June 9, 2016

Prof. dr. ir. A. Stein
Editor in Chief
Spatial Statistics

Ms. Ref. No.: **SPASTA-D-16-00008**
Title: 'Grid-spacing and the quality of abundance maps for species that show spatial autocorrelation and zero-inflation'

Dear Professor Stein,

We are grateful for the very useful review of our manuscript which you recently supplied. We have followed it to revise our work which we now offer for consideration as a revised manuscript. While addressing issues raised by the first reviewer we discovered an error in the procedure which is now corrected. All results were updated.

Reviewer 1

General comments:

Comment: Page 4, eq'n 3: Is it reasonable to assume the $\eta_{B,i}$ and $\eta_{P,i}$ are independent? If $\eta_{B,i}$ is very large, that may indicate that there is some unmeasured factor other than silt or altitude (I cant make any suggestions, as I'm a statistician) which, at this site, increases the chance that there are no organisms at all. If there in fact are organisms, then maybe the expected number would be smaller. This suggests a negative correlation between $\eta_{B,i}$ and $\eta_{P,i}$. In any event, a scatterplot of the appropriate residuals from the fitted model should be done to check that this assumption is OK.

Response: We agree with the reviewer that this assumption need to be checked and we did so. We found a weak correlation between $\eta_{B,i}$ and $\eta_{P,i}$ (0.3). Scatterplot did not reveal any patterns. We have added this new information into the text on page 4.

Comment: Page 5: I found paragraphs 3 - 4 extremely confusing. First, I didn't understand 'The SIBES data are used to calibrate a ZIP model' (line 15). To a statistician, calibration either refers to estimating precise values using imprecise measurements (as in machine calibration), or possibly evaluating the accuracy of estimates.

Response: Here we mean model calibration that involves systematic adjustment of model parameter estimates so that model outputs more accurately reflect initial estimates of the hyperparameters.

Comment: 'The fitted hyperparameters are then used to predict the Bernoulli signal (S_B) and Poisson signal (S_P) at the nodes of a very fine square grid with a spacing of 100 m covering the study area. This prediction grid is extended with 1000 randomly selected validation locations in between the grid-points. The prediction method is simple kriging with an external drift, using silt, the squared values of silt, and altitude as external drift variables (predictors).' If you are only using the fitted hyperparameters, then you can generate (not predict) a set of Bernoulli signals by simulating a realization of the random field $\eta_{B,i}$, taking into account the dependence structure (which is problematic for such a large number of points, since you are not assuming a CAR structure - but I believe there are approximations). I don't see what this has to do with kriging, unless youre using the observations Y_i as well. But then you need to explain a bit about how you estimate the signals π_i corresponding to the observed Y_i. I don't think this would involve kriging; kriging would be for using the $\hat{\pi}_i$ at the observed sites to estimate the π_j at the unobserved sites on the finer grid. Or perhaps youre using the fitted hyperparameters, together with estimates π_i from the SIBES data set (and if so, how were the π_i obtained?) to estimate (using kriging) the π_j at the points on the finer grid?

Response: The reviewer is perfectly right, i.e. that the signals should be simulated, not predicted (by kriging). We now simulated signals on the 100 m grid, conditional on the simulated signals at the SIBES locations by sequential Gaussian simulation, which is well-known in geostatistics. The text is changed at many places to correct this.

Comment: Page 5, line 51: If I understand the previous paragraph correctly, you are now describing the first step in simulating of fields with count data, using the estimated signals. So what do you need this step for? You

just need to go to step 2. Regarding Step 1, if you add white noise with variance $\hat{\tau}_B^2$ to the already-estimated signals on the finer grid, then from equation (4) they will now have a nugget effect of $2\,\hat{\tau}_B^2$

Response: The reviewer again is perfectly right. As said above, originally we predicted signals instead of simulating them, which was wrong. The sentence saying that we added white noise was not correct either. We removed this text, and now signals are simulated by sequential Gaussian simulation, see reply above.

Comment: And I have no idea what you mean by 'The resulting fields are back-transformed by second-order Taylor expansion to give 100 simulated fields of the prevalence parameter π of the Bernoulli distribution'. If you have, in fact, generated the signals, all you need to do is apply the inverse of the link functions in equation (2) to get the fields of π_i and μ_i.

Response We agree with the reviewer, the inverse of the link functions should be used, the sentence is now corrected on p 6. The calculations are adjusted correspondingly.

Comment: Page 8, paragraph 3: Why do you base your estimate of the true MSE on the mean of N = 100 simulations? How accurate is this estimate? A way to address this question is as follows: Denote the random variable in equation (5) by V (the π_i are assumed fixed), and let $\theta = E(V)$. Since computing θ is too complicated, we are going to simulate observations $V_1, \ldots, V_N$ and estimate θ by $\hat{\theta} = (V_1 + \ldots + V_N)/N$. We know that $\hat{\theta} \approx N(\theta, v/N)$ where $v = Var(V)$. If you use the approximations $Var(V) \approx (1/n^2)\sum_{i=1}^{n} Var[(\hat{\pi}_i - \pi_i)^2]$ (actually, this might be an upper bound, as the $\hat{\pi}_i$ are probably positively correlated) and $\hat{\pi}_i \approx N(\pi_i, u_i)$, together with properties of the Chi-squared distribution, then a straightforward calculation gives $Var(V) \approx 2\theta/N$. You can use this to choose N large enough to get a desired relative error.

Response:

We think that the reviewer is referring to 100 grid-samples which we used in prediction with parent-hyperparameters. We agree, that this is not ideal, but unfortunately due to heavy computations involved it is not feasible to re-calculate parameters at each grid spacing. We also agree that sampling with a small grid spacing will give fewer combinations than sampling with a large grid spacing. To account for this we would have to rewrite programming code. Instead, we decided to keep the number of grid-samples large enough at all grid spacing to ensure that all possible combinations are covered.

Minor comments:

Comment: Page 3, line 40: Why do you supplement the grid with the randomly chosen locations? What will you do with them, and why do you describe this here?

Response: The grid was supplemented with the randomly chosen locations in order to calculate small-scale autocorrelation structure. This was designed by Bijleveld et. al (2012), and this information is mentioned here as a part of the methods of data collection.

Comment: Page 4, eq'n 1: The left-hand side should be $P(Y_i = y|\eta i)$.

Response: We agree with the reviewer. Eq. 1 is now corrected to include random effects that are spatially dependent.

Comment: Page 4, line 27: Are the η_i 'residuals of the spatial trend' or error terms which express the spatial dependence?

Response: η_i are residuals of the spatial trend that are spatially dependent.

Comment: Page 5, line 9: 'So note that the aim is not to predict the species abundance counts, but the expected counts, conditional on covariates that define a spatial trend and the observed counts at the sample to be evaluated.' should read 'The objective is to use the observed counts in the sample to estimate, at the desired sites i, the expected counts conditional on the values $\mathbf{x_i}$ of the covariates and the random effects η_i that express spatial dependence.'

Response: Corrected.

Comment: Page 6, line 13: 'We used as summary statistics the fraction of zeroes, the mean of the Poisson counts, and the variance of the Poisson counts. This leads to three difference values per field. These differences values were standardized by dividing through their standard deviation.' What is a 'difference value'? Do you mean that you have three criteria, namely (for each summary statistic) the difference between its value for the field and for the SIBES data? And the plural should not be 'differences values'.

Response: True, we have decided to remove the whole paragraph.

Comment:
Page 9: What are the boxes on the plots? If you have no results for the first grid-spacing, then it should not appear on the plots.

Response: This is explained in the caption of the Figure 2. The boxes on the plots are MSEs obtained with fixed parent-hyperparameters. The first grid-spacing (200 m) with hyperparameters estimated from a sample could not be calculated due to memory constraints. Since we compare fixed parent-hyperparameters with hyperparameters estimated from a sample, we show both on the graph (even when there are no result).

Reviewer 2

General comments:

Comment: First, the authors do not examine the trade-off between sample spacing and prediction accuracy given various levels of residual spatial autocorrelation. They examine only one pseudo reality with a realistic level of spatial autocorrelation for a benthic species, and they find that beyond a spacing of 1600 m, the mean MSE increases quickly. I expect that had the authors looked at a range of residual spatial structures from weakly correlated to strongly correlated, their findings would be similar to Bijleveld et al.' s. So their finding that MSE increases beyond a spacing of 1600 m is not particularly insightful without being able to understand what the trade-off might be (i.e., the trade-off between prediction accuracy and grid spacing) for rare species with weak or strong residual correlation structures. I can't help suspecting that mean MSE increases at a much faster rate beyond a grid-spacing that matches the range of their covariance functions. I do understand that there are significant computational challenges involved in simulating and analyzing the results for multiple species, but I do think that such an analysis would make the study more widely applicable.

Response: We strongly agree with the reviwer on this comment. Unfortunately, as reviewer has pointed out an analysis of multiple species (and hence different correlation strength) is not feasible due to the computing time involved in estimating the hyperparameters. However, we do believe that our study demonstrates a useful methodology that can be readily extended to any species that show similar distributional properties.

Comment: Second, one way to easily alleviate my suspicions that the ideal grid spacing matches or almost matches (i.e., is somewhat below) the range of the covariance function would be to show the values of these parameters in a table. These values could also be published in the body of the text. Perhaps a table of the hyperparameter values used to create the pseudo reality would be a nice addition? i.e. with partial sill, nugget, range, and trend parameters.

Response: We agree that a table of the hyperparameter values used to create the pseudo-reality is useful. Such a table, however, is already presented as Table 1 in Lyashevska et al. (2016) which we closely followed in simulating fields with zero-inflated, spatially autocorrelated count data. Additionally, we have checked the hyperparameters values estimated from a sample (red stars on Figure 2) and found out that the range of the covariance function was alway bigger than a given grid spacing. For example, for the largest grid spacing of 3200 m the range of covariance function was 4760 m.

Specific comments:

Comment: P1, L1: Environmental niche models and even habitat suitability models are special types of species distribution models (SDM's) with a particular purpose and theoretical foundation. I do not consider them to be equivalent to other species-environment models or SDM's. Please revise for greater clarity and descriptive accuracy.

Response: We agree with the reviewer that this sentence should be clarified. We now revised it to indicate that environmental niche models and habitat suitability models are special types of SDM.

Comment: P1, L9: The sentence beginning with, 'Knowledge of the .' should be revised. This is the first time the authors have introduced either accuracy or cost, and both deserve further development. Please separate this sentence into two sentences.

Response: We agree with the reviewer, the sentence was corrected.

Comment:
P2, L11: The authors write that effects of sample size on accuracy was evaluated by few, but they go on to list several studies just in the last little while and I'm sure that this list does not represent an exhaustive search. I think they should start the sentence by saying something like, Several studies evaluated effects of sample size

on SDM accuracy (e.g.,).

Response: We agree with the reviewer, the sentence was corrected.

Comment: P2, L24: Please expand very briefly (1 sentence) on why real datasets are problematic for studying effects of sampling design and sample size on accuracy.

Response: Sentence was expanded.

Comment:
P2, L25: Sentence starting with 'These studies' The word, 'these' in this context is ambiguous and could refer to either the Perner & Schueler etc studies or the studies that use real data. Although it is obvious from the content of the sentence, a better writing style would be to begin the sentence with something like, 'Perner & Schueler (or whomever) recommend an updated approach whereby first ' Or, 'An updated approach is to first simulate a spatial'

Response: Sentence was corrected.

Comment: P3, L53. 'Only few' is redundant. Begin with 'Few studies'

Response: Sentence was corrected.

Comment: P2, L60: Provide a reference or two to back up claim that most species will show the two properties of zero inflation and spatial autocorrelation.

Response: A reference has been added.

Comment: Is there a threshold number of zeros beyond which you would consider a species distribution to be zero inflated? Are there any rules of thumb about this?

Response: As far as we are aware there is no rule of thumb for this, although the general manifestation of zero-inflation is that the incidence of zero counts is greater than expected for the Poisson distribution.

Comment: Why was the grid square supplemented by exactly 578 locations? Is there anything significant about the number 578?

Response: We agree with the reviewer that it is unclear from the present description how the numbers were derived. We now explain in the text that the total sample size of 4029 locations can be calculated as a sum of permanent locations (3451) and supplemented locations (578). The number 578 was taken as it was practically achievable for a given area. This was added to the text on page 3.

Comment: P6, L142: What is the text in the parenthesis about? (remind that the aim was to predict the model parameters..)

Response: The text in the parenthesis was removed.

Comment: P9: L231: You write: "The difference in the sampling distributions of MSE reflects the contribution of the uncertainty about the hyperparameters to uncertainty about the predictions due to sampling errors in the estimated hyperparameters" I don't follow what the authors mean by this sentence. Please clarify.

Response: This is explained in step 4 on p 8 of 'Overview of evaluation method'.

Comment: Please go through the paper carefully and get rid of all superfluous 'the's' A good example sentence is as the one at L78 on P2. Please also report what you did in past tense.

Response: We thank reviewer for pointing this out. We have now reviewed the final version and corrected all grammatical issues.

Comment: The map is really impossible to read in greyscale and it would be better to use squares rather than circles in the legend b/c these are grids.

Response: We thank the reviewer for pointing this out. To adress this issue we created a palette (using bpy.colors(), package sp) with colors that differ in brightness/lightness as these are the only aspects that are conserved in greyscale. Circles in the legend were replaced by squares to reflect the fact that these are grids.