remap: Regionalized Models with Spatially Smooth Predictions
Response to Review

The authors would like to thank this reviewer for their helpful comments. Addressing these comments has improved both the quality of the manuscript as well as the quality of the software. This document provides a point by point response to the reviewer comments. Reviewer comments are indicated by indentations to the left margin. Responses are provided using standard margins. Changes are reflected in both the manuscript, as well as version 0.3.0 of the package which is currently available on CRAN.

> I have tested all of the example code and everything works as expected. Perhaps the only issue
> is that in some instances a warning message is printed as coordinates specified in the example
> data are lat/long, but some of the operations used assume planar coordinates. I think in the
> examples here that isn't an issue, but perhaps something in the text could explain the warnings
> - or alternatively sp_transform could be used to work with a different map projection.

Recent changes to the `sf` package have fully integrated Google's S2 geometry library. This has resulted in a change in the default behavior for several `sf` function when applied to unprojected coordinates. We have altered the coding examples to work with the latest version of the `sf` package, which has eliminated all of the lat/long warning messages referenced by the reviewer.

The example included additional warnings related to issues within the autoKrige function of the automap package. We were able to rewrite the example avoid the use of autoKrige and circumvent the warnings while still retaining exactly identical results. The example scripts now run with no warnings.

> The method itself is well presented, and its reasons for use are well set out. Perhaps one issue
> I do have, is that although the examples are well set out, not all of the results shown have the
> appropriate R code. Since this article is intended for the R Journal, it would be good if there
> were a clear correspondence between the analysis examples and the code given. Currently
> there are distinct sections for applications and example code - and at the start of the code
> section it is stated that it shows 'the code used to create some of the regional models described
> in the Applications section' (my italics). I think it would be better to give code for all of the
> examples, and to do supply the code alongside the individual examples, all in the Applications
> section.

We have eliminated the code section and integrated coding examples into the applications section. This new consolidated application section can be seen on Pages 8 - 16 of the updated manuscript. Several minor formatting changes are made to the text to accommodate the consolidation, but the examples and majority of the text remain largely unchanged.

Another notable change in response to this comment includes the expansion of the code provided within the manuscript for the examples in the application section. Previously we had only included code for GAM and kriging models on the snowload dataset. Now, the text contains examples for linear, GAM, and kriging models for both the snowload dataset and snowpack dataset (pages 9 to 12 and 15 to 16). These examples cover all of the remap modeling applications discussed in the text.

As in the previous submission, code for most figures along with the cross validation and speed test are not included in the text due to computational and readability constraints, but the complete code continues to be available at `https://github.com/jadonwagstaff/remap_manuscript_code`, as is referenced in the text.

> One further thing I wondered is whether estimates of standard errors are possible for this kind
> of prediction merging. Although the article demonstrates the predictive ability of the method,
> it would be useful, if practically possible, to include SEs. Even if this isn't viable, or would
> need further work, maybe it is worth mentioning this in the conclusions?

The estimation of standard errors (SE) for remap model predictions in the smoothing zones is difficult for multiple reasons. First and foremost, the arbitrary selection of a smoothing zone ultimately influences the blended SE calculation, yet the smoothing zone decision is one that is made for practical, rather than theoretical, reasons. Further, the assumptions underlying SE estimates for some models clash with remap's assumptions. For example, estimates of SE from ordinary least squares regression assume independent model residuals, which violates the spatial autocorrelation assumption implicit to remap. On the other hand, estimates of Kriging variance rely on the assumption of fixed Covariance structures within well-defined spatial boundaries, which contradicts remap's goal of blurring boundary lines in order to provide spatially smooth transitions.

Another difficulty with SE estimation within remap is the fact that individual model predictions along spatial boundaries are likely to be highly correlated, especially when the selected buffer zones allow for large amounts of data overlap in model creation. It may be possible to estimate empirical correlations by generating large lists of co-located predictions for all relevant model pairs. However, such an approach would undermine remap's "partition and conquer" approach that fits regional models in isolation to facilitate parallelization and/or to limit computer RAM usage. All of these challenges prevent us from providing precise estimates of prediction SE in a remap framework.

All this in mind, the authors still see the utility of providing a reasonable and conservative approximation of the prediction SE for interested users. This could be accomplished by making an appealing to definition of variance for a sum of random variables and applying the Cauchy–Schwarz inequality, i.e.

$$\begin{aligned} Var\left(\sum_i w_i X_i\right) &= \frac{\sum_i \sum_j w_i w_j Cov(X_i, X_j)}{\left(\sum_k w_k\right)^2} \\ &= \frac{\sum_i w_i^2 Var(X_i) + \sum_i \sum_{j<i} 2 w_i w_j \rho_{ij} \sqrt{Var(X_i) Var(X_j)}}{\left(\sum_k w_k\right)^2} \\ &\leq \frac{\sum_i w_i^2 Var(X_i) + \sum_i \sum_{j<i} 2 w_i w_j \sqrt{Var(X_i) Var(X_j)}}{\left(\sum_k w_k\right)^2}. \end{aligned}$$

This allows us to provide an upper bound for the SE calculation while eliminating the need to calculate ρ_{ij}, which is not viable in the current iteration of remap. The updated version of the manuscript includes this estimation of the variance in the national snowload GAM example (pages 11 to 12). Additionally, some commentary explaining the difficulty of SE calculations as well as the rationale behind this conservative approximation has been added to the text (see Section "Standard error approximations" on page 5).

> I think the paper is suitable for the R Journal, but would strongly recommend that the above
> amendments are made before it is published.

The authors appreciate the recommendation. The updates to the `sf` package referenced previously in our response necessitated some additional changes to the manuscript beyond those recommended by the reviewer. Most notably, we have updated Table 2 of the manuscript to show speed tests for both the old and new default distance calculation options provided in the `sf` package. The new default distance calculations use Google's S2 geometry library, and are computationally much faster than the previous default options, as in shown in the updated Ta-

ble 2. The code to perform this updated speed test has been provided in the GitHub repository referenced in the manuscript.

Additionally, we have created a docker image that will exactly reproduce all of the example, figure, and table code provided in this manuscript regardless of any future changes to ancillary packages. A link to this docker image is now included in the manuscript (see Section "Computer code availability and competing interests", page 3). Further, we have included an unintentionally omitted reference to the masters thesis upon which this software package is based (see Section "Computer code availability and competing interests", page 3) as well as an Acknowledgements section at the end of the manuscript. A few minor changes to formatting have been applied to bring the manuscript more in line with the R Journal style guide. Lastly, we have confirmed that the examples work with the most recent versions of relevant ancillary software packages. These software package versions have been updated in the bibliography accordingly.