# *Interactive comment on* "Can river temperature models be transferred between catchments?" *by* Faye L. Jackson et al.

**Anonymous 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. However, 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 modeling 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.

2. What was the sampling design? Were sites selected randomly within some pre-defined 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.

3. 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).

4. 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.

5. 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).

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.

6. 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?

7. In the conclusion, the authors suggest that further work should investigate the modeling 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).

8. 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.

9. 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.

References

Isaak, D. J., and W. A. Hubert. 2001. A hypothesis about factors that affect maximum summer stream temperatures across montane landscapes. Journal of the American Water Resources Association 37: 351–366.

Luce, C.H., Staab, B., Kramer, M., Wenger, S., Isaak, D., and McConnell, C. 2014. Sensitivity of summer stream temperatures to climate variability in the Pacific Northwest. Water Resources Research 50 (4), 3428-3443, doi: 10.1002/2013WR014329

Moore, R.D., Nelitz, M. and Parkinson, E. 2013. Empirical modelling of maximum weekly average stream temperature in British Columbia, Canada, to support assessment of fish habitat suitability. Canadian Water Resources Journal 38: 135-147, DOI:10.1080/07011784.2013.794992.

Scott, M. C., G. S. Helfman, M. E. McTammany, E. F. Benfield, and P. V. Bolstad. 2002. Multiscale influences on physical and chemical stream conditions across Blue Ridge

landscapes. Journal of the American Water Resources Association 38: 1379–1392.

Tague, C., M. Farrell, and G. Grant. 2007. Hydrogeologic controls on summer stream temperatures in the McKenzie River Basin, Oregon. Hydrological Processes 21 (24): 3288–3300. doi: 10.1002/hyp.6538.

Wehrly, K. E., T. O. Brenden, and L. Wang. 2009. A comparison of statistical approaches for predicting stream temperatures across heterogeneous landscapes. Journal of the American Water Resources Association 45: 986–997. doi: 10.1111/j.1752-1688.2009.00341.x.