

Re-Submission Cover Letter

To the Editor(s),

Please find enclosed the resubmission of our manuscript, entitled “A Spatial Framework for Understanding Population Structure and Admixture, for re-consideration for publication in PLoS Genetics.

We appreciate the opportunity to resubmit a re-revised version of the manuscript. The comments provided by the Guest Editor and two anonymous reviewers continue to be very helpful, and we believe the manuscript is substantially improved as a result of their critiques.

Our detailed responses to those comments are inset below (in bold), but, in brief, we have carried out a simulation study that explores the effects of uneven sampling on SpaceMix in a more quantitative way, and used those simulations as the basis for an explicit (and again, more quantitative) comparison with PCA. We have also edited the text for greater clarity.

We hope that the paper is now acceptable for publication.

Thank you for your consideration,
Gideon Bradburd (for all of the authors).

Guest Editor's summary

Both reviewers feel that the paper is much improved but reviewer 2 still thinks that improvements are needed. The paper can be accepted if the authors can satisfy reviewer 2's remaining objections. In particular, the authors should do a more substantive treatment of uneven spatial sampling, the relationship to PCA, and model choice. All of those issues will be important to people who want to use the method.

Overall this paper is good and the topic is of wide enough interest to be appropriate for PLoS Genetics. However, the reviewers, particularly 1 and 2, made several points about the paper that would need to be addressed before it could be accepted. The reviewers noted a number of deficiencies in the paper that will require major revision. The most important points of the reviewers are as follows.

Reviewer 1

I applaud the authors on their consideration of the reviewers' points and the reworking done to the manuscript. I especially like the discussion, and that the points made by reviewers were addressed quite well in it, giving a nice context to the utility of SpaceMix versus existing methods.

- I have only one small change - in line 634 the authors refer to "hierarchical" clustering-based methods and cite Pritchard et al. 2000. I would like to see this changed to "model-based clustering methods".

Fixed.

Reviewer 2

In the resubmitted version, the article is more detailed on many points and globally improved. The authors answered all the questions, or comments. Some answers are convincing, some are not.

- Potential effect of LD: I think the question related to LD is now clearly addressed in the article.
- Uneven Sampling: The effect of uneven sampling is not convincingly answered. The point of the question was to investigate, by the means

of the simulations, how much does the geogenetic map is changed by uneven sampling or subsampling, compared to a full sampling scheme (Figure 1). The figures S3a and S3b show two particular cases of down-sampled grids, without evaluating the mean squared difference (or another measure) of geogenetic coordinates between a fully sampled grid and a downsampled grid.

We appreciate this comment. To better demonstrate the effects of uneven sampling on SpaceMix’s inference, we have simulated a series of datasets across a range of “uneven-ness” of sampling. We ran SpaceMix on each of these simulated scenarios, then Procrustes-transformed the output of the analyses and measured the median distance between the SpaceMix geogenetic map and the true sample configuration for each scenario. We also ran a PCA on each simulated scenario, created a map by plotting PC1 against PC2, did a Procrustes-transform on that map, and measured the median distance of each PC map to the true sample configuration. The description of these simulations and analyses is given in Lines 230-239 of the revised manuscripts, and the results are shown in Figures S3-S10. Briefly, however, we find that PCA biplots become more and more distorted relative to the true sample configuration as uneven-ness of sampling increases. Over the same range of uneven-ness of sampling, SpaceMix becomes more uncertain about the placement of samples in a geogenetic map, but in all cases the SpaceMix geogenetic map is more faithful to the true sample configuration than the PC map.

- Comparisons with TreeMix and PCA: The comparison with Treemix is quite convincing, and shows how taking IBD pattern into account helps finding admixture events. Simulations show that Spacemix is useful in cases where Treemix is not well suited for. I just wish it was included in the main text and figures. The comparison with PCA in FigS2 is very short (only homogeneous migration rates and one admixture event), and it would have been useful to detail how and why the expectations of PC plots and geogenetic maps are different.

We have now included an explicit comparison of the performance of SpaceMix and PCA over a number of simulated scenarios (see comment above). In addition, we have substantially re-written the comparison of SpaceMix and PCA in the Discussion (Lines 644-689) to more explicitly build an intuition for when and why the methods give different answers. We appreciate the point about including the TreeMix comparison in the main body of the text, but we feel that the manuscript is already quite long, and we trust that interested readers will refer to the Supplementary Materials.

- MCMC: I understand that the authors keep the MCMC algorithm to estimate the credibility intervals. I think geogenetic maps are way more informative when credibility intervals are displayed such as in figS3. Especially, if displaying this ellipses makes the visualization too difficult (because of overlaps), it might mean that the geogenetic map is not reliable.

We agree that visualizing the credible ellipses on geogenetic location parameters is very useful. In the current implementation SpaceMix R package, the default visualization function produces these ellipses, and we have also included a vignette in the R package explaining the usefulness of visualizing the uncertainty in the location parameters.

- Model choice: The authors now detail the model choice question. However, l.804 “formally test for the presence of admixture in the sampled dataset.” I still don’t see a formal test for admixture, or model choice procedure. It would have been useful to detail this part.

We have not implemented a model choice procedure, and were merely stating that such a procedure could be used as a general test for, e.g., admixture in the dataset. However, we realize that this was confusingly worded, and we have now re-written these sections (Lines 296-297 and 821-824) for greater clarity.

- Despite the questions that are (in my opinion) not convincingly answered, I think the article presents an interesting tool for population

geneticists, given with caveats to help with the interpretation or over-interpretation of the results.

We hope that the revised manuscript more convincingly answers these concerns.