A JSDM with zero-inflation to increase the ecological relevance of analyses of species distribution data.

Jeanne Tous[1], Julien Chiquet[2], Amy E. Deacon[3], Ada Fontrodona-Eslava[4], Douglas F. Fraser[5], Anne E. Magurran[6]

[1]MIA 518, AgroParitech/INRAE, Universite Paris-Saclay, Palaiseau, France, jeanne.tous@inrae.fr - corresponding author.

[2]MIA 518, AgroParitech/INRAE, Universite Paris-Saclay, Palaiseau, France, julien.chiquet@inrae.fr.

[3]Department of Life Sciences, The University of the West Indies, St. Augustine, Trinidad and Tobago, amy.deacon@uwi.edu

[4]Centre for Biological Diversity, School of Biology, University of St. Andrews, St. Andrews, UK, afe1@st-andrews.ac.uk

[5]Department of Biology, Siena College, Loudonville, NY 12211, USA fraser@siena.edu

[6]Centre for Biological Diversity, School of Biology, University of St. Andrews, St. Andrews, UK, aem1@st-andrews.ac.uk

Statement of authorship :
JT led the investigation
JT & AEM designed the study
JC designed the ZIPLN-network model and provided technical expertise on its use.
JT ran the analyses
AEM, AED, A F-E & DFF provided ecological expertise to support the interpretation of the results
JT wrote the initial draft
All authors reviewed and approved the final manuscript

Data availability statement:
Data and code can be found on the public git repository
jeannetous/JSDM_with_zero_inflation_for_increased_ecological_relevance_in_species_distribution_analysis

Acknowledgements:
This work was supported by public grants from the Fondation Mathématique Jacques Hadamard, AgroParisTech, the DESSE INRAE and the Erasmus+ program. We thank Stéphane Robin and Wilfried Thuiller for the fruitful discussions about this work.

Conflicts of interest:
The authors declare that they have no conflict of interest.

Abstract

Deciphering the mechanisms that explain species distributions has been a long-term goal in ecology and is critical in predicting the responses of ecological communities to environmental change. To this end, Joint Species Distribution Models (JSDM) assess the effects of abiotic factors on these communities and produce species association networks that describe statistical dependencies between species abundances. Here we propose a novel JSDM, called ZIPLN-network, that directly models count data while accounting for zero-inflation. We illustrate the utility of the ZIPLN-network approach using data from tropical freshwater fish communities. By dealing with the zero-inflation that is a near-universal feature of species occurrence data sets, our approach overcomes a previous roadblock in analyses of species distribution data, and makes it possible to extract more of the ecological signal. A further advantage of this method is that it increases our understanding of ecological mechanisms that structure species distributions in a landscape, and in doing so helps formulate testable hypotheses.

The data and code used for this paper can be found in the git repository jeannetous/JSDM_with_zero_inflation_for_increased_ecological_relevance_in_species_distribution_analysis

Keywords: Joint Species Distribution Models (JSDM) ; zero-inflation ; species association networks ; assemblage mechanisms

1. Introduction

Species distribution data describe the distribution of a set of species in a collection of sites, often based on counts of individuals at each site. Quantifying the patterns of these distributions, and deciphering the processes that underpin them has been a long term goal in ecology [Connor and Simberloff, 1979, Diamond, 1975, Gotelli and Graves 1996, Ovaskainen et al. 2010]. Three main factors underlie these distributions [Guisans and Thuiller, 2005, Poggiato et al., 2021]: abiotic factors through environmental filtering (whether the environment is suitable for a given species [Cadotte and Tucker, 2017, Kraft et al., 2015]), biotic interactions (how species' presences influence one another) and dispersal (whether a species has been able to reach a site). Understanding how these mechanisms structure biodiversity patterns is critical for predicting ecosystem responses to environmental stressors and to design effective conservation strategies [Heinen et al. 2020]. However, the complicated ways in which these mechanisms operate, their intermingled influences, and the way their relative effects can depend on scale [Götzenberger et al. 2012, Viana and Chase 2019] makes species distribution data both information-rich and hard to interpret.

One of the major challenges in this quest is that rare taxa often predominate in local assemblages with the result that the $species \times sites$ dataframes that are the target of species distribution analyses contain many zeros. This problem is compounded by 'unseen' species, species that are present but have not been sampled, or species that occur only occasionally in a particular site but were absent when the sampling occurred.

One set of methods compares species distributions with those predicted by null models [Connor and Simberloff, 1979, Ulrich and Gotelli, 2010] to determine if observed patterns can be attributed to chance. These approaches often consider the species assemblage globally and typically do not distinguish between different explanations for non-random patterns.

Species Distribution Models (SDM) [Elith and Leathwick, 2009, Miller, 2010] are a more recent approach. This method relates species distributions to abiotic drivers using, for example, generalised linear models or neural network approaches. These models have wide application, but only account for abiotic effects.

To further disentangle the basket of mechanisms involved (unobserved abiotic effects, biotic interactions, dispersal), one needs to go beyond direct abiotic effects when analysing data. Specifically, one can study species associations, that is statistical dependencies (positive or negative) between each pair of species' abundances, after having accounted for observed covariates and for the effects of all other species. Together, these pairwise associations form a species association network in which the network's nodes are the species. An edge (line) is then drawn between two nodes if the two species' abundances display an association.

Associations are statistical relationships; while they may result from ecological interactions, they could also be explained by unobserved abiotic factors or indirect effects of unobserved species [Poggiato et al., 2021]. Some ecological interactions produce indirect or emergent effects that do not manifest as statistical dependencies. None-the-less, with these constraints acknowledged, we find that association networks can help identify patterns in species distributions related to underlying ecological processes. This is because association networks reveal how previously identified ecological interactions can shape species distribution patterns and uncover statistical dependencies in species abundances. They can also be used as an exploratory tool to unveil potential interactions and inform hypotheses that will need further exploration through modelling or field studies.

Joint Species Distribution Models (JSDM) [Pollock et al., 2014, Tikhonov et al., 2020, Chiquet et al., 2021] are a recent and widely used class of models used to infer association networks from species distribution data. Like SDMs, they identify abiotic effects but also account for residual statistical

correlations between abundances to obtain a species association network. Since species distribution data are often count data and are thus challenging to model, many JSDMs model presence-absence data [Wilkinson et al., 2019]. However, reducing count data to presence-absence constitutes an important loss of information.

Other methods, such as the Hierarchical Modelling of Species Communities (HMSC) approach [Tikhonov et al., 2020] or the Multivariate Poisson log-normal model (PLN) [Aitchison and Ho, 1989, Chiquet et al., 2019, Chiquet et al., 2021] can directly model count data using Poisson distributions. Both models use a latent Gaussian layer to model statistical associations between species. However, species distribution data often include numerous zeros, some of which cannot be explained by the effects of covariates or that of other species. Indeed, some species might not be detected, or they may never be found in a particular site. This absence may be due to the fact that these species are hard to detect in certain conditions, or may never have reached a given site in the first place due to dispersal limitations. Falsely attributing these zeros to abiotic effects or to statistical associations with other species can lead to the detection of spurious effects or associations.

To overcome this issue, we have developed a Zero-Inflated PLN-network approach (ZIPLN-network). This model builds on both the PLN-network [Chiquet et al., 2019] and ZIPLN approaches [Batardière et al., 2024]. It is a generalised multivariate mixed model that uses a Poisson distribution to directly model count data with a latent Gaussian layer to represent the dependency structure corresponding to the association network (further explained in 2.1 below) and another latent variable to account for the zero-inflation in the data. The model is specifically designed to infer a statistically meaningful species association network while accounting for covariates or potential differences in sampling efforts. Indeed it uses a penalised, LASSO-like approach [Friedman et al., 2008] to select the most meaningful association.

The goal of this paper is to introduce the ZIPLN-network approach and highlight its methodological advantages as a model that accounts for zero-inflation in species distribution data. We explain how this feature of the model influences the resulting association network and show that the analysis of the zero-inflation probabilities inferred by the model facilitates more targeted hypotheses about the underpinning ecological mechanisms and strengthens ecological interpretations of species distribution data. To exemplify the use of the ZIPLN-network approach we apply it to species distribution datasets of tropical freshwater fishes sampled in the Caribbean island of Trinidad.

Our approach is based on the assumption that accounting for the zero-inflation will impact the association network inferred by the model. Some associations are likely to disappear because they are related to common zero-inflated patterns between species. We argue that in doing so the approach helps identify ecologically meaningful relationships.

Below, we first introduce the ZIPLN-network model and, using a hypothetical example, explain how it works. Next, we apply it to a tropical fish data set and discuss the results including how their interpretation is influenced by the choice of covariates. We compare the results with the non-zero-inflated PLN-network approach, and also compare the networks obtained with several sets of covariates. Finally we provide recommendations for users of the ZIPLN-network approach.

2. Material and methods

2.1 ZIPLN-network

In the model, n denotes the number of samples considered (a sample corresponding to a pair $(geographical\ site,\ sampling\ date)$ in a dataset) and p the number of species. Y_{ij} is the number of individual counts of species j found in sample i so that Y_i is a p-dimensional vector of integers. The Poisson log-normal (PLN) model [Aitchison and Ho, 1989] models observations Y using a latent variable Z with:

$$\text{Latent layer: } Z_i \sim \mathcal{N}(0, \Omega^{-1})$$

$$\text{Observed layer: } Y_{ij} \mid Z_{ij} \sim \boldsymbol{\mathcal{P}}(exp(X_i^T B_j + Z_{ij}))$$

X_i denotes the covariates used for sample i , B_j the corresponding regression coefficients for species j. Just as in a regular generalised linear model, the $X_i^T B_j$ term accounts for abiotic effects on the species' abundances.

The latent variable Z allows the modelling of a statistical dependency structure between the species. Indeed, Z_i follows a Gaussian distribution, with variance-covariance matrix Ω^{-1}**.** Ω is the precision matrix of the distribution, so that Z_{-j} and Z_{-k} (the $j - th$ and $k - th$ columns of Z) are conditionally independent given all other Z_{-l} if and only if $\Omega_{jk} = 0$. It ensues that Ω gives an association network between the species. To infer a sparse network and avoid the detection of spurious associations, the PLN-network inference procedure [Chiquet et al, 2018] adds an ℓ_1 penalty on the off-diagonal values of Ω.

Selecting which penalty to apply can be done using different statistical criteria. In our study, we use the StARS method (Stability Approach to regularisation Selection) [Liu et al., 2010]. This approach uses resampling to compute multiple association networks and keep only the associations that are present in a high enough proportion of the networks, a proportion that is fixed at 80 % here.

So far, we have described a model that assumes all observations arise from a Poisson distribution (with an over-dispersion arising from the latent variable). This means that any 0 observed in the data is explained by the effects contained in X and / or the effects of other species through the latent correlations. However, a particular species might be absent from a sample for other reasons, either because it is hard to detect despite being present, or because it can never be found in a given site due to natural barriers, such as waterfalls in the case of freshwater fish.

To account for such effects, we introduce the Zero-Inflated PLN-network (ZIPLN-network). This model uses an additional binary latent variable W that follows a Bernoulli distribution. If $W_{ij} = 1$, Y_{ij} is automatically equal to 0 and if $W_{ij} = 0$, Y_{ij} is drawn from a Poisson distribution as with the PLN-network approach:

$$\text{Latent layer: } Z_i \sim \mathcal{N}(0, \Omega^{-1}), W_{ij} \sim B(\rho_{ij})$$

$$\text{Observed layer: } Y_{ij} | W_{ij}, Z_{ij} \sim \mathcal{P}(exp(X_i^T B_j + Z_{ij})) \text{ if } W_{ij} = 0$$

$$Y_{ij} | W_{ij}, Z_{ij} = 0 \text{ if } W_{ij} = 1$$

The parameters of W's distribution can depend on the species, the sample or the covariates (through a linear effect, with species-specific regression coefficients). This means that it is possible to account for the fact that some species are more likely to be absent from a sample than others, but also that the probability of having a zero for a given species can depend on the environmental conditions.

The ZIPLN-network and other PLN models [Chiquet et al., 2021] are implemented in the publicly available R package PLNmodels [https://cran.r-project.org/package=PLNmodels], and in the python package pyPLN [Batardière et al., 2024].

Accounting for the zero-inflation that appears in a dataset is key to accurate analysis. We use an example with simulated data to illustrate this point. The code for this example can be found in the git repository that comes with this paper. We consider three fictional species A, B and C, whose abundances are simulated under a PLN model. We assume that all three of them are more abundant when the value of an abiotic variable X is lower. We further assume a strong positive association between B and C (meaning that both species tend to be found together more than expected with the abiotic effect only), a weak positive A - B association and a weak negative A - C association (see Figure 1.1 for the corresponding association network). Finally, we consider two zero-inflated versions of the simulated data, with either species-dependent or covariate-dependent zero-inflation.

We compare the association networks inferred from the data with PLN models both with and without accounting for the zero-inflation effect. Figure 1 shows that adding zero-inflation makes the association network harder to retrieve. When the species-dependent zero-inflation is not accounted for, a spurious positive association is detected between species A and C, and the strength of the positive B-C association is largely under-estimated. Using ZIPLN-network to account for the zero-inflation produces a network much closer to the real one. Although less striking with the covariate-dependent zero-inflation, the difference between the two retrieved networks remains significant, the ZIPLN-network one being closer to the ground-truth network, especially when it comes to the strong B - C positive association.

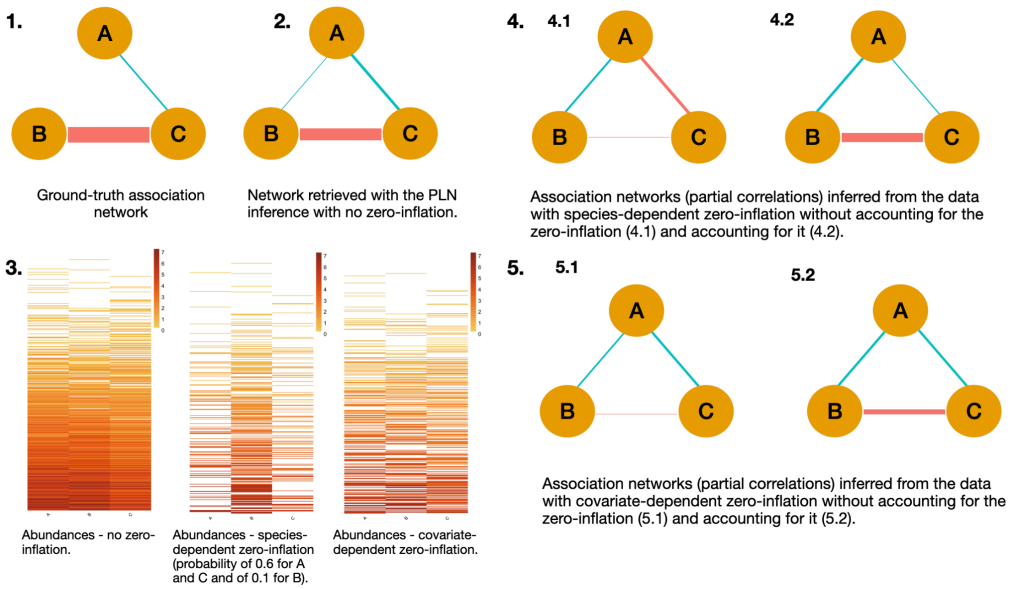

Figure 1: Not accounting for existing zero-inflation in the data can significantly bias the inference of the association network. Association networks here correspond to partial correlations between the three species' abundances. Pink edges correspond to positive associations whereas blue ones are negative associations. The edges' thickness is proportional to the associations' importance. (1) Real association network. (2) Association network inferred with the PLN approach, when non-zero-inflated data is provided. (3) Heatmaps of the simulated abundances in log-scale without zero-inflation, with species-dependent zero-inflation and with covariate-dependent zero-inflation. (4) Association networks inferred when using the PLN approach, i.e. not accounting for the species-dependent zero-inflation [4.1] or with the ZIPLN approach i.e. accounting for it [4.2]. (5) Association networks inferred when using the PLN approach, i.e. not accounting for the covariate-dependent zero-inflation [5.1] or with the ZIPLN approach i.e. accounting for it [5.2].

2.2 Data

To illustrate how the ZIPLN-network approach can be used to study species distribution data, we used a freshwater fish survey from Trinidad. Trinidad is an island in the Southern Caribbean, located 12km from Venezuela. It is home to a rich freshwater fish fauna of both Antillean and South American origin that has been extensively investigated [Phillip et al., 2013].

The survey took place in Trinidad's Northern Range, a mountain system in the North of Trinidad where a series of parallel rivers are found. Fish were surveyed at 16 sites across eight Northern Range rivers over a period of 5 years from December 2010 to August 2015. In each river, two sites were surveyed, one considered as "disturbed" due to human recreational activities with the other being "undisturbed" [Magurran et al., 2018]. Each site was surveyed four times a year, at the beginning and end of both the dry (January to May) and wet (June to December) season [Deacon et al., 2015, Magurran et al., 2018]; the first sample was removed due to differences in sampling efforts [Magurran et al., 2018]. Hence 19 surveys are considered for each site. The dataset contains 20 species, after filtering out those that had been observed less than twice in the dataset. For the sake of readability, we use shortened versions of the species names on the figures; their complete names can be found in Appendix A. As shown on Figure 2, the dataset contains numerous zeros, whose presence is stream-dependent.

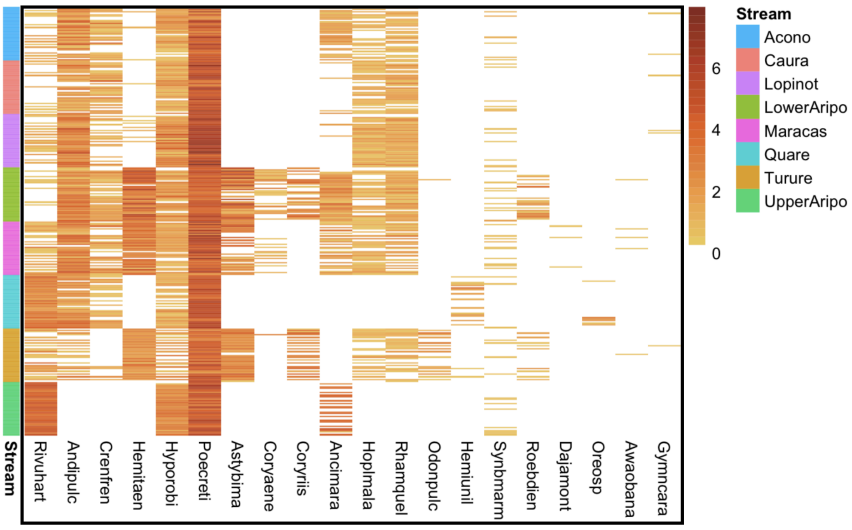

Figure 2: Species log-abundances (log(1 + species count), in the dataset, sorted by stream. Blank cells indicate species absences. Each column corresponds to a species, each row corresponds to a combination (site, time step). The rows are grouped by stream (there are 2 sites per stream).

2.3 Choice of covariates

To account for the abiotic effects, we ran a principal component analysis (PCA) on a set of descriptive variables. The variables used were: width, depth and volume of the stream, altitude, the number of pieces of garbage found at a site, water conductivity, dissolved oxygen concentration, pH, temperature and turbidity, as well as several quantitative descriptors of the substrates found in the river and a visual estimate of the canopy cover. We only kept the first two principal components, denoted *PC1* and *PC2*; using more did not statistically improve or significantly modify the results. We included these covariates to account for environmental effects and thus avoid them impacting the inferred association network.

We additionally accounted for two categorical variables: disturbance (binary variable, i.e. whether the site was a priori labeled as disturbed) and river identity. Together, these two variables characterise each geographical site. Accounting for the stream ID in this way allowed us to include a dispersal-related effect in the model: if a species was never found in a particular stream or in nearby sites, the model adjusted the coefficient of B corresponding to that stream, without attributing such an effect to the abiotic covariates.

For the PLN model, we assumed that all the covariates we included influence the abundances whereas for the ZIPLN-network model, we included the stream effect in the zero-inflation part of the model only, and used PC1, PC2 and the disturbance variable for the abundance.

The models are described with the following formula structure:

$$\text{PLN}(\text{Abundance} \sim 1 + X_1 + X_2 \ldots)$$

$$\text{ZIPLN}(\text{Abundance} \sim 1 + X_1 + X_2 \ldots \mid 1 + X_3 + X_4 + \ldots)$$

The first formula corresponds to PLN-network run with an intercept ("1 + …") and covariates X_1, X_2 … The second formula corresponds to ZIPLN-network run with, on the one hand, an intercept and covariates X_1, X_2… for the abundance part and, on the other hand, an intercept and covariates X_3, X_4

…. for the zero-inflation part. The same covariate can be used for the zero-inflation and for the count part of the model. The intercepts translate into additional, covariate-independent species effects on the abundance / zero-inflation.

In the end, we kept six different models and compared the results with and without zero-inflation. For the covariates, our choice is based on the models' statistical results and interpretation. We kept the first two principal components from the PCA on the environmental covariates in order to account for abiotic effect while avoiding redundancies. We also assessed the effect of a site being disturbed on the models' results so that we included the *disturbance* too. Finally, it is reasonable to expect different streams to host different species (owing to a dispersal mechanism for instance), so we additionally included the *stream* variable.

The first three models we ran were PLN-network models (no zero-inflation), the first one with *PC1* and *PC2* covariates only, the second one with the addition of the *disturbance* covariate, and the last one with the addition of both the *disturbance* and *stream* covariates. "Abundance" corresponds to the counts matrix Y in the model equations.

PLN(Abundance ~ PC1 + PC2)

PLN(Abundance ~ PC1 + PC2 + disturbance)

PLN(Abundance ~ PC1 + PC2 + disturbance + stream)

We also ran three ZIPLN-network models. The first model included only *PC1* and *PC2* as covariates for the abundance, whereas the other two models additionally considered a *disturbance* effect. In terms of zero-inflation, the first two assumed a species-dependent zero-inflation (each species has its own zero-inflation probability that is the same on all sites), whereas the third one assumed this probability to also depend on the steam (that is each species has a zero-inflation probability that is different for each stream).

ZIPLN(Abundance ~ PC1 + PC2 | 1)

ZIPLN(Abundance ~ PC1 + PC2 + disturbance | 1)

PLN(Abundance ~ PC1 + PC2 + disturbance + stream | stream)

2.4 Cross-validation

During the inference process, both the PLN-network and ZIPLN-network models optimised the model's parameters as well as variational parameters [Chiquet et al., 2021] to fit the data, yielding a resulting fit. However, this differs from the abundances the models would predict when encountering new data, that is when being given new sets of streams and covariates and having to predict the abundances of each species.. To assess their predictive abilities, a 5-fold cross-validation approach was used: the dataset was divided into 5 subsets, and the model was repeatedly fitted on 4 of them before assessing the prediction on the 5th one. ZIPLN-network predicted a 0 when it found a probability higher than 0.5 that there was one, and otherwise used the Poisson part of the model to predict the abundance. This method gives a measure to assess the model in terms of predictive abilities.

For instance, let us consider a dataset with n sites and p species. For each model (e.g. PLN(Abundance ~ PC1 + PC2)), we first fit the model on the whole dataset. The inferred parameters give a fitted value for each $i \in [1\,;\,n]$, for each $j \in [1\,;\,p]$ and the root mean squared error (RMSE) between the real and fitted values can be computed. However, these fitted values are obtained on observed data and we also want to assess what abundances the model would predict on new data (that is, given new sites and covariates, what abundances would the model predict on each site, for each species). To assess its predictive abilities, we divide the n sites into 5 subsamples. For each subsample, we fit the model on the other 4 and test the predictions on the 5th one, computing the RMSE between the predictions and the real data. We then use as a metric the median of the five RMSE thereby obtained.

3. Results

Here we compare the output of the PLN-network and ZIPLN-network in terms of Bayesian Information Criteria (BIC), goodness-of-fit, abiotic effects and network associations, depending on which covariates are included.

The models are described in Table 1, along with their BIC and goodness-of-fit. Two measures are used for the goodness-of-fit. The first one is the root mean squared error (RMSE) between the real data and the fit produced during inference. The second one is the median RMSE obtained from a 5-fold cross-validation as described in the methods. Table 1 shows that in all cases, accounting for a stream effect significantly improves the BIC; it also improves the cross-validation RMSE, except for the ZIPLN-network model. When the stream covariate is not included, the cross-validation RMSE is dramatically improved with the zero-inflation but this effect disappears when the stream effect is included in the PLN model.

Model	**BIC** (the higher, the better)	**Fitting root of the RMSE** (the lower, the better)	**Cross-validation RMSE** (the lower, the better)
PLN(Abundance ~ 1 + PC1 + PC2)	-9800	0.42	95.2
PLN(Abundance ~ 1 + PC1 + PC2 + disturbance)	-9778	0.42	95.4
PLN(Abundance ~ 1 + PC1 + PC2 + stream + disturbance)	-8988	0.56	48.5
ZIPLN(Abundance ~ 1 + PC1 + PC2 | 1)	-10044	0.51	56.31
ZIPLN(Abundance ~ 1 + PC1 + PC2 + disturbance | 1)	-10073	0.52	53.4
ZIPLN(Abundance ~ 1 + PC1 + PC2 + disturbance | 1 + stream)	-9049	0.52	62.2
Poisson regression - Abundance ~ 1 + PC1 + PC2	-131 938	64	54.7
Poisson regression - Abundance ~ 1 + PC1 + PC2 + disturbance	- 121 348	62.2	51.5
Poisson regression - Abundance ~ 1 + PC1 + PC2 + stream + disturbance	-71 861	50.5	44.9
Log-normal regression - Abundance ~ 1 + PC1 + PC2	-9936	79.3	57.8
Log-normal regression - Abundance ~ 1 + PC1 + PC2 + disturbance	-9889	62.2	51.5
Log-normal regression - Abundance ~ 1 + PC1 + PC2 + stream + disturbance	-8098	79.2	47.1

Table 1: BIC and goodness-of-fit for the PLN-network, ZIPLN-network and two simpler methods with different covariates combinations. The higher the BIC, the better the model from a statistical point of view.

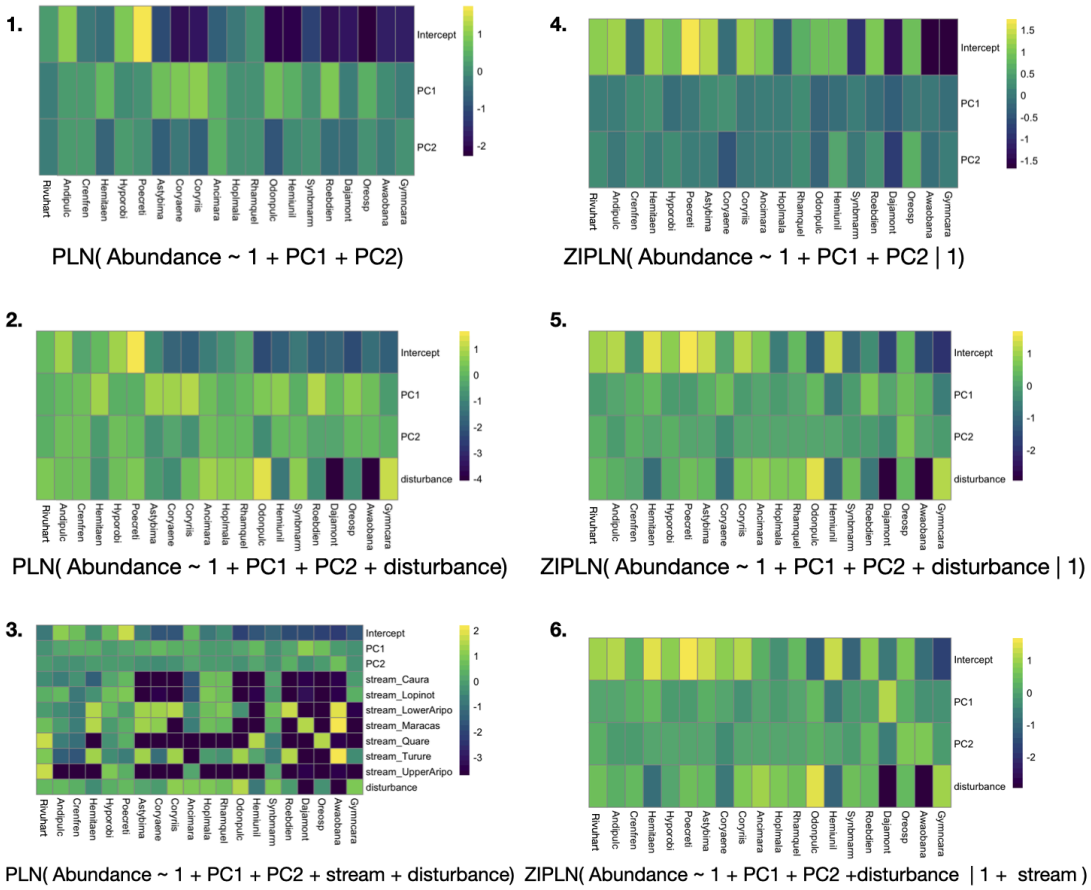

Figure 3: Regression coefficients for abiotic effects on the abundances (log-scale).

Figure 3 shows the coefficients corresponding to the abiotic effects each model infers. This corresponds to matrix B in the model's notations described above. The effects of the environment-related principal components display little variance, with a stronger intercept effect, whereas Figure 3.3 shows an important variance in the stream effect (for the ZIPLN-network model, the stream effect is only included in the zero-inflation coefficients, see Figure 4).

ZIPLN-network infers a zero-inflation probability for each pair $(species,\ sample)$ depending on the covariates used in the model. Here, for the first two ZIPLN-network models, this probability is only species-dependent whereas for the third one it also depends on the stream considered. Figure 4 shows the resulting probabilities. We see that when the probabilities are made stream-dependent, for a given species, they can vary a lot from one stream to another..

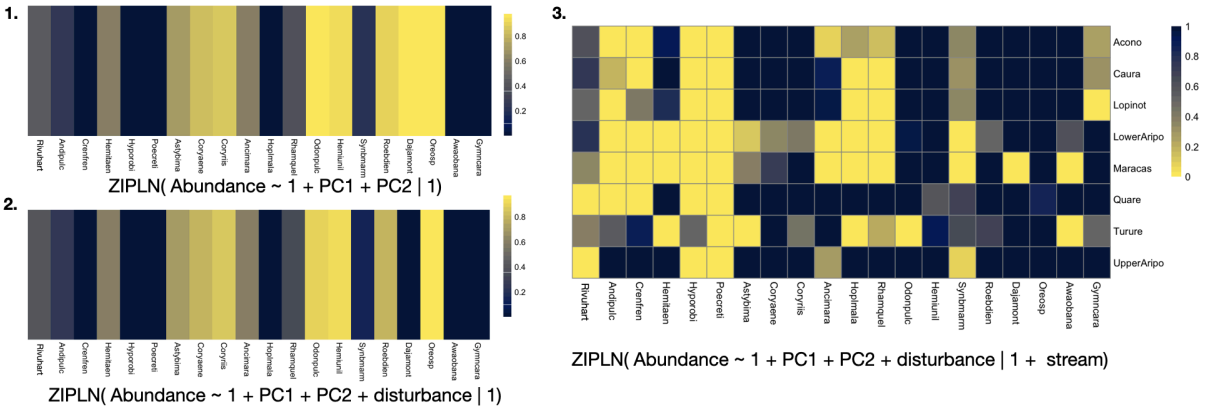

Figure 4: zero-inflation probabilities inferred by the ZIPLN-network approach with the different covariates combinations.

Figure 5 shows the association networks inferred by PLN-network and ZIPLN-network using various combinations of covariates. Accounting for a zero-inflation effect significantly modifies the association networks. For both models, adding a disturbance effect in the abundance only marginally modifies the network. However, accounting for a stream effect, be it in the abundance for the PLN model or in the zero-inflation for the ZIPLN model removes a certain number of associations. When the stream is accounted for, the PLN and ZIPLN association networks are quite similar, but the associations in the former appear to be much weaker than in the latter. Some associations are only present in the non-zero-inflated model, such as the positive *Roeboides dientonito* - *Corynopoma riisei* association, the negative associations of *Poecilia reticulata* with both *Astyanax bimaculatus* and *Hypostomus robinii*. Other associations are only visible in the ZIPLN network, such as the positive associations of *Poecilia reticulata* with *Hoplias malabaricus* and *Rhamdia quelen* or the negative one with *Ancistrus maracasae*.

Several associations of interest can be noted. We find a strong negative association between *Hoplias malabaricus* and *Rivulus hartii* (also known as *Anablepsoides hartii*), present with both approaches, which disappears when the stream variable is accounted for. Positive associations between *Astyanax bimaculatus* and *Hemibrycon taenerius* on the one hand, and *Poecilia reticulata* and *Andinoacara pulcher* on the other hand, persist in all the networks.

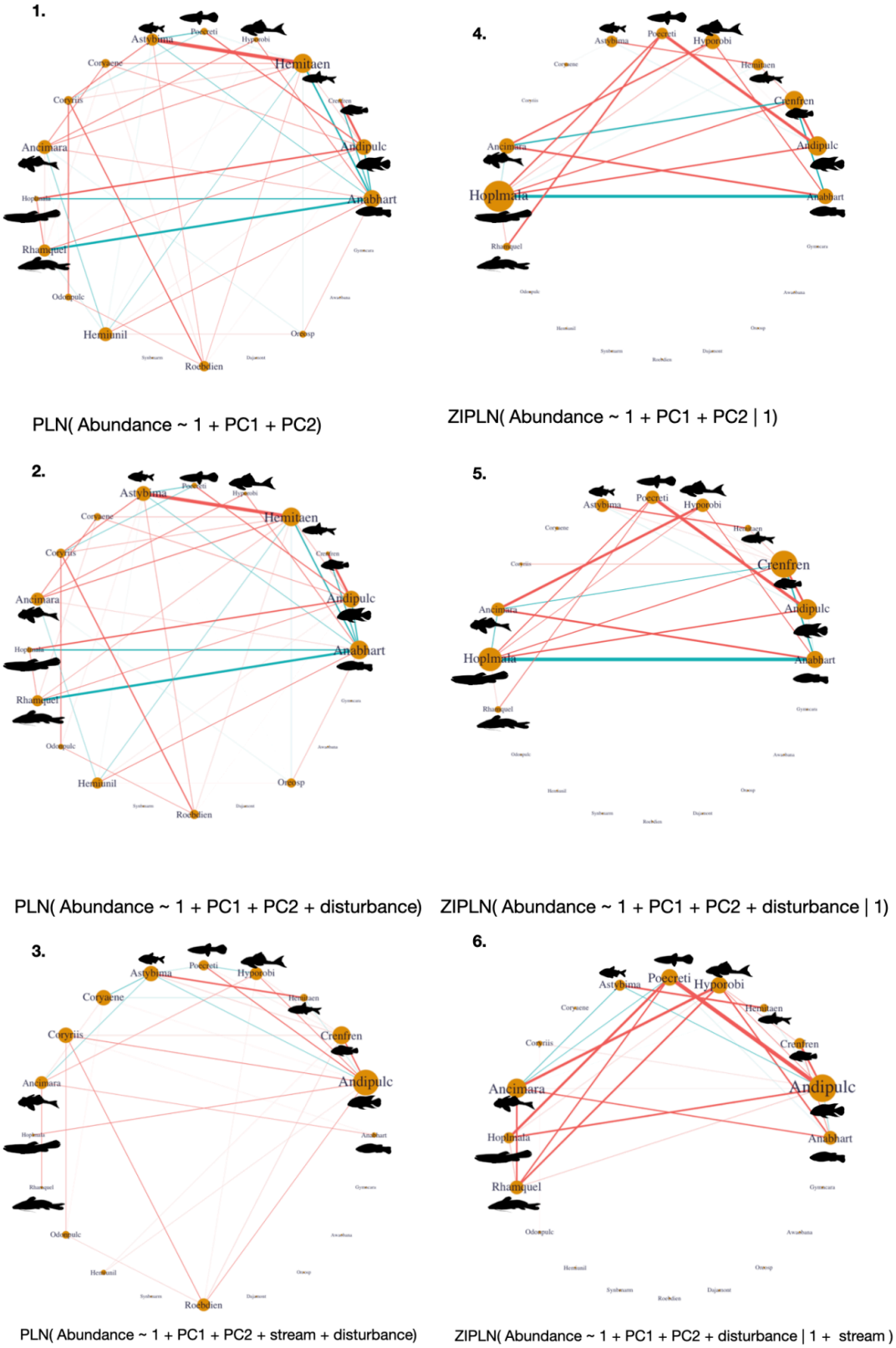

Figure 5: Association networks obtained with PLN-network and ZIPLN-network for different combinations of covariates. Blue and pink edges represent negative and positive associations, respectively. The thicker the edge, the stronger the association. The size of each species' name is proportional to the sum of absolute values of associations it is involved in. License details for the species silhouettes are given in the references.

4. Discussion

We have introduced the new ZIPLN-network approach to analyse species distribution data. This approach directly models abundance data while using zero-inflation to account for the absence of a species owing to factors that invariably prevent its occurrence, e.g. barrier waterfalls, in the first place, or for the fact that a species has a high probability of not being sampled. Accounting for the zero-inflation is a long-standing issue in the modelling of species abundances [Cunningham and Lindemayer, 2005, Nolan et al., 2021, Wenger and Freeman, 2008]. Using an approach that takes zero inflation into account makes it possible to formulate specific hypotheses regarding abiotic effects and species associations. Moreover, the ZIPLN-network allows users to specifically analyse zero-inflation probabilities, and thus learn more about the reasons for a species not being found.

In this section we first analyse the association network inferred with the ZIPLN-network approach, and consider how it can be used to build ecological hypotheses regarding the mechanisms underlying the observed distribution. We then address the choice of covariates and the consequences of this choice for the interpretation of the model's results. Finally, we provide recommendations for using the ZIPLN-network approach to analyse species distribution data.

Inferences from the ZIPLN-network approach

Table 2 draws on the literature to identify the key interspecies relationships we could expect to retrieve as statistical relationships in the species association networks inferred by the models. We use this as a benchmark to analyse the associations identified by the models, beginning with the network obtained with ZIPLN-network using *PC1*, *PC2* (the first two principal components from a PCA run on the covariates that describe each site) and disturbance level as covariates for the abundance, with the stream as a zero-inflation covariate (Figure 5.6).

We find multiple associations that are consistent with existing ecological knowledge of species interactions in these freshwater systems. For example, positive associations between putative

predators and prey emerged. For instance, guppies are positively associated with *Andinoacara pulcher* (blue coscorobs), and *Hoplias malabaricus* (guabines). Several hypotheses could explain this situation. As guabines are primarily nocturnal predators [Fraser et al., 2006], their coexistence with guppies could reflect temporal niche sharing [Fraser et al., 2004]. Further, the behavioural response to predation threat by guppies is less in the form of habitat shifts and more likely shoaling and avoidance during the most vulnerable crepuscular periods [Fraser et al., 2004]. Guppies and blue coscorobs co-occur in many stream habitats in the Northern Range, with the latter an invertivore, rather than a piscivore as was sometimes assumed in the earlier literature.

A positive association that was anticipated is found between *Astyanax bimaculatus* (two-spot sardine), and *Hemibrycon taenerius* (mountain stream sardine). These species can benefit from adaptive advantages due to heterospecific schooling behaviour [Krause and Ruxton, 2002]. Interestingly, while there is a strong association between these species, there are no others within the group of shoaling characids (locally referred to as 'sardines') that also includes *Corynopoma riisei* (swordtail sardine) and *Roeboides dientonito* (hunchback sardine). It could be interesting to further explore this pattern to study whether it is explained by different schooling preferences or by other mechanisms.

While our analysis uncovers statistically robust associations, not all of these necessarily correspond to ecological interactions, nor have all potential interactions been identified. For instance the predator-prey interaction between *Crenicichla frenata* (pike cichlids) and guppies is not detected in the network, although a large literature documents the evolutionary interactions between these species. An explanation for this is that behavioural (e.g., shoaling [Magurran and Seghers 1991, Seghers, 1974]) and morphological (e.g., color patterns [Endler 1987]) adaptations in guppy populations, which covary with predation risk, enable coexistence in localities where the two species naturally co-occur.

Species 1	Species 2	Interaction that could be expected to translate into a statistical association	Value in ZIPLN-network (FIgure 5.6)
Rivulus hartii	*Poecilia reticulata*	*P. reticulata* predates young *A. hartii*. The second is known to be less abundant in presence of the first [Walsh et al., 2011, Gilliam et al., 1993, Fraser and Lamphere 2013, Goldberg et al., 2022]	**0**
Andinoacara pulcher	*Poecilia reticulata*	Weak predator-prey relationship [Magurran, 2005].	**+** **0.35**
Astyanax bimaculatus	*Hemibrycon taeniurus*	The two species may school together and benefit from this interaction: potential positive association [Goodale et al., 2010, Krause and Ruxton 2002].	**+** **0.15**
Crenicichla frenata	*Rivulus hartii*	Strong predator-prey interaction [Tran et al., 2024].	**+** **0.05**
Crenicichla frenata	*Poecilia reticulata*	Strong predator-prey interaction [Magurran, 2005].	**0**
Hoplias malabaricus	*Rivulus hartii*	Strong predator-prey interaction. The presence of the first is known to reduce that of the second: potential negative association [Fraser and Gilliam, 1992, Gilliam et al. 1993]	**0**
Hoplias malabaricus	*Poecilia reticulata*	Strong predator-prey interaction [Winemiller 1989]	**+** **0.2**
Rhamdia quelen	*Rivulus hartii*	Potential weak predator-prey relationship. Known to take smaller fish [Winemiller, 1989]	**0**
Rhamdia quelen	*Poecilia reticulata*	Potential weak predator-prey relationship. Known to take smaller fish [Winemiller, 1989]	**+** **0.12**

Table 2: Comparison between interactions that we could expect to translate into associations and those that are actually found in the ZIPLN(Abundance ~ 1 + PC1 + PC2 + disturbance | 1 + stream) network.

There were also examples of unexpected associations in the network. In some cases, mechanisms other than interactions, such as similar habitat preference, could be responsible. For instance, *Rhamdia quelen* (silver catfishes) and *Hypostomus robinii* (tetas) (positive association of 0.15) are both benthic dwellers (they live near river beds) [Phillip and Ramnarine, 2001] and may prefer sites with overhangs along the banks. Pike cichlids and blue coscorobs (association of 0.16 in the Northern Range network) both tend to prefer clearer and deeper waters [Phillip and Ramnarine, 2001].

In the inferred network, a third of the species display no associations at all. This might partly be a statistical artefact, a phenomenon that can be explained by the small populations and limited distributions of many species. This means that there is less signal to be exploited for these species and hence makes it harder to identify patterns in their distributions that could unveil underlying mechanisms. Some species may also be very localised and therefore have little chance of displaying statistical correlations with more widespread species. This relates to the difficulties SDMs and JSDMs can have more generally in modelling rare species [Pollock et al., 2014, Zhang et al., 2019].

Thus, we have shown how, in this example, the ZIPLN-network analysis allows one to make several hypotheses regarding the biotic and abiotic mechanisms that shape community structure. It can also be used to understand how pre-identified biotic interactions relate to species distributions..

Influence of covariates

How does the choice of covariates influence the results? In JSDMs, covariates are generally used to account for abiotic effects. In our case, including the streams as covariates also makes it possible to add a dispersal effect since the model then accounts for the fact that some species may never be found in a given stream.

As might be expected, Figures 3 and 5 demonstrate that changing the covariates of the model means that the signal is captured differently. For instance, Figures 3.2 and 3.3 show that the variance of the regression coefficients variance for *PC1* and *PC2* decreases when adding a stream covariate (by 23 %

for PC1, and 34 % for PC2). Thus, in this system, the signal is clearer when the stream effect is accounted for, than with just the abiotic parameters.

Similarly Figure 5 shows that the association networks change when the set of covariates is modified. However, most of the strongest associations seem robust to the addition of new covariates. This stability confirms that the PLN and ZIPLN models are robust; it supports the fact that these associations correspond to strong patterns in the observations.

One striking change is the disappearance of the strong negative association between the apex predator guabine and its prey *Rivulus hartii* (jumping guabine) when a stream covariate is added. This association probably reflects strong predator-avoidance behaviour. Indeed, Fraser and Gilliam (1992) analysed the predator-prey relationship between both species and found that the presence of guabines suppressed jumping guabines' egg production by 50 %. Jumping guabines have excellent jumping abilities (hence their common name), and can survive out of water for extended periods, owing to cutaneous breathing [Fraser, pers. obs.] and strong jumping abilities. They can easily disperse overland on wet nights to colonize remote bodies of water free of piscivorous fish [Fraser et al. 1995]. These strategies allow them to concentrate into sites that are inaccessible to piscivorous fishes. A hypothesis that explains the absence of this association when accounting for a stream effect is the fact that the predator-prey dynamics induce a concentration of jumping guabines in sites where guabines cannot be found, a pattern that is caught in the zero-inflation, stream-related part of the model. Looking back on the abundance heatmap (Figure 2), it indeed appears that jumping guabines seem to thrive more in sites where the guabines are completely absent. This example shows that the same effect can be represented in different ways by the model depending on which covariates are included.

Therefore, while statistical criteria can be used to define a "best" choice of covariates, ecological knowledge plays a crucial role in the interpretation of the resulting networks. Taking account of both

abiotic effects and association networks strengthens ecological understanding and helps formulate ecologically relevant hypotheses to explain distribution patterns.

Using the ZIPLN-network approach

When and how should the ZIPLN-network model be used? Species occurrence data sets often include many zeros. Some of these absences can be explained by abiotic effects or by the presence of other species; both mechanisms are included in existing JSDM [Ovaskainen et al., 2027, Pollock et al., 2014]. However, other absences could be due to failure to detect a species that is actually present , or because of natural barriers that prevent a species from reaching a site that is otherwise suitable. Including a zero-inflation statistical effect is an efficient way to deal with these different reasons why a dataset suggests a species is not present.

For instance Figure 4.3 shows that several species have a high probability of being absent from certain streams, regardless of their abiotic conditions. While PLN-network can potentially catch these effects through high negative stream coefficients (Figure 5.3), using the zero-inflation allows the user to make a clearer distinction between effects that induce a low abundance of a given species, and mechanisms that lead it to be rarely or never found in a given place.

The comparison of association networks (Figure 5) shows that accounting for zero-inflation filters out several low-intensity associations. It is likely that some of the associations thus removed were due to similar (for positive associations) or antagonistic (for negative associations) zero-inflation patterns. For instance, the swordtail sardine and hunchback sardine have a positive association in the PLN networks that disappears when adding zero-inflation (Figure 5). Figures 1 and 4.3 show that these species tend to be absent from the same streams suggesting that their common presence in the remaining streams might explain the positive association. However, the ZIPLN networks do not show this association. This means that once the fact they are present in only the same subset of streams has been accounted for, their abundances do not seem further correlated.

This does not mean that the ZIPLN network associations are necessarily more representative of biotic interactions. Indeed, one cannot rule out the hypothesis that two species having a common zero-inflation pattern could be explained by a biotic mechanism. However, comparing the networks obtained with and without the zero-inflation allows one to make more precise interpretations since it shows that it is only the two species' presence that seem correlated and not their abundances, as the PLN association network could imply.

We conclude that the ZIPLN-network is a useful tool for analysing species distribution data as it models both presence-absence with zero-inflation, along with species abundances, with the Poisson distribution modelling discrete counts. We have shown that comparing the results obtained with PLN-network and ZIPLN-network helps better disentangle the assemblage mechanisms at stake. Finally, it can help distinguish dispersal effects from other mechanisms by accounting for the absence of species in sites they did not reach.

We have shown how the ZIPLN-network approach can be used to produce more informed analyses of species distribution data. Accounting for zero-inflation and using both presence-absence and count data can filter out spurious associations and help distinguish different mechanisms. We argue that such a model is better used in combination with its non-zero-inflated counterpart so as to better identify putative ecological mechanisms. Similarly, comparing the results obtained with different combinations of covariates allows a more nuanced analysis. As with JSDMs in general, the ZIPLN-network approach can help ecologists decipher community assembly rules, but at the same time underlines the importance of fieldwork.

References

John Aitchison and CH Ho. The multivariate poisson-log normal distribution. Biometrika, 76(4):643–653, 1989.

Bastien Batardière, Julien Chiquet, François Gindraud and Mahendra Mariadassou. Zero-inflation in the multivariate poisson lognormal family. 2024.

Bastien Batardière, Joon Kwon and Julien Chiquet. pyPLNmodels: A Python package to analyze multivariate high-dimensional count data. Journal of Open Source Software 9.104 (2024): 6969, 2024.

David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. Journal of the American statistical Association, 112(518):859—877, 2017.

MS Botham, CJ Kerfoot, Vasilis Louca, and J Krause. The effects of different predator species on anti predator behavior in the trinidadian guppy, poecilia reticulata.Naturwissenschaften, 93:431–439, 2006.

Marc W Cadotte and Caroline M Tucker. Should environmental filtering be abandoned?Trends in ecology & evolution,32(6):429–437, 2017.

Julien Chiquet, Mahendra Mariadassou, and Stéphane Robin. Variational inference of sparse network from count data. Proceedings of Machine Learning Research, 97:1162–1171, 2019.

Julien Chiquet, Mahendra Mariadassou, and Stéphane Robin. The Poisson log-normal model as a versatile framework for the joint analysis of species abundances. Frontiers in Ecology and Evolution, 9:588292, 2021.

Edward F. Connor, and Daniel Simberloff. The assembly of species communities: chance or competition? – Ecology 60: 1132–1140, 1979.

Gabor Csardi and Tamas Nepusz. “The igraph software package for complex network research.” _InterJournal_, *Complex Systems*, 1695. <https://igraph.org>, 2006.

Ross B. Cunningham and David B. Lindenmayer. Modeling count data of rare species: some statistical issues. Ecology, 86(5), 1135-1142, 2005.

Manuela D'Amen, Heidi K. Mod, Nicholas J. Gotelli, and Antoine Guisan, 2018. Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co‐occurrence. Ecography, 41(8), 1233-1244.

Amy E. Deacon, Faith A. M. Jones and Anne E. Magurran. Gradients in predation risk in a tropical river system. Current Zoology, 2018, vol. 64, no 2, p. 213-221.

Amy E Deacon, Hideyasu Shimadzu, Maria Dornelas, Indar W Ramnarine, and Anne E
Magurran. From species to communities: the signature of recreational use on a tropical river ecosystem. Ecology and Evolution, 5(23):5561–5572, 2015.

Jared M. Diamond. Assembly of species communities. – In: Cody, M. L. and Diamond, J. M. (eds), Ecology and evolution of communities. Harvard Univ. Press, pp. 342–344, 1975.

Jane Elith and John R Leathwick. Species distribution models: ecological explanation and prediction across space and time.Annual review of ecology, evolution, and systematics, 40(1):677–697, 2009.

Jeremy Fox. Mathematical constraints in ecology and evolution, part 1: species can't all covary negatively (updated).
https://dynamicecology.wordpress.com/2016/10/05/mathematical-constraints-in-ecology-part-1-species-cant-all-covary-really-negatively/,2016. Accessed: 2024-11-13.

Douglas F. Fraser and James F Gilliam. Non lethal impacts of predator invasion: facultative suppression of growth and reproduction. Ecology, 73(3):959–970, 1992.

Douglas F. Fraser, James F. Gilliam, and Trevor Yip‐Hoi. Predation as an agent of population fragmentation in a tropical watershed. Ecology, 76(5), 1461-1472, 1995.

Douglas F. Fraser, James F. Gilliam, Jenna T. Akkara, Brett W. Albanese, & Sunny B. Snider, S. B. Night feeding by guppies under predator release: effects on growth and daytime courtship. Ecology, 85(2), 312-319, 2004.

Douglas F. Fraser, James F. Gilliam, Brett W. Albanese & Sunny B. Snider . Effects of temporal patterning of predation threat on movement of a stream fish: evaluating an intermediate threat hypothesis. Environmental Biology of Fishes, 76, 25-35, 2006.

Douglas F. Fraser, and Bradley A., & Lamphere. Experimental evaluation of predation as a facilitator of invasion success in a stream fish. Ecology, 94(3), 640-649, 2013.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso.Biostatistics, 9(3):432—441, 2008.

James F Gilliam, Douglas F Fraser, and Mary Alkins-Koo. Structure of a tropical stream fish community: a role for biotic interactions. Ecology, 74(6):1856–1870, 1993.

E. Goodale, G. Beauchamp, R. D., Magrath, J. C. Nieh and G. D. Ruxton. Interspecific information transfer influences animal community structure. Trends in ecology & evolution, 25(6), 354-361, 2010.

Lars Götzenberger, Francesco de Bello, Kari Anne Brathen, John Davison, Anne Dubuis, Antoine Guisan, et al., Ecological assembly rules in plant communities—approaches, patterns and prospects. Biological reviews, 87(1):111–127, 2012.

Gotelli, N. J. and Graves, G. R. Null models in ecology. – Smithsonian Inst., Washington, 1996.

Antoine Guisan and Wilfried Thuiller. Predicting species distribution: offering more than simple habitat models.Ecology letters, 8(9):993–1009, 2005.

Caryl P Haskins, Edna F Haskins, John JA McLaughlin, Richard E Hewitt, Polymorphism and population structure in Lebistes reticulatus, an ecological study.Vertebrate speciation, 320:395, 1961.

Julia Helena Heinen, Carsten Rahbek, and Michael Krabbe Borregaard. Conservation of species interactions to achieve self-sustaining ecosystems.Ecography, 43(11):1603–1611, 2020.

Jan Johansson, Håkan Turesson, and Anders Persson. Active selection for large guppies, poecilia reticulata, by the pike cichlid, crenicichla saxatilis.Oikos, 105(3):595–605, 2004.

Faith AM Jones, Mike G Rutherford, Amy E Deacon, Dawn AT Phillip, and Anne E Magurran. Quantifying regional biodiversity in the tropics: A case study of freshwater fish in Trinidad and Tobago. Biotropica, 51(5):700–708, 2019.

Nathan JB Kraft, Peter B Adler, Oscar Godoy, Emily C James, Steve Fuller, and Jonathan M Levine. Community assembly, coexistence and the environmental filtering metaphor. Functional ecology, 29(5):592–599, 2015.

Jens Krause and P. Krause. Living in Groups, 2002.

M. K. Lau, S. R. Borrett, B. Baiser, N. J. Gotelli, and A. M. Ellison. Ecological network metrics: opportunities for synthesis. Ecosphere, 8, 2017.

Steffen L Lauritzen.Graphical models, volume 17. Clarendon Press, 1996.

Han Liu, Kathryn Roeder, and Larry Wasserman. Stability approach to regularisation selection (stars) for high dimensional graphical models. Advances in neural information processing systems, 23, 2010.

Anne E Magurran.Evolutionary ecology: the Trinidadian guppy. Oxford University Press, USA, 2005.

Anne E Magurran, Amy E Deacon, Faye Moyes, Hideyasu Shimadzu, Maria Dornelas, Dawn AT Phillip, and Indar W Ramnarine. Divergent biodiversity change within ecosystems.Proceedings of the National Academy of Sciences, 115(8):1843–1847, 2018.

Jennifer Miller. Species distribution modeling.Geography Compass, 4(6):490–509, 2010.

Raphaëlle Momal, Stéphane Robin, and Christophe Ambroise. Accounting for missing actors in interaction network inference from abundance data.Journal of the Royal Statistical Society Series C: Applied Statistics, 70(5):1230–1258, 2021.

Victoria Nolan, Francis Gilbert & Tom Reader. Solving sampling bias problems in presence–absence or presence‐only species data using zero‐inflated models. Journal of Biogeography, 49(1), 215-232, 2022.

Robert O'Hara and Johan Kotze. Do not log-transform count data. Nature Precedings, pages 1–1, 2010.

Otso Ovaskainen, Jenni Hottola, and Juha Siitonen. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. – Ecology 91: 2514–2521, 2010.

Dawn A. T. Phillip and Indar W. Ramnarine. An illustrated guide to the freshwater fishes of Trinidad and Tobago.Darwin initiative, 2001.

Dawn A. T. Phillip. Biodiversity of freshwater fishes of Trinidad and Tobago, West Indies. PhD thesis,University of St Andrews, 1998.

Dawn A.T. Phillip, Donald C Taphorn, Erling Holm, James F Gilliam, Bradley A Lamphere, and Hernan Lopez-Fernandez. Annotated list and key to the stream fishes of trinidad & tobago.Zootaxa, 3711(1):001–064, 2013.

Giovanni Poggiato, Tamara Münkemüller, Daria Bystrova, Julyan Arbel, James S Clark, and Wilfried Thuiller. On the interpretations of joint modeling in community ecology. Trends in ecology & evolution, 36(5):391–401, 2021.

Laura J Pollock, Reid Tingley, William K Morris, Nick Golding, Robert B O'Hara, Kirsten M Parris, et al. Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (jsdm).Methods in Ecology and Evolution, 5(5):397–406, 2014.

Gleb Tikhonov, Øystein H Opedal, Nerea Abrego, Aleksi Lehikoinen, Melinda MJ de Jonge, Jari Oksanen, and Otso Ovaskainen. Joint species distribution modelling with the r-package hmsc.Methods in ecology and evolution, 11(3):442–447, 2020.

Stephanie M. Tran, Kaitlyn J. Howell & Matthew R. Walsh,. (2024). Increased eye size is favoured in Trinidadian killifish experimentally transplanted into low light, high competition environments. Journal of Evolutionary Biology, 37(8), 960-966.

Werner Ulrich and Nicholas J Gotelli. Null model analysis of species associations using abundance data.Ecology, 91(11):3384–3397, 2010.

Alfonso Valiente-Banuet, Marcelo A Aizen, Julio M Alcántara, Juan Arroyo, Andrea Cocucci, Mauro Galetti, et al. Beyond species loss: the extinction of ecological interactions in a changing world.Functional Ecology, 29(3):299–307, 2015.

Duarte S Viana and Jonathan M Chase. Spatial scale modulates the inference of metacommunity assembly processes. Ecology, 100(2):e02576, 2019.

Martin J Wainwright, Michael I Jordan, et al. Graphical models, exponential families, and variational inference.Foundations and Trends in Machine Learning, 1(1–2):1–305, 2008.

Matthew R Walsh and David N Reznick. Phenotypic diversification across an environmental gradient: a role for predators and resource availability on the evolution of life histories.Evolution, 63(12):3201–3213, 2009.

Matthew R Walsh, Douglas F Fraser, Ronald D Bassar, and David N Reznick. The direct and indirect effects of guppies: implications for life-history evolution in rivulus hartii. Functional Ecology, 25(1):227–237, 2011.

David I Warton, Guillaume Blanchet, Robert B O'Hara, Otso Ovaskainen, Sara Taskinen, Steven C Walker, and Francis KC Hui. So many variables: joint modeling in community ecology.Trends in ecology & evolution, 30(12):766–779, 2015.

Seth J. Wenger, & Mary C. Freeman, M. C. Estimating species occurrence, abundance, and detection probability using zero‐inflated distributions. Ecology, 89(10), 2953-2959, 2008.

David P. Wilkinson, Nick Golding, Gurutzeta Guillera‐Arroita, Reid Tingley and Michael A. McCarthy "A comparison of joint species distribution models for presence–absence data." Methods in Ecology and Evolution 10.2 (2019): 198-211.

Kirk O. Winemiller. Ontogenetic diet shifts and resource partitioning among piscivorous fishes in the Venezuelan ilanos. Environmental Biology of fishes, 26, 177-199, 1989.

Chonglian Zhang, Yong Chen, Binduo Xu, Ying Xue, Y.iping & Ren. How to predict biodiversity in space? An evaluation of modelling approaches in marine ecosystems. Diversity and Distributions, 25(11), 1697-1708, 2019.

Damaris Zurell, Laura J. Pollock, and Wilfried Thuiller. "Do joint species distribution models reliably detect interspecific interactions from co‐occurrence data in homogenous environments?." Ecography 41.11: 1812-1819, 2018.

License details for species silhouettes that are not in the public domain:

For Ancistrus maracasae, Andinoacara pulcher, Hemibrycon taeniurus, Hoplias malabaricus, Hypostomus robinii, Roeboides dientonito,: credits to Grupo de Ictiología de la Universidad de Antioquia. https://creativecommons.org/licenses/by/4.0/

For Rhamdia quelen, Crenicichla frenata: credits to Cesar Julian.
https://creativecommons.org/licenses/by/3.0/

Appendices

Appendix A: Correspondence between full species names and their shortened versions.

Shortened name	Full name	Common name
Ancimara	*Ancistrus maracasae*	Jumbie teta
Andipulc	*Andinoacara pulcher*	Blue coscarob
Angurost	*Anguilla rostrata*	American eel
Arconudu	*Arcos nudus*	Padded clingfish
Astybima	*Astyanax bimaculatus*	Two spot sardine
Awaobana	*Awaous banana*	River goby
Cichtaen	*Cichlasoma taenia*	Brown coscarob
Coryaene	*Corydoras aeneus*	Bronze cory
Coryriis	*Corynopoma riisei*	Swordtail sardine
Crenfren	*Crenicichla frenata*	Pike cichlid
Dajamont	*Dajaus monticola*	Mountain mullet
Eleopiso	*Eleotris pisonis*	Spinycheek sleeper
Gastster	*Gasteropelecus sternicla*	Silver hatchet fish
Gobidorm	*Gobiomorus dormitor*	Bigmouth sleeper
Gymncara	*Gymnotus carapo*	Banded knifefish
Hemitaen	*Hemibrycon taeniurus*	Mountain stream sardine
Hemiunil	*Hemigrammus unilineatus*	Featherfin tetra
Hopllitt	*Hoplosternum littorale*	Cascadu
Hoplmala	*Hoplias malabaricus*	Guabine
Hyporobi	*Hypostomus robinii*	Teta
Odonpulc	*Odontostilbe pulchra*	Tetra
Oreosp	*Oreochromis* species (information given at the genus level)	Tilapia
Poecboes	*Poecilia boesemani*	Maraval toothcarp
Poecreti	*Poecilia reticulata*	Guppy
Poecvivi	*Poecilia vivipara*	Eye spot toothcarp
Polyscho	*Polycentrus schomburgkii*	Guyana leaffish
Rhamquel	*Rhamdia quelen*	Silver catfish
Rivuhart	*Rivulus hartii (= Anablepsoides hartii)*	Jumping guabine

Roebdien	*Roeboides dientonito*	Hunchback sardine
Sicypunc	*Sicydium punctatum*	Tri Tri
Steiarge	*Steindachnerina argentea*	Stout sardine
Synbmarm	*Synbranchus marmoratus*	Zangee