Reviewer 1

1. Does the introduction provide sufficient background and include all relevant references?

☐ Yes  ☐ Can be improved  ☑ Must be improved  ☐ Not applicable

1. Is the research design appropriate?

☐ Yes  ☐ Can be improved  ☑ Must be improved  ☐ Not applicable

1. Are the methods adequately described?

☑ Yes  ☐ Can be improved  ☐ Must be improved  ☐ Not applicable

1. Are the results clearly presented?

☐ Yes  ☐ Can be improved  ☑ Must be improved  ☐ Not applicable

1. Are the conclusions supported by the results?

☐ Yes  ☐ Can be improved  ☑ Must be improved  ☐ Not applicable

The manuscript titled “ensembleDownscaleR: R package for Bayesian ensemble averaging of PM2.5 geostatistical downscalers”, by Maden et al., describes an R package to estimate PM2.5 concentration based on chemical transport model and satellite-derived AOD. Despite a provided example of PM2.5 estimation at Los Angeles metropolitan area, the manuscript is mainly focused in describing the R package functions and no proper analysis of results are provided. It doesn’t seem suitable for “Remote Sensing” journal and my recommendation is reject it. More appropriate journals could be “R journal” or “Journal of statistical software”. Some recommendations to improve the manuscript are highlighted in the attached document

The remainder of Reviewer 1’s comments are contained in text boxes in the pdf “reviewer\_1\_comments.pdf”.

We thank the reviewer for their thoughtful and detailed comments. We found them very useful in improving many aspects of our work.

We would like to clarify to the reviewer that this work is submitted as a Technical Note and not a research article, which is why the work focuses on describing a statistical workflow and does not provide as much depth to the analyses included as would a typical research article. We believe this work is more suited to Remote Sensing than “R journal” or “Journal of statistical software” because the R package we present is specifically motivated by the application of exposure modeling with satellite-derived AOD. In our experience, statistical software journals are suited for more general methodological packages. Our analyses and results are aimed to be illustrative of our workflow and not original scientific contributions, though in response to the reviewer’s comments we add more details to the datasets used.

We also make the following updates in response to the specific comments left on the included annotated pdf:

Line 24: We agree with the comment and add “aerodynamic” to “in aerodynamic diameter”

Line 39: We agree with the comment and clarify that CTM does not always have finer spatial resolution than AOD

Line 40: We agree with the comment and add that transformation is required for AOD, prior to bias-correction

TODO Line 70: Wants to know why PM2.5 is averaged – Howard though I was referring to CTM

Line 92: We agree with the comment and remove the statistical formula.

Line 95: We agree with the comment and clarify that this step produces predictions for all grid cells

Line 108: We agree with the comment and provide additional description of the dataset in the "Case Study Data" section. We also include the grid cells in the Figure 1 study area plot.

Line 118: We note to the reviewer that there is a distinction between the “spatial-temporal” random effects, and the “spatio-temporal” covariates. The “spatial-temporal” random effects are separate sets of spatial effects for different temporal windows, while the “spatio-temporal” covariates are covariates collected over space and time. We agree that “spatial-temporal” should be homogenized with “spatial-time”, and change all occurrences of latter to the former.

Line 128: We clarify to the reviewer that the model and software do allow for space-time interaction (equation 5), by allowing the spatial effects to vary by discrete time periods (e.g. weeks, months, seasons), which can be specified using the “spacetime.id” parameter in the grm function (line 207).

Line 132: We agree with the comment and update “D” distance matrix to “d” distance scalar, to homogenize notation.

Line 133: We agree with the reviewer and add a citation and further explanation/justification for the ridge regression equivalence of the normal prior.

Line 160: We agree with the comment and add citations and further details.

Line 211: In response to this comment, we include AOD missingness information in the data description.

Line 220: We agree with this comment and add units to the figure 2 measurements and replace "observed CTM" with CTM PM2.5 Simulations"

Line 342: We note that the spatial (and other variants) of CV typically do perform worse than regular CV, but provide an indication of how well the spatial interpolation components of the model perform. We also include all CV types to give the reader an indication of the functionality included in the R package. While we agree that further analysis of AOD measurement availability and how it affects the improvement of the ensemble result would be useful, we think that this is outside the scope of this technical note, as our primarily contribution is introducing a user-friendly statistical workflow rather than providing analyses for a full research article.

Reviewer 2

1. Does the introduction provide sufficient background and include all relevant references?

☐ Yes  ☑ Can be improved  ☐ Must be improved  ☐ Not applicable

1. Is the research design appropriate?

☐ Yes  ☐ Can be improved  ☐ Must be improved  ☑ Not applicable

1. Are the methods adequately described?

☑ Yes  ☐ Can be improved  ☐ Must be improved  ☐ Not applicable

1. Are the results clearly presented?

☑ Yes  ☐ Can be improved  ☐ Must be improved  ☐ Not applicable

1. Are the conclusions supported by the results?

☑ Yes  ☐ Can be improved  ☐ Must be improved  ☐ Not applicable

Authors present a statistical and modelling procedure using R package software  aimed to modelling and forecast PM2,5 concentrations.  .

The paper is a technical report so it describes the used procedure. The manuscript is clear and is presented in a well-structured manner.  Figures are relevant and clear

Cited references are relevant but not recent and they need to be updated.

Supplementary file is important to understand this technical report

We thank the review for the positive comments and agree that additional recent references will strengthen the paper. To this aim we have added several more recent references related to PM2.5 prediction (using statistical/predictive methods for either CTM, AOD, or both CTM and AOD).

Review 3

1. Does the introduction provide sufficient background and include all relevant references?

☐ Yes  ☐ Can be improved  ☑ Must be improved  ☐ Not applicable

1. Is the research design appropriate?

☐ Yes  ☑ Can be improved  ☐ Must be improved  ☐ Not applicable

1. Are the methods adequately described?

☐ Yes  ☑ Can be improved  ☐ Must be improved  ☐ Not applicable

1. Are the results clearly presented?

☐ Yes  ☑ Can be improved  ☐ Must be improved  ☐ Not applicable

1. Are the conclusions supported by the results?

☐ Yes  ☐ Can be improved  ☑ Must be improved  ☐ Not applicable

This study provides an R package for Bayesian ensemble averaging of PM2.5 geostatistical downscaling. As an article type of Technical Note, the paper is technically correct, but some key issues and shortcomings should be further detailed as follows.

1. Section 1: Introduction: It is important to incorporate an introduction that evaluates the strengths and weaknesses of the existing research to enhance the quality of your work.

We agree with this comment and add a more thorough review in the introduction that evaluates strengths/weaknesses in relation to existing research. We specifically note that while machine learning methods exist that outperform in terms of prediction accuracy, our approach is cutting-edge in terms of uncertainty quantification, which is crucial for downstream exposure health analyses. We also direct the reader to Murray et al. (2019) for a more thorough comparison.

1. It is suggested to add a flowchart to illustrate the specific steps and key processes of downscaling.

We agree that a flowchart will help the reader better navigate the full Bayesian ensemble fitting process. We add Figure 2 flowchart, which visually denotes stages, tasks, and outputs.

1. Lines 77-80: A brief description of data sources should be added.

We agree with this comment and have added a brief description of the data sources for each covariate, following lines 77-80.

1. How to obtain the optimal parameters of the grm model?

We add clarification at the end of the “Stage 1 code example” which clarifies that the user may access all posterior parameter draws from the MCMC sample contained within the fitted model object (the output of the grm() function). One could calculate the mean of these parameter draws – the maximum a posteriori (MAP) estimate – if they want a single optimal parameter value.

1. Section 3: Since the time and space effects are included in the program, it is recommended to add accuracy validation for PM2.5 downscaling under various scenarios.

We believe the reviewer is suggesting that we examine model predictive performance for different model specifications (e.g., different temporal random effect scales, removing spatial and/or temporal random effects). However, we believe this analysis is beyond the scope of this Technical Note, which aims to demonstrate how the R package can be used. Moreover, our example model implementation is based on previous analyses, where we have demonstrated that removing model components can lead to poorer performance (Chang et al. 2013).

1. Lines 108-112: Considering that the study area is not particularly large, the time of 11.29 hours seems a bit long, indicating that the program may not be applicable to relevant studies.

We agree that there are some studies for which this method is not appropriate due to computation time. However, there are a subset of studies for which the method is applicable – namely studies for which both principled quantification of uncertainty and accurate predictions are important, and for which the study area is not prohibitively large. This is often the case when model outputs are used in downstream exposure analyses. One such current use of our modeling framework is by the Multi-Angle Imager for Aerosols (MAIA) project to estimate daily PM2.5 in multiple large population centers around the world (Diner et al. 2018). We also note that we use a conservatively large number of MCMC iterations in our tutorial analysis (25,000). Users may specify a smaller number of iterations, at least for preliminary model assessment, which could substantially decrease computation time.

1. The conclusion section is missing. It might be helpful to combine the discussion and conclusion into one section.

We note that the conclusion section is indicated as "not mandatory but may be added to the manuscript if the discussion is unusually long or complex", per the Remote Sensing author instructions.

References

Chang HH, Hu X, Liu Y. Calibrating MODIS aerosol optical depth for predicting daily PM2. 5 concentrations via statistical downscaling. Journal of exposure science & environmental epidemiology. 2014 Jul;24(4):398-404.

Diner, D.J.; Boland, S.W.; Brauer, M.; Bruegge, C.; Burke, K.A.; Chipman, R.; Di Girolamo, L.; Garay, M.J.; Hasheminassab, S.;Hyer, E.; et al. Advances in multiangle satellite remote sensing of speciated airborne particulate matter and association with adverse health effects: from MISR to MAIA. Journal of Applied Remote Sensing 2018, 12, 042603–042603.

Murray, Nancy L., et al. "A Bayesian ensemble approach to combine PM2. 5 estimates from statistical models using satellite imagery and numerical model simulation." *Environmental research* 178 (2019): 108601.