This is a preprint. A revised version of this article has been accepted for publication in Ecography:

Article type: Forum paper

Title: Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments?

Authors: Damaris Zurell*, Laura J. Pollock², Wilfried Thuiller²

Affiliations:
1 Swiss Federal Research Institute WSL, Dept. Landscape Dynamics, Birmensdorf, Switzerland
2 Univ. Grenoble Alpes, CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France

* Corresponding author: damaris.zurell@wsl.ch, ORCID: 0000-0002-4628-3558
Abstract:

Whether species interactions influence species response to environment and species ranges has always been a central question in Ecology. Joint species distribution models (JSDMs) simultaneously model the species-environment relationships of multiple species and the residual correlation between these species. These residual correlations are assumed to depict whether species co-occur less or more often than expected by the modelled species-environment relationships, which could ultimately be attributed to species interactions, or hidden environmental information. Here, we propose to specifically test the capacity of JSDMs to detect species interactions, at different aggregation scales of data. Using a recently published point-process model, we simulated equilibrium co-occurrence patterns of species pairs by varying the strength and type of interactions (e.g. competition, predator-prey, mutualism) as well as the prevalence of the interacting species in homogeneous environments (assuming the environment does not influence the species responses and co-occurrence). Then, we fitted JSDMs without environmental predictors, and compared the estimated residual correlations against the known interaction coefficients. JSDMs detected competition and mutualism well, but failed with predator-prey interactions. For the latter, JSDMs predicted both negative and positive residual correlations for these kinds of interactions, depending on the prevalence of the interacting species. Interestingly, the degree of residual correlation estimated by JSDMs was strongly influenced by species’ prevalence and can thus not be translated to interaction strength. At increasingly coarser data resolution, the signals of negative and positive interactions became indiscernible by JSDMs, but – reassuringly – were rarely confounded. Our simple theoretical exercise demonstrates that when species are modelled at the scale at which
species interactions occur, JSDMs are able to depict symmetric interactions (+/+, −/−) but fail for asymmetric interactions between trophic levels. However, when JSDMs are applied at coarse resolutions, they are unlikely to give insights on the underlying biotic interaction mechanisms.
**Introduction**

Determining the factors that influence species’ ranges is a prerequisite for deriving reliable biodiversity scenarios (Urban, et al. 2016). The importance of abiotic factors has long been recognized (Chase and Leibold 2003, Pulliam 2000, Soberón and Nakamura 2009), and correlative species distribution models (SDMs) that relate species’ occurrence to prevailing environmental factors have become the most widely used tools in biogeography (Elith and Leathwick 2009, Guisan and Thuiller 2005, Guisan and Zimmermann 2000). However, it is now generally accepted that interspecific interactions likely influence species’ distributions at larger spatial and temporal scales than previously thought (Blois, et al. 2013, Svenning, et al. 2014). Ignoring these interactions could thus hamper the ability of SDMs to predict current and future species ranges and communities (Leathwick and Austin 2001, Meier, et al. 2010, Zurell, et al. 2016a). Hence, over the last couple of years, several approaches have been discussed or proposed to incorporate interspecific interactions between multiple species into distribution models (Kissling, et al. 2012, Pellissier, et al. 2013, Thuiller, et al. 2013, Wisz, et al. 2013).

patterns of co-occurrence that are unexplained by the environmental information given
to the model. The temptation is then to attribute this residual correlation to biotic
interactions, such as competition and facilitation. Although this residual correlation can
reflect intuitive ecological (Latimer, et al. 2009, Ovaskainen, et al. 2010) or evolutionary
processes (Pollock, et al. 2015), there could also be many non-biological explanations
such as missing environmental variables or poor model fit. Theoretically, JSDMs should
predict positive residual correlation between a pair of species if these co-occur more
often than expected given the environment (or by chance in a homogenous
environment) and predict negative residual correlation when the species co-occur less
often than expected given the environment (or by chance in a homogenous
environment). However, to date, no study exists that explicitly tests JSDM’s ability to
identify the signal and the strength of species interactions and the effects of
complicating factors such as species’ prevalence and spatial scale. Also, the limits of
JSDMs to detect asymmetric (positive-negative) interactions such as predator-prey
relationships have not been tested. For example, a recent study by Araújo and Rozenfeld
(2014) showed that range overlaps of species pairs resulting from predator-prey
relationships could resemble both co-occurrence patterns of competition (-/-) and
mutualism (+/+). The authors also showed that at increasingly coarser resolution, the
signal of negative interactions gets lost earlier than the signal of positive interactions.

Here, we build on the point-process model of Araújo and Rozenfeld (2014), and estimate
JSDMs from the co-occurrence patterns resulting from different types of interspecific
interactions (-/-, +/-, +/+) in homogeneous environments and from different species’
prevalence levels. That way no other reason like history or missing environmental
variables could explain the residual patterns. With that, we aim to answer the questions
of whether JSDMs can detect different interspecific interactions ranging from negative to
positive, how JSDMs cope with positive-negative interactions such as predator-prey
relationships, and how the ability to detect interactions is influenced by the resolution of
the data used for modelling (i.e. scale-dependence).

Methods

Simulating co-occurrence patterns

We used the point-process model developed by Araújo and Rozenfeld (2014) to simulate
co-occurrence of species at steady state across all possible interaction types (+/+; +/−, −/
+/−, +/0, −/0) and all possible combinations of interaction strengths (0 ≤ |ls| ≤ 1). If there
is no interaction between species, then the expected probability of co-occurrence P(A
and B) of a species pair is simply given by the product of their prevalence P(A) x P(B).
This corresponds to the null expectation of co-occurrence. With interspecific interaction
present, the probability of co-occurrence is a function of species’ prevalence and the
strength of their interaction, which could be repulsive or attractive. Mutualistic
interactions (+/+ will cause the species to co-occur more often than expected under the
null model, whereas competitive interactions (−/−) will cause them to co-occur less often
than expected. In case of predator-prey relationships (+/−), both positive and negative
interactions will cause deviation from the null model. Whether this results in higher or
lower co-occurrence than expected by the null model, depends on both the relative
strength of positive and negative interactions and on the prevalence of the species. For a
more detailed description of the model formulation, please see Araújo and Rozenfeld
(2014). For simplicity, the environment is assumed homogenous and does not influence
species’ ranges and prevalence.
Here, we simulated co-occurrence patterns of two species A and B in homogeneous environments of 100x100 cells for all combinations of potential interactions. Species’ occurrence in space was random without any spatial autocorrelation. Araújo and Rozenfeld (2014) showed that the co-occurrence patterns and their scale dependence were comparable between spatially correlated and spatially uncorrelated landscapes. We varied the interaction coefficients $I$ from -1 to 1 in 0.1 increments. This resulted in 441 simulations per prevalence level. As an extension to previous results (Araújo and Rozenfeld 2014), we systematically varied prevalence levels. Specifically, in the first set of five scenarios, species A and B had equal prevalence of varying magnitude [0.1, 0.2, 0.3, 0.4, 0.5]. In a second set of two scenarios, species A and B had unequal prevalence [A=0.1 and B=0.5; A=0.2 and B=0.4].

To test for scale dependence in co-occurrence patterns and in the estimation of JSDMs, we aggregated the original landscape grid to two increasingly coarser resolutions by aggregating 2x2 cells and 4x4 cells, respectively. In these cases, species interactions still happen at the initial resolutions, but JSDMs are fitted at coarser resolution mimicking large scale and coarse resolution data such as atlas data (Gotelli, et al. 2010, Zurell, et al. 2016b), or continental and global datasets (Jetz, et al. 2012, McGill 2010).

All simulations were carried out in R (R Core Team 2015), and the code is provided in the Supplementary Material Appendix 1.

**Estimating joint species distribution models**

The spatial distributions of species A and B as simulated by the point-process model over the full biotic interaction space served as input for JSDMs. We fitted JSDMs using the code provided by Pollock, et al. (2014). These jointly estimate the occurrence
probability of multiple species with a hierarchical multivariate probit regression, and
the residual correlation between those species' presences by means of an unstructured
covariance matrix. As we were mainly interested in the ability of JSDMs to detect
underlying interactions, we assumed homogeneous environments and fitted the
environmental response with an intercept only model that controls for prevalence. The
JSDMs were run with JAGS from within R (R Core Team 2015) with 3 chains. As the
simulated data were very simple and without noise, for most combinations of
interaction strengths (-1 < I < +1) a comparably low number of 10000 iterations with a
burn-in of 5000 and a thinning rate of 20, were sufficient for reaching convergence with
rhat values below 1.1 for all parameters. For extreme interactions strengths (with an
interaction coefficient I of -1 or +1), convergence was not achieved under 10000
iterations and we subsequently tested stability of parameters values for increasing
numbers of iterations. The residual correlation between species' presences was given by
the mean of the posterior distribution and was judged as significant when the range
between the 5% and 95% posterior quantiles did not include zero.

Results

Different prevalence levels produced different co-occurrence patterns across the biotic
interaction space (Fig. 1a-f). The co-occurrence index was defined as the number of cells
where both species occur together divided by the total number of occupied cells. The
null probability of co-occurrence P(A and B) is given by the product of species'
prevalence P(A) x P(B), and thus the co-occurrence value corresponding to the null
expectation is given by ( P(A) x P(B) ) / ( (P(A) + P(B) – P(A) x P(B)) ) where the dividend
is the proportion of cells occupied by both species together and the divisor is the overall
proportion of occupied cells. Hence, the exact co-occurrence value corresponding to the
null expectation, where both species co-occur simply by chance, varied with prevalence (Fig. 1a-f). Co-occurrence was always lower than the null expectation for competitive interactions (-/-) and for amensalism (-/0) and always higher than the null expectation for mutualism (+/+ and commensalism (+/0). However, the exact asymmetric interaction coefficients (+/-), at which co-occurrence was not different from the null expectation, strongly depended on prevalence (of the more common species of the species pair). When prevalence was low in both species, then co-occurrence was higher than the null expectation for most combinations of positive-negative (+/-) interactions (Fig. 1b), while co-occurrence was mostly lower than the null expectation if at least one species in the predator-prey relationship was more common (Fig. 1f; Supplementary material Appendix 1, Fig. A1).

JSDMs correctly assigned negative residual correlations to competitive interactions, and correctly assigned positive residual correlations to mutualistic interactions. Specifically, the null expectation of co-occurrence determined the residual correlation estimated by JSDMs. Negative residual correlations were estimated when species co-occurred less often than the null expectation (meaning the co-occurrence value corresponding to the case when both species co-occur only by chance), and positive residual correlations when the species co-occurred more often than the null expectation. However, the magnitude of residual correlations varied across scenarios with different prevalence of the interacting species (Fig. 1g-k). For example, for two species A and B with an intermediate and symmetric negative interaction ($I_A = I_B = -0.5$), JSDMs estimated a residual correlation of approx. -0.19 if both species had a prevalence of 0.1, and a residual correlation of approx. -0.71 if both species had a prevalence of 0.5. These results were robust, as we obtained convergence with rhat values below 1.1 for most
combinations of interaction strengths except for the extreme interaction coefficients of $|I|=1.0$ (Fig. A2). For these extreme cases, convergence was achieved for larger numbers of iterations, whereby the model parameters were stable across different numbers of iterations (Fig. A3). Also, prevalence levels were correctly predicted in all cases.

JSDMs were not able to separate predator-prey relationships from competitive or mutualistic interactions, and assigned both negative and positive residual correlations to such asymmetric (+/-) interactions (Fig. 1). Here, residual correlations from JSDMs exhibited the same sensitivity to prevalence as co-occurrence patterns, meaning that the exact asymmetric interactions coefficients (+/-) that divided positive from negative residual correlations estimated by JSDMs depended on prevalence (Fig. 1g-k; Fig. A1).

Last, JSDMs had difficulties at depicting the underlying co-variation between species at increasingly coarser resolution of the data (Fig. 2a-f). As shown by Araújo and Rozenfeld (2014), co-occurrence patterns were scale-dependent and especially negative interactions became indiscernible at increasingly coarser resolution of the data. This scale dependency of co-occurrence patterns led to more insignificant residual correlation estimates in JSDMs (meaning that the parameters credible interval between the 5% and 95% quantiles of the posterior distribution included zero; Fig. 2 g-j).

Generally, chains mixed well and parameter convergence was well achieved. Thereby, more prevalent species were more sensitive to scaling (Fig. 3, Fig. A4-A10).

Furthermore, for competitive interactions, increasingly coarser resolution of the data led to lower estimates of absolute residual correlation than at the original resolution (where the interaction takes place) while this effect of smaller estimates was much less pronounced for facilitative interactions (Figs. 2-3). For intermediate prevalence levels, negative and positive correlations between species only rarely got confounded (Fig. 2j).
However, for rare species, the probability of confounding negative and positive residual correlations seemed to increase at increasingly coarser resolution of the data (Fig. 3a-b).

**Discussion**

In this paper, we propose to systematically test the capacity of JSDMs to retrieve species interactions from co-occurrence patterns in the case of simple bi-partite interactions within a homogenous environment, and test whether this capacity was influenced by the resolution at which species were modelled. The key results are that (i) JSDMs reliably predict negative residual correlations for competitive interactions and positive residual correlations for mutualistic interactions. However, the magnitude of residual correlation does not directly relate to interaction strength because it is mostly driven by species prevalence. (ii) Positive-negative interactions such as predator-prey relationships can be problematic for model interpretation because JSDMs cannot disentangle these from competitive and mutualistic interactions. (iii) At increasingly coarser spatial resolution of the data, both the signals of negative and positive interactions become indiscernible by JSDMs, but reassuringly the signals rarely were confounded. Overall, our results have important implications for the interpretation of JSDMs, and pinpoint important aspects that will need further model development. At the same time, we want to stress that these results were obtained under idealized conditions, and real-world applications will be complicated by even more factors. If JSDMs fail to detect interspecific interactions in very simple cases such as the one used here, they is no reason why they should work better at detecting interactions (occurring within local communities) in more complex real world applications.
First, we tested JSDMs’ ability to detect negative and positive interactions from co-
occurrence data. And indeed, JSDMs reliably detected competitive and mutualistic
interactions in our simulated data, which is promising. However, interpretation of model
residuals as interaction coefficients proved non-trivial. Importantly, prevalence strongly
affected the magnitude of residual correlations estimated by JSDMs. This implies that
residual correlations from JSDMs do not directly relate to nor can be interpreted as
interaction strength, and also they cannot be compared across different species (with
different prevalence levels). In real-world applications, this may be further confounded
by the species’ environmental response, detection bias, and by (direct and indirect)
interactions between multiple species, which should be further tested.

Second, we asked whether JSDMs could distinguish competitive and mutualistic
interactions from asymmetric interactions such as predator-prey relationships, which is
important when dealing with taxa from multiple trophic levels. As shown previously by
Araújo and Rozenfeld (2014), asymmetric positive-negative interactions can result in
coopercurrence patterns that could be either higher or lower than expected by chance.
Here, we showed that this also depends on interaction strength between species A and B
and their prevalence. Specifically, predator-prey relationships are more likely to
produce positive residual correlations when prevalence of both species is low. By
contrast, when prevalence of at least one species is high, then predator-prey
relationships are more like to produce negative residual correlations in JSDMs. As a
consequence, JSDMs are not able to tease apart predator-prey relationships from
competitive or mutualistic interactions, which has been discussed previously but never
been shown explicitly (Morales-Castilla, et al. 2015, Warton, et al. 2015). These results
suggest that a useful next step would be to incorporate directional conditionality into
future joint species distribution modelling, extending their capacity beyond pairwise symmetric correlations (Hui 2016, Pollock, et al. 2014, Warton, et al. 2015). Recent examples are beginning to consider asymmetric and more complex interactions for small datasets (e.g. Harris 2016, with up to 20 species). These advances would be extremely useful not only for distinguishing predator-prey relationships but also for detecting asymmetric interaction strengths between competing or facilitating species (Harris 2016, Kissling, et al. 2012). Still, it remains to be tested in how far co-occurrence data will allow estimation of asymmetric residual correlation or under which circumstances co-abundance data will be needed (Cazelles, et al. 2015). Incorporating trait information could also help to distinguish competitive and mutualistic interactions from predator-prey relationships, for example by simple plausibility checks (Morales-Castilla, et al. 2015).

Last, we tested for scale dependence of JSDM estimates by analysing the co-occurrence data at increasingly coarse resolution while the interactions took place at the original, fine spatial resolution. We found that with increasingly coarser resolution, both the signals of positive and negative interactions became indiscernible by JSDMs, whereby the signal of negative interactions was more sensitive to scale. This is in line with previous findings that co-occurrence patterns are highly scale-dependent (Araújo and Rozenfeld 2014) and also that the effects of local interspecific interactions vanish at coarser spatial scales (Thuiller, et al. 2015). Thus, JSDMs are unlikely to give any insight on the potential interactions between species if the scale of the data does not match the process scale. At best, they might give better and more reliable models and the residual correlation might shed light on missing environmental variables and historical factors. Also, the species lists from monitoring schemes or atlas data are often at comparably
coarse resolution, for example for many animals, such that the signal of interspecific interactions and sub-scale environmental heterogeneity can get easily confounded (Zurell, et al. 2016b). When we apply JSDMs to vegetation plots, the analysis of the residual correlation in respect to interspecific interactions might thus be more meaningful since the interactions between species, for instance for light competition, might here be at the right scale (e.g. a 10x10m plot). Empirical analyses testing whether the residual correlations from JSDMs are related to known interactions between species might thus be of particular interest. Furthermore, our results showed that less prevalent species are less sensitive to mismatches between process scale and data scale. At the same time, at increasingly coarse resolution negative and positive interactions were confounded easier. Thus, we recommend interpreting JSDMs cautiously when modelling rare species.

In summary, our simulation study indicates that JSDMs might be reliable in inferring simple symmetric species interactions if the species are modelled at the resolution of the interactions. However, even in our idealized and error-free data, JSDMs could easily confound predator-prey relationships with either competitive or mutualistic interactions, they do not give an estimate of interactions strengths, they are not easily comparable across species, they are affected by prevalence and by scale. Thus, we should be very cautious when interpreting JSDM results as interactions (Clark, et al. 2014, Pollock, et al. 2014). Nevertheless, they have the ability of enhancing our analyses and allow us to form hypotheses about potential interspecific interactions when used cautiously (Ovaskainen, et al. 2010). Our results provide a first test of potentially important or confounding factors in JSDM analyses. In the future, more efforts are needed to test JSDMs in more complicated settings, for example to test their ability to
disentangle environmental response and residual correlation for multiple interacting species, for cases when species interact directly and indirectly, and for other confounding factors such as detection bias and missing environmental covariates. Thereby, further theoretical tests under controlled conditions could be useful, but we also emphasise the need for more empirical tests, for example comparing JSDM estimates against known interactions (e.g. from experiments). Hence, JSDMs constitute an important step forward for testing hypotheses of how interspecific interactions affect species distributions. Nevertheless, there are important limits to their capability of detecting interspecific interactions, for example for trophic and indirect interactions, highlighting the need for further model improvements in this respect.

**Acknowledgements**

DZ acknowledges funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. 624958, and from the Swiss National Science Foundation SNSF (grant: PZ00P3_168136/1). WT acknowledges support from the European Research Council (ERC-2011-StG-281422-TEEMBIO). LJP acknowledges funding from the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. 659422.
References


Supplementary material (Appendix EXXXX at <www.oikosoffice.lu.se/appendix>).

Appendix 1.
Figure captions

Figure 1. Expected co-occurrence (A-F) and residual correlation (estimated by JSDM; G-K) across the biotic interaction space of two species A and B and for different prevalence levels. (A) Species A and B can have negative to positive effects on each other. The co-occurrence index is defined as the number of cells where species A and B occur together divided by the total number of occupied cells. The null probability of co-occurrence \( P(A \text{ and } B) \) is given by the product of species' prevalence \( P(A) \times P(B) \), and the corresponding co-occurrence value is given by \( (P(A) \times P(B)) / ((P(A) + P(B) - P(A) \times P(B)) \) (white line). (B-F) show the co-occurrence values obtained at different prevalence levels (0.1, 0.2, 0.3, 0.4, 0.5) with equal prevalence for both species. Again the null expectation of co-occurrence is shown in white. (G-K) show the corresponding residual correlations estimated by JSDMs, with grey areas indicating non-significant residual correlations.

Figure 2. Scale dependence of co-occurrence patterns and corresponding residual correlations estimated by JSDMs. Both species A and B have prevalence of 0.3. (A-C) show the spatial distribution of species at increasingly coarser resolution resulting from intermediate competition and intermediate mutualism (species A in black, species B in grey, species A and B co-occurring in red). (D-F) show the resulting co-occurrences across the biotic interaction space of species A and B, and (G-I) the corresponding residual correlations estimated by JSDMs (please refer to Fig. 1 for legend description). (J) shows the relationship between (significant) residual correlations estimated at the original resolution and at coarser resolution.
Figure 3. Scale dependence of JSDMs for different prevalence levels. Shown is the significant residual correlation estimated by JSDMs at the original scale against significant residual correlation estimated at coarser resolution (block size 2x2 in black, block size 4x4 in red) across biotic interaction space. Scale dependence of JSDMs increases with prevalence from left to right such that no significant residual correlations were estimated for the 4x4 block size in (E).
Figures

Figure 1
Figure 2
Figure 3