1 Tw_{max} responses to landscape covariates

The single catchment LS models contained similar covariates with comparable effect sizes and response shapes which suggested that transferability between catchments could be reasonably successful. This was confirmed by the lack of significant interactions with Catchment in the multi-catchment model. However, when there are inter-catchment differences in mean temperature, the models might only be good predictors of relative values of Tw_{max} within a new catchment (i.e.

areas of higher or lower Tw_{max}) rather than absolute values. It is also unclear how well the models would perform in years with differing hydro-climatic characteristics. This study was conducted in a single year with relatively low temperatures and high flows. In a hotter, drier year it might be expected that between site differences would be greater. Under such circumstances the current models may not provide accurate predictions of absolute temperatures or inter-site differences without refitting.

All of the Tw_{max} responses to landscape covariates (across all models) were physically plausible and hence broadly transferable (Smith et al., 2016). UCA (which was in all the models) is a proxy for discharge, water volume and thermal capacity (Chang and Psaris, 2013; Hannah et al., 2008). Higher UCAs are generally associated with larger water volumes which have a greater thermal capacity, taking longer to warm but also retaining heat for longer (Chang and Psaris, 2013; Imholt et al., 2011). Elevation reflects adiabatic lapse rates which reduces temperatures with increasing altitude (Hrachowitz et al., 2010, Jackson et al ~~2016a~~2017). The negative relationship between Tw_{max} and %RW woodland occurs because riparian shading reduces the amount of incident shortwave radiation reaching the river during daylight hours (Garner et al., 2014; Hannah et al., 2008; Moore et al., 2005). The positive relationship between Tw and HS is consistent with greater Tw in locations with lower shading effects and greater direct shortwave contributions (illumination). Tw was greatest in channels characterised by a north/south orientation which typically experience maximum exposure to incoming radiation (Malcolm et al., 2004). Increasing Tw with distance from the coast, reflected continentality and the differing specific heat capacities of land and sea, specifically thermal buffering of relatively cooler sea during summer months (Chang and Psaris, 2013; Hrachowitz et al., 2010).

4.2 Tw ~ Ta relationships

In contrast to the LS models, one LS_Ta model included a physically implausible relationship that would not be expected to transfer well (Smith et al., 2016). Specifically, an inverse modal relationship between Tw_{max} and Ta_{max} in the Bladnoch model ~~transferred particularly poorly~~. This relationship could have arisen due to the relatively small air temperature range (1.7°C) observed in the Bladnoch which provided only limited contrast between sites. However, it is also possible that this reflected systematic spatial variability in other controls (described by our covariates or otherwise) that influence local Tw_{max} ~ Ta_{max} relationships e.g. hydrogeology or landuse. Nevertheless, ~~E~~even where the relationships between Tw_{max} and Ta_{max} relationships were plausible, they were inconsistent between catchments in terms of effect size, as indicated by the varying responses in the single catchment models and the interaction with Catchment in the multi-catchment model.

Given the number of previous studies that have predicted Tw from Ta, both within sites over time (temporal models) and between sites (spatial models), ~~-~~it may appear surprising that Ta_{max} was such a poor predictor of between-catchment temperature variability in this study. However, previous spatial models of Tw incorporating air temperature as a predictor (e.g. Wehrly et al., 2009; Moore et al., 2013) have focussed on the ability of these models to predict within the data space (interpolate), while this study investigated the ability of models to predict outside of the data space (extrapolate). Indeed, within our multi-catchment model it would have been possible to force a single Tw_{max} ~ Ta_{max} relationship that

reflected an average response across catchments. However, this would result in biased estimates of Tw_{max} within individual catchments. ~~However, this study was only concerned with spatial variability in a temporally static Tw metric. For~~

The ability of Ta_{max} ~~-~~ to ~~-~~ predict spatial variability in Tw_{max} ~~have been a useful predictor in these models would have required~~ is likely to degrade where the temporal relationships between Tw and Ta vary~~a~~ spatially, ~~-consistent -relationship between -~~within and between catchments. ~~Tw_{max} -and Ta_{max} -~~ It is expected that within catchment (between site) variability in the temporal relationships between Tw and Ta would add noise to any spatial relationships making them harder to detect and reducing the overall precision of any predictions. Systematic differences in Tw~Ta relationships between catchments would result in biased predictions when models are transferred between rivers or regions.

Many studies have shown that relationships between Tw and Ta can be highly variable (Arismendi et al., 2014; Arora et al., 2016; Fellman et al., 2014; Luce et al. 2014; Mayer, 2012; Tague, 2007) across a range of spatial scales~~, with the relationships~~ depending on hydrological and landscape controls (Tague, et al., 2007; Chang and Psaris, 2013). For example, Arismendi et al. (2014) investigated Tw~Ta relationships at 25 sites across the Western US using linear regression and reported that the slope of the relationship varied between 0.32 and 1.01, while Fellman et al. (2014) observed slopes of between -0.180 and 1.282 across 9 watersheds in Alaska depending on glacial influence. Similarly, Tague et al., (2007) observed systematic regional differences in Tw~Ta relationships in Western Oregon that depended on local hydrogeology and concluded that under such circumstances air temperature alone would be unlikely to explain river temperature variability. ~~Even at smaller spatial scales there can still be considerable inter-site variability in Tw~Ta relationships. For example, Johnson et al. (2014) developed linear and logistic Tw~Ta models for 36 sites across the Manifold and Dove catchments, English Peak District (131 and 75 km2 respectively) and found that parameter estimates varied substantially between sites, so the models were effectively site specific.~~ Given the reported spatial variability in Tw~Ta relationships and importantly, that these relationships can vary systematically between catchments depending on other controls (e.g. hydrogeology), it is unsurprising that Ta_{max} does not substantially improve predictions of the spatial variability in Tw_{max} and that transferred models result in biased Tw predictions. If models including Ta are to provide substantially better predictions then it is likely that they would need to~~without~~ include~~ing~~ greater model complexity e.g. ~~by~~ allowing for interactions between Ta_{max} and landscape covariates (e.g. Mayer, 2012).

4.4 The importance of RNS

The performance of the single catchment LS and LS_Ta models in this study compared favourably to regional models of Tw_{max} (Moore et al. 2013; Roberts et al. 2013; Wehrly, et al. 2009) when predictions were made for the catchment in which the models were developed (i.e. interpolation). For the models that included a RNS, RMSE (0.7-0.9 °C) was approximately half that reported by previous studies, although it should be noted that these studies were conducted at considerably larger spatial scales (Moore et al. 2013; Roberts et al. 2013; Wehrly, et al. 2009). The RMSE of models without a RNS (0.9-1.8 °C) was generally similar or slightly better than reported by other studies.

The landscape covariates included in the models in this study explained large (catchment) scale trends in Tw_{max}, but were less successful at explaining variability at finer spatial scales. For example, the ca. 20% variance explained by UCA in the Spey and Dee models is consistent with the 18-25% of Tw variability explained by discharge in Arora et al. (2016). Smaller scale variability tends to reflect drivers such as water residence time (and heat advection), water sources (Brown et al., 2006; Brown and Hannah, 2008), channel incision, gradient (Jackson et al., ~~2016a~~2017) and land use (Imholt et al., 2013) which are harder to accurately characterise from spatial datasets. In the absence of accurate local scale characterisation of landscape controls, smaller scale spatial variability is modelled by the RNS. However, whilst the RNS improves predictions within catchments, it is not transferable so does nothing to help predictions between catchments.

4.5 Extending predictions

The inclusion of the Catchment main effect in both multi-catchment models showed differences in mean Tw_{max} between catchments (that were not accounted for by the covariates). This sometimes led to substantial bias when transferring single catchment models to new catchments. Accounting for between-catchment differences in mean Tw will be necessary to improve between-catchment predictions of Tw. The multi-catchment models in this study used a simple categorical variable to allow the intercept (and hence mean Tw_{max}) to differ between catchments. However, to predict to new catchments, it would be necessary to extend the modelling approach so that the intercept can be predicted from surrounding catchments. One approach could be to allow the intercept to vary smoothly between catchments using a Gaussian Markov Random Field (Cressie, 1993), so the intercept in unmonitored catchments could be estimated from nearby monitored catchments. This approach has been developed in other contexts (Millar et al., 2015, 2016) and offers promise in the context of large-scale Tw modelling.

An alternative approach could involve modelling Tw as a function of Ta over shorter time periods (days or weeks) and then allowing this relationship to interact with landscape covariates or location. Such an approach could have additional benefits, allowing the inclusion of temporally incomplete data (e.g. Letcher et al., 2016) or data from temporally inconsistent locations. Where sufficient resources were available it may be possible to supplement the existing network with sites that are monitored for shorter time periods to expand spatial coverage although the consequences of such deployments for assessing inter-annual temperature variability would need to be investigated. Finally, the development of spatio-temporal models, where temporal variability was driven by Ta or discharge, could potentially allow for fore- or hind-casting of river temperature which wasn't possible using the approaches presented in this paper.

5 Conclusions and future work

This study demonstrated that landscape covariates can explain broad scale patterns in Tw_{max} and that such relationships are transferable between catchments, at least to predict relative levels of Tw_{max}. It was necessary to use a RNS to characterise and predict finer scale spatially correlated variation, so predictions of spatial temperature variability were better within

catchments than between catchments. Ta_{max} was not a transferable predictor of Tw_{max} and could result in poor predictions when the relationship was implausible or transferred outside the range observed in the donor catchment. It would be unwise to use a Tw~Ta relationship to predict spatial variability in Tw without also including meaningful (process relevant) interactions between Ta and landscape covariates, something that was not possible in this study due to data constraints.

Mean Tw_{max} also varied between catchments (having adjusted for the landscape covariates). Future work that looks to predict to new catchments should investigate how to understand and predict these between catchment differences. A large scale correlated spatial smoother (e.g. regional effect) offers potential in this respect. Finally, some of the local scale processes represented in this study (e.g. effect of riparian shading) may benefit from improved characterisation using finer scale spatial datasets or remotely sensed data. Improved process representation could lead to both better within and between catchment model predictions.

Data availability

Some map features are based on digital spatial data licensed from the Centre of Ecology and Hydrology, NERC © Crown Copyright and database right (2016), all rights reserved, Ordnance Survey License number 100024655. Catchment boundaries were from SEPA (2009). The digitised river network is from the CEH and includes Scottish Environmental Protection Agency (SEPA) coding. Catchment boundaries were from SEPA (2009) and Salmon rivers from Marine Scotland (2008). The gridded Ta dataset was from UKCP09: Daily gridded air temperature dataset (2015) UK MET Office. Summary Tw data used in the study will be made available on the Marine Scotland Science webpages, upon acceptance of the article.

Author contributions

I.A.Malcolm and D.M.Hannah secured funding for the project. The authors conceived the study. F.L.Jackson, carried out the data analysis with support from I.A.Malcom and R.J.Fryer. F.L.Jackson, I.A.Malcolm, R.J.Fryer and D.M.Hannah interpreted the results and prepared the manuscript.

Competing interests

The authors declare that they have no conflict of interest

Acknowledgements

This project is part of Marine Scotland Science (MSS) Freshwater Fisheries Laboratories Service Level Agreement FW02G and it is supported by a NERC Open CASE studentship (University of Birmingham and MSS) awarded to Faye Jackson; grant reference NE/K007238/1. The authors also thank the Scotland River Temperature Monitoring Network

collaborators for their contribution to data collection (http://www.gov.scot/Topics/marine/Salmon-Trout-Coarse/Freshwater/Monitoring/temperature/Collaborating).

References

Anon: Sun or Moon Altitude/Azimuth Table, Data Serv. Washington, DC U.S. Nav. Obs. Astron. Appl. Dep. [online] Available from: http://aa.usno.navy.mil/data/docs/AltAz.php (Accessed 26 March 2014), 2001.

Arismendi, I., Safeeq, M., Dunham, J. B. and Johnson, S. L.: Can air temperature be used to project influences of climate change on stream temperature?, Environ. Res. Lett., 9, doi:10.1088/1748-9326/9/8/084015, 2014.

Arora, R., Tockner, K. and Venohr, M.: Changing river temperatures in Northern Germany: trends and drivers of change, Hydrol. Process., 30, 3084–3096, doi:10.1002/hyp.10849, 2016.

Bivand, R. and Rundel, C.: rgeos: Interface to Geometry Engine - Open Source (GEOS). R package version 0.3-17., [online] Available from: https://cran.r-project.org/package=rgeos, 2016.

Brown, L. E. and Hannah, D. M.: Spatial heterogeneity of water temperature across an alpine river basin, Hydrol. Process., 22, 954–967, doi:10.1002/hyp, 2008.

Brown, L. E., Hannah, D. M., Milner, A. M., Soulsby, C., Hodson, A. J. and Brewer, M. J.: Water source dynamics in a glacierized alpine river basin (Taillon-Gabiétous, French Pyrénées), Water Resour. Res., 42(8), doi:10.1029/2005WR004268, 2006.

Chang, H. and Psaris, M.: Local landscape predictors of maximum stream temperature and thermal sensitivity in the Columbia River Basin, USA., Sci. Total Environ., 461–462, 587–600, doi:10.1016/j.scitotenv.2013.05.033, 2013.

Comte, L., Buisson, L., Daufresne, M. and Grenouillet, G.: Climate-induced changes in the distribution of freshwater fish: Observed and predicted trends, Freshw. Biol., 58(4), 625–639, doi:10.1111/fwb.12081, 2013.

Cressie, N.: Statistics for spatial data:Wiley series in probability and statistics, Wiley-Interscience, New York., 1993.

Csardi, G. and Nepusz, T.: The igraph software package for complex network research, InterJournal, Complex Sy, 1695 [online] Available from: http://igraph.org, 2006.

Dobbie, M. J., Henderson, B. L. and Stevens, D. L.: Sparse sampling: Spatial design for monitoring stream networks, Stat. Surv., 2, 113–153, doi:10.1214/07-SS032, 2008.

Elliott, J. M. and Elliott, J. A.: Temperature requirements of Atlantic salmon Salmo salar, brown trout Salmo trutta and Arctic charr Salvelinus alpinus: Predicting the effects of climate change, J. Fish Biol., 77(8), 1793–1817, doi:10.1111/j.1095-8649.2010.02762.x, 2010.

Fellman, J. B. J., Nagorski, S., Pyare, S., Vermilyea, A. W., Scott, D. and Hood, E.: Stream temperature response to variable glacier coverage in coastal watersheds of Southeast Alaska, Hydrol. Process., 28, 2062–2073, doi:10.1002/hyp.9742, 2014.

Garner, G., Malcolm, I. A., Sadler, J. P. and Hannah, D. M.: What causes cooling water temperature gradients in a forested stream reach?, Hydrol. Earth Syst. Sci., 18, 5361–5376, doi:10.5194/hess-18-5361-2014, 2014.

Gurney, W. S. C., Bacon, P. J., Tyldesley, G. and Youngson, A. F.: Process-based modelling of decadal trends in growth, survival, and smolting of wild salmon (Salmo salar) parr in a Scottish upland stream, Can. J. Fish. Aquat. Sci., 65, 2606–2622, doi:10.1139/F08-149, 2008.

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, doi:10.1002/hyp, 2008.

Hijmans, R. J.: raster: Geographic Data Analysis and Modeling. R package version 2.5-2., [online] Available from: https://cran.r-project.org/package=raster, 2015.

Hill, R. A., Hawkins, C. P. and Carlisle, D. M.: Predicting thermal reference conditions for USA streams and rivers, Freshw. Sci., 32(1), 39–55, doi:10.1899/12-009.1, 2013.

Ver Hoef, J. M. and Peterson, E. E.: A Moving Average Approach for Spatial Statistical Models of Stream Networks, J. Am. Stat. Assoc., 105(489), 6–18, doi:10.1198/jasa.2009.ap08248, 2010.

Ver Hoef, J. M., Peterson, E. E., Clifford, D. and Shah, R.: SNN: An R package for spatial statistical modeling on stream networks, J. Stat. Softw., 56(3), 2014.

Ver Hoef, J. M., Peterson, E. and Theobald, D.: Spatial statistical models that use flow and stream distance, Environ. Ecol. Stat., 13, 449–464, doi:10.1007/s10651-006-0022-8, 2006.

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, doi:10.1002/hyp.7756, 2010.

Imholt, C., Soulsby, C., Malcolm, I. A. and Gibbins, C. N.: Influence of contrasting riparian forest cover on stream temperature dynamics in salmonid spawning and nursery streams, Ecohydrology, 6(3), 380–392, doi:10.1002/eco.1291, 2013.

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, doi:10.1002/rra, 2011.

Isaak, D. J., Luce, C. H., Rieman, B. E., Nagel, D. E., Peterson, E. E., Horan, D. L., Parkes, S. and Chandler, G. L.: Effects of climate change and wildfire on stream temperatures and salmonid thermal habitat in a mountain river network, Ecol. Appl., 20(5), 1350–1371, 2010.

Isaak, D. J., Peterson, E. E., Ver Hoef, J. M., Wenger, S. J., Falke, J. A., Torgersen, C. E., Sowder, C., Steel, A. E., Fortin, M.-J., Jordan, C. E., Ruesch, A. S., Som, N. and Monestiez, P.: Applications of spatial statistical network models to stream data, WIREs Water, 1(3), 227–294, doi:10.1002/wat2.1023, 2014.

Isaak, D. J., Wenger, S. J., Peterson, E. E., Hoef, J. M. Ver, Hostetler, S., Luce, C. H., Dunham, J. B., Kershner, J., Roper, B. B., Nagel, D., Horan, D., Chandler, G., Parkes, S. and Wollrab., S.: NorWeST: An interagency stream temperature

database and model for the Northwest United States. U.S. Fish and Wildlife Service, Great Northern Landscape Conservation Cooperative Grant. www.fs.fed.us/rm/boise/AWAE/projects/NorWeST.html., 2011.

Isaak, D. J., Wollrab, S., Horan, D. and Chandler, G.: Climate change effects on stream and river temperatures across the northwest U.S. from 1980–2009 and implications for salmonid fishes, Clim. Change, 113(2), 499–524, doi:10.1007/s10584-011-0326-z, 2012.

Jackson, F. L., Hannah, D. M., Fryer, R. J., Millar, C. P. and Malcolm, I. A.: Development of spatial regression models for predicting summer river temperatures from landscape characteristics: implications for land and fisheries management, Hydrol. Process., 31, 1225-1238, doi:10.1002/hyp.11087, ~~2016a~~2017.

Jackson, F. L., Malcolm, I. A. and Hannah, D. M.: A novel approach for designing large-scale river temperature monitoring networks, Hydrol. Res., 47(3), 569–590, doi:10.2166/nh.2015.106, 2016~~b~~.

~~Johnson, M. F., Wilby, R. L. and Toone, J. A.: Inferring air water temperature relationships from river and catchment properties, Hydrol. Process., 28(6), 2912–2928, doi:10.1002/hyp.9842, 2014.~~

Jonkers, A. R. T. and Sharkey, K. J.: The Differential Warming Response of Britain's Rivers (1982–2011), PLoS One, 11(11), e0166247, doi:10.1371/journal.pone.0166247, 2016.

Jonsson, B. and Jonsson, N.: A review of the likely effects of climate change on anadromous Atlantic salmon Salmo salar and brown trout Salmo trutta, with particular reference to water temperature and flow, J. Fish Biol., 75(10), 2381–2447, doi:10.1111/j.1095-8649.2009.02380.x, 2009.

Letcher, B. H, Hocking, D. J, O'Neil, K, Whiteley, A.R, Nislow, K.H, O'Donnell, M.J.: A hierarchical model of daily stream temperature using air-water temperature synchronization, autocorrelation, and time lags. PeerJ, 4:e1727, https://doi.org/10.7717/peerj.1727, 2016

Luce, C., Staab, B., Kramer, M., Wenger, S., Isaak, D. and McConnell, C.: Sensitivity of Summer Stream Temperatures to Climate Variability in the Pacific Northwest, Water Resour. Res., 50(4), 3428–3443, doi:10.1002/2013/WR014329, 2014.

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, doi:10.5194/hess-8-449-2004, 2004.

Malcolm, I. A., Soulsby, C., Hannah, D. M., Bacon, P. J., Youngson, A. F. and Tetzlaff, D.: The influence of riparian woodland on stream temperatures: implications for the performance of juvenile salmonids, Hydrol. Process., 22(7), 968–979, doi:10.1002/hyp, 2008.

Marine, K. R. and Cech, J. J. J.: Effects of High Water Temperature on Growth , Smoltification , and Predator Avoidance in Juvenile Sacramento RiverChinook Salmon, North Am. J. Fish. Manag., 24(1), 198–210, doi:10.1577/M02-142, 2004.

Mayer, T. D.: Controls of summer stream temperature in the Pacific Northwest, J. Hydrol., 475, 323–335, doi:10.1016/j.jhydrol.2012.10.012, 2012.

McCullough, D., Spalding, S., Sturdevant, D. and Hicks, M.: Summary of Technical Literature Examining the Physiological Effects of Temperature on Salmonids., 2001.

MET Office: Regional values - August 2015, [online] Available from: http://www.metoffice.gov.uk/climate/uk/summaries/2015/august/regional-values (Accessed 20 October 2016), 2016.

Millar, C. ., Millidine, K. ., Middlemass, S. . and Malcolm, I. .: Development of a Model for Predicting Large Scale Spatio-Temporal Variability in Juvenile Fish Abundance from Electrofishing Data, Scottish Mar. Freshw. Sci. Rep., 6(6), 33, doi:10.7489/1616-1, 2015.

Millar, C. P., Fryer, R. J., Millidine, K. J. and Malcolm, I. A.: Modelling capture probability of Atlantic salmon (Salmo salar) from a diverse national electrofishing dataset: Implications for the estimation of abundance, Fish. Res., 177, 1–12, doi:10.1016/j.fishres.2016.01.001, 2016.

Millidine, K. J., Malcolm, I. A. and Fryer, R. J.: Assessing the transferability of hydraulic habitat models for juvenile Atlantic salmon, Ecol. Indic., 69, 434–445, doi:10.1016/j.ecolind.2016.05.012, 2016.

Moore, R. D., Nelitz, M. and Parkinson, E.: Empirical modelling of maximum weekly average stream temperature in British Columbia, Canada, to support assessment of fish habitat suitability, Can. Water Resour. J., 38(2), 135–147, doi:10.1080/07011784.2013.794992, 2013.

Moore, R. D., Sutherland, P., Gomi, T. and Dhakal, A.: Thermal regime of a headwater stream within a clear-cut, coastal British Columbia, Canada, Hydrol. Process., 19(13), 2591–2608, doi:10.1002/hyp.5733, 2005.

O'Donnell, D., Rushworth, A., Bowman, A. W., Scott, M. E. and Hallard, M.: Flexible regression models over river networks, J. R. Stat. Soc. Ser. C (Applied Stat., 63(1), 47–63, doi:10.1111/rssc.12024, 2014.

Pebesma, E. J. and Bivand, R. S.: Classes and methods for spatial data in R., [online] Available from: http://cran.r-project.org/doc/Rnews/, 2005.

Perry, M. and Hollis, D.: The development of a new set of long-term climate averages for the UK, Int. J. Climatol., 25(8), 1023–1039, doi:10.1002/joc.1160, 2005a.

Perry, M. and Hollis, D.: The generation of monthly gridded datasets for a range of climatic variables over the UK, Int. J. Climatol., 25(8), 1041–1054, doi:10.1002/joc.1161, 2005b.

Peterson, E. E. and Ver Hoef, J. M.: STARS: An ArcGIS Toolset Used to Calculate the Spatial Information Needed to Fit Statistical Models to Stream Network Data, J. Stat. Softw., 56(2), 2014.

Peterson, E. E. and Urquhart, N. S.: Predicting water quality impaired stream segments using landscape-scale data and a regional geostatistical model: a case study in Maryland., Environ. Monit. Assess., 121(1–3), 615–38, doi:10.1007/s10661-005-9163-8, 2006.

Peterson, E., Ver Hoef, J. and Scopel, C.: SSN and STARS : Tools for Spatial Statistical Modeling on Stream Networks, , 2–4, 2013.

R Core Team, ..: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., [online] Available from: http://www.r-project.org/, 2015.

Roberts, J. J., Fausch, K. D., Peterson, D. P. and Hooten, M. B.: Fragmentation and thermal risks from climate change interact to affect persistence of native trout in the Colorado River basin. Glob Change Biol, 19, 1383–1398. doi:10.1111/gcb.12136, 2013

Ruesch, A. S., Torgersen, C. E., Lawler, J. J., Olden, J. D., Peterson, E. E., Volk, C. J. and Lawrence, D. J.: Projected Climate-Induced Habitat Loss for Salmonids in the John Day River Network, Oregon, U.S.A., Conserv. Biol., 26(5), 873–882, doi:10.1111/j.1523-1739.2012.01897.x, 2012.

Rushworth, A. M., Peterson, E. E., Ver Hoef, J. M. and Bowman, A. W.: Validation and comparison of geostatistical and spline models for spatial stream networks, Environmetrics, (26.5), 327–338, doi:10.1002/env.2340, 2015.

Smith, T., Hayes, K., Marshall, L., McGlynn, B. and Jencso, K.: Diagnostic calibration and cross-catchment assessment of a simple process-consistent hydrologic model, Hydrol. Process., doi:10.1002/hyp.10955, 2016.

Som, N. A., Monestiez, P., Ver Hoef, J. M., Zimmerman, D. L. and Peterson, E. E.: Spatial sampling on streams: principles for inference on aquatic networks, Environmetrics, doi:10.1002/env.2284, 2014.

Sowder, C. and Steel, E. A.: A note on the collection and cleaning of water temperature data, Water, 4, 597–606, doi:10.3390/w4030597, 2012.

Steel, E. A., Sowder, C. and Peterson, E. E.: Spatial and Temporal Variation of Water Temperature Regimes on the Snoqualmie River Network, JAWRA J. Am. Water Resour. Assoc., 1–19, doi:10.1111/1752-1688.12423, 2016.

Strahler, A.: Quantitative analysis of watershed geomorphology, Trans. Am. Geophys. Union, 38, 913–920, 1957.

Tague, C., Farrell, M., Grant, G., Lewis, S. and Rey, S.: Hydrogeologic controls on summer stream temperatures in the McKenzie River basin, Oregon, Hydrol. Process., 21, 3288–3300, doi:10.1002/hyp, 2007.

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), 902–918, doi:10.1002/hyp, 2008.

Wehrly, K. E., Brenden, T. O. and Wang, L.: A Comparison of Statistical Approaches for Predicting Stream Temperatures Across Heterogeneous Landscapes, JAWRA J. Am. Water Resour. Assoc., 45(4), 986–997, doi:10.1111/j.1752-1688.2009.00341.x, 2009.

Wood, S. N.: mgcv: GAMs and generalized ridge regression for R., R-News, 1(2), 20–25, 2001.

Wood, S. N.: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation, , 1–273 [online] Available from: https://cran.r-project.org/web/packages/mgcv/mgcv.pdf, 2016.

Table 1 Covariate calculations. All calculations were in R, version 3.2.3 (R Core Team, 2015) except where specified.

Covariate	Process and associated packages	Datasets
Elevation	'extract' function in the 'raster' package (Hijmans, 2015)	OS. Terrain 10m DTM, CEH DRN
Gradient	'extract' function in the 'raster' package (Hijmans, 2015) to get elevations of the node and a location 1km upstream. The difference in these elevations divided by the length between the two nodes provided Gradient. The length upstream was calculated using 'SpatialLinesLengths' from 'sp' (Pebesma and Bivand, 2005)	OS. Terrain 10m DTM, CEH DRN
Orientation	Standard trigonometry based on the x and y locations of the node and associated upstream points 1km upstream lengths. The lengths upstream was calculated using 'SpatialLinesLengths' from 'sp' (Pebesma and Bivand, 2005)	CEH DRN
Upstream Catchment Area (UCA)	Arc Hydro Tools (ArcGIS 10.2.1) was used to 'burn in' the DRN to the DTM and then calculate an UCA raster.	OS. Terrain 10m Digital Terrain Model, DTM; CEH DRN
Hillshading/ Illumination (HS)	'terrain' and 'hillShade' functions in the 'raster' package (Hijmans, 2015) were used to create a hillshade layer for every hour the sun was above the horizon. These layers were then summed to create a single layer of maximum potential exposure. HS values for the nodes were an average of the raster grid cells in the 1km river polygon. Raster grid cells were weighted by the proportion of the cell within the buffer.	CEH DRN , OS. Terrain 10m DTM; Solar azimuth and altitude values from the U.S. Naval Observatory Astronomical Applications Department (Anon, 2001)
Percentage riparian woodland (%RW)	The percentage of woodland in a buffer 50m wide and 1km long (upstream) provided %RW. Areas were calculated using 'gArea' from 'rgeos' (Bivand and Rundel, 2016) and lengths the 'SpatialLinesLengths' from 'sp' (Pebesma and Bivand, 2005).	OS MasterMap, CEH DRN
Width	Width was calculated by finding the area classified as water within the 1km upstream and dividing this by the distance upstream. Areas were calculated using 'gArea' from 'rgeos' (Bivand and Rundel, 2016) and lengths the 'SpatialLinesLengths' from 'sp' (Pebesma and Bivand, 2005).	OS MasterMap, CEH DRN
Distance to coast (DC)	gDistance' from the 'rgeos' R package (Bivand and Rundel, 2016).	CEH DRN, OS Panorama coastline
River distance to sea (RDS)	"shortest.paths" function from the igraph R package (Csardi and Nepusz, 2006)	CEH DRN, OS Panorama coastline
Highest 7-day average maximum August Ta (Ta_{max})	Take the Ta value, from daily maximum predicted Ta matrix, for each cell containing a SRTMN site. Use these daily values to calculate rolling averages then select the highest, for each site.	Gridded UKCP09 predicted Ta dataset (UK MET Office)

5

Table 2 LS model and LS_Ta model for each catchment, with the % variance explained by the model (all terms) and the same model but with the RNS omitted (covariates).

Catchment	Model	AICc / BIC	df	% variance	
				all terms	covariates
LS model					
Dee	UCA + RNS	137.0	8.8	80.0	18.3
Tweed	s(UCA) + %RW + RNS	100.1	7.6	85.6	71.7
Spey	s(UCA) + RNS	69.8	6.8	85.5	21.8
Bladnoch	s(UCA)	55.5	2.9	76.6	76.6
LS_Ta model					
Dee	Ta_{max} + UCA + s(Orientation) + RNS	131.8	11.7	85.1	19.9
Tweed	Ta_{max} + s(UCA) + %RW + s(HS) + Orientation	98.5	7.7	85.3	85.3
Spey	UCA + DC + RNS	69.3	7.4	85.1	19.9
Bladnoch	Ta_{max} + UCA	53.9	4.8	85.9	85.9

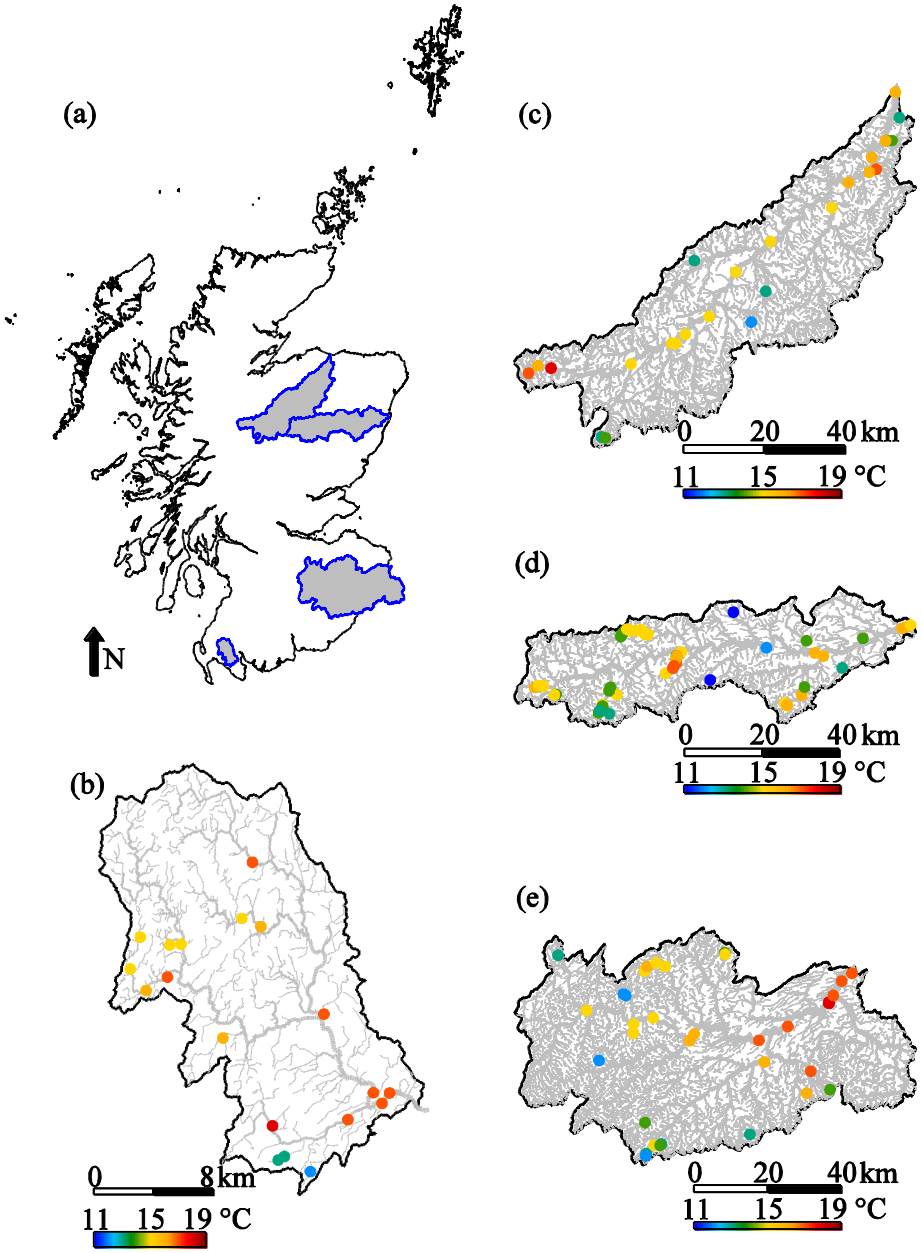
Table 3 Transferability of LS and LS_Ta models. The values in normal font are for predictions using only covariates (any RNS information is ignored). The values in italics are for predictions when the target and donor catchments are the same and all model terms are used (both covariates and the RNS).

	Donor catchment	Target catchment			
		Dee	Tweed	Spey	Bladnoch
LS models					
RMSE	Dee	1.3 *(0.8)*	1.2	1.3	2.3
	Tweed	1.1	1.1 *(0.9)*	1.3	1.9
	Spey	1.3	1.3	1.4 *(0.9)*	2.5
	Bladnoch	2.2	1.9	2.4	0.9 *(0.9)*
Bias	Dee	-0.6 *(0.1)*	-0.6	-0.1	-2.0
	Tweed	0.2	-0.1 (*0.0)*	0.3	-1.7
	Spey	-0.6	-0.8	-0.2 *(-0.0)*	-2.3
	Bladnoch	1.9	1.4	2.0	0.0 *(0.0)*
Standard Deviation	Dee	0.8 *(0.8)*	1.1	1.3	1.2
	Tweed	1.1	1.0 *(0.9)*	1.2	1.0
	Spey	1.1	1.1	1.4 *(0.9)*	1.0
	Bladnoch	1.1	1.3	1.3	0.9 *(0.9)*
LS_Ta models					
RMSE	Dee	1.7 *(0.7)*	0.9	1.9	1.9
	Tweed	1.2	0.9 *(0.9)*	2.3	1.5
	Spey	1.5	2.0	1.8 *(0.8)*	4.2
	Bladnoch	8.4	4.2	5.2	0.9 *(0.9)*
Bias	Dee	-1.1 *(0.1)*	-0.3	0.0	-1.4
	Tweed	-0.7	-0.0 *(-0.0)*	-0.1	-1.0
	Spey	-0.1	-1.4	-0.7 *(0.0)*	-4.1
	Bladnoch	7.6	3.2	4.1	-0.1 *(-0.1)*
Standard deviation	Dee	1.3 *(0.7)*	0.9	1.9	1.2
	Tweed	1.0	0.9 *(0.9)*	2.3	1.1
	Spey	1.5	1.4	1.6 *(0.8)*	1.1
	Bladnoch	3.7	2.7	3.3	0.9 *(0.9)*

Table 4 Multi-catchment LS and LS_Ta model, with the % variance explained by the model (all terms) and when the RNS is omitted (covariates).

Model	BIC	df	% variance	
			all terms	covariates
Multi-catchment LS model				
Catchment + s(UCA) + %RW + Elevation + RNS:Catchment	379.3	24.8	84.4	51.9
Multi-catchment LS_Ta model				
Catchment + Ta_{max}:Catchment + s(UCA) + %RW + RNS:Catchment	395.4	25.7	83.2	57.2

Figure 1. Study catchments and spatial patterns of Twmax for August 2015 a) Catchment positions in Scotland b) River Bladnoch catchment, c) River Spey catchment, d) River Dee catchment, e) River Tweed catchment



5

Figure 2. Box and whisker plot of Tw_{max} across sites by catchment for August 2015.

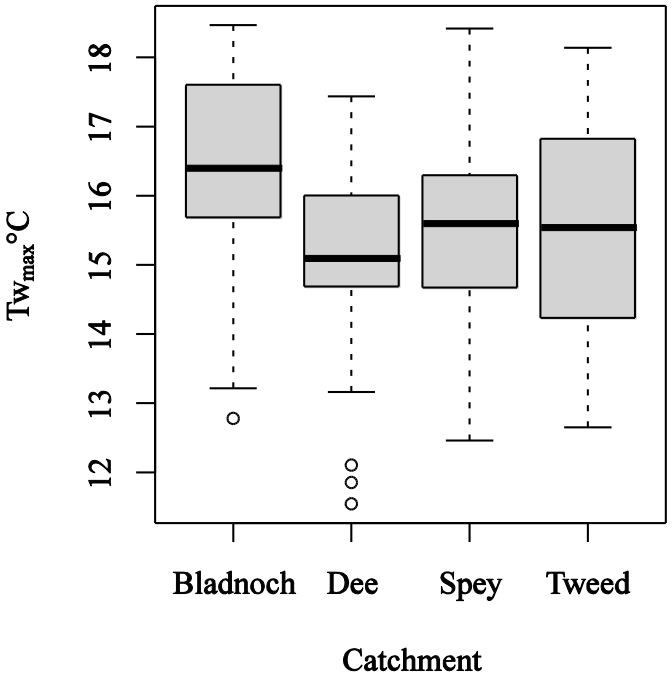


Figure 2 Distributions and inter-relationships between Tw_{max} and covariates. Scatter plots of the relationships are shown below the diagonal, kernel density plots of the individual covariates in the diagonal (scaled to have the same maximum value) and correlation coefficients above the diagonal. Numbers in black indicate the correlation coefficients where data is pooled across all catchments.

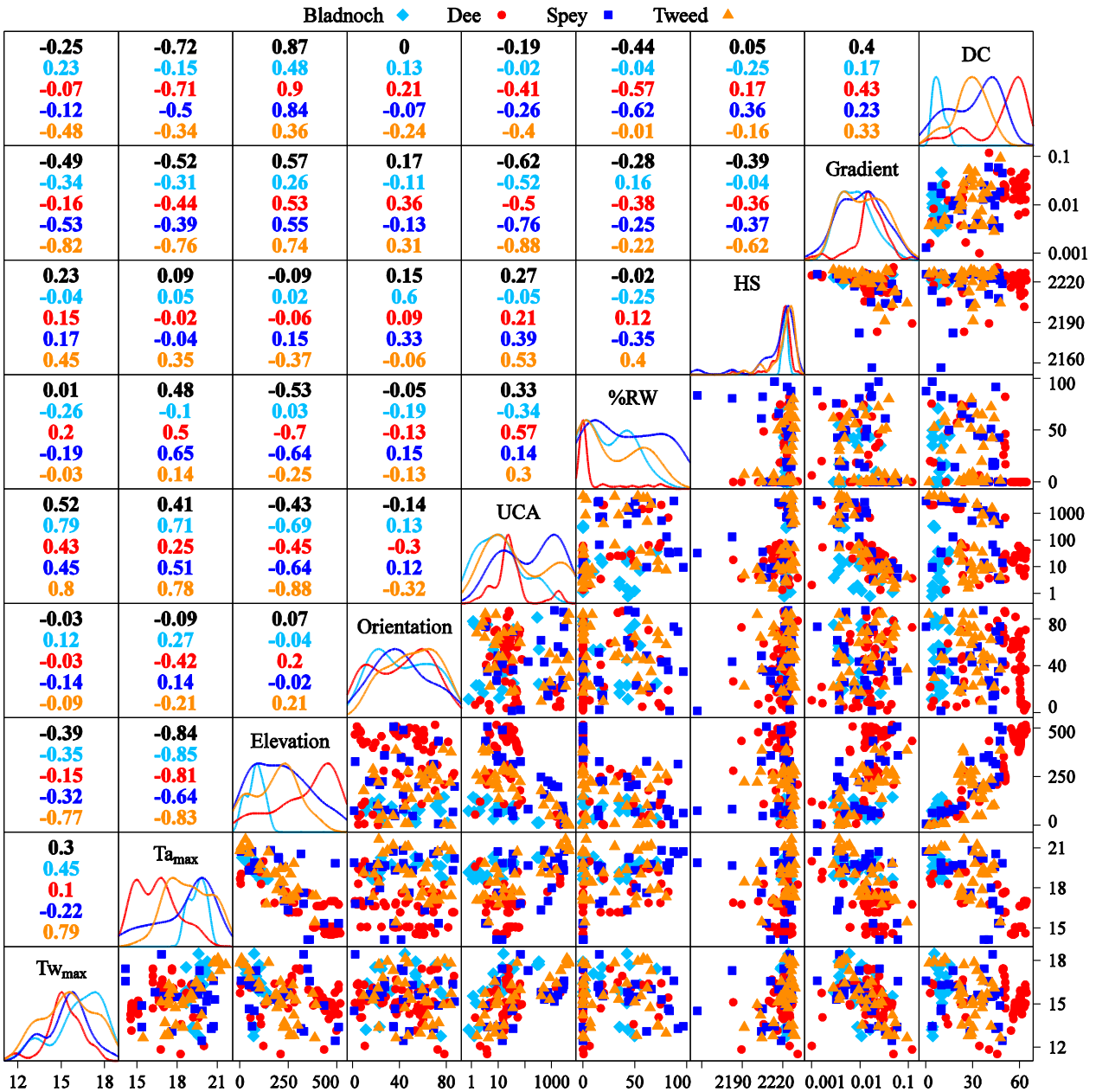


Figure 3. LS model effects with pointwise 95% confidence bands: a) Dee UCA, b) Tweed UCA, c) Spey UCA, d) Bladnoch UCA, e) Tweed %RW.

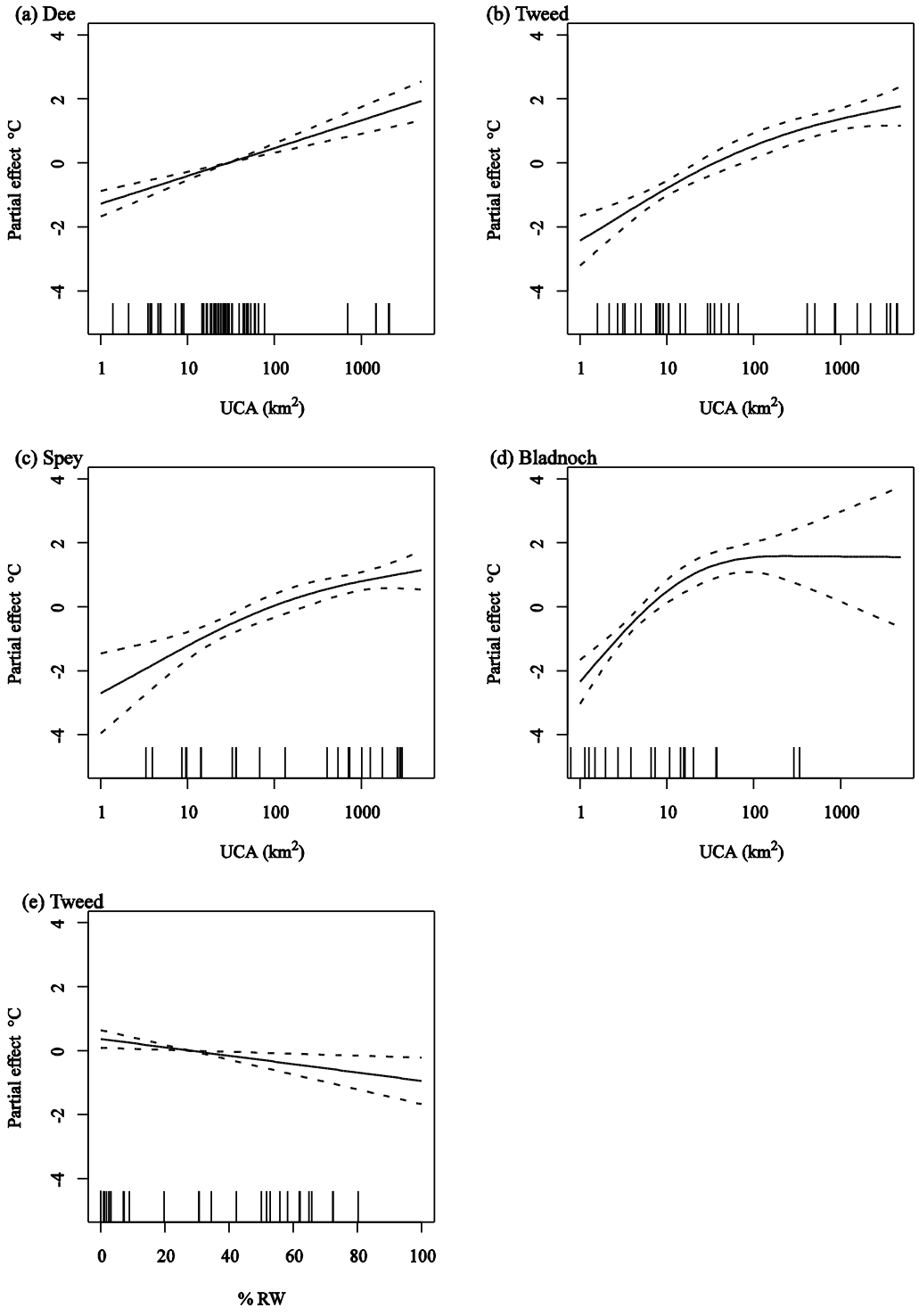


Figure 4. LS_Ta model effects with pointwise 95% confidence bands. Each column corresponds to a catchment and each row to a covariate. a) Dee UCA, b) Tweed UCA, c) Spey UCA, d) Bladnoch UCA, e) Dee Ta_{max}, f) Tweed Ta_{max}, g) Bladnoch Ta_{max}, h) Dee Orientation, i) Tweed orientation, j) Tweed %RW, k) Tweed hillshading, l) Spey DC.

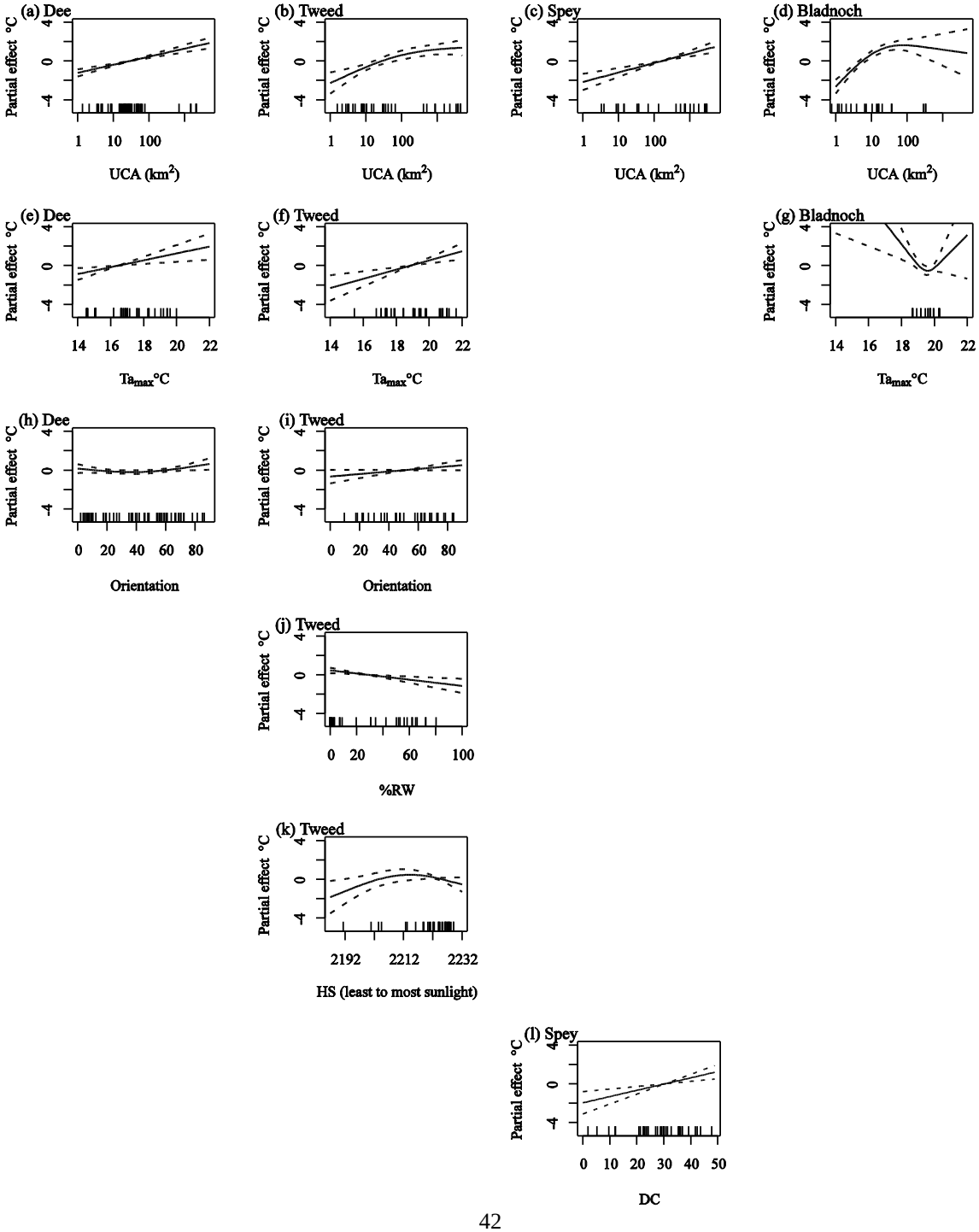


Figure 5. LS model transferability. Panels a, b, c, d show predicted Tw_{max} when the target catchment is the Dee, Tweed, Spey and Bladnoch respectively. The colours and symbols indicate the donor catchment: Dee (red circles), Tweed (orange triangles), Spey (dark blue squares) and Bladnoch (light blue diamonds). Filled (open) symbols indicate sites in (out) of the environmental range of the donor catchment. When the target and donor catchments are the same, the coloured points are based on predictions using only covariates; the grey symbols show the corresponding predictions based on the covariates and the RNS. The dashed lines are robust regression lines of observed against predicted values. Models which transfer well have points falling close to the 1:1 line.

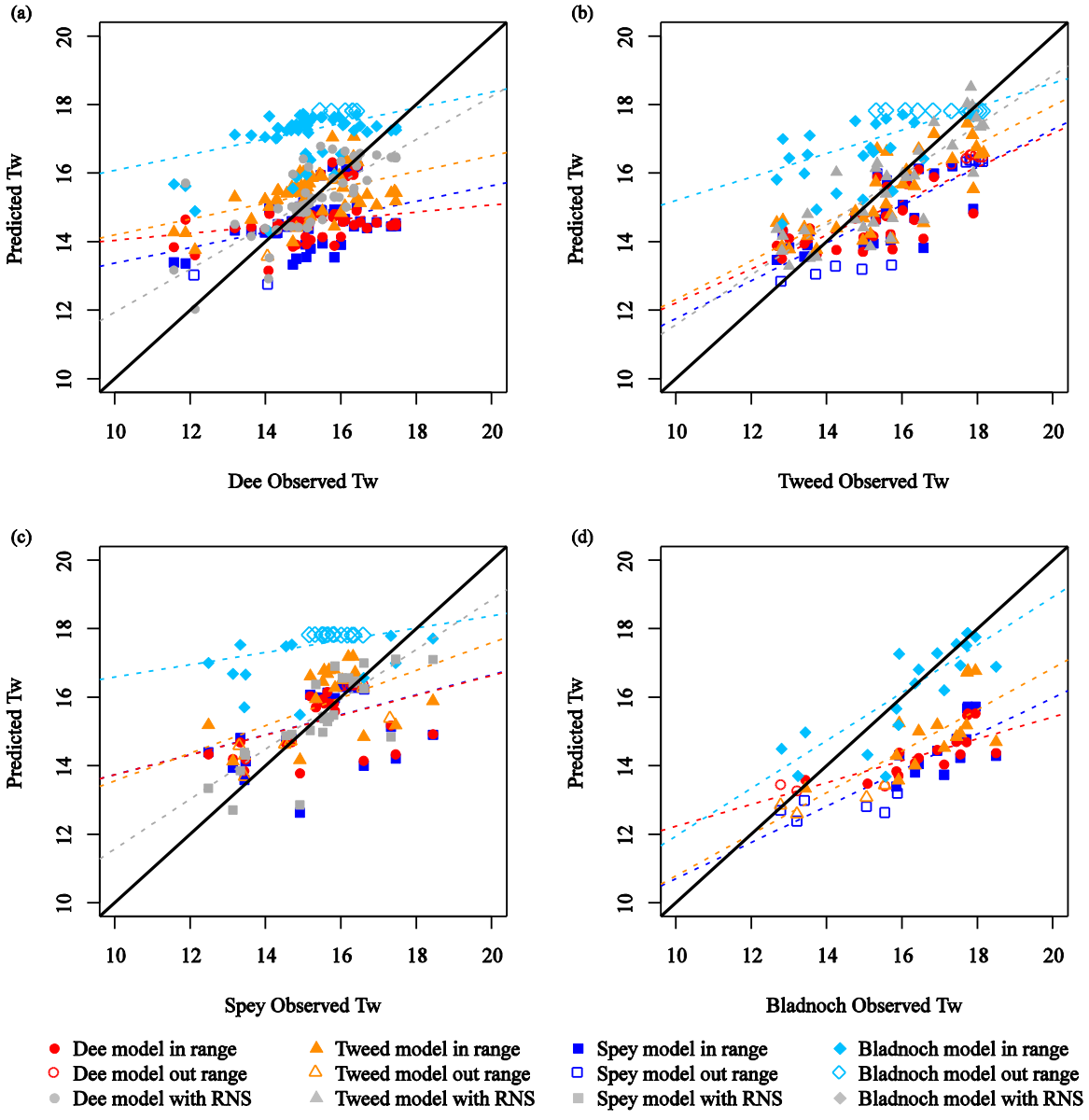


Figure 6. LS_Ta model transferability. Panels a, b, c, d show predicted Tw_{max} when the target catchment is the Dee, Tweed, Spey and Bladnoch respectively. The colours and symbols indicate the donor catchment: Dee (red circles), Tweed (orange triangles), Spey (dark blue squares) and Bladnoch (light blue diamonds). Filled (open) symbols indicate sites in (out) of the environmental range of the donor catchment. When the target and donor catchments are the same, the coloured points are based on predictions using only covariates; the grey symbols show the corresponding predictions based on the covariates and the RNS. The dashed lines are robust regression lines of observed against predicted values. Models which transfer well have points falling close to the 1:1 line.

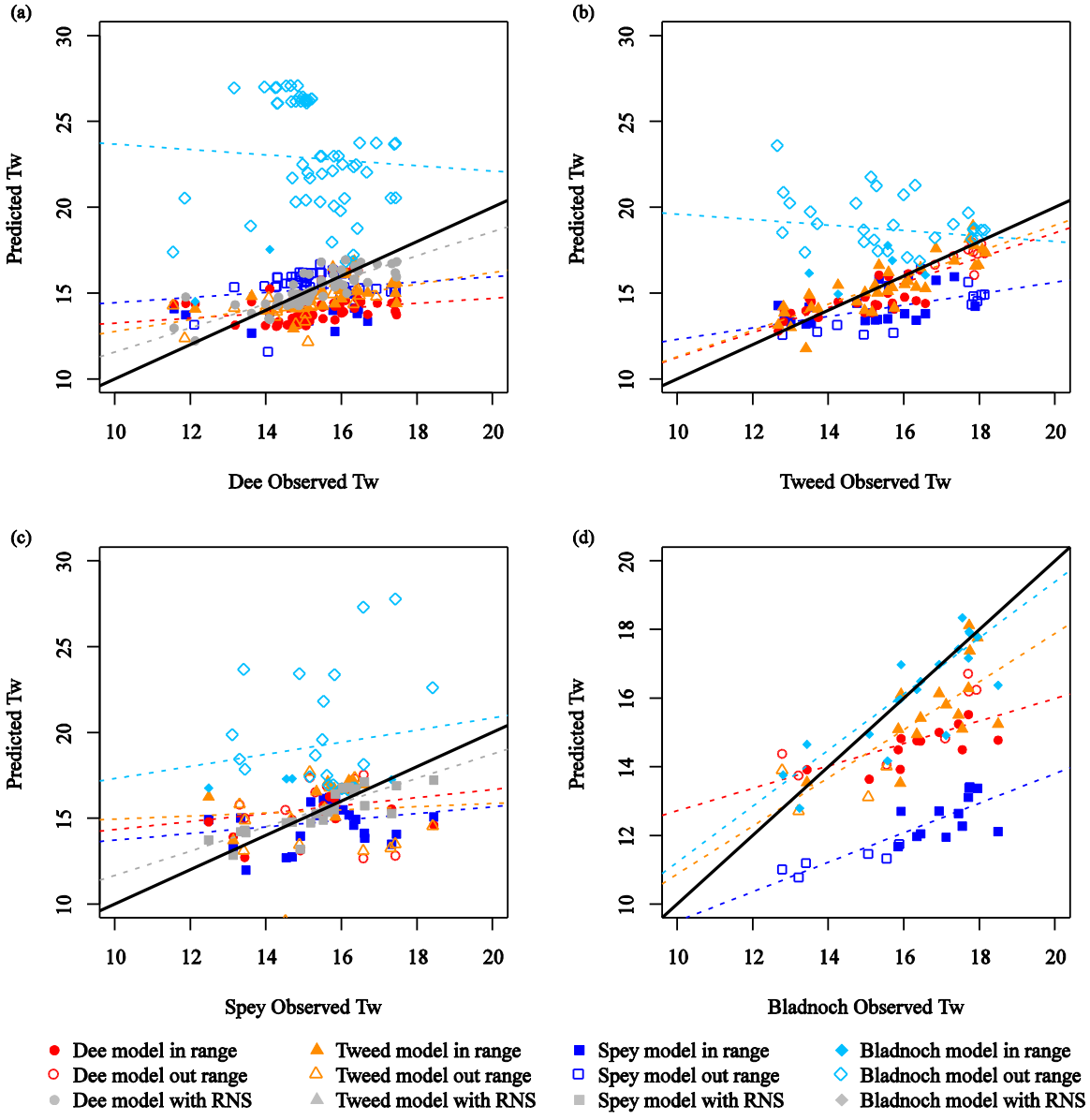


Figure 7. Multi-catchment LS model effects with pointwise 95% confidence bands: a) UCA, b) %RW, c) Elevation, d) Catchment.

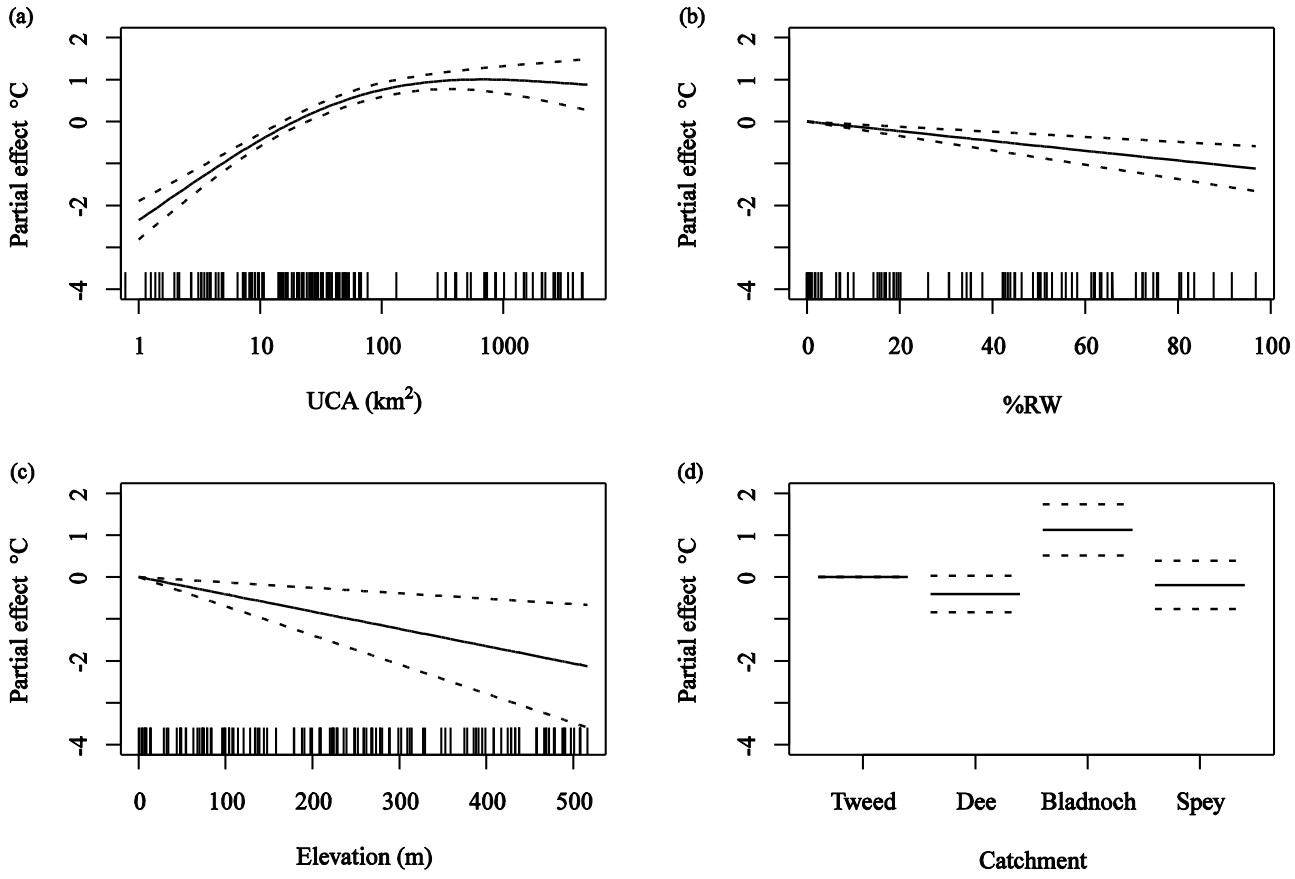
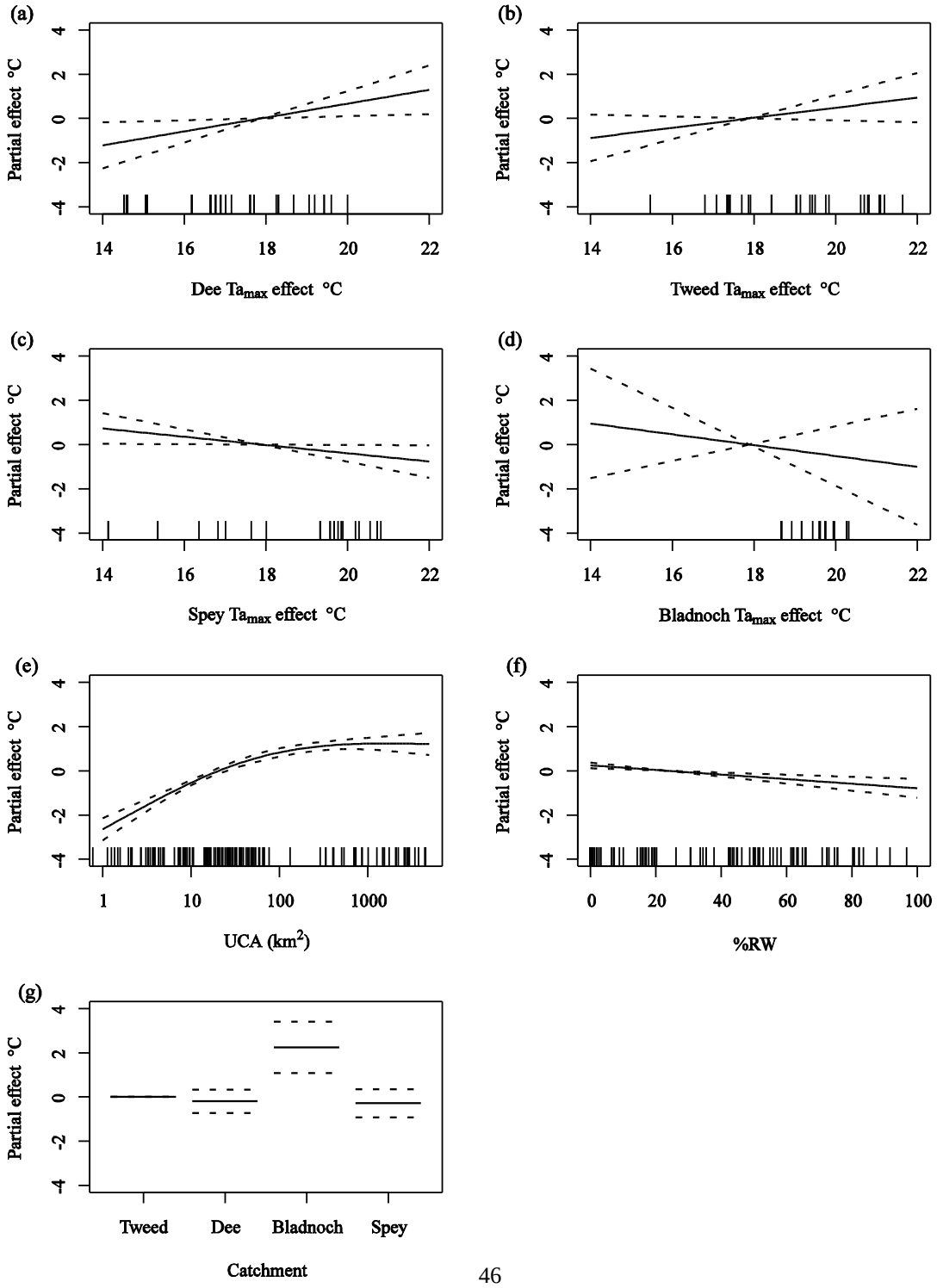



Figure 8. Multi-catchment LS_Ta model effects with pointwise 95% confidence bands: a) Dee Ta_{max}, b) Tweed Ta_{max}, c) Spey Ta_{max}, d) Bladnoch Ta_{max}, e) UCA, f) %RW, g) Catchment.