# 1 Responses to Editor's comments

- The reviewers and I appreciate the work you have accomplished. Based on the reviews, we are willing to consider a revised version for publication in the journal, assuming that you are able to modify the manuscript according to the recommendations.

    **Response:** Thank you.

1. I would like to see a more conventional presentation of a new statistical method: state what kind of data needs to be modeled, present the model and estimation method, and then present the simulations and/or examples. This would mean placing your "Simulations" section after the section that should be labeled "Estimation" (rather than "Inferring..."), so that it comes as an evaluation of the method you have just explained. Actually what you have stated is a maximum likelihood estimation method and you have not addressed inference, i.e. how to decide uncertainty or significance.

    **Response:** I've re-ordered the material as you suggested (model, then estimation method, then examples). I have also made a clearer distinction between parameter estimation and hypothesis testing throughout the manuscript (e.g. lines 185-191 discuss point estimates and inferential statistics separately).

2. On page two please go more directly to the method. As written it sounds like the issue is somehow to adapt partial correlations to non-Gaussian data, and I think this is a rather indirect and confusing explanation. What you have are multivariate 0/1 data and what you are presenting is a joint probability model for such data. This would provide a more direct path toward explaining to your contribution.

1

<sup>22</sup> **Response:** I now treat the partial correlations as an approximation to the

<sup>23</sup> discrete network (lines 141-154), rather than treating the discrete network as a

<sup>24</sup> generalization of partial correlations and introduce the method as a way to model

<sup>25</sup> the probability of multivariate presence-absence data under different parameter

<sup>26</sup> sets (lines 66-111).

<sup>27</sup> 3. Most of lines 84-96 are redundant and could be removed. This would save you space

<sup>28</sup> (and the reviewers may or may not have realized you were hewing to the 20-page limit

<sup>29</sup> for Statistical Reports).

<sup>30</sup> **Response:** Removed.

<sup>31</sup> 4. Please tread lightly on physics analogies and language and where possible use the

<sup>32</sup> corresponding statistical language. For example, the "partition function" is a

<sup>33</sup> normalizing constant.

<sup>34</sup> **Response:** I now refer to the partition function as a normalizing constant (line

<sup>35</sup> 386) and no longer discuss the energy function.

<sup>36</sup> 5. I think you could use some better paragraph choices, but I am not sure since you have

<sup>37</sup> no paragraph indentations. Please do not do that again.

<sup>38</sup> **Response:** Indentations have been added. It may be worth updating the

<sup>39</sup> Society's page on "Preparing LaTeX manuscripts for ESA journals" (e.g. to

<sup>40</sup> suggest adding \setlength{\parindent}{1cm} at the beginning of the

<sup>41</sup> manuscript and \setlength{\parindent}{0cm} before the references and figure

<sup>42</sup> captions) so that authors know to do this in the future.

43  6. Provide distinct labels for the two correlation-based methods to carry through the

44      paper (line 137).

45      **Response:** The distinctions between the different methods (including the ones

46      added in response to reviewer comments) should now be clear throughout the

47      paper (e.g. lines 192-203).

48  7. 24 simulations for a scenario is not very much.

49      **Response:** I now perform 50 simulations per landscape size (150 total). The

50      bootstrap resampling performed in Appendix 3 indicates that the estimates

51      reported in the Results section are stable with this many replicates (e.g. the 95%

52      bootstrap CI indicates that Markov networks performed between 14 and 18

53      percent better than the next best method, which is a narrow enough range to

54      have some confidence in the results).

55  8. The interpretation of your more complicated simulation and estimation exercise is

56      fairly convoluted. Normally one would like to see a set of scenarios with fixed

57      parameter values, but you are integrating over arbitrary distributions of parameter

58      values in a way that a Bayesian might appreciate but really makes it difficult to know

59      if the relative performance of the methods depends on the parameters. You could

60      provide more elaborate results in the appendix. Please attempt to disentangle some of

61      the results to provide greater understanding.

62      **Response:** The simulation and estimation procedure should now be much

63      clearer. It could be thought of as simulating 150 ecologists that each collect data

64      from a different archipelago, and then asking whether their results are consistent

3

65     with the underlying biology and with one another. There is no integration over

66     any distribution, nor is there anything particularly Bayesian about the approach.

67     The bootstrapping analysis in Appendix 3 in particular produces frequentist

68     confidence intervals showing that the relative performance of the different

69     methods is insensitive to changes in parameter values.

70  9. Your use of a prior distribution for alpha and beta is a serious concern. It appears to

71     be highly informative towards the range of true parameters that you drew from. On

72     the other hand, I am not sure what you did because you stated you used a logistic

73     distribution, but the paper you cited discusses use of t distributions for parameters of

74     logistic regression. A potential drawback of your method is the number of parameters

75     to estimate, so use of an informed prior is troubling. You must give results without

76     this prior or present a very solid justification for using it or come up with another

77     approach. Given that the reviewers and I find your method potentially valuable, then

78     even if it turns out that estimation is difficult without informative priors, that would

79     be an important result for readers to see clearly.

80     **Response:** I've removed the citation, since I agree it wasn't clear. That paper

81     justified the $t$ prior in part as an approximation to the logistic, but that wasn't

82     clear from my previous submission. The new discussion on lines 128-137 is more

83     directly on-point. I've also substantially weakened the regularizer substantially

84     (well beyond what was done in the Faisal et al. paper pointed out by Reviewer 2)

85     and clarified its likely influence on the results.

86 10. At lines 163-165, please explain why maximum likelihood estimates should give a

87     model with "exactly the observed occurence frequencies and co-occurrence frequencies".

88    **Response:** This is now clarified (lines 105-107).

89    11. Lines 174-175. Typically statisticians talk about the two steps of estimation followed

90    by inference for a given method. In this sentence you talk about "fitting" (estimation),

91    by which you mean your method, and then "inferring", by which you actually mean

92    other methods for comparison. And you have not actually stated how you would do

93    inference with the new method (presumably likelihood ratios or such). Please re-write.

94    **Response:** This distinction should now be clearer throughout the manuscript.

95    The new manuscript now discusses inferential statistics in two places, using

96    approximate confidence intervals derived from the Hessian matrix (lines 168-169)

97    and the bootstrap distribution Appendix 4.

98    12. Lines 178-181 and 185-190 need explaining, with rationale and sufficient details for a

99    reader to repeat the work. This can go in the appendix if you prefer.

100    **Response:** This material is now explained in the first three Appendices, along

101    with code that a reader could use to replicate the analyses exactly. The

102    regularizer discussed on lines 178-181 of the old text is now chosen automatically

103    by the `corpcor` package (lines 146-148 of the new text). The description of the

104    "Pairs" model's *Z*-score (lines 157-158) should now be clearer as well.

105    13. I believe I followed lines 194-202 but they will be confusing for most readers. Again

106    you can use space in the Appendix to clarify.

107    **Response:** The section on model evaluation has been rewritten, and should be

108    much clearer (see lines 170-181). Additionally, the first three Appendices include

109    heavily-commented code for replicating the analysis.

5

14. In lines 207-211, it is unclear for each method whether you are reporting a percentage of "significant" results or simply the percentage of correct signs (positive or negative). This is confusing because on the one hand the randomization methods aim explicitly for significance, but on the other hand you haven't stated how you would do a hypothesis test for your method (although clearly you could). As a result I am not sure if you are comparing apples to oranges here.

    **Response:** I now discuss the sign of the point estimate separately from its statistical significance. See lines 183-191.

15. Please consider carefully if "Markov networks" is the label you want to provide for these models. At first glance it seems a trivial or ironic usage of the concept, since by assumption all vertices (species) in the graph are connected, so the Markovian notion of depending only on graph neighbors carries no impact. I am not sure if there is a more precise or otherwise better label, but I ask you to give this some consideration.

    **Response:** I've considered this, and I can't find a better name from the literature. "Markov random field" still has "Markov" in the name; the term "Ising model" is not helpful to non-physicists and often implies a grid-structured network; "Boltzmann machine" is not helpful outside of one area of machine learning and often implies the inclusion of latent random variables, which are absent here. My reading of the literature indicates that the terms "Markov network" and "Markov random field" are routinely used to describe fully-connected graphs (e.g. "Learning Factor Graphs in Polynomial Time and Sample Complexity", by Pieter Abbeel, Daphne Koller, and Andrew Y. Ng, bottom of page 1746).

6

133 16. Section 1 of the Appendix needs explicit equations

134 **Response:** This whole Appendix has been replaced.

135 17. Your simulations seems to have a curious feature: you are drawing from distributions

136 conditioned on having all species present. It is conceivable that this could introduce

137 bias to your results. The conditional distribution you are actually sampling from may

138 have different correlations (and hence parameter sampling distributions) than the full

139 distribution. Please address this concern.

140 **Response:** I no longer condition on the number of species; some landscapes now

141 have fewer than 20 species.

142 18. Both reviewers note that you have missed substantial relevant literature.

143 **Response:** This literature is now addressed more directly (see responses to

144 reviewers below).

## 145 Responses to Reviewer 1's comments

146 • I very much liked the manuscript: it has a clear point to make. As somebody who has

147 used covariance matrices to make inferences about interactions, I agree that their

148 interpretation can be spurious, for reasons pointed out by the author.

149 **Response:** Thank you.

150 • My first major critical comment is that - if I understood correctly - the data are

151 generated by the same Markov networks that are fitted to the data as one of the

152 competing approaches. Thus it is not at all surprising that the Markov network

7

¹⁵³ approach works best in recovering the correct parameter values - the comparison

¹⁵⁴ between the approaches is not fair. A more convincing demonstration that the Markov

¹⁵⁵ network approach is able to infer interactions (not just associations) would be to use a

¹⁵⁶ more mechanistic model to generate the data. For example, a time series model of

¹⁵⁷ interactive species where the interactions would happen preferably at the individual

¹⁵⁸ level (as they do in nature; predators eat prey, etc.). Or at least a population level

¹⁵⁹ model where the current abundances of the species influence their growth rates.

¹⁶⁰ **Response:** I now simulate my landscapes from a time series model where each

¹⁶¹ species' birth and death rates depend on the abundances of the other species on

¹⁶² the landscape (see lines 118-124 and Appendix 2).

¹⁶³ • My second major comment is that the ms seems to ignore a large amount of recent

¹⁶⁴ work in community ecology that has been focused on estimating co-occurrence

¹⁶⁵ [examples of such paper include Pollock, L.J. et al. (2014) Understanding

¹⁶⁶ co-occurrence by modelling species simultaneously with a Joint Species Distribution

¹⁶⁷ Model (JSDM). Methods in Ecology and Evolution 5, 397-406. Clark, J.S. et al. (2014)

¹⁶⁸ More than the sum of the parts: forest climate response from Joint Species

¹⁶⁹ Distribution Models. Ecological Applications 24, 990-999. Ovaskainen, O., Hottola, J.

¹⁷⁰ and Siitonen, J. 2010. Modeling species co-occurrence by multivariate logistic

¹⁷¹ regression generates new hypotheses on fungal interactions. Ecology 91, 2514-2521].

¹⁷² Thus I think the author should do a better job in describing the state of the art.

¹⁷³ **Response:** I now include a JSDM's correlation matrix as a competing estimator

¹⁷⁴ (lines 159-164, Figure 3) and reference Pollock et al. (2014) in a number of places

¹⁷⁵ (lines 33, 64, 162-163, and 263).

176 • My third major comment relates to the fact that all partial covariances can be derived

177 simply by inverting the covariance matrix (which I think is worth mentioning in the

178 paper, now I think it is mentioned only in the Appendix). As many of the earlier

179 methods (e.g. listed above) provide the full posterior distribution of the covariance

180 matrix, it seems highly feasible to link to (some of) those earlier methods by using

181 them to estimate the posterior distribution of the covariance matrix, and then inverting

182 the covariance matrix. That would yield the full posterior distribution of the partial

183 covariances, without the technical problem mentioned in the ms. Such comparison

184 would seem better connected to recent literature than the regularized ridge regression.

185 **Response:** I now include partial correlations calculated from the JSDM as well,

186 inverting 1000 samples from the posterior distribution over correlation matrices

187 (lines 163-164, Figure 3).

188 • I further note that while partial covariances may not have been used much in

189 community ecology, they have been used e.g. in gene and protein interaction networks -

190 I think the author should make links to this literature.

191 **Response:** This is now noted on lines 145-146.

## 192 Responses to Reviewer 2's comments

193 • The problem of species interactions is an important one, and Markov networks do

194 indeed present a very promising avenue of research. The technical competency of the

195 work appears strong.

196 **Response:** Thank you.

1) I'd really like to see a real data example in here. The simulations are fine at an introductory level. But I think adding a real example would help illustrate the true worth of the methodology. A real data application (which doesn't have to be large) could show how using Markov networks produces a result which tallies well with our current understanding of interactions between species - in much the same way as the hypothetical example is presented.

**Response:** I agree that this would be an important step, however I do not have access to any co-occurrence data sets where the signs and strengths of the interactions between species had been independently estimated. I also don't have enough space to fit such an analysis in the short format for the journal. Fortunately, the Azaele et al. (2010) paper I cite already includes an application of the method to a real data set, which should make it less important to do so here.

2) I do think there are some key references missing. The following are but two examples of inferring species interaction networks from data: [Aderhold et a. 2012, *Ecological informatics* and Faisal et al. 2010, *Ecological Informatics*]. The reader needs to know how the current paper compares with the above, and other similar works.

**Response:** The manuscript now cites Faisal et al. (2010) on lines 142 and 151, and makes comparisons two several of the methods proposed in that paper (regularized generalized linear models and gaussian models [in the form of partial correlations]), in addition to the Markov network, a null model, the correlation matrix, and two versions of a latent variable model (see responses to Reviewer 1 above). I did not find the spaially-explicit methods in Aderhold et al. (2012) to

220    be as relevant for the kinds of spatially-implicit model comparisons being made

221    in the paper, although it is possible that I misunderstood something.

222    • Of course, it is also perhaps worth noting that Markov networks are common for

223    models of spatial correlation in Ecology.

224    **Response:** This is now noted on line 53.

225  3) The beta symbol in Fig 2A should be beta_12; this then helps connect the specific

226    example in the figure with the formula at the top of page 5.

227    **Response:** The symbol has been modified as suggested.

228  4) Lines 172-173; the value of 0.2 for the ridge parameter should be justified.

229    **Response:** The paper no longer uses a pre-specified penalty. Instead, it relies

230    on the `corpcor` package to determine the appropriate shrinkage for each matrix

231    (line 148).

232  5) Some comments on sensitivity would be welcome - especially with regard to Figure 3.

233    We really should see something more on the uncertainty (or stability) of the estimated

234    networks.

235    **Response:** The revised manuscript includes several new analyses to address this

236    issue. First, the uncertainty in Figure 3's results is estimated with bootstrap

237    resampling of the 50 simulated landscapes in each size class (Appendix 3).

238    Second, the sensitivity to sampling error is estimated in Appendix 4, using

239    bootstrap resampling of the sites in one landscape. Third, the manuscript now

240    mentions the approximate confidence intervals that can be produced using the

241    Hessian matrix (lines 168-169).