

Resubmission Cover Letter
Genetics

C. J. Battey,
Peter Ralph,
and Andrew Kern
Friday 6th December, 2019

To the Editor(s) –

We are writing to submit a revised version of our manuscript, “Space is the Place: Effects of Continuous Spatial Structure on Analysis of Population Genetic Data”. We thank the AE and reviewers for constructive comments, which we have addressed in detail, below. The main changes we have made have to do with our choice of model. We have added some more discussion of the biological motivation for the particular model of continuous space we use (with references to the ecological literature, where it appears) to the Methods (around (p. 3, l. 131)), and in the Discussion (at (p. 21, l. 744)). We have included more explanation of some of the more opaque consequences of a more realistic demographic model, around (p. 8, l. 343). Finally, we have added a substantial comparison to the discrete stepping-stone model, most of which is in the Appendix.

We have included a PDF of the revised version as well as a latex diff showing changes relative to the previous submission.

Sincerely,

C. J. Battey, Peter Ralph, and Andrew Kern

Reviewer AE:

The manuscript admirably explores a lot of consequences of isolation-by-distance in the context of a novel model that is easily amenable to forward simulation; however, given that this model may be used in a lot of future studies based on the precedent set here, there is some concern about the model and its support. Reviewers 2 and 3 highlight this in particular (it underlies the main 2 points of reviewer 2's review, and the core of Reviewer 3's comment), and I agree. Whatever can be done to strengthen the standing of this model, and/or connect it to more thoroughly studied models, will be helpful for the manuscript. The concern would be that there are peculiarities of this model that do not generalize well. A new supplemental section or opener to the results section establishing the model more thoroughly would make the strongest response.

This is a good point. There are many ways to go in the quest for demographic realism, and the fundamental question – what regulates population density in real populations – is still very much an open question (or rather, a question with a great many answers). As highlighted below, we have included further discussion of our particular choice of local, density-mediated control on mortality, along with pointers into the literature, at several places in the manuscript; see in particular in the Methods when we introduce the model (p. 3, l. 140), and (p. 21, l. 744).

(AE.1) Line 35: *Also cite Wilkins and Wakeley, Genetics 2002; Wilkins 2004*

Reply: Done. (p. 1, l. 36)

(AE.2) (p. 3, l. 140) *“Such models have been used extensively in ecological modeling but rarely in population genetics”: Detailing these previous uses via citations and elaboration may help alleviate the major concern about the provenance of this model and its unique behaviors (see general comments above and R2 and R3 comments).*

Reply: Good idea – we have added some historical discussion and a few more citations to this section, (p. 3, l. 140). Also see the additional discussion in the Discussion (p. 21, l. 744).

(AE.3) (p. 4, l. 187) *Please describe computation time needed per replicate*

Reply: We have added a figure (Figure 3) and short discussion (p. 9, l. 360) of run times.

(AE.4) (p. 22, l. 806) *I read the acknowledgement to the Hearth and Falling Sky Brewing with a sense of familiarity in feeling of gratitude to my own favorite cafes and breweries, but I it's not a great precedent for Acknowledgements to be filled this way. Please cut.*

Reply: Good point; we have done this.

(AE.5) Figure 5: *Show random-mating expectation*

Reply: This is done.

(AE.6) Figure 4A, S5: *Perhaps more revealing to show on log-log scale?*

Reply: Good suggestion – the SFS in Fig 4A is now on a log-log scale, which shows the slight decrease in low frequency SNPs a little better. We've left Fig S5 as-is.

(AE.7) Figure S3: *Caption seems to be missing detail*

Reply: Thanks for catching this - we have revised this caption to add details including the simulation parameters. (Supplemental Figure S6).

Reviewer 1:

Overall, the authors explore an important but often neglected source of bias that can affect inference in many population-based studies (in medical genetics, evolutionary biology and ecology). This study can be of interest to a broader audience of readership, and I have only minor comments to improve clarity and increase accessibility for readers:

We thank the reviewer for their very constructive comments! Responses follow below:

(1.1) *When neighbourhood size is small (10-100), the mean number of IBS tracts $> 2bp$ (nIBS as in Table S1) is elevated similar to Wright's inbreeding coefficient, but mean of the distribution of pairwise IBS (mean(IBS)) is decreased. What could be the source of this discrepancy? How exactly mean(IBS) was calculated?*

Reply: The mean IBS is simply the mean of the vector of lengths of IBS tracts (p. 6, l. 236). The short answer is that if one splits a chromosome up into more pieces (increasing nIBS), the mean length of those pieces must necessarily be smaller. At low neighborhood sizes, spatially distant individuals tend to have longer coalescence times, leading to more (and hence smaller) IBS tracts. This can be seen in Supplemental Figure S4. Note that skew(IBS) and number of long ($> 1e6$) IBS tracts reflect other aspects of the distribution.

(1.2) *The authors use K to denote both carrying capacity (p. 4, l. 165) and population density (p. 4, l. 168). It might be better to use a different notation for these quantities since carrying capacity is fixed while density is an emergent quantity in the non-Wright-Fisher model. Use of K to denote carrying capacity and density is a bit confusing. For example, on (p. 7, l. 320) it is said that 'the "population density" (K) and "mean lifetime" (L) parameters were the same in all simulations'. Here K seems to indicate carrying capacity rather than density? The latter is an emergent quantity and varies across simulation runs?*

Reply: We agree and have adjusted our language to underscore the point that K is a parameter that controls population density, rather than being equal to it, at (p. 4, l. 168) and (p. 5, l. 205) and (p. 7, l. 320).

(1.3) *Concerning the non-Wright-Fisher model used, it would be helpful to emphasize that some of the parameters are emergent in contrast to Wright-Fisher model. For example, on Page 11, lines 306-308, the author's goal was to look at census size variation and variation in other quantities. This would be better understood if to emphasize that these parameters are emergent properties in the non-Wright-Fisher model used.*

Reply: We have added to the text at the beginning of the results to emphasize that this analysis is necessary because these parameters are emergent rather than fixed (p. 7, l. 319).

(1.4) *(p. 6, l. 255) Perhaps 'Demographic Inference' might better reflect the content of this section.*

Reply: Good suggestion – we have changed the section heading to “demographic inference” (p. 6, l. 255).

(1.5) *(p. 7, l. 286) This sentence with 'Gaussian noise with mean zero and standard deviation 10' is confusing since it was mentioned earlier that the modelled phenotype must vary as human height across Europe, and human height varies 2 standard deviations. Only after reading the whole paragraph it becomes clear that 'standard deviation 10' here refers to unit of height. Please consider rephrasing this sentence.*

Reply: We have revised this sentence to clarify that we aim to produce a variation in mean phenotype of two standard deviations across the landscape (p. 7, l. 286).

(1.6) (p. 7, l. 308) In the sentence, 'We also examined p values for systemic inflation' I think the authors meant 'systematic inflation'.

Reply: Whoops; thanks. Fixed.

(1.7) Please correct the legend in Figure 2: must be 'spatial model' and 'random mating' model.

Reply: Thanks for catching our confusing legend title placement! We have moved "model" to after "spatial" as suggested.

(1.8) Optional: a dashed line in Figure 2 that shows the total carrying capacity of $50 \times 50 \times 5 = 12500$ would be helpful.

Reply: This is a good suggestion, but we decided to not include this as we don't have straightforward expectations for the other parameters shown.

(1.9) (p. 9, l. 369) The phrase 'affect summaries of variation' is better to replace with 'summaries of genetic variation'.

Reply: Done. (p. 9, l. 369)

(1.10) Please add or correct references to supplementary figures: For example, Figure S2 was probably meant to accompany Figure 3A, while Figure S1 Figure 3B, but references in the text are absent. In fact, the first reference is made to Figure S3 on page 15.

Reply: Thanks for catching this, should be fixed now.

(1.11) There are also several typos and errors in the text. For example, (p. 7, l. 320); (p. 19, l. 656).

Reply: Thank you for noting these – they have been corrected.

Reviewer 2:

Batthey et al. use spatially explicit population genetic simulations to analyze the effects of spatial structure on (i) the estimation of key population genetic parameters, in turn used to (ii) make inferences about population history, and on (iii) confounding in genome-wide association studies (GWAS). I liked the paper a lot. It's interesting, well-written and addresses an important question - the effect of spatial population structure on population genetic statistics and inference-and I enjoyed reading it. The most positive aspects were:

1. It nice to actually see spatially explicit simulations and I'm happy that forward simulation is now fast enough that you can do this sort of thing.
2. The paper is very clear and well-written, easy to understand the motivation and most of the details. That's not always the case for this sort of paper.
3. I felt that the section about the effect on GWAS was the most interesting and novel part of the paper and gave me some intuition that I hadn't had before.

I don't have any major criticisms. There were a few aspects that I thought might warrant some additional discussion, and a few specific questions below. The general questions I had after reading it were:

Thanks very much for the encouraging words!

(2.1) *To what extent are any of the results dependent on the exact method of simulation. There are a number of choices about the exact details of the simulations (e.g. the way the overlapping generations are handled, the edge effects and, particularly, the form of Equation 1 - see below). It's not so much that these are non-standard (since I don't think there is a standard) and they all sort of make sense heuristically, and I was left wondering whether these sorts of choices actually make a difference. Do the authors have some thoughts/intuition/results about that? Given that the results in Figure 4 seem quite consistent with expectations, I suspect that on some level it doesn't make much difference but then there are intermediate results like Figure 2 which seem a bit counter-intuitive and I wonder if those aspects depend on the simulation scheme.*

Reply: This is a good point also raised by other reviewers. We have added some more discussion of the choice of demographic model ((p. 3, l. 140) and (p. 21, l. 744)).

(2.2) *Related to the first point, to what extent are the results qualitatively different to those that would be obtained in a stepping-stone model? My interpretation is that they are actually very similar, but I didn't see whether that was explicitly discussed. In some sense, it's still easier to do large simulations in a stepping-stone model so it would be nice to be reassured that that's still ok.*

Reply: We have added an appendix comparing our model to a reverse-time stepping stone simulation on several relevant summary statistics (p. 26, l. 1025). For some of these statistics (e.g., θ_π), the behavior of the stepping stone model becomes more similar (although not identical) to the continuous model as the resolution of the landscape increases, but other statistics are more strongly affected by discretization. We are also curious how all the other analyses we perform would be affected using other simulation schemes, but hope to explore that aspect in future work. Furthermore, discrete-deme coalescent simulations are certainly faster, but come with significant issues either when neighborhood size is lower than the population of a deme, or when demes are small enough that the sample size approaches local N_e .

(2.3) *The source of equation (1) is not obvious to me. I sort of see how it makes sense, but a little but more intuition or a brief derivation or an illuminating either in the main text or the supplement, would be helpful.*

Reply: Additional discussion of the model is now provided in the appendix (see "Demographic model") (p. 28, l. 1065)

(2.4) *The authors use a scaling factor in equation (2) to counteract the increase in fitness of individuals at the edges. Can they provide a figure showing that this is the case. What does "roughly" mean on line 164. Perhaps a heatmap of the fitness of individuals across the grid with and without the scaling factor?*

Reply: Good suggestion – we have added a supplemental figure (Figure S2) to show the distribution of individual fitness across the landscape with and without our edge-scaling approach.

(2.5) *It would be helpful provide the figure showing that generating mutations during the forward simulations in SLiM is equivalent to applying mutations using msprime on pre-generated trees (p. 5, l. 194)? It sounds like this procedure would underestimate the variance in the number of mutations, since you remove the effect of random generation time. Is this effect small?*

Reply: Theory says that any difference here would be extremely small and would affect only the variance, not the mean. Nonetheless, we have added a figure showing sample site frequency spectra generated from a subset of simulations run with SLiM mutations, and then using msprime to apply

mutations to the same tree sequences with our generation-time scaling approach (Figure S3). These approaches yield extremely similar spectra.

(2.6) *Can the authors provide a bit more intuition behind the patterns of variation seen in generation time, census population size, and variance in the number of offspring with respect to neighborhood size seen in Figure 2? For example, it is not obvious to me why the census population size, for example, should decline systematically with respect to neighborhood size. Presumably this isn't just due to the local demographic stochasticity. Could the authors briefly interpret the observed patterns or cite appropriate literature?*

Reply: We agree that this is hard to intuit: there are a lot of factors at work. We have added some more discussion of these phenomena (p. 8, l. 343).

(2.7) *Figure 8D: I am surprised by the extent to which the observed values of $-\log_{10}(p)$ fall below the $y = x$ line. Particularly in the lower right panel for large neighbourhood sizes. I would expect that to be close to panmictic - why are the P-values underdispersed? That seems like a potential bug, or else something weird is going on.*

Reply: We have checked the code to the best of our abilities and did not find a bug causing the underdispersion, but we have switched our visualization in Figure 8D to now show a subset of qqplots as lines, rather than the hexagonal binning over all simulations. This lets us see the density in addition to the range of values, and shows that relatively few points fall below the 1:1 lines. Those that do seem to reflect overcorrection in the regression when using PC coordinates as covariates – the PCA is capturing some information about the spatial genetic variation which itself covaries only weakly with the phenotype, and as a result we see anomalously low $-\log_{10}(p)$ when regressing genotype against phenotype.

(2.8) (p. 20, l. 712) *It might be worth citing (Haworth et al. 2019), who do the proposed test (GWAS for birth location) in UK Biobank to illustrate the population structure.*

Reply: Done - thank you for pointing us to this study. (p. 20, l. 712)

(2.9) *The analysis and discussion around the effect of GWAS is focused on PCA correction. Do mixed models help at all?*

Reply: We are also very interested to know how mixed models perform here, but think that adding a second GWAS method would make this section too large for the current paper. We have added a note to (p. 19, l. 696) specifically citing mixed models as alternate methods that may perform better.

(2.10) *The github link to the code didn't work for me. I assume it will be made public later, but at this point I can't tell whether the code is available/useable.*

Reply: We apologize – we had forgotten to make the repository public. This has been corrected.

Reviewer 3:

The present study deals with a “hot topic” in spatial population genetics. Most inferential and descriptive methods in statistical spatial population genetic rely on a discrete approximation of space and it is not clear what impact this approximation may have when individuals migrate along a continuum instead. Spatial patterns in sampling is also another major issue which is often simply dismissed, mainly because of the paucity of statistical methods to deal with it. This work touches on these important issues in a timely manner.

We are glad you agree this is an important topic to explore!

Although I was enthusiastic about the topic, I was quite disappointed with the core of the study, i.e., the forward-in-time simulation of populations in continuous space. The field has been struggling with this issue for decades – examples of spectacular failures like the Wright-Malecot model (see Felsenstein’s “pain in the torus” article, 1975) or, more recently, the “mugration” or “discrete trait analysis” model in phylodynamics (see De Maio et al. 2015) have probably mostly harmed our research field – that one cannot make the economy of using a sound probabilistic model for generating geo-referenced genetic data. It does not seem to be the case here unfortunately.

We were surprised that the reviewer seems to be saying that local, density-mediated control of mortality doesn’t constitute a “sound probabilistic model”, but hope that our additional discussion of the choice, and citations to the history of this sort of model, help to better motivate our choice.

(3.1) *First, the simulation starts with individuals distributed uniformly at random in space. Is there any indication that the three-step algorithm used here maintains this distribution during the course of evolution? If it does not, then is there any stationary regime and how many generations does one need to wait before reaching it? I do appreciate that the competitive interaction term was introduced in order to avoid seeing the “clumping” of individuals that hampers the Wright-Malecot model. Yet, just because there are no such clusters does not mean that the spatial distribution of individuals reaches a stable regime and that the distribution reached, if any, is reasonable from a biological perspective.*

Reply: This is a natural question that we asked ourselves while developing the simulation using the built-in visualization tools of the SLiM GUI. We have now added a supplementary figure (Figure S1) showing the distribution of individuals in our density-dependant spatial model and a continuous-space Wright-Fisher simulation without density dependence. We encourage the reviewer to run our SLiM recipe using the GUI at different parameter values to see for themselves how quickly equilibrium is attained – we are tempted to include movies of this as supplementary material, but that seems excessive. Also note that we did *not* include competitive interaction to avoid clumping – we included it to make the model biologically realistic (so that, for instance, the population does not grow without bound), and lack of clumping is a *consequence* of this choice. We have clarified this in the text. (p. 3, l. 138)

(3.2) *Second, the demographic process used here involves birth and death of individuals. Does the population survive asymptotically or, like any birth-death process, eventually dies with probability one? In fact, one needs to know a little about the dynamics of the population size to decide whether the corresponding process is reasonable from a biological standpoint.*

Reply: As in the last point, as part of sanity checking our models we monitored the population size over time while with SLiM. All simulations maintained asymptotic populations, and none of the runs we started ended because populations crashed. We agree that the population model will die out eventually with probability 1 (it is a Markov process with a single accessible absorbing state (death), and since it isn’t a branching process, doesn’t grow without bound). However, it will take an extremely long time to hit that state, so that extinction is not a practical consideration. We have added a clarifying note about this. (p. 3, l. 137)

(3.3) *Third, it is not clear what the relationship between the expected lifespan and the probability of survival is. The expected lifespan, L , is first defined as the inverse of the expected number of offspring produced by a parent. The authors also define the probability of survival of a given individual at a given point in space, p_i . Hence, the expected lifespan at a point in space (and time) is the mean of a geometric distribution with parameter p_i , i.e., $1/p_i$. Now, it is far from being obvious what the relationship between these two approaches for defining the expected lifespan actually is.*

Reply: We explained the relationship here, in a new Appendix (‘Demographic Model’; (p. 28, l. 1065)).

(3.4) Also, the web page <https://github.com/petrelharp/spaceness> does not seem to exist so that I was not able to experiment with the forward-in-time generator used here unfortunately.

Reply: We apologize – we had forgotten to make the repository public. Now it is.

(3.5) *All in all, more efforts need to be made here in my opinion to show that the forward-in-time simulations generate sensible outcomes. Sensible in terms of the behavior of the population demography at equilibrium (provided such equilibrium indeed exists) along with that of the spatial distribution of individuals. The authors could provide some guarantee of the good behavior of their model as evidenced from simulations using a broad range of parameter values for generating data. Alternatively, they could elect to use the spatial-Lambda-Fleming-Viot model for their simulations, which, in my opinion would seem the most sensible option given that (1) it is possible to run backward-in-time simulations under this model, thereby saving a lot of computation time and (2) it is a well-studied model with good mathematical and biological properties and (3) it is implemented in a publicly available software program (<https://github.com/jeromekelleher/discsim>).*

Reply: Hopefully, the additional plots and discussion around demographic modeling (notably, the discussion at (p. 21, l. 744)) helps to increase the transparency in this model. We agree that demographic realism is a very important consideration, and due to the often hard-to-intuit nature of spatial demography, simulations need to be carefully sanity checked. (Furthermore, what is “realistic” for one species will not be for another!) We have added an Appendix comparing our model to a stepping-stone (p. 26, l. 1025), because we think these are the most familiar and widely used class of spatial models. We find that many features of our model are well approximated by stepping-stone models, and that for statistics like θ_π the stepping stone model results approach our continuous space model as the number of demes used to describe the landscape increases.

We definitely think that a comparison to the spatial Lambda-Fleming-Viot would be an extremely useful thing to do, for the reasons the reviewer mentions, but this is beyond the scope of the current study. The SLFV model is only known to be an approximation of one particular biologically explicit model (of patchy extinction-recolonization) and is only conjectured to be a good approximation for spatial models more generally. To reiterate, the definition of the SLFV was explicitly motivated by a desire to approximate spatial models, especially Bolker-Pacala models like ours. Therefore, our study does important preliminary work for simulation study of the spatial Lambda-Fleming-Viot, since it works through some of the issues in simulating biologically realistic spatial models, and describes some of the patterns.

(3.6) **Figure 2:** *I do not understand why the neighborhood size varies to the same extent in the random mating model as it does for the spatial model. For the random mating model, I would have expected the neighborhood size to be equal to the census size since all individuals have the same probability of being a parent of any given offspring. From [the paragraph at (p. 5, l. 211)], it is clear that the spatial model would converge to the random mating model when the mean parent-offspring distance tends to infinity only if we were to ignore the impact of range edges. I am thus wondering whether the variation of neighborhood size one observes in Fig 2 for the random mating model is just a consequence of border effects. If that is the case, then the authors should state it clearly and try to justify it from a biological perspective.*

Reply: It’s a good point that neighborhood size, as defined to be “number of possible parents”, should indeed be the census size. However, we are using N_W defined in the same way for both spatial and random mating simulations as a way of comparing spatial to “nonspatial” models keeping other aspects of demography (almost) the same. This is indeed part of what’s going on. We have added some discussion of the census size scaling at (p. 5, l. 211) and (p. 8, l. 343), but haven’t added a detailed discussion of whether or not most species experience decreased habitat quality at range edges, since that seems like it is getting too far into the weeds.

(3.7) (p. 21, l. 730) *“Many more species occur in a middle range of neighborhood sizes between 100 and 1000 - a range in which spatial processes play a minor role in our analyses [...]”. Do the authors think that the spatial processes would still play a minor role when neighborhood sizes exceed 100-1000 if the habitat was larger than that taken in the present simulations? It would also probably be useful to mention that neighborhood sizes given in Table 1 should be compared with extreme caution since the size of the corresponding habitats vary across species. More generally, I suspect that the size of the habitat has a substantial impact on the vast majority of statistics examined in this study. Indeed, the mean parent-offspring distance, which is at the core of the definition of Wright’s neighborhood size, is only small or large relative to the size of the habitat.*

Reply: This is a good point. Wright’s work (Wright 1943) suggests some aspects of genetic variation such as variance in allele frequencies and inbreeding coefficients can be estimated by looking only at what he would later (Wright 1946) call “neighborhood size”, but certainly other aspects like the number of segregating sites will also depend on total landscape size. We now note on (p. 21, l. 721) that we have evaluated only one landscape size, and have added a sentence to the discussion noting that exploration of these patterns in varying landscape sizes is an important avenue for further research (p. 22, l. 785).

(3.8) (p. 22, l. 781) *Please add a reference to Guindon, Guo and Welch (2016). This study clearly shows that population density and dispersal parameters are identifiable and can indeed be estimated in practice under the spatial Lambda-Fleming-Viot model.*

Reply: Done. Thank you for pointing us to this study. (p. 22, l. 781)

Reviewer 4:

The manuscript by Battey et al explores the consequence of a well-known violation to population genetic models: the fact that populations are spatially structured and mate along a geographical cline, rather than randomly. This topic is important, particularly in light of recent work describing how spatially correlated genetic and environmental impacts can confound some population genetic insights, such as positive selection for height in Europe. The analyses and investigations presented here are thorough and sensible, and my comments are primarily intended to broaden accessibility for this interesting topic.

(4.1) Introduction. *The discussion is very clear, articulating the three primary goals of the project: the impact of failing to model spatial population structure on 1) population genetic summary statistics, 2) inference on demographic history from population genetic data, and 3) impacts on GWAS summary statistics. I found the discussion a bit easier to follow than the introduction and would suggest streamlining and introducing the topic a bit more. Since the paper follows the flow described in the discussion, it might help orient readers by introducing these topics in the same order.*

Reply: Thank you for this suggestion. We have slightly revised the introduction and hope it is now clearer; however since we want to cover a little history and motivation for our continuous model vs stepping-stone approaches in the intro it does have a different flow from the discussion.

(4.2) *I agree that most modern work describes structure as discrete populations connected by migration. However, some methods/studies have explicitly modeled spatial structure, e.g. especially in ecology or using methods like dadi (diffusion approximations). Highlighting some examples of previously identified structure not possible to infer without modeling geography would be helpful to contextualize this work.*

Reply: We have expanded our citations of some of the relevant ecology literature (p. 3, l. 131), which we hope helps to contextualize the study better. We haven’t cited dadi, which works with discrete populations.

(4.3) *There is some reference to spatial models using grids (e.g. Rousset 1997). Some additional discussion contextualizing more recent methods like EEMS that also construct demes and model migration through divergence between neighboring demes would be helpful and interesting.*

Reply: We agree that not noting EEMS was an oversight. We have added the most recent EEMS paper to the citations at (p. 2, l. 39), and have added an Appendix giving a more thorough comparison with stepping-stone models.

(4.4) *Demographic modeling. Both approaches tested, stairwayplot and SMC++, are most sensitive to older demographic events, and consequently are very noisy and underestimate effect population sizes, especially in smaller neighborhood sizes. Models that consider haplotype structure are much better suited to this time period. It would be helpful to either 1) discuss the varying time sensitivities of different classes of demographic inference methods and how spatial patterns of genetic variation would influence these inferences, or 2) apply a method of this class (many options, e.g. DoRIS, IBDNe, Tracts, Globetrotter, etc) and show how it performs.*

Reply: We now discuss haplotype methods in the relevant discussion section (p. 18, l. 629). However though these methods should be more accurate for recent events it is not clear that this will improve performance per se. The dips in recent inferred N_e from stairwayplot are not just prediction noise, but actually reflect an underlying genealogy in which terminal branches are shorter than expected from a constant-size random-mating population (see e.g. figure 4A and 5). The interpretation error is that these short branches are generated by spatial structure rather than changes in population size over time – a point also made in the (Mazet *et al.* 2015) paper we discuss in the introduction and discussion.

(4.5) *GWAS mixed models. To what extent can spatial signals (e.g. corner, patchy) be corrected with mixed models, e.g. with PCs and PC-adjusted GRM as in Conomos et al, 2016 using PC-AiR and PC-Relate)? Is patchiness related to dispersal? I'm curious how this relates to the predictive ability of GWAS phenotypes with some spatial association that may or may not be associated with environmental effects.*

Reply: Good question – we are also interested to know how mixed models perform here, but think that to properly test that we would want to change our design to generate phenotypes from simulated genotypes. This would allow us to evaluate false-negatives in addition to false-positives. This is important because, if mixed models do provide stronger control for stratification they are also likely to remove true signals of causal SNPs if those SNPs covary with spatial structure. We now point to these methods explicitly in the discussion (p. 19, l. 696), but think that incorporating that study here would make this paper too long and we plan to explore this timely issue in the future. We also think the PC results are still quite relevant as the method is quite population in current studies.

(4.6) *Code availability. This github link doesn't work, but is important to be able to evaluate for review: <https://github.com/petrelharp/spaceness>.*

Reply: Apologies, it was accidentally set to private. The link should work now (and has been replaced by the primary link, <https://github.com/kern-lab/spaceness>).

(4.7) *Definitions and interpretations. There are quite a large number of metrics discussed in Figure 4, and it's a lot to take in. It might be helpful to have a table with a reminder of what the metric is, its interpretation, and how it is computed.*

Reply: We have included a table describing the summary statistics in Table S1.

(4.8) (p. 4, l. 157) *Notation: "Offspring disperse a Gaussian-distributed distance away from the parent with mean zero and standard deviation σ in both the x and y coordinates. Each offspring is produced with a mate selected randomly from those within distance 3σ , with probability of choosing a neighbor*

at distance x proportional to $\exp(-x^2/2\sigma^2)$." I think x may be overloaded here, or I'm confused. Clarify?

Reply: We have clarified the notation for the second instance (referring to the distance among individuals) to d . (p. 4, l. 157)

(4.9) When introducing the "spatial model" as opposed to this "random model," the more concrete illustration in Figure 1 hasn't yet been referenced, which makes it harder to follow. It would be helpful to introduce this figure with the model. Additionally, when Figure 1 is introduced, the order is from right to left (random, then point, then midpoint). It would be helpful to rearrange the figure to mirror what's in the text.

Reply: We have rearranged the figure as suggested.

(4.10) (p. 7, l. 294) Not sure I follow this example: "Concretely, an individual at position (x, y) in a 50×50 landscape has mean phenotype $100 + 2x/5$."

Reply: We have clarified by switching to "the phenotype p for an individual at location (x, y) is then $p = 100 + 2x/5$ ". (p. 7, l. 294)

(4.11) (p. 7, l. 320) Minor typo (*through* vs *though*): This occurs because, even through the "population density" (K) and "mean lifetime" (L) parameters...

Reply: Thanks, this sentence has been revised and fixed.

(4.12) Define NS abbreviation in Figure 6.

Reply: Done.