One Toolbox, Many Tools: A Practitioner's Guide to Model-Based Ordination for Community Ecology

Audun Rugstad, 0009-0003-7084-4831[1*], Bert van der Veen, 0000-0003-2263-3880[1,3], Robert Brian O'Hara, 0000-0001-9737-3724[1], and Anne Catriona Mehlhoop, 0000-0002-0748-7515[2]

[1]Department of Mathematical Sciences, Norwegian University of Science and Technology, NO
[2]Norwegian Institute for Nature Research, NO
[3]University of Bayreuth, DE

July 31, 2026

This article presents the case for model-based ordination by way of Generalized Linear Latent Variable Models (GLLVMs) as a go-to choice of statistical method for any community ecologist wanting to tackle a range of present-day ecological research questions. Model-based ordination brings tools and capabilities from classic (mixed-effects) regression models to multivariate community analysis, providing a number of novel ways to tailor models specifically to one's study questions and data properties not available when using non-model-based multivariate methods. In order to facilitate further adoption of these methods by community ecologists, we provide 1) a practitioner-focused and practical overview of the advantages model-based ordination brings to the table when addressing different core ecological questions, 2) a number of concrete suggestions for how model-based ordination best can be incorporated into the analytical workflow of community ecologists, and 3) two illustrative worked examples of this workflow in action on real-world data.

*Corresponding author. Contact: audun.rugstad@ntnu.no

Keywords: Ecological modelling, Multispecies data, Community ecology, Community modelling, Ordination, Data exploration, Model selection, Model-based workflow, Latent variable modelling, Invasive species, Ecological restoration

1 Introduction

Using different types of data to improve our understanding of the nature and dynamics of ecological communities is becoming increasingly important in a range of real-world scenarios. Examples include assessing restoration success (Ribeiro et al., 2023), the impacts of invasive species (Souza-Alonso et al., 2022; Herrmann et al., 2022), and the modelling of community responses to climate change (Sahade et al., 2015). In all of these cases, how well one's ecological research questions can be addressed depends not only on data, but also on the selection of appropriate tools and methods for analysis. And while the statistical toolbox available to ecologists today is large, it is also fragmented, which can make it difficult to choose a set of methods to address the relevant research questions in a way that is coherent, streamlined and reproducible.

One important example of this is the fact that community ecologists today often find themselves juggling two quite different methodological frameworks when addressing different ecological questions. On the one hand, univariate ecological phenomena such as predation rates, breeding success, or the abundance of a single species in different habitats is typically studied with model-based methods within the overarching framework of Generalized Linear Mixed Models (GLMM) (Bolker et al., 2009; Zuur et al., 2009). On the other hand, the same type of model-based approach has historically not been available to perform multivariate analysis on species community data. Instead, questions about community composition and -structure have typically been addressed through classical ordination methods, due to their ability to effectively condense and visualize patterns in multivariate data (ter Braak and Šmilauer, 2015). By classical ordination is meant either algorithmic methods such as Non-Metric Multidimensional Scaling (NMDS), or eigenvalue-based methods such as Principal Component Analysis (PCA), Correspondence Analysis (CA), or Principal Coordinate Analysis (ter Braak and Prentice, 2004), used to visualise differences in species composition in a low-dimensional space. Some ordination methods, so-called constrained ordination, as opposed to unconstrained ordination, also allow for the incorporation of observed environmental variables. The most notable examples being Canonical Correspondence Analaysis (CCA) and Redundancy analysis (RDA) (ter Braak and Šmilauer, 2015), in which ordination axes are constrained to be linear combination of the measured environment. In this way, the methods can also facilitate more direct study of multivariate species-environment relationships.

Both unconstrained and constrained classical ordination methods have historically proved to be exceedingly useful tools for community ecology. However, they also come with some key limitations. In general, classical ordination do not provide the same capabilities for inference and diagnostics as GLMMs applied to univariate data analysis in community ecology (Warton

et al., 2015). This means that they cannot be use to make predictions in time, space or on novel environmental conditions, provide estimates of uncertainty for any of the components of the ordination (such as site scores or species loadings), or provide reliable tools for diagnosing lack-of-fit to the data. The latter point especially is made all the more relevant by the fact that classical ordination methods are known to be vulnerable to artefacts and distortions, and often fail to account for key properties of ecological data (Warton et al., 2012; Goodall, 1954; Podani and Miklós, 2002; ter Braak and Šmilauer, 2015). A more conceptual limitation of many ordination methods is also the fact that they are a "one way street" for the observed species data, in the sense that the species data is collapsed (typically into a distance matrix) in order to construct the ordination. This makes ecological inference back to the species level from these methods harder, and nearly impossible to validate. For all of the above reasons, many have argued that the use of classical ordination for answering ecological questions outside of data exploration and hypothesizing is limited (Warton and Hui, 2017; Warton et al., 2015; Jupke and Schäfer, 2020).

In recent years, several efforts have been made to bridge the gap between univariate and multivariate analysis in ecology by developing new methods for analysing community ecological data using the same model-based framework that exists for univariate data (e.g. Hui et al., 2015; Niku et al., 2019; Ovaskainen et al., 2017). One of the main products of this effort has been the development of Generalized Linear Latent Variable Models (GLLVMs), based on Generalized Linear Mixed Models (GLMMs), as a model-based counterpart to classical ordination methods. As the name suggests, GLLVMs allow users model ecological communities with latent variables, i.e. by assuming that patterns of species co-occurrence observed in the data are explained by a few underlying *latent*, or unobserved, explanatory variables, which correspond to ordination axes in the classical terminology. Estimated values for the latent variables together with the estimated species responses to the latent variables, conventionally called the species *loadings*, then form an ordination, which in turn informs the linear predictor for the abundance or occurrence of each species. In this way, GLLVM fits into a long history of latent variable-based thinking in community ecology, in which both constrained and unconstrained classical methods have been shown to be approximations of the maximum likelihood solutions for specific latent variable models in special cases (Jong and Kotz, 1999; Braak, 1985; Ter Braak and Barendregt, 1986; van der Veen, 2022; ter Braak, 1987; Reinsel et al., 2022; Yee, 2004).

The fact that GLLVMs situate model-based ordination in the context of other established regression models designed to predict species occurrence or abundances not only addresses many of the limitations with classical methods, but also bring many new, fundamental advantages to community analysis and inference. Firstly, GLLVMs allow for a greater flexibility and possibilities in the *types* of ordinations that can be fitted to the data. As indicated by Figure 1 and discussed in more detail in Section 3, latent variables can be estimated both freely (i.e. unconstrained); that is, based solely on the species co-occurrences in the sites (Hui et al., 2015), as well as fully or partially estimated based on environmental variables through canonical coefficients (van der Veen et al., 2023). The fact that all of these different ordination types are fitted within the same model-based framework, rather than through separate types of algorithms (for example, an unconstrained NMDS ordination versus a constrained RDA or

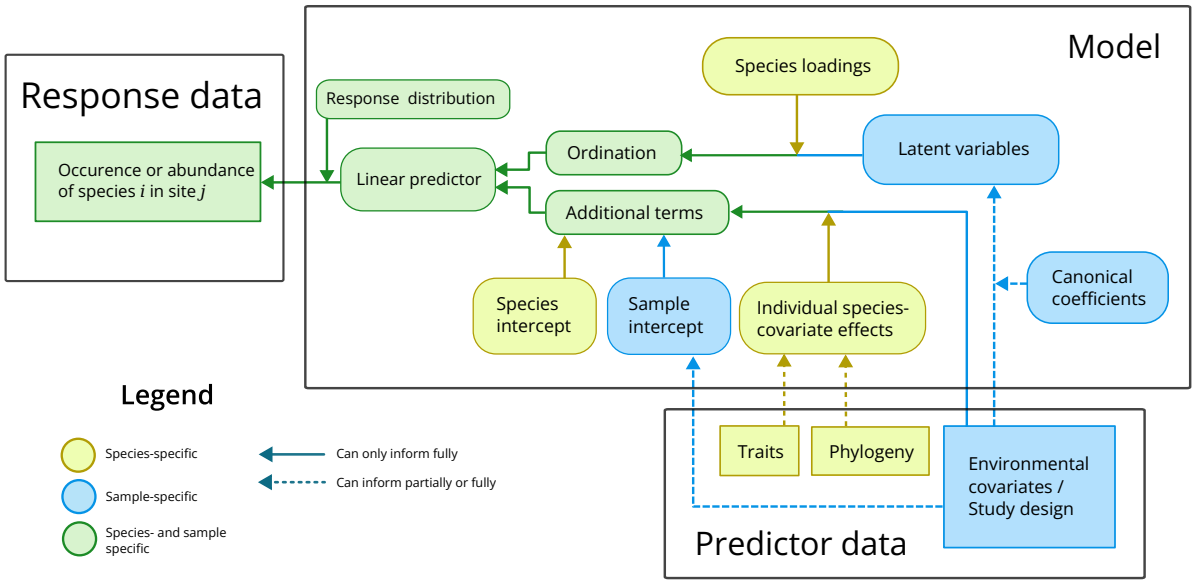

Figure 1: Graphical overview of the model structure of a Generalized Linear Latent Variable Model (GLLVM), as implemented in the `gllvm` R package. Data and model components are colour coded based on whether they are species specific, sample specific or both species- and sample specific. "Partially informed" means that the model component can either be fully determined based on the data, be determined by the data in addition to freely estimated (random effect) variation, or only determined by freely estimated variation. For example, the sample intercept can be either estimated completely freely as a random effect, be determined completely as a response to a study design variable (e.g. block in a block design), or be a combination of both. Conversely, fully determined means that the model component can only be determined solely as a response to the data, if included (e.g. the individual species-covariate effects). With "additional terms" is meant all model effects that are not part of an ordination. The figure is inspired by Figure 4 from Ovaskainen et al. (2017).

CCA ordination), adds to the interpretability and flexibility of GLLVMs as a one-stop toolbox for community ecology.

Secondly, GLLVMs give users a large amount of flexibly and freedom to combine model-based ordinations with other model components which can be specified and combined in several different ways to produce many different types of models for species communities (see the overview of the model structure in Figure 1). These additional terms can include everything from simple species- and site intercepts in order to standardize the ordination to modelling species composition, as well as complex species-specific (i.e. full-rank) effects that can potentially also incorporate phylogenetic correlation and environment-trait interaction (i.e. "fourth corner") terms (Niku et al., 2021; van der Veen and O'Hara, 2025).

GLLVMs are implemented in several software packages. The `gllvm` `R` package (Niku et al., 2025) is aimed at community ecologists, and contains by far the richest toolbox for model-based ordination. The `R` package `hmsc` (Tikhonov et al., 2025) is focused on GLLVMs for Joint Species Distribution Modeling, and has a similarly rich toolbox for this purpose.. Other noteable R software implementations include `ecoCopula` (Popovic et al., 2019) and `boral`. (Hui, 2025). Different types of GLLVMs might also be fitted using more general-purpose GLM(M) packages such as `VGAM` (Yee, 2025) and `glmmTMB` (McGillycuddy et al., 2025), as well as packages for Bayesian hierarchical models like `Stan` (Stan Development Team, 2026b) and `nimble` (de Valpine et al., 2017).

Despite the availability of user-friendly software, as well as several examples of model-based ordination through GLLVMs being used successfully in the ecological literature (see e.g. Lam-Gordillo et al., 2025; Daudt et al., 2025; Wong et al., 2026), the uptake of model-based ordination methods in community ecology has so far been relatively slow. Classical ordination methods are still the dominant methodology for multivariate analysis; as indicated by the number of monthly downloads for the `vegan` `R`-package being much more than an order of magnitude greater than for packages that focus on model-based ordination, and seeing little change over the past five years (Appendix S1, however note that the `vegan` `R`-package is more multipurpose; ordination is only a part of its features). In our experience, two potential barriers for improved uptake seem especially important. The first is a lack of accessible arguments and evidence for why model-based ordinations make it possible to obtain better and more reliable ecological inference from one's data as compared to classical ordination methods. The second is a lack of instructive examples on real ecological data.

This article sets out to help remove these two barriers by providing a focused and practically oriented guide to the tools and capabilities of the GLLVM framework for model-based ordination, aimed at the types of ecological questions that may be especially relevant to current users of classical ordination methods. The text is divided into four parts: 1) An overview of what we consider to be the most important benefits of using model-based ordination for statistical analyses in community ecology, 2) how the methods can be used more concretely to address different types of ecological questions; both with and without observed environmental covariates, 3) a suggestion for a modelling workflow when using model-based ordination to address these questions, and 4) a demonstration of this workflow on two relevant, real-world data sets. Throughout the paper, the phrase "model based ordination" will always refer to ordinations fitted using GLLVMs.

2 Fundamental benefits of model-based ordination

Because GLLVMs are an extension of Generalized Linear Mixed-effects models, model-based ordinations offer the same capabilities as GL(M)Ms for specifying, fitting, interpreting and comparing models. Here, we highlight six of the most substantial benefits that this brings to

the analysis of multi-species community data. These benefits are applicable regardless of the specific ecological questions asked.

1. **Accounting for different types of data** Model-based ordination lets community ecologists analyse data as is, without data transformation or manipulation. As with GLMMs, this is done by specifying a suitable response distribution for the data, and by specifying the model's structure to match the study system or experimental design at hand. Most model-based ordination software includes a variety of different response distributions, making it possible to model data recorded as presence-absence, counts, percentage cover, cover classes, biomass, and more. See Table A1 in Korhonen et al. (2025) for an overview of the response distributions available in the `gllvm` package.

2. **Assessing model fit to the data** Sound ecological inference requires modelling assumptions to be met. Because of the distribution specification, model-based ordinations can leverage well-established tools for diagnostics to evaluate whether the chosen response distribution is a good fit to the data (See point 2) to improve the ordination. The fit of any model-based ordination can be assessed using diagnostic plots and metrics familiar from the GLMM framework, such as residual versus fitted plots or Q-Q plots. In the `gllvm` package, the metrics used are randomized quantile residuals, as in the `DHARMa` package (Dunn and Smyth, 1996; Hartig, 2024). Residual diagnostics can for example be used to check if there are systematic trends, zero-inflation or over-/underdispersion in the data not accounted for by the model. For example, when the observed data type are counts, residual or QQ-plots will indicate if a Poisson assumption is applicable. If the model predicts too few zeros relative to the data, it might be more reasonable to switch to a zero-inflated Poisson distribution or a negative-binomial distribution. Conversely, the validity of one's assumptions cannot be checked simply by looking at the ordination. A good model may generate a poor looking ordination for perfectly valid reasons, or a poor model may produce a good looking ordination for invalid reasons. That is to say that model misspecification may not affect the ordination, although there is a body of evidence to suggest that it certainly can (see Warton et al., 2012; Warton and Hui, 2017, for the case of the mean-variance relationship and NMDS/DCA)

3. **Accounting for sampling properties and study designs** In general, model-based ordinations offer the same tools as GLMMs to account for properties of the sampling and study design, such as block- and hierarchical sampling designs, or differences in the read depth of samples in the case of DNA meta-barcoding data, which are not available for traditional multivariate methods. This can be done through fixed and random effects, nesting of effects, offsets or other changes to the model's structure. For example, blocks in a randomized block design can be included as a random effect outside of a model-based ordination, to filter its effects from patterns of interest in the ordination.

By adjusting the model and not the data, model-based ordinations remove the need for data transformation prior to analysis. As an example, DNA sequencing data often suffers from variable read depths as well as double-zeros. In a model-based ordination,

the variable read depths can be accommodated by augmenting the model with offset terms and/or row effects. The classic "double-zero" problem is implicitly addressed by operating in a model-based framework, as zeros of two species at a site are differently weighted by the model, through their mean-variance relationships. For more on different types of model specifications, see also Figure 1 and section 3.

4. **Model comparison** Model-based ordination allows for the use of a range of different goodness-of-fit statistics to compare the relative fit and predictive power of different models for species composition. For ecologists, Information Criteria like AIC and BIC, or area under the curve (AUC), will perhaps be the most familiar of these. Depending on the goal of the analysis, AIC or BIC can be used to determine the ideal set of observed predictor variables, or to determine the number of unobserved latent variables that best represent the data, given that the model is valid (see Point 2 in this section, as well as Section 4). Classical counterparts to this are e.g. the use of stress to determine the number of dimensions in an NMDS ordination Dexter et al. (2018), or the use of pseudo-AIC in methods such as Canonical Correspondence Analysis (CCA) and Redundancy Analysis (RDA).

5. **Estimation and visualisation of uncertainty** Because model-based ordinations are fitted using either (marginal) Maximum Likelihood estimation or with Bayesian methods, all parameters and fitted values estimated the model will have a measure of uncertainty associated with them, which are always estimated jointly as part of the model fitting itself. This includes uncertainties associated with site scores, species loadings, canonical coefficients and other model terms. These uncertainties can then be visualized, e.g. by plotting confidence or prediction regions in an ordination diagram or confidence intervals in a coefficient plot, and be used to make statements about statistical significance, or alternatively, the "strength of evidence", of different model components (see Muff et al., 2022).

   Visualising uncertainties rather than simply reporting significance values from a statistical test may however help paint a more complete and truthful picture of the model results in general (Ho et al., 2019), and the fact that they are analogous to standard errors, confidence intervals and prediction intervals in other GL(M)Ms, helps make them interpretable and comparable across studies. Another strength of uncertainty quantification in model-based ordination is that is it also always directly translatable into uncertainties for the prediction of individual species occurrences in the data.

6. **Prediction** As statistical models, model-based ordinations can also be used to predict or forecast, with associated uncertainty. This opens up many new possibilities for community ecologists, not available when applying traditional ordination methods. It allows practitioners to predict different species' abundances based on covariates included in the model, or based on an unconstrained ordination axis.

   Predictions can be used both to validate the goodness-of-fit of the model for different species in the input data, validate how well the predicted species community of a given

habitat type fits with newly collected data, or to predict changes in composition under new environmental conditions, such as different climate change scenarios or time since disturbance (see Section 5.2).

3 Using the framework to answer ecological questions

The main strength of the model-based ordination framework for ecologists lies in its capability to provide in-depth answers to questions about the composition and structure of ecological communities. This includes questions about which species co-occur and which factors (habitat types, climatic variables, time etc.) best explain observed patterns of composition or co-occurrence. Among the most important tools to help researchers address these questions are the many options to effectively visualize the outputs of model-based ordinations. Depending on the model and the goals of analysis, these can combine information from environmental-, species- and sample- specific parameters related to the latent variables. Figure 2 provides a general overview of the most relevant types of visualisations of the different ordination-related model parameters shown in Figure 1.

This section is grouped into two parts: The first part focuses on questions that can be addressed by models only considering species observations, the second section focuses on questions involving measured environmental variables and ecological communities. However, it is important to bear in mind that contemporary community ecology studies often address multiple ecological questions simultaneously, sometimes by including both analyses on species composition alone and species composition in combination with environmental predictors. As such, the methods in the literature examples given between Section 3.1 and Section 3.2 will sometimes overlap.

3.1 Species composition data

When information on the environment is absent, model-based ordination can be a powerful tool for exploring basic patterns in a multispecies dataset. As with traditional methods, an unconstrained model-based ordination can be fitted to the species data alone, and patterns can be inferred from visualisation of the results. This basic model will return site scores and species loadings. These can then be used to make inferences about site conditions, transitions between community types, and which species associations drive these patterns. Conceptually, if we view the latent variable(s) as estimates of unobserved environmental gradients, the species loadings represent the slopes, or the species response, of each species to the gradient(s) represented by the latent variables, similar to their response to predictor variables in a regression. The site scores then represent the specific values of these unobserved predictor variables, calculated for each sample. As such, the latent variables are analogous to observed measures of the environment, e.g. pH or soil moisture; the difference being that they are estimated from the data rather than being measured in the field (Niku et al., 2019). Species' responses to the

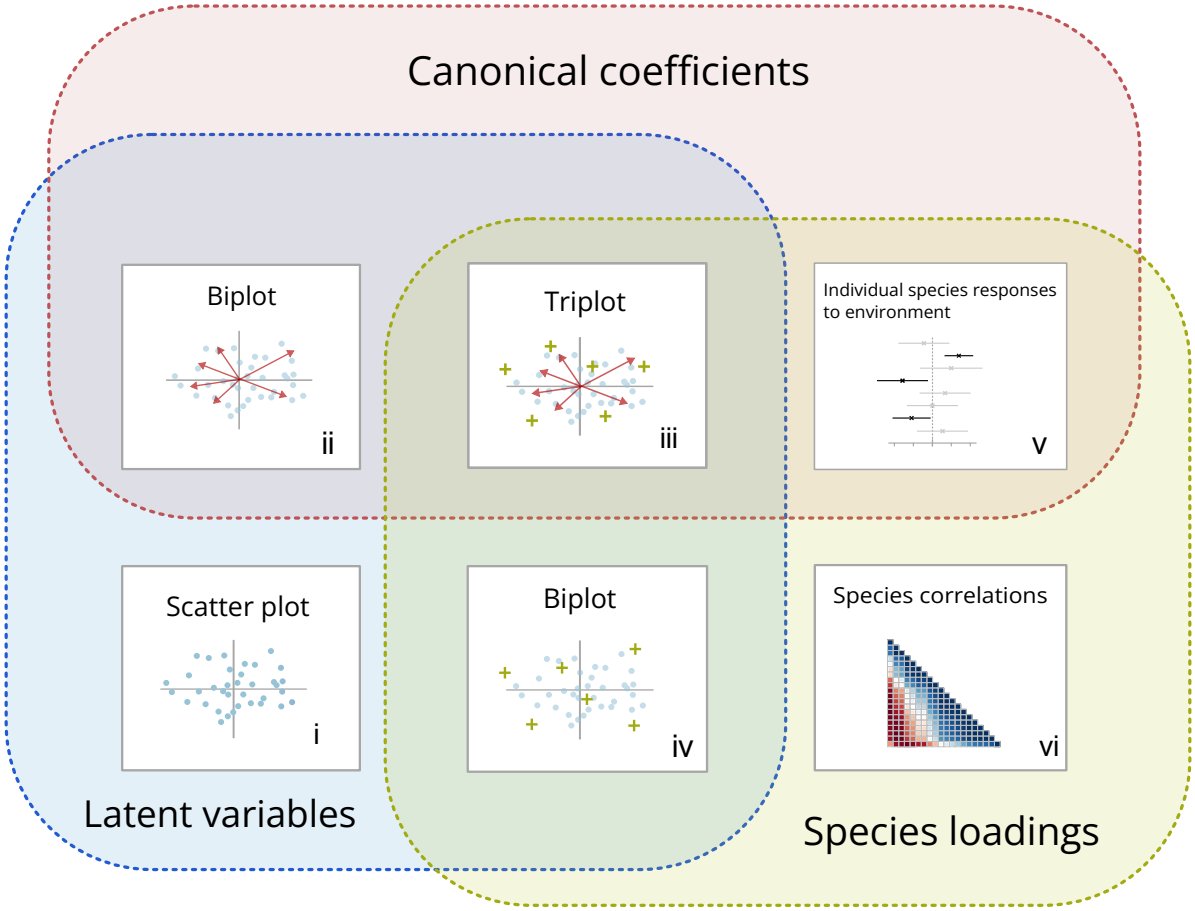

Figure 2: Overview of the different visualisations available for key model components of a model-based ordination. The coloured areas represent the different model components related to latent variables introduced in Figure 1. Overlapping shaded areas indicate which component(s) of the model is used to produce the visualizations. In the figures, latent variable site scores are drawn as blue circles, species loadings as yellow crosses and canonical coefficients as red arrows.

latent variables can be specified as both linear and quadratic. The latter meaning that species have an estimated optimum value along the latent variable, corresponding to the concept of the species niche (van der Veen et al., 2021).

Visual inspection of site scores and species loadings or optima can be done the same way as with the results produced by classical unconstrained ordination methods, using scatter plots of the site scores (Figure 2i) or biplots (Figure 2iii). In addition, model-based ordinations have two other important tools for visual inference which traditional ordination methods lack.

The first tool is a correlogram, a plot of pairwise correlations between species. The sum of the square of the species loadings in a model-based ordination are statistical estimates of the overall correlation between pairs of species in the data, which can be visualized in a correlogram (see Figure 2vi). Correlograms can be effective (albeit somewhat unwieldy) tools to construct an overview of species co-occurrence patterns in one's data (Ovaskainen et al., 2017), although ordination plots make it possible to also visualize the distance of species loadings and site scores.

The second tool is uncertainty estimates — i.e. prediction and confidence intervals — for both the site scores and species loadings. These allow researchers to meaningfully evaluate the statistical strength of evidence for the patterns observed in the data. For instance, if the prediction intervals of two site scores are clearly separated, it can be interpreted as the model being confident that the species compositions at these two sites are in fact different, and are expected to remain so if both sites were to be re-surveyed. The same logic holds for the species loadings, where uncertainties can be used to determine if two species are in fact expected to co-occur or not.

These two tools, together with options for combining unconstrained ordination with additional terms (see the discussion of partial ordination in Section 3.2), allow a number of exploratory community ecology questions to be addressed in a single model-based framework. A selection of examples are presented in Table 1, although some may be considered exploratory before they are tackled by using information about the environment directly in the model. In both Tables 1 and 3, "fundamental questions" is meant to refer to basic research, and "applied questions" to management or impact assessment related studies. However we acknowledge that the distinction between these two is somewhat arbitrary and unclear, and that a single study will often have mixed objectives.

Table 1: Examples of ecological questions that can be investigated in an exploratory manner using unconstrained ordination. The questions are broadly divided into fundamental (F) and applied (A) questions. Recent examples refer to studies in which these questions have recently been addressed using traditional methods for unconstrained ordination.

Question	Recent examples
F1: Does species composition change along one or more biotic or abiotic gradients (e.g. elevation, forest age, water salinity)	Handegard et al. (2024); Maunsell et al. (2013); Mulders et al. (2022)
F2: Are there seasonal patterns in community composition within a habitat?	Li et al. (2022); Naz et al. (2024)
F3: Are there characteristic clusters of species that tend to occur together in different sites, that can be interpreted as distinct communities?	Shembo et al. (2024); Lourenço et al. (2024)
F4: Are there associations between species in the community that are independent of associations accounted for by environmental predictors, which can be interpreted as biotic interactions?	Suárez-Tangil and Rodríguez (2023); Wang et al. (2025)
A1: How does the species composition of communities differ between different habitats or land management practices?	Larson et al. (2024); Fanfarillo et al. (2022); Graser et al. (2025); Pedley et al. (2023); Hu et al. (2024)
A2: Is there a difference between species composition of sites undergoing different ecological restoration treatments, and between those sites and undisturbed reference vegetation?	Brasil Neto et al. (2025); Helbing et al. (2023); Reis et al. (2022); see also worked example 2
A3: How do alien species occur together with native species in an invaded community?	Hejda et al. (2023); Lanta et al. (2022); Reeve et al. (2022); see also worked example 1

3.2 Explaining species composition data using environmental predictors

When environmental predictors are available, fitting model-based ordination with GLLVMs offers even more tools to make inference about species-environment relationships. One approach is to use the environmental predictors to explain the distribution of each species individually, with the latent variables modelling any residual co-variation between the species, as is typical in joint species distribution modelling (Ovaskainen et al., 2017), but equivalent to *partial* ordination. However, with large numbers of species, especially species that occur infrequently,

this approach will quickly involve too many parameters to accurately estimate. Another more parsimonious approach in line with ecological theory, is to assume that species' distributions are explained by a few underlying latent variables that are, in turn, explained by environmental predictors (ter Braak and Prentice, 1988; Legendre and Legendre, 2012).

The core model in this case is *concurrent ordination*, where the latent variables depend on both the effect of environmental predictors and additional variation unexplained by the predictors (van der Veen et al., 2023, see Table 2). Concurrent ordination works by estimating canonical coefficients (cf. Figure 1 and 2 ii, iv, vi), that explain how a change in the latent variable (and thus the species composition) is associated with a change in each environmental variable. Specifically, it is an estimate of how much a latent variable changes following a one-unit change in a given environmental variable, all other variables being equal. In addition to the canonical coefficients, the latent variables estimated by the models may also consist of a residual, or unexplained, component (for more detail see van der Veen et al., 2023). This means that the model can provide estimates not only of the degree to which the main patterns of species composition are explained by the environmental factors, but also to what degree there may be additional unobserved factors driving species composition. The relative importance of the environmental and unobserved factors can then be disentangled by variance partitioning. In this regard, concurrent ordination addresses a longstanding problem with the use of constrained ordination for exploratory analysis (Økland, 1996; ter Braak and Šmilauer, 2015), as it does not disregard variation in species composition unexplained by the predictors. Whereas specifying the concurrent ordination to have no residual variation, i.e. assuming that the latent variables are completely explained by the environmental predictors, corresponds to a *constrained* ordination in the classical sense.

In general, we recommend concurrent ordination as the default starting point for exploratory analysis with model-based ordinations (see Section 4). However a constrained ordination may potentially be used to simplify the model and make the model fitting faster if a concurrent ordination has estimated the residual variation explained by the axis near-zero (see Section 5.2). If the residual variation is large relative to the canonical coefficients, it should be taken as a sign that there are patterns in species composition in the data that are unexplainable by the predictors. Which patterns of species associations this residual variation describes can then be further investigated using a correlogram. If the residual variation is very large, and the effect of the predictors is negligible, it may be taken as a sign that the included predictor variables are not informative in terms of explaining species composition.

Modeling communities with constrained or concurrent model-based ordinations presents a number of additional features and tools for statistical inference over traditional methods: (1) As in the unconstrained case, the canonical coefficients will have an uncertainty, and thus a confidence interval, associated with them. These confidence intervals can be used to make inference about the strength of evidence for the effect of different environmental predictors, site scores and species loadings in the model. (2) Although the predictors affect the latent variables, they can be easily translated to predictor effects for individual species, making it straightforward to connect movement along environmental gradients to changes in individual

species' abundances. As shown in Figure 2, the individual species effects, extracted from the model, are typically plotted using a caterpillar plot, while the latent variable coefficients are typically represented in an ordination biplot or triplot. (3) Predictor effects – both fixed and random – can be specified both outside and inside the latent variables, easily allowing for combining partial ordination with constrained or concurrent latent variables (see Table 2). For example, if one is interested in investigating the environmental factors driving local vegetation change on a mountainside, while controlling for the larger-scale effect of the elevational gradient, one could do this by including elevation as a full-rank effect outside the ordination, and all other relevant environmental variables as predictors (canonical coefficients) for a concurrent ordination; in other words taking the effect of the elevation gradient out of the ordination. (4) The relative importance of the different model components in explaining the responses of the different species can be assessed through variance partitioning. This can include assessing the importance of residual variation of the unexplained part of the latent variable(s) in a concurrent ordination, the effects of predictor variables both within and outside ordinations, and other model components, such as site intercepts, traits etc. (see Figure 1) in explaining the linear predictor for each species. Proportions of variance can be calculated to estimate the relative contributions of each model component in explaining each species' response.

Table 2: Summary of the different types of latent variables currently implemented with model based ordination

Latent variable type	Description
Unconstrained	Latent variables are the freely estimated arrangement of sites that, through the species loadings, best explain observed species-co-occurrences in the data
Constrained	Latent variables are linear combinations of observed environmental covariates that, through the species loadings, best explain observed patterns of species co-occurrences in the data
Concurrent	Each latent variable has a constrained and unconstrained (i.e. residual) component, meaning they represent the *combination* of predictor effects and freely estimated additional variation that best explain explain observed species-co-occurrences in the data
Partial	Species-specific (i.e. full rank) effects of predictor variables are included outside the ordination. This can be done in combination with any of the aforementioned ordination types, to produce partial unconstrained (i.e., residual) ordination, or partial constrained or concurrent ordination.

Table 3 outlines some examples of ecological questions where concurrent or constrained ordination would be relevant to answer ecological questions, as well as examples from the recent literature where they have been approached using primarily classical methods.

Table 3: Examples of ecological questions that can be investigated using latent variable models with predictors, divided into fundamental (F) and applied (A) questions.

Question	Recent examples
F1: How do different environmental gradients (e.g. elevation, climate, water depth) explain differences in the community composition between sites?	Cheng et al. (2023); Young et al. (2022); Askeyev et al. (2023); Matavelli et al. (2022)
F2: Are specific species in a community indicators of changing environmental conditions?	Andrew-Priestley et al. (2022); Korolyuk et al. (2024)
A1: What is the effect of antropogenic vs. non-antropogenic factors in terms of explaining community composition?	Christman et al. (2022), Sanchez et al. (2023)
A2: Do certain environmental factors explain the prevalence of alien species in an ecosystem?	Kalusová et al. (2019), see also worked example 1
A3: How does a community respond to different restoration treatments?	Crouch et al. (2022), see also worked example 2
A4: How will the composition of a community shift in response to changing climate?	Forte et al. (2024)

4 Guidelines for a model-based ordination workflow

Guidelines for other model-based analyses have been outlined by Warton et al. (2015), Zuur et al. (2010) and Zuur and Ieno (2016), among others, and the same recommendations generally hold for model-based ordinations. Based on these, we present a five-step workflow, specifically geared toward the effective and sound application of model-based ordination in community ecology. The workflow outline is primarily adapted from Warton et al. (2015), and is summarized in Figure 3. Section 5 demonstrates the workflow on two relevant real-world data sets.

4.1 Formulate the biological question as a statistical question

After the biological and ecological questions of the study are clarified, the first step in any model-based workflow should be to formulate them as concretely as possible in statistical terms. This means clarifying why a model-based ordination is the right tool for the problem, and how exactly the model will be used to answer the ecological questions. For example, which parameters should be included in the model, and how are they assumed to be predictors of the observed data.

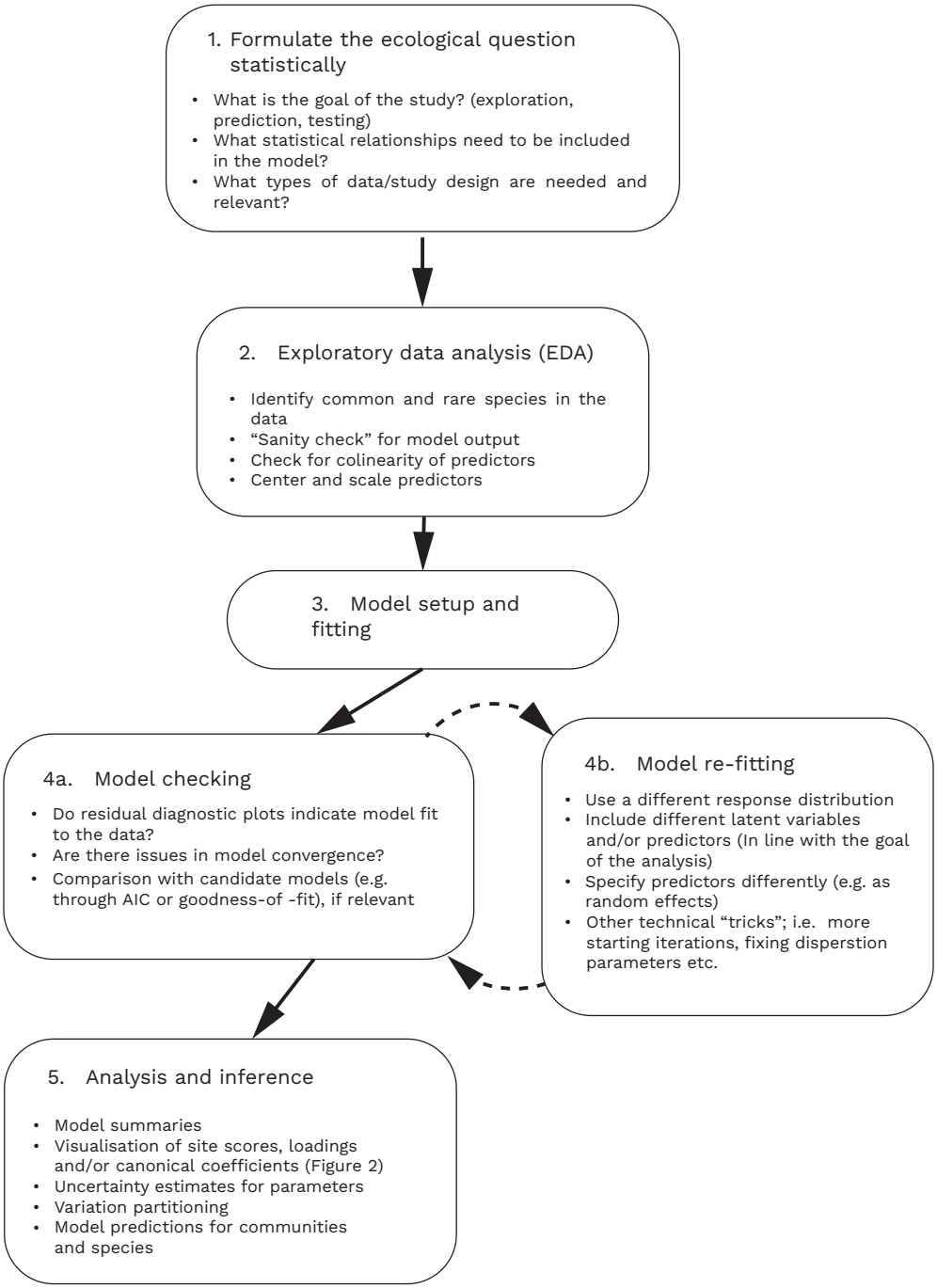

Figure 3: Visual representation of the analytic workflow suggested for modelling ecological communities with model-based ordination. Adapted from Figure 1 in Warton et al. (2015).

Ideally this first step should be undertaken before collecting data, in order to make sure that the study design and sampling strategies are geared towards getting the data needed to answer the ecological questions of interest (for a further discussion of this, see Warton et al., 2015). For example, if the main interest of a study is in making inference about how a species community changes along a temperature gradient, care should be taken to sample the environmental variables along that gradient so that they capture enough variation in the environment to meaningfully answer that question. Similarly, if the goal is to investigate the response of one or more specific focal species within the community to environmental and biotic changes, one should make sure to collect data on a wide enough range of conditions where they might be expected to occur and not occur (i.e., their niches should be well-sampled), in order to actually obtain enough data to make meaningful statistical inferences about their relationship to the environment and/or other species (see also worked example 1, as well as the Section 6). These considerations might occasionally also need to be balanced with strategies for ensuring sample representativeness, for example by deploying sampling methods that have some way of quantifying detectability (see e.g. Jeliazkov et al., 2022), as long as it is is consistent with the broader objectives of the study. This step also includes considering which type of model is best suited to answer one's research questions and represent the ecological relationships of interest. For example, if gathering data on environmental or habitat-type variables is part of the study, representing these in a concurrent ordination will often be a natural choice.

Clarifying whether the objective of one's study is primarily explanatory, confirmatory or predictive is arguably another important part of this step (Shmueli, 2010), for guiding choices around the inclusion of predictor variables and model selection. If the goal is prediction, i.e. to find the model-based ordination with the combination of predictor variables that most accurately predicts either community composition or the occurrence of specific focal species, optimizing one's model for this purpose through model selection, using e.g. AIC or similar tools, can be a meaningful strategy. However, if the goal of the analysis is rather to explore or make inference about the species community or communities in the data, as in the example above, variable selection for prediction could lead to biased inference and should in general be avoided (Sainani, 2014). Instead, all variables that are of interest should as far as possible be included in the model as the best statistical representation of the ecosystem, and the results of the fitted model should be explored as is. The concurrent ordination framework is robust enough In this case, AIC or BIC might still be useful for determining the number of latent variables that best fit the data. Model selection of predictor variables based on optimizing for prediction should however be avoided, especially if the aim of the study is more clearly *confirmatory*, i.e. testing specific hypotheses about ecological relationships rather than exploring them more generally.

In general, our modeling philosophy is the same as that of Ovaskainen et al. (2017): as far as possible, the aim should be to fit a single comprehensive model which can be used to address all relevant research questions, rather than analysing many different models in parallel. This helps to streamline and making the analysis more reproducible, as well as preventing data dredging and ensuring that uncertainties are handled correctly.

Still, the reality is that many studies in community ecology will often have mixed goals and

uneven data collection, with study questions that are both exploratory, predictive and/or confirmatory. In these cases, setting up the analysis requires careful thought. For example, as in worked example 1, data related to one study question may be collected for a significantly smaller subset of the data than another, and as such two distinct model-based ordinations might have to be fit to the data rather than one, in order to utilise the full dataset in the analysis. Or in a case like worked example 2, where there are study questions that are both predictive and exploratory, a decision is made about which one is more fundamental or important, and which is secondary, and the model specified accordingly. Again, the most important point is that all the goals of the analysis are clarified, and a plan made to address them, before model fitting begins (and ideally also before data collection), and that the goals are not conflated with the methods. Above all, a strategy of carelessly throwing many different models at the data just to see what "sticks", should be avoided at all cost.

4.2 Exploratory data analysis (EDA)

After collecting data, and before fitting a model-based ordination, exploratory inspection and visualisation of the raw data should always be done in order to get a better understanding of the dataset and to act as a sanity check on the model output. Relevant dataset properties to consider for model-based ordination are largely the same as for other models in the GLM family, and we generally recommend the same strategies proposed by Zuur et al. (2010).

When dealing specifically with the types of multivariate species data considered here, we will also recommend a few additional exploratory strategies as good practice. The first is to simply get a broad-scale overview of the data by creating a table or histogram of how many samples (rows) each species (column) is observed in, as well as the inverse (how many different species are observed in each sample). This makes it possible to get a sense of how the data is spread over the samples, which e.g. can be seen in context with the sampling design, or to identify potentially data-deficient species (see discussion in Section 6). When species data are quantitative (i.e. not simply presence/absence), visualizing the relationship between species' prevalence in the data and their average abundance in each site with an Abundance-Occupancy (AO) plot can also be a helpful tool in this regard, making it possible to see whether the data follows the classic positive relationship commonly found in ecological data sets or not (Gaston et al., 2000), and whether some species deviates notably from others in terms of their AO-relationship – either due to factors do to the sampling design or the ecological dynamics of the system, which sometimes can be challenging to untangle (Russell et al., 2005; Gaston and Blackburn, 2003), but which in any case may provide important context for interpreting the results of a model fit.

Depending on the goals of the study, fitting a simple unconstrained ordination to the data – either through an unconstrained model-based ordination or a classical method like PCA – could also be a part of this exploratory phase, to be used as a simple summary of the main species co-occurrence patterns in the data, before a model specifically geared towards one's research objectives is specified in step 3.

As for the EDA of predictor variables, visualizing their pairwise co-linearity using a correlogram or similar is a good general-purpose tool for informing decisions about predictor inclusion in the model. However, as predictor collinearity is typically associated either with properties of the study design or inherent properties of the study system (e.g. the relationship between temperature and altitude), the question of whether to include or discard predictors due to collinearity should be informed by the goals of the study, study design and one's *a priori* knowledge of the study system, rather than numerical rules of thumb. Note also that while including highly co-linear predictors in a model-based ordination might lead to increased uncertainty in the coefficient estimates and potentially convergence issues, it should not in principle lead to different inference about the underlying gradients that drive species composition. Scaling and centering of the predictors is also recommended here as a standard procedure to improve coefficient estimation and convergence of the model.

4.3 Model setup and fitting

Following from steps 1 and 2, the relevant model(s) should have been identified, and can now be fitted to the data. Important parts of the model to specify are (a) the response distribution for the species abundances/occurrences, (b) row (i.e. site) effects to explain the total abundance of individuals in the samples (i.e. predictors that effect the abundance of all species equally), (c) the number of latent variables (of different types) fitted to the data, and (d) model formulae for latent variables and species effects. It is important to note again that transforming, scaling or otherwise changing the species response variables in order to give them more desirable statistical properties is, in general, not in line with the model-based ordination philosophy.

4.4 Model checking and re-fitting

After a model has been fitted to the data, it should be evaluated thoroughly. If there are issues with the model fit, these should be addressed and the model re-fit, as illustrated in the flowchart in Figure 3. Below, we detail a number of the specific issues that practitioners may encounter while fitting model-based ordinations, with some advice on how to address them.

4.4.1 Checking model validity

As with classic GLMMs, it is important to check that the data meet the model assumptions. This is most readily done by visualizing the residuals in diagnostic plots (see also Section 2, point 2). If residual diagnostics indicate that the distributional assumptions are violated, the most obvious response to this is to change the response distribution. Testing the fitted model for zero-inflation over- or under-dispersion can be tools to indicate which alternative distributions make sense to fit to the data (see Section 2 for more on different response distributions).

4.4.2 Dealing with convergence issues

This latter point needs some consideration, as it can sometimes be difficult to get good convergence and numerical stability when fitting GLLVMs, and correctly diagnosing and dealing with convergence issues in GLLVMs can be a bit tricky. Below, we list strategies that can give an indication of the degree of model convergence:

For models fitted through (marginal) maximum likelihood estimation, as in the `gllvm` package, inspecting the gradient vector of the likelihood function to see if it is close to zero is a good general-purpose indication of model convergence. If some values are very large compared to the rest, it is an indication of poor convergence. Exactly what the cutoff should be is impossible to answer generally, and must be seen in context with other features of the model fit. Still, values larger than 1 for the gradient vector of a parameter can usually be taken as a sign that the parameter estimate has not converged. Checking for negative estimates for the standard error of parameter estimators is also an indicator of lack of convergence. In `gllvm` version 2.0.10, this is done automatically and produces a warning if detected. Other model issues, such as singular fits and optimisation algorithms converging to infinity, will also produce errors in most software applications.

Convergence of GLLVMs estimated through bayesian estimation can be assessed through standard tools for assessing MCMC convergence and mixing. These include potential scale reduction factors, effective sample size estimation and rank- or trace plots. For further information see e.g. Roy (2020); Vehtari et al. (2021); Stan Development Team (2026a).

For both types of estimation, visualisation of model estimates and uncertainties might also be helpful, e.g. if some species have "exploding" species loading estimates or uncertainties. This may happen when some species occur very infrequently in the dataset, and/or exhibit complete separation: That is, they are associated with a subset of predictors (e.g. a species only occurs in one habitat, and habitat is included as a categorical predictor).

In order to deal with convergence and numerical stability, a number of "tricks" and adjustments can be tried, such as: 1) Scaling and standardising numerical predictor variables if that is not already done. 2) Increasing the number of starting iterations (for maximum likelihood estimation). If time and computing power allows, models should always be fitted through enough iterations to ensure that a stable result is achieved, or at the least to rule out lack of computation time as the problem. The same goes for the number of MCMC iterations or chains. 3) Reordering the species data frame. In most implementations of model-based ordination, the latent variables are "anchored" by the species loadings of the first number of species equal to the latent variables. This means that the ordering of the remaining species is relative to this anchoring. Species that are data poor (i.e., "rare" species) usually occur on the limits of the observed gradients, so that if data poor species are placed first in the data, the model can only respond by placing all other species in extreme locations of the ordination instead. Generally, a good heuristic is then to instead place species that are rich in data first in the dataset, so that the anchoring is based on species that can be expected to occur in the middle

of the gradient, stabilising estimation of sites scores and species loadings. If there are *a priori* hypotheses about species that are indicative of different changes in community structure, this can also be a meaningful way to select the anchors, given that the species are well sampled. 4) Filtering out infrequent or very unevenly sampled "problem" species from the data. However, this needs to be considered carefully in the context of the study. See Section 6 for a further discussion on this. 5) Having certain parameters be fixed or shared across species. A good first strategy can be to specify shared dispersion parameters for the response distributions of all or a subset of (infrequent) species.

Still, some aspects of convergence may not need a fix. For example, complete separation may be an inherent property of the data, i.e. if one samples very different habitats. In this case, the solution is to acknowledge the existence of the problem, and ensure it does not affect ecological inferences. Other times it might result more artificially from adding a large amount of categorical predictors to the ordination relative to the data, in which case the model formulation should probably be revised. For a thorough and slightly more technical discussion of convergence issues in likelihood-based GLLVMs, we recommend consulting the `gllvm` package vignette (Niku, 2026), as well as other articles discussing how to deal with model convergence in mixed models, and the details of the estimation methods; see e.g.: Bolker et al. (2009); Lavielle and Aarons (2016); Wood (2017), or Du et al. (2022); Gabry et al. (2019); Plummer et al. (2006) for bayesian estimation.

4.4.3 Model selection

Assessing and comparing models with respect to prediction or selection, depending on the goal of the study, can be done in a number of ways. Information criteria like AIC or BIC are perhaps the most well-known. As these two criteria have slightly different interpretations (Aho et al., 2014), which criterion to use will depend on the objective of the study (see Section 4.1). Other measures of model predictive quality can also be used, e.g. root-mean square error of the prediction, or cross-validation.

However, this should only be done after checking model validity and convergence. In principle, AIC, other information criteria or goodness-of-fit measures should not be used to compare models that are not converged or whose assumptions are not met. Even for valid models, model comparison assessment tools should be used with care, caution and common sense and never trusted blindly as context-less measures of "model quality". This is especially true as there are still some questions about the validity of Information criteria for mixed-effect type model comparison (Müller et al., 2013).

4.5 Visualisation and inference

After step 4 is completed, the model can finally be explored to make inferences about the relevant ecological questions of the study. We refer here primarily to Section 3 for a discussion

of the different tools that can be used to make inferences from model-based ordinations in terms of different ecological questions, as well as the worked examples.

5 Worked examples

In this section, we demonstrate how model-based ordination can be applied in real-world settings, using two relevant case studies from the recent ecological literature. The case studies are selected to demonstrate the tools and questions discussed in Section 3.1 and Section 3.2.

In order to demonstrate different paths to visualizing the output of model-based ordinations, visualisations in Example 1 (Figure 4) are produced primarily using the native plotting functionality from the `gllvm` package, using the base `R` plotting interface, while three of the four visualisations in Example 2 (Figure 5) are constructed using the `ggplot2` package with extracted model components. Walk-throughs of the complete data analyses and visualisations, including figures for model diagnostics, are available in Appendix S2.

5.1 Example 1: Invasive trees in Argentina

In the first case study, we reanalyse data from Fernandez et al. (2021). Here, the researchers were interested in how the presence and abundance of an invasive tree species, the broad-leaf privet (*Ligustrum lucidum*), impacts the native tree community in an Argentinian second-growth subtropical forest.

Data on the tree community was recorded by measuring the basal area of 20 common species (including *L. lucidum*) in 164 forest monitoring plots. In a subset of 44 of these plots, samples of four physical-chemical characteristics of the soil: soil carbon content, nitrogen content, carbon to nitrogen ratio, and soil humidity, were collected as well.

For the purposes of this article, and to best showcase the capabilities of model-based ordination, we have condensed the ecological questions from Fernandez et al. (2021) into the following two research questions: 1) How is the abundance of *L. lucidum* in an area associated with the composition of other (native) tree species, and 2) Are some soil properties associated with increased abundance of *L. lucidum* specifically, compared to the native species?

5.1.1 Formulating the statistical question

The aim of the analysis is clearly exploratory, rather than confirmatory or predictive. No specific hypotheses about species-species or species-environment relationships are tested, and the goal is to make inferences about the ecosystem rather than finding the model that best predicts abundances of *L. lucidum*. This suggests we should aim to model the data in a way

that includes all relevant predictors of interest, and that extensive model selection beyond finding the optimal number of latent variables is not relevant.

However, the fact that environmental predictors (i.e. soil properties) are only available for a small subset of the vegetation plots, does present a challenge. In order to make the most of this data, we decide to veer slightly from our ideal workflow, and fit two different model-based ordinations to the data: (1) A model with only unconstrained (i.e. not predictor informed) latent variables fitted to the full dataset; this will be used to make inferences about the patterns of co-occurrence between *L. lucidum* and the other species, and (2) a model with predictor informed latent variables (i.e., a concurrent ordination), fitted to the subset of plots with environmental variables recorded, using all 4 recorded soil properties as predictors. This second model will be used primarily to answer research question 2, make inferences about potential relationships between soil conditions and the co-occurrence of *L. lucidum* with native species. If predictor variables had been available for all plots, we could most likely have addressed all of these questions with a single concurrent ordination.

As the original study does not contain or consider explicit information about the study design, we will treat each sample (i.e. site) as independent. We do this by adding random intercepts for each row in the response data (see paragraph four in Section 2) to ensure the latent variables only account for composition rather than total abundance at each site.

5.1.2 Exploratory data analysis

Aggregating and visualizing the number of occurrences of all species in the full dataset (see Appendix S2, Section 3.2.1.), we see that every species appears in more than three plots. Of the 164 plots, only five contain just a single species, and the vast majority contains three or more species. Based on this, we assume that we have enough information in our data to avoid removing samples or species.

When selecting only the subset of the plots where soil variables were measured, however, two species were absent from all of these plots, and one species only occurred once. We thus excluded these three species from model 2, as they do not hold information, and keeping them will likely hurt model convergence.

Other than filtering the data, and centering and scaling all predictor variables to mean zero and unit variance, as discussed in Section 4, no further pre-processing was done for the data.

5.1.3 Model setup

Because our observed response variables are recorded as the area of each species in a plot, we decide to fit both models using a Tweedie distribution (Jørgensen, 1987). The Tweedie distribution arises as a Poisson sum of Gamma random variables. In other words, we assume that the number of observed individuals follows a Poisson distribution, and the area of each

individual follows a Gamma distribution. As well as having an intuitive derivation, the distribution can accommodate species with zero area (unlike, for example, gamma and log normal distributions), and is also appropriate for data that follow Taylor's law (Kendal, 2004).

For both of the proposed models (the unconstrained and the concurrent), we intend to find the optimal number of latent variables which best fit the data. As discussed in Section 4, we decide to do this by finding the number of latent variables with the lowest information criterion that also fitted the data. In this case we will use AIC, as it is primarily recommended for exploratory analyses (Aho et al., 2014). It is also important to stress that in this case we only selected for the number of latent variables, not the predictors, due to the exploratory nature of the study.

We fit the models using the `gllvm()` function, with the syntax shown below, commented for clarity. We initially fit the models with one latent variable each, and proceed to add latent variables to find the AIC minimum, checking the diagnostics of each new model as we went. See Appendix S2, Section 3.3. for the full model fitting code, with explanatory comments.

5.1.4 Model checking and refitting

The diagnostic plots for both the unconstrained and constrained models did not indicate any violations of the model assumptions, and the addition of more latent variables to each model did not change this (see Sup. Figures 2.3, 2,4 and 2.6). The only caveat to this is that there seemed to be a slight structure in the residuals-versus-fitted plots, where the most prevalent species had slightly more negative residual than would be expected.

In the case of the unconstrained model, there was an AIC minimum for a model with five latent variables (see Sup. Table 2.1.). However, this was not as well converged as the model with three latent variables. Because of this, and partially in order to make the analysis as parsimonious as possible for demonstration purposes, we decided to continue with the model with three latent variables for the analysis (see Appendix S2, Section 3.4.1.). For the concurrent model, there was a clear AIC optimum at the model with two latent variables, and as such, we decided to continue with this model for visualisation and inference for the second part of the example.

5.1.5 visualisation and inference

Looking at the visualized species loadings of the unconstrained ordination (model 1) in Figure 4a, we see that *L. lucidum* is a clear outlier among all the other species. The predicted abundance of *L. lucidum* is primarily summarised by the first latent variable after rotating in the direction of maximum variance, as the position along the second latent variable (the vertical axis) is close to zero. As such, we might inspect the other species' responses to the first latent variable (the horizontal axis), for indications of their co-occurrence with the invasive species. The fact that only three other species have a positive loading along the first latent variable, and most other species are associated with the other end of the diagram, clearly indicates that an increased

presence and basal area of *L. lucidum* is associated with fewer occurrences and lower basal area of most other tree species. This is also supported by the confidence intervals of the species loadings, in which the C.I. of *L. lucidum* overlaps with almost no other species.

These co-occurrence patterns are also clearly supported by Figure 4b, albeit more nuanced, as the correlation plot uses information from all three latent variables. The correlation of *L. lucidum* with the other species resulting from the species scores are all estimated to be negative, except in three cases. The ecological interpretation of this first model is that *L. lucidum* seems to either displace most native species where it occurs, or that its environmental tolerance or preference is different from most other species in our data, thus thriving in conditions that are not favourable to other species. It could also be a combination of both scenarios, as Fernandez et al. (2021) hypothesize, in which *L. lucidum* alters the soil chemical properties where it establishes itself, making it more favorable for itself and less for the native species.

The concurrent ordination that includes environmental predictors (model 2) suggests that the observed environmental variables explain a significant portion of the community structure. As Figure 4c shows, the species scores of *L. lucidum* are separated from the others along the horizontal axis, as in the first model. Additionally, we see that it is clearly negatively associated with increasing soil moisture content, and positively associated with a larger soil carbon-to-nitrogen ratio. This is made even more clear when looking at the species-specific predictor effects in Figure 4d. *L. lucidum* is the only species which is estimated to decrease its abundance with higher soil moisture, while all other species respond either neutrally or positively to moisture. An inverse association seems to exist for the C:N ratio, although less pronounced, being shared by a few other species, as well as associated with higher uncertainties for all species (Sup. Figure 2.10). Variation partitioning also revealed soil moisture to be the variable explaining the highest mean proportion of variance for the species in the second model (Sup. Figure 2.11). However the variance partitioning, as well as the model summary (Appendix S2, Section 3.5.2), also indicates that about 30% of the variation in the species composition was not explained by the environmental covariates, and is therefore an indication that there might be other important environmental predictors – or other dynamics in the community – that influence the species composition and which were not included in the model.

In summary, the ecological conclusion to draw from these two models seems relatively clear: the models provide a strong indication that in the ecosystem where it appears as an invasive species, *L. lucidum* is associated with a lower abundance of most other native common tree species. Secondly, this effect can largely be explained by *L. lucidum* either preferring, better tolerating, or even facilitating drier soils, supporting the initial hypothesis from Fernandez et al. (2021).

5.2 Example 2: Roadside Restoration in Norway

The second case study is based on data and ecological questions from Mehlhoop et al. (2022). Their aim was to assess the impact of different restoration efforts on roadside vegetation, in

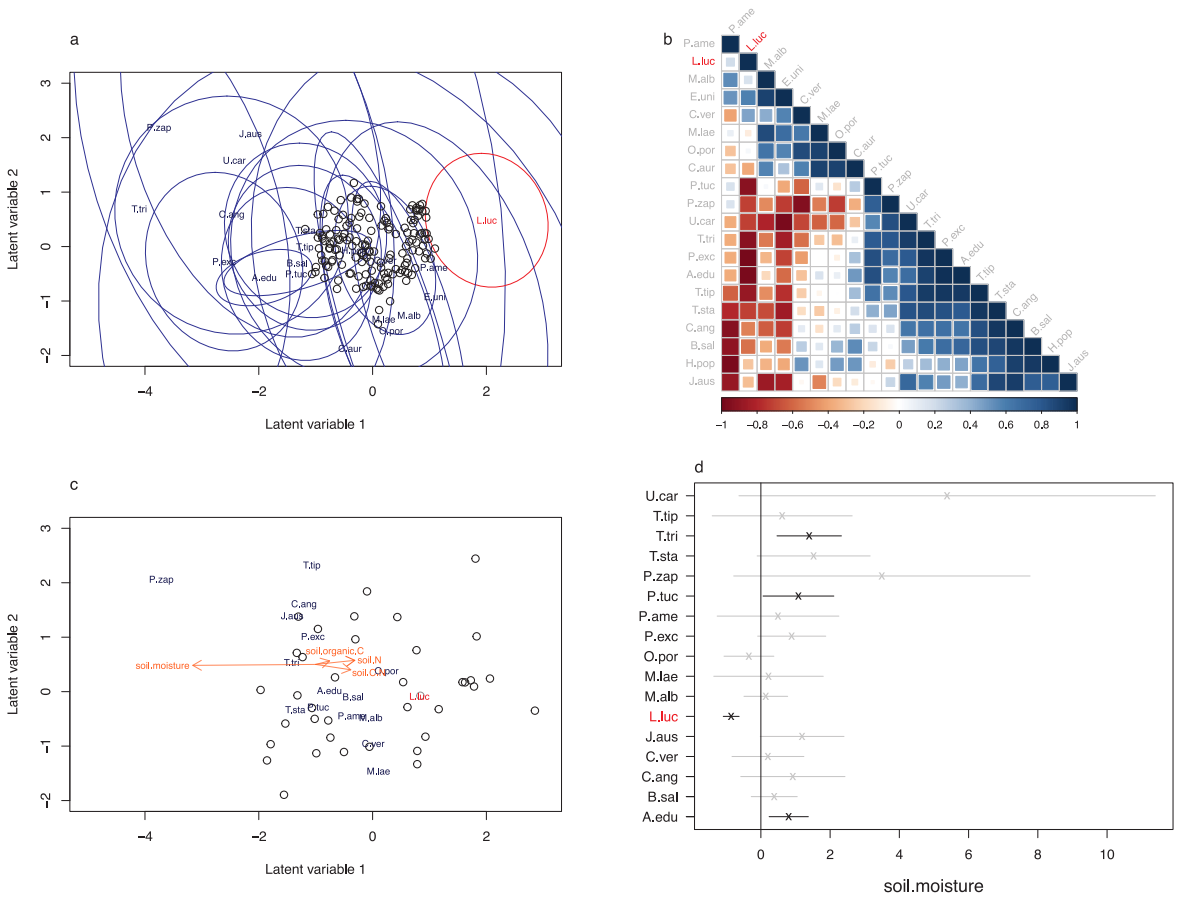

Figure 4: Selected visualisations of the estimates from the unconstrained (a,b) and concurrent (c,d) latent variable models. L.luc = *Ligustrum lucidum*, is indicated in red text in all figures. a) Site scores (black) and species scores (blue, red) for the unconstrained model with three latent variables (model 1) Ellipses represent prediction intervals for species scores. Species and site score and uncertainty ellipses of the three latent variables are all rotated in the directions of maximum variance to produce latent variable 1 and 2 using singular value decomposition, similar to a PCA rotation of an NMDS ordination. b) Correlation plot of the between-species correlations estimated from the species scores of the unconstrained model. c) Ordination diagram of site scores (black points), species scores (blue, red text) and environmental coefficients (red arrows) of the concurrent (predictor-informed) latent variable model using a subset of 40 points. Light red indicates that the 95% confidence interval of the latent variable predictor includes 0 for one or more of the latent variables, while the converse is true for the dark red arrows. d) Species specific coefficients (slopes) for the effect of soil moisture content on the abundance on the different species in the model (on the link scale), ordered from lowest to highest. Cross = coefficient estimate, line = 95% confidence intervals. Confidence intervals that cross 0 are indicated in grey.

order to mitigate the effects of road construction. The dataset consists of the percentage cover of 164 different vascular plant species at 282 roadside plots across 3 regions in southern Norway. At each site, plots were subject to one of three restoration treatments: Re-seeding using commercial seed mixes, planting with native vegetation, or natural, i.e. unassisted re-vegetation. In addition, plots in intact reference vegetation were also sampled. Other variables, including the time since restoration (for the non-reference plots), as well as biological and environmental variables like soil organic matter content, canopy cover and grain size, were recorded at each plot to account for potential environmental factors that may influence species composition not directly related to restoration.

The primary research question of Mehlhoop et al. (2022) was how effective the three different restoration treatments were in bringing the vegetation of the impacted sites closer to the assumed natural vegetation in the reference sites. Ideally, this knowledge can then be used to inform future restoration efforts in similar nature types. As a secondary goal, chosen specifically to further showcase the capabilities of model-based ordination, we also ask how the vegetation in the restored sites is expected to change over the next 20 years.

5.2.1 Formulating the statistical question

The main goal of the analysis is to understand the relationship between a set of predictor variables (restoration method and time) and species occurrences in the data, and not to test any specific hypothesis. However, in contrast to the first example, the samples themselves (i.e. the restoration sites), rather than the species, are the primary unit of interest. Although the secondary goal is prediction oriented, the primary goal is explanatory in nature. As such, we decide to base our prediction on whichever model serves the explanatory purpose of the study best, rather than the other way around, even if that model might not predict optimally.

Consequently, fitting a concurrent ordination for the species composition including all potentially relevant predictors (treatment, time since restoration and the environmental variables) best aligns with the goals of this study. As the effect of time on species composition might be different for different restoration treatments, and because this potential difference is central to the ecological question in this case, we decide to include an interaction effect between the restoration treatments and time since restoration.

It is also necessary to think about how to account for the influence of our study design. In particular, there could be potential differences in the overall prevalences of species between the three different study regions that we want to separate out from the effect of restoration. To address this, we included region as a fixed row effect in the model, with additional random species-specific intercepts for each region. Within regions we might also want to account for differences in the sampling intensity between sites and plots. To do this, we add an additional random row effect for each site to account for potentially confounding differences in the total sample abundance between sites. In other words, we condition the ordination on the study

design, and thus remove information about the effect of the regions and sites on the species community from the ordination.

5.2.2 Exploratory data analysis

Of the 164 species in the data, more than 50 only appear in a single plot and almost 40 appeared only two or three times. In order to reduce the chance that the final model is unduly influenced by data-deficient species, and because the focus of the study was the effect on restoration on the overall compositional differences between sites, rather than a focus on any particular species, we decided to exclude the species with three or fewer occurrences. This was done primarily to lessen the impact we would expect species complete separation (i.e. species only appearing once or twice a single treatment group) to have on the model fitting, as we are dealing with several categorical predictors in this case. However, we acknowledge that this cutoff is somewhat arbitrary, and done primarily to ease model fitting for demonstration purposes. If this analysis was to be done in a real research setting, more careful consideration of which species to include from an ecological point of view, including a more in-depth more exploratory analysis of species' distributions between treatments, should be a part of the workflow (see Section 6). Consequently we did not exclude any sites from our data. See Section 6 for a further discussion on the handling of data-deficient species in model-based ordinations. We also scaled and centered all numeric predictor variables.

5.2.3 Model setup

The data in this case is proportional and continuous. As such, the beta distribution would be a natural choice for the response distribution, as it is a very flexible probability distribution for continuous proportions (Ferrari and Cribari-Neto, 2004). In accordance with the model-based ordination philosophy, and as pointed out by others (Geissinger et al., 2022), it allows one to model proportional data directly, rather than relying on data transformations to enforce normality which obscures the link between data and model, which have otherwise been common for proportional ecological data (Warton and Hui, 2011). However, we also need to accommodate the fact that our species data has several full absences (0) and some cases of 100% cover (1). To do this, we elect to fit an ordered beta model to the data (Kubinec, 2023; Korhonen et al., 2024). The ordered beta model is related to the cumulative logit regression typically used for ordinal data (also implemented in the `gllvm` package), in that it estimates two cut-off points for the linear predictor for the species abundances. Below the first, the species is fully absent, above the second, it is fully present, and between the two, it is distributed according to a beta distribution with a mean determined by the linear predictor. As such, it is considered a robust parsimonious model for data in which the presence, percentage and "saturation" of a species in a plot are all derived from the same underlying ecological process (Kubinec, 2023).

After selecting the response distribution, we then set up the initial model following the structure outlined in the beginning of this section. To include the interaction effect between restoration

treatment and time (and exclude a time interaction with the reference category), a custom model matrix was constructed where only interaction effects between the treatments and time were included (see Appendix S2, Section 4.2.). As in worked example 1, we decide to use information criteria to determine the optimal number of latent variables, conditional on model convergence. Code for the model fitting, with comments, can be found in Appendix S2, Section 4.3.

5.2.4 Model checking and re-fitting

Unlike in example 1, fitting the concurrent ordination specified above presented some numerical challenges, as models with both one, two, and three latent variables struggled to converge. We thus changed the fitting method to Extended variational approximation (Korhonen et al., 2023), and changed the ordering of the species in the input data, placing the most abundant species first. This helped to stabilise fitting of the models with one and two latent variables, however the model with three was still not able to converge. And while the diagnostic plots for both the one- and two latent variable models looked good, the model with two latent variables still showed some potential convergence issues. In particular, many of the variances of the parameter estimators calculated by the model were negative, which makes the model fit hard to interpret.

As the model summaries of both models also indicated that the residual variation in the latent variables (i.e. the unexplained part), was consistently negligible (variance <e-7), we tried instead to fit a simpler model with constrained (i.e. fully predictor determined) latent variables, to make both the fitting and the inference easier. However, lack of convergence persisted for the constrained models with two and three latent variables, which could be seen both the values of the gradients, negative variances of parameters estimates, and the fact that many species loadings were severely "blown up" and linearly correlated in the ordination loadings, which also indicated that the lack of convergence was likely related to overfitting. As such, we ultimately decided to move forward with the model with one constrained latent variable for our analysis. The fact that the AICc was lower for the two-variable model for both the concurrent and constrained models (Sup. Table 2.1), was therefore deemed less relevant due to lack of convergence.

5.2.5 Visualisation and inference

Our one-dimensional constrained latent variable model indicates that the different restoration treatments are the most important factor separating the species composition of the different sites (Figure 5a), with the reference vegetation sites clustering on one end of the scale, the naturally re-vegetated sites in the middle, and the planted and seeded sites on the other side. This is supported by the species loadings of the model, showing that the species most associated with the sites in the reference vegetation (left part of axis 1) are the european blueberry (*Vaccinium myrtillus*), may lily (*Maianthemum bifolium*) and oak (*Quercus robur*),

all species characteristic of Norwegian south boreal forests, in which the reference plots were placed. Other tree species such as Norway spruce (*Picea abies*) were also strongly associated with the left-hand side. On the other side of the restoration axis, the plots undergoing seeding and planting were mostly associated with grasses such as red fescue (*Festuca rubra*), timothy (*Phleum pratense*) and small-reed (*Calamagrostis stricta*), plants more typical of typical of roadside vegetation and early succession, as well as some commercial seed mixes (Mehlhoop et al., 2022).

The estimated interaction between restoration treatment and time since restoration (i.e. how the effect of the restoration on species composition changes with time) is also different between treatments (Figure 5a, see also model summary in Appendix S2, Section 4.5.). Natural sites had a moderate trend towards the reference sites, while the effect was much smaller for the planted sites and absent for the seeded sites. The other measured environmental variables mostly have weak associations with the latent variable, exception for soil organic matter, which has a moderate correlation with the species composition of the intact sites.

The effect of the study design (Figure 5b) was also pronounced, explaining around 26% of the total variation in species responses according to variance partitioning (see Appendix S2, Section 4.5.). Among other things, we see that Region 3 was in general more species-rich than the other regions. This indicates that not accounting for the study design might have led to a different inference about the effect of restoration, because the distribution of the treatment groups and time since restoration is not equally distributed among the study regions (Mehlhoop et al., 2022). We also see that the variance explained by site is distributed differently between species. For instance, the variance in predicted abundance of some forest-associated species like *Calluna vulgaris* and *Equisetum sylvaticum* (the two left species in Figure 5b) were explained almost fully by the site, not by restoration treatment, while other forest-associated species like *V. myrtillus* were explained primarily by restoration. In other words, the confounding could have lead to regional differences being modelled as treatment effects in the ordination. Our results also clearly show that after accounting for this 26%, almost all of the variation in species occurrences are attributed to the effect of restoration vs. reference in the ordination.

Forecasting 20 years in the future, assuming that all other environmental variables in the sites remain the same, the model predicts that the composition of the natural re-vegetated sites will have caught up to the composition of the reference forest, while sites in the other restoration treatment groups will have changed little (Figure 5c). Forecasting for two species which could potentially be used as indicator species, based on their species loadings and pre-existing knowledge about their ecology, *F. rubra* and *P. abies* (Figure 5d), underpins this by showing a marked difference between the different restoration treatments.

The main takeaway from this analysis is that the roadside vegetation sites that were left to naturally re-vegetate, were closer in terms of species composition to the forest reference than the sites that had been artificially seeded. This vegetation treatment also showed a stronger response to time, which can be interpreted as a faster succession than the other treatments. Because no other variables in the model explained differences in species composition to the same degree, potential confounding effects of the study design were accounted for. Finally,

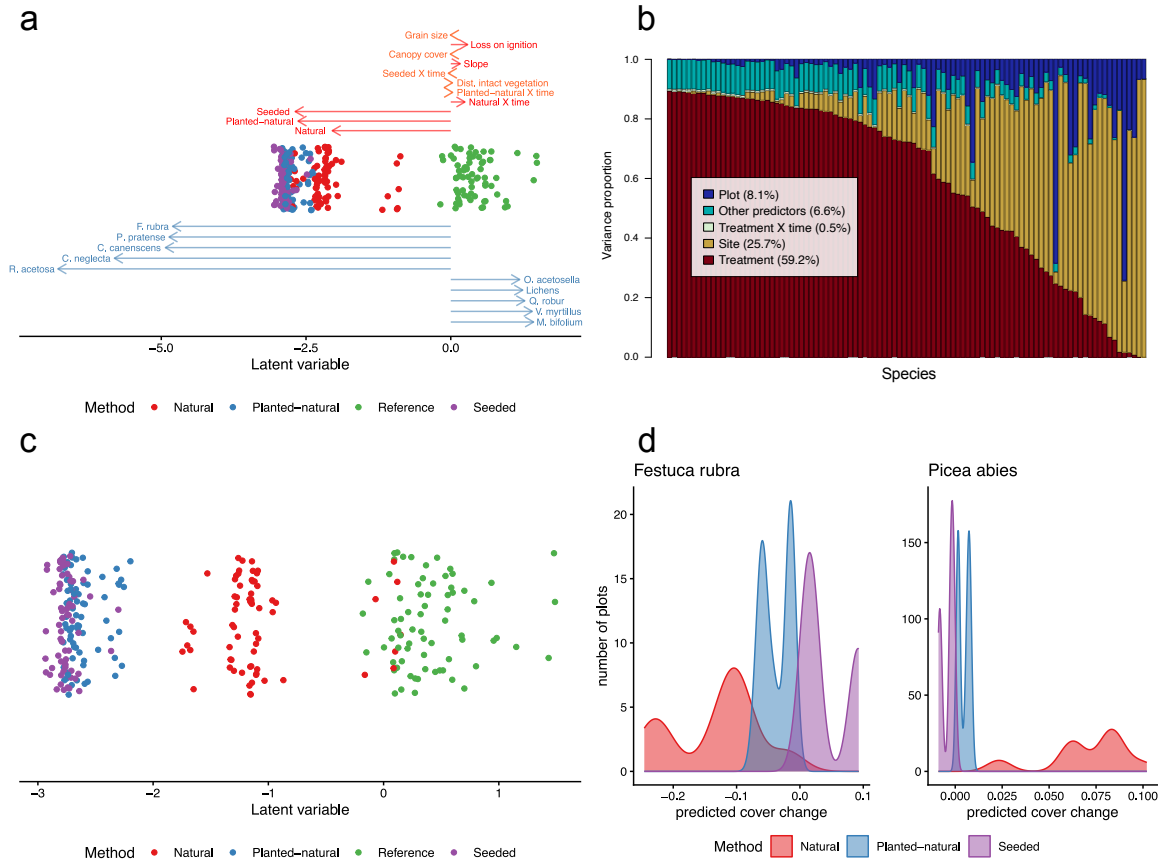

Figure 5: Visualisations from the constrained latent variable model of the roadside vegetation. (a) One-dimensional diagram of the constrained latent variable. Sites are coloured by restoration treatment; red arrows indicate the canonical coefficients of the model (X denotes an interaction effect). Dark red indicates that the 95% confidence interval of the predictor does not cross 0. Blue arrows show the species loadings of the five most positively and five most negatively associated species with the latent variable. (b) Variance partitioning of the model components of the model. Each vertical bar represents one species, sorted by the amount of variation explained by restoration vs reference. "Other predictors" encompass all variables outside the restoration treatment. Plot and Site are site-specific and species-specific random effects, respectively. Percentages indicate mean variance explained across all species (c) Predicted site scores for the latent variable with 20 years added to the site coefficients. (d) Density plot of predicted change in cover in different treatment groups for two potential indicator species, *Festuca rubra* and *Picea abies*.

because we did not estimate residual variation in the latent variables, we can be reasonably confident that there were no unaccounted for additional unobserved factors that influenced the species composition.

6 Discussion

In this article, we have provided an overview of model-based ordination with Generalised Linear Latent Variable Models and a practical introduction to a range of their uses in community ecology. We have shown that a fully model-based methodology and workflow can produce models of ecosystems and communities that are both feature-rich and highly statistically and conceptually interpretable. This includes a number of novel benefits to community ecological analysis such as prediction and uncertainty quantification.

We have attempted to make model-based ordination more apporoachable to community ecologists by demonstrating applications of the framework on two real world examples, which includes novel ecological inference. The Supplementary material containing the full code and workflows should also be useful in this regard. In the first worked example, we showed how the impact of an invasive species on a community could be described using both unconstrained and concurrent model-based ordination. In contrast to classical ordination methods, we could look at species loadings with uncertainties, as well as their associated between-species correlation estimates to indicate how strongly these associations were supported by the data. This let us paint more a comprehensive picture of what the data say about the associations between the native and invasive species in the community. The same was true for the effect of the predictors describing species co-occurrence in the communities, where visualizing the effect of covariates on individual species was able to identify the predictor most associated with the negative association between *L. lucidum* and the native species. This species-centered way of using model-based ordinations will be readily transferable to a number of other ecological questions, such as identifying indicator species related to specific environmental variables of habitats, or identifying distinct clusters of species associations in a community (see references in Tables 1 and 3).

In the second worked example, we demonstrated how a concurrent ordination can be used to estimate the effect of different ecological restoration treatments on community composition, while accounting for a spatially grouped study design. Within a single model we could include the effect of the different treatments with parameter uncertainty, accounting for the study design, and forecast how the communities will change in the future; examples of capabilities model-based ordinations offer that is not possible to do in a comprehensive way with traditional methods. The use of the methods demonstrated in this example will hopefully serve as a relevant template for other sample-focused research questions. For instance, assessing the effect of different management practices on community composition, or which level of a hierarchical habitat classification system that best explains variation in community structure (see again also references in Tables 1 and 3).

6.1 A note on species filtering

In both examples we explored the number of occurrences per species and site, which lead to removal of species in the second example. It is important to clarify that it is not strictly necessary to remove data deficient species prior to fitting a model-based ordination, but it can at times make the modelling process easier. Data deficiency can cause difficulties with model convergence, presenting results, or drawing inference, without improving the inference overall. For example, a species that only has one or two occurrences, which by chance are on top of a mountain may exhibit extreme clustering in a (constrained) ordination diagram when the ordination axis represents elevation. Here, the model will interpret the data as the species not occurring at lower elevations at all, thus placing the species at the far end of the ordination axis. This is natural as the model has not seen any other information. But if the lack of data is because the species is hard to observe, the results may be an artefact of the sampling process. Still, some extra species can add valuable information on the end points of an ecological gradient (i.e. serve to better inform the positions of site scores), so that removal is not always advisable. Such data deficiency of species is often used as an argument for analysing data in collapsed form, with species identities masked (e.g. as in NMDS), and the data are analysed on the basis of sites only. However, we argue that there is nothing inherently more complex to model-based ordination that makes it less suitable for the analysis of data deficient or rare species. The important thing is rather to distinguish between species that have few occurrences in data because they have been insufficiently sampled, so that they cannot be correctly placed in the environment, versus species that are rare for other reasons. Ideally, the pool of species being studied is clearly defined prior to data collection, so that a survey can be expanded to ensure data sufficiency for all species when necessary.

6.2 Extensions to model-based ordination

It is also important to stress that the model-based ordination framework presented here can be extended in several ways for for modelling community data outside of the main use cases presented in this article. This includes the possibility of including other sources of information, such as species traits, in the models. Currently, traits can be integrated with model-based ordinations in two main ways: (1) by fitting a fourth-corner model that estimates trait-environment interactions outside of the ordination (Niku et al., 2021; Abrego et al., 2025) (see "additional terms" in Figure 1) or 2) by swapping species and sites and fitting an ordination. Fourth-corner models based on GL(M)Ms are well established in ecology already (Brown et al., 2014; Jamil et al., 2013), and incorporating them as part of a GLLVM (see the "additional terms" in Figure 1) allows for an even larger degree of flexibility in their specification: Making in possible to simultaneously include both different community wide effects, species intercept, individual (random) species variations in environmental responses, even data on species phylogeny (van der Veen and O'Hara, 2025) together with latent variables modelling additional unexplained species correlations (see the discussion on partial ordination, Table 2). However, these models might become quite parameter-heavy when many species,

traits and environmental covariates are included. In 2), by reversing the sites and species in a concurrent or constrained ordination, the canonical coefficients can be based on traits instead of environmental variables. In other words, one would model change in community composition as a function of species' traits rather than site-specific environmental predictors. The main limitation of this model is that environmental predictors cannot be included alongside traits in informing the ordination axis. Integrating functional traits and environmental predictors together into the concurrent ordination framework, similar to approaches that have been developed for other ordination methods (ter Braak et al., 2018),is currently an active area of development for GLLVMs (see e.g. O'Hara and Veen, 2024). Alternatively, model-based ordinations can been used to look at how traits covary between species, by letting species act in the place of a site, and traits as species. In this case, the GLLVM would be used to model a hypothetical lower-dimensional community trait space (Laughlin, 2014).

Model-based ordination also offer ways to account for temporal and spatial correlation between samples. Spatio-temporal correlations in model-based ordinations can either take the form of species-specific or community-wide effects being spatio-temporally correlated, or the latent variables themselves: with proximity in the ordination reflecting proximity in space and time to a degree determined by one or more correlation coefficients estimated by the model. As such, the models may be used to disentangle spatial or temporal proximity effects from other predictors of interest not possible with classical methods. For more literature on this, see e.g. (Ovaskainen et al., 2017; Thorson et al., 2015; Ovaskainen et al., 2017; Korhonen et al., 2025; Tikhonov et al., 2019), as well as the `gllvm` vignettes (Niku, 2024). Development of spatio-temporal ordination models is currently an active area of development for GLLVM researchers.

Model-based ordination also open up a range of other avenues for modeling community ecology. This includes the possibility of using latent variables to quantify species niches (Ovaskainen et al., 2016) and niche overlap (van der Veen et al., 2024), including for models with different niche sizes of species along latent variables (van der Veen et al., 2021). Modelling species with different data types, e.g. where some species are recorded as counts and others as cover or presence absence, is now also possible with modern model-based ordination software like `gllvm`.

There has also been work done to develop methods and protocols for using pilot studies to determining the sampling effort and amount of data required to confidently answer specific ecological questions using GLLVM-type models (Maslen et al., 2023), which could potentially have a profound impact on resource- and time management when planning community ecology research. Developing new types of model-based ecological indicators or classification schemes that may be useful in management settings, would be another interesting path to explore in this regard.

6.3 When should classical methods still be used?

There are some situations where model-based ordinations will be challenging to fit to one's data, and in which classical multivariate methods might be useful as a more time-effective way of data visualisation and exploration. This would typically be the case when one is dealing with very large, high-dimensional data, like high-throughput meta-barcode datasets with several thousand OTU reads, and where access to computing time and resources are limited. In these cases, we would as a general rule recommend the use of simple, mathematically well-defined data transformations such as the classic PCA for data exploration. Even though its implicit assumptions of linearity and normality is not realistic for most ecological data, its simplicity in our view makes it easier to interpret, and to identify when it does not fit the data, than other methods. If there are ample species-specific observations, and community structure or species co-occurrences are not central to the study questions, and they can in fact be answered by a simpler GLM(M), there may no need to fit an ordination to the data at all.

Still, if the goal *is* to test an hypothesis or make quantifiable inferences about species communities, we do think model-based ordinations should be tried, even for very large and sparse datasets. Despite the fact that the models might take some time to run, and require some iterations to get well converged, the added analytical benefits will be substantial.

7 Conclusion

Applying model-based ordinations in community ecology does not have to require a change in one's research questions, theoretical frameworks or data. Rather, it is a broad and robust toolbox that gathers a wide range of methodological tools in community ecology under the same statistical roof. It can both help you "do what you were already doing", only in more powerful and informative ways, and also to address ecological questions in new ways. As we have demonstrated, this makes the framework relevant for a number of research topics. As model-based ordination becomes more widely adopted within community ecology, researchers will no doubt also discover new uses for the methods that their developers did not think of, which may again lead to further development of the framework. In our view, a constructive two-way collaboration between statistical developers and ecological practitioners will therefore be crucial to address the main ecological challenges of the 21st century.

References

Abrego, N., P. Niittynen, J. Kemppinen, and O. Ovaskainen. 2025. Joint species-trait distribution modeling: The role of intraspecific trait variation in community assembly. *Ecology 106*(9): e70174. https://doi.org/10.1002/ecy.70174 .

Aho, K., D. Derryberry, and T. Peterson. 2014. Model selection for ecologists: The worldviews of AIC and BIC. *Ecology 95*(3): 631–636. https://doi.org/10.1890/13-1452.1 .

Andrew-Priestley, M., K. Newton, M.E. Platell, L. Le Strange, H. Houridis, M. Stat, R.M.K. Yu, C. Evans, Z. Rogers, J. Pallot, J. Van Den Broek, and G.R. MacFarlane. 2022. Benthic infaunal assemblages adjacent to an ocean outfall in Australian marine waters: Impact assessment and identification of indicator taxa. *Marine Pollution Bulletin* 174: 113229. https://doi.org/10.1016/j.marpolbul.2021.113229 .

Askeyev, O., S. Monakhov, I. Askeyev, A. Askeyev, and T.H. Sparks. 2023. Fish assemblages in lakes along environmental gradients at the eastern edge of Europe. *Environmental Biology of Fishes 106*(6): 1265–1276. https://doi.org/10.1007/s10641-023-01414-0 .

Bolker, B.M., M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M.H.H. Stevens, and J.S.S. White. 2009. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution 24*(3): 127–135. https://doi.org/10.1016/j.tree.2008.10.008 .

Braak, C.J.F.t. 1985. Correspondence Analysis of Incidence and Abundance Data: Properties in Terms of a Unimodal Response Model. *Biometrics 41*(4): 859–873. https://doi.org/10.2307/2530959 .

Brasil Neto, A.B., G. Schwartz, N.C. Noronha, M.A.P. Gama, G.C. Ferreira, E.J.M. Carvalho, and N.M.d.Q.X. Brasil. 2025. Seedling planting to restore ecosystems after bauxite mining in Brazilian amazon: The need of monitoring and reviewing techniques and procedures. *Plant Ecology 226*(8): 907–920. https://doi.org/10.1007/s11258-025-01539-5 .

Brown, A.M., D.I. Warton, N.R. Andrew, M. Binns, G. Cassis, and H. Gibb. 2014. The fourth-corner solution – using predictive models to understand how species traits interact with the environment. *Methods in Ecology and Evolution 5*(4): 344–352. https://doi.org/10.1111/2041-210X.12163 .

Cheng, Z., T. Aakala, and M. Larjavaara. 2023. Elevation, aspect, and slope influence woody vegetation structure and composition but not species richness in a human-influenced landscape in northwestern Yunnan, China. *Frontiers in Forests and Global Change* 6. https://doi.org/10.3389/ffgc.2023.1187724 .

Christman, M.E., L.R. Spears, J.P. Strange, W.D. Pearse, E.K. Burchfield, and R.A. Ramirez. 2022. Land cover and climate drive shifts in *Bombus* assemblage composition. *Agriculture, Ecosystems & Environment* 339: 108113. https://doi.org/10.1016/j.agee.2022.108113 .

Crouch, C.D., B.O. Knapp, S.A. Cohen, J.S. Glitzenstein, J.L. Walker, and G.G. Wang. 2022. Longleaf pine restoration on hydric sites: Understorey plant community responses to site preparation through 15 years. *Applied Vegetation Science 25*(1): e12637. https://doi.org/10.1111/avsc.12637 .

Daudt, N.W., G. Loh, K.I. Currie, M.R. Schofield, R.O. Smith, E.J. Woehler, L. Bugoni, and W.J. Rayment. 2025. Changing species occurrences in seasonal seabird assemblages at the Subtropical Frontal Zone. *Estuarine, Coastal and Shelf Science* 323: 109405. https://doi.org/10.1016/j.ecss.2025.109405 .

de Valpine, P., D. Turek, C.J. Paciorek, C. Anderson-Bergman, D.T. Lang, and R. Bodik. 2017, April. Programming With Models: Writing Statistical Algorithms for General Model Structures With NIMBLE. *Journal of Computational and Graphical Statistics 26*(2): 403–413. https://doi.org/10.1080/10618600.2016.1172487 .

Dexter, E., G. Rollwagen-Bollens, and S.M. Bollens. 2018. The trouble with stress: A flexible method for the evaluation of nonmetric multidimensional scaling. *Limnology and Oceanography: Methods 16*(7): 434–443. https://doi.org/10.1002/lom3.10257 .

Du, H., Z. Ke, G. Jiang, and S. Huang. 2022, November. The Performances of Gelman-Rubin and Geweke's Convergence Diagnostics of Monte Carlo Markov Chains in Bayesian Analysis. *Journal of Behavioral Data Science 2*(2): 47–72. https://doi.org/10.35566/jbds/v2n2/p3 .

Dunn, P.K. and G.K. Smyth. 1996. Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics 5*(3): 236–244. https://doi.org/10.2307/1390802 .

Fanfarillo, E., D. Calabrese, C. Angiolini, G. Bacaro, S. Biagiotti, P. Castagnini, S. Loppi, T. Martellini, and S. Maccherini. 2022. Effects of conventional and organic management on plant and insect communities in a traditional elephant garlic crop. *Community Ecology 23*(3): 417–427. https://doi.org/10.1007/s42974-022-00091-w .

Fernandez, R.D., P. Castro-Díez, R. Aragón, and N. Pérez-Harguindeguy. 2021. Changes in community functional structure and ecosystem properties along an invasion gradient of Ligustrum lucidum. *Journal of Vegetation Science 32*(6): e13098. https://doi.org/10.1111/jvs.13098 .

Ferrari, S. and F. Cribari-Neto. 2004, August. Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics 31*(7): 799–815. https://doi.org/10.1080/02664760420002145 01 .

Forte, T.G.W., M. Carbognani, A. Vannini, G. Chiari, and A. Petraglia. 2024. Short-term vegetation shifts in an alpine grassland under current and simulated climate change. *Journal of Vegetation Science 35*(6): e70000. https://doi.org/10.1111/jvs.70000 .

Gabry, J., D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman. 2019. Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 182*(2): 389–402. https://doi.org/10.1111/rssa.12378 .

Gaston, K.J. and T.M. Blackburn. 2003. Dispersal and the interspecific abundance-occupancy relationship in British birds. *Global Ecology and Biogeography 12*(5): 373–379. https://doi.org/10.1046/j.1466-822X.2003.00054.x .

Gaston, K.J., T.M. Blackburn, J.J. Greenwood, R.D. Gregory, R.M. Quinn, and J.H. Lawton. 2000. Abundance–occupancy relationships. *Journal of Applied Ecology 37*(s1): 39–59. https://doi.org/10.1046/j.1365-2664.2000.00485.x .

Geissinger, E.A., C.L.L. Khoo, I.C. Richmond, S.J.M. Faulkner, and D.C. Schneider. 2022. A case for beta regression in the natural sciences. *Ecosphere 13*(2): e3940. https://doi.org/10.1002/ecs2.3940 .

Goodall, D. 1954, December. Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Australian Journal of Botany 2*(3): 304–324. https://doi.org/10.1071/BT9540304 .

Graser, A., M. Georg, J. Kallmayer, A. Marten, C. Pertl, H. Rumpf, C. Senf, and J. Kamp. 2025. Large-scale forest disturbance and associated management shape bird communities in Central European spruce forests. *Journal of Applied Ecology 62*(2): 329–343. https://doi.org/10.1111/1365-2664.14849 .

Handegard, E., I. Gjerde, R. Halvorsen, R. Lewis, K.O. Storaunet, M. Sætersdal, and O. Skarpaas. 2024. How important is Forest Age in explaining the species composition of Near-natural Spruce Forests? *Forest Ecology and Management* 569: 122170. https://doi.org/10.1016/j.foreco.2024.122170 .

Hartig, F. 2024. *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models*. R package version 0.4.7. https://doi.org/10.32614/CRAN.package.DHARMa.

Hejda, M., J. Čuda, K. Pyšková, L.C. Foxcroft, K.V. Nkuna, A. Novoa, and P. Pyšek. 2023. Impacts of invasive alien species on riparian plant communities in South African savanna. *Journal of Tropical Ecology* 39: e39. https://doi.org/10.1017/S0266467423000299 .

Helbing, F., T. Fartmann, C. Morkel, and D. Poniatowski. 2023. Rapid response of vascular plants and insects to restoration of montane grasslands. *Frontiers in Ecology and Evolution* 11. https://doi.org/10.3389/fevo.2023.1148266 .

Herrmann, A., K. Grabow, and A. Martens. 2022. The invasive crayfish Faxonius immunis causes the collapse of macroinvertebrate communities in Central European ponds. *Aquatic Ecology 56*(3): 741–750. https://doi.org/10.1007/s10452-021-09935-5 .

Ho, J., T. Tumkaya, S. Aryal, H. Choi, and A. Claridge-Chang. 2019, July. Moving beyond P values: data analysis with estimation graphics. *Nature Methods 16*(7): 565–566. https://doi.org/10.1038/s41592-019-0470-3 .

Hu, Z., M. Fernández-Martínez, Q. He, Z. Xu, L. Jiang, G. Zhou, J. Chen, M. Nie, Q. Yu, H. Feng, Z. Huang, and S.T. Michaletz. 2024. Fungal composition associated with host tree

identity mediates nutrient addition effects on wood microbial respiration. *Ecology 105*(8): e4375. https://doi.org/10.1002/ecy.4375 .

Hui, F.K. 2025. *boral: Bayesian Ordination and Regression AnaLysis*. R package version 2.0.3. https://doi.org/10.32614/CRAN.package.boral.

Hui, F.K., S. Taskinen, S. Pledger, S.D. Foster, and D.I. Warton. 2015. Model-based approaches to unconstrained ordination. *Methods in Ecology and Evolution 6*(4): 399–411. https://doi.org/10.1111/2041-210X.12236 .

Jamil, T., W.A. Ozinga, M. Kleyer, and C.J. ter Braak. 2013. Selecting traits that explain species–environment relationships: a generalized linear mixed model approach. *Journal of Vegetation Science 24*(6): 988–1000. https://doi.org/10.1111/j.1654-1103.2012.12036.x .

Jeliazkov, A., Y. Gavish, C.J. Marsh, J. Geschke, N. Brummitt, D. Rocchini, P. Haase, W.E. Kunin, and K. Henle. 2022. Sampling and modelling rare species: Conceptual guidelines for the neglected majority. *Global Change Biology 28*(12): 3754–3777. https://doi.org/10.1111/gcb.16114 .

Jong, J.C. and S. Kotz. 1999. On a Relation Between Principal Components and Regression Analysis. *The American Statistician 53*(4): 349–351. https://doi.org/10.2307/2686055 .

Jørgensen, B. 1987. Exponential Dispersion Models. *Journal of the Royal Statistical Society. Series B (Methodological) 49*(2): 127–162. https://doi.org/10.1111/j.2517-6161.1987.tb01685.x .

Jupke, J.F. and R.B. Schäfer. 2020. Should ecologists prefer model- over distance-based multivariate methods? *Ecology and Evolution 10*(5): 2417–2435. https://doi.org/10.1002/ece3.6059 .

Kalusová, V., N. Čeplová, M. Chytrý, J. Danihelka, D. Pavel, K. Fajmon, O. Hájek, V. Kalníková, P. Novák, V. Řehořek, J. Těšitel, L. Tichý, T. Wirth, and Z. Lososová. 2019. Similar responses of native and alien floras in European cities to climate. *Journal of Biogeography 46*(7): 1406–1418. https://doi.org/10.1111/jbi.13591 .

Kendal, W.S. 2004. Taylor's ecological power law as a consequence of scale invariant exponential dispersion models. *Ecological Complexity 1*(3): 193–209. https://doi.org/10.1016/j.ecocom.2004.05.001 .

Korhonen, P., F.K. Hui, J. Niku, and S. Taskinen. 2023. Fast and universal estimation of latent variable models using extended variational approximations. *Statistics and Computing 33*(1): 26. https://doi.org/10.1007/s11222-022-10189-w .

Korhonen, P., F.K.C. Hui, J. Niku, S. Taskinen, and B. van der Veen. 2024. A comparison of joint species distribution models for percent cover data. *Methods in Ecology and Evolution 15*(12): 2359–2372. https://doi.org/10.1111/2041-210X.14437 .

Korhonen, P., F.K.C. Hui, J. Niku, S. Taskinen, and B. van der Veen. 2025. Gllvm 2.0: Fast fitting of advanced ordination methods and joint species distribution models. *PeerJ* 13: e20338. https://doi.org/10.7717/peerj.20338 .

Korolyuk, A.Y., H.F. Shomurodov, B.S. Khabibullaev, and Z.S. Sadinov. 2024. Composition and Structure of Tugai Communities in the Indication of Ecological Conditions in the Lower Amu Dar'ya. *Contemporary Problems of Ecology 17*(1): 106–111. https://doi.org/10.1134/S1995425524010074 .

Kubinec, R. 2023, October. Ordered Beta Regression: A Parsimonious, Well-Fitting Model for Continuous Data with Lower and Upper Bounds. *Political Analysis 31*(4): 519–536. https://doi.org/10.1017/pan.2022.20 .

Lam-Gordillo, O., G.L. Petersen, S.F. Hailes, K. Carter, N.H. Salmond, L. McCartain, M. Ferries, B. Greenfield, E.J. Douglas, R. Hattingh, T. Drylie, B. Shanahan, and A.M. Lohrer. 2025. Extreme weather event causes contrasting macrobenthic disturbance-recovery dynamics in two New Zealand estuaries. *Next Research 2*(3): 100726. https://doi.org/10.1016/j.nexres.2025.100726 .

Lanta, V., P. Liancourt, J. Altman, T. Černý, M. Dvorský, P. Fibich, L. Götzenberger, O. Hornych, J. Miklín, P. Petřík, P. Pyšek, L. Čížek, and J. Doležal. 2022. Determinants of invasion by single versus multiple plant species in temperate lowland forests. *Biological Invasions 24*(8): 2513–2528. https://doi.org/10.1007/s10530-022-02793-8 .

Larson, D.L., M. Simanonok, A. Landsman, J.L. Larson, C. Davies, and C.R.V. Otto. 2024. Bee Habitat, but Not Bee Community Structure, Varies Across Grassland Management in Four National Parks in the Mid-Atlantic, USA. *Ecology and Evolution 14*(12): e70719. https://doi.org/10.1002/ece3.70719 .

Laughlin, D.C. 2014. The intrinsic dimensionality of plant traits and its relevance to community assembly. *Journal of Ecology 102*(1): 186–193. https://doi.org/10.1111/1365-2745.12187 .

Lavielle, M. and L. Aarons. 2016, February. What do we mean by identifiability in mixed effects models? *Journal of Pharmacokinetics and Pharmacodynamics 43*(1): 111–122. https://doi.org/10.1007/s10928-015-9459-4 .

Legendre, P. and L. Legendre. 2012. Chapter 11 - Canonical analysis, In *Numerical Ecology*, eds. Legendre, P. and L. Legendre, Volume 24 of *Developments in Environmental Modelling*, 625–710. Elsevier. https://doi.org/10.1016/B978-0-444-53868-0.50011-3.

Li, S., Z. Qian, J. Yang, Y. Lin, H. Li, and L. Chen. 2022. Seasonal variation in structure and function of gut microbiota in Pomacea canaliculata. *Ecology and Evolution 12*(8): e9162. https://doi.org/10.1002/ece3.9162 .

Lourenço, Á., C.V. Souza, A.F. Mendonça, G.G. Reis, P.F. Linhares, R.P. Moura, and E.M. Vieira. 2024. Increasing fire severity alters the species composition and decreases

richness of seeds potentially dispersed by small mammals. *Biotropica 56*(3): e13318. https://doi.org/10.1111/btp.13318 .

Maslen, B., G. Popovic, M. Lim, E. Marzinelli, and D. Warton. 2023. How many sites? Methods to assist design decisions when collecting multivariate data in ecology. *Methods in Ecology and Evolution 14*(6): 1564–1573. https://doi.org/10.1111/2041-210X.14094 .

Matavelli, R., J.M. Oliveira, J. Soininen, M.C. Ribeiro, and J. Bertoluci. 2022. Altitude and temperature drive anuran community assembly in a Neotropical mountain region. *Biotropica 54*(3): 607–618. https://doi.org/10.1111/btp.13074 .

Maunsell, S.C., R.L. Kitching, P. Greenslade, A. Nakamura, and C.J. Burwell. 2013. Springtail (Collembola) assemblages along an elevational gradient in Australian subtropical rainforest. *Australian Journal of Entomology 52*(2): 114–124. https://doi.org/10.1111/aen.12012 .

McGillycuddy, M., D.I. Warton, G. Popovic, and B.M. Bolker. 2025. Parsimoniously fitting large multivariate random effects in glmmTMB. *Journal of Statistical Software 112*(1): 1–19. https://doi.org/10.18637/jss.v112.i01 .

Mehlhoop, A.C., A.B. Skrindo, M. Evju, and D. Hagen. 2022. Best practice—Is natural revegetation sufficient to achieve mitigation goals in road construction? *Applied Vegetation Science 25*(3): e12673. https://doi.org/10.1111/avsc.12673 .

Muff, S., E.B. Nilsen, R.B. O'Hara, and C.R. Nater. 2022. Rewriting results sections in the language of evidence. *Trends in Ecology & Evolution 37*(3): 203–210. https://doi.org/10.1016/j.tree.2021.10.009 .

Mulders, Y., K. Filbee-Dexter, S. Bell, N.E. Bosch, A. Pessarrodona, D. Sahin, S. Vranken, S. Zarco-Perello, and T. Wernberg. 2022. Intergrading reef communities across discrete seaweed habitats in a temperate–tropical transition zone: Lessons for species reshuffling in a warming ocean. *Ecology and Evolution 12*(1): e8538. https://doi.org/10.1002/ece3.8538 .

Müller, S., J.L. Scealy, and A.H. Welsh. 2013. Model Selection in Linear Mixed Models. *Statistical Science 28*(2): 135–167. https://doi.org/10.1214/12-STS410 .

Naz, F., M. Arif, T. Xue, and L. Changxiao. 2024. Seasonal dynamics of soil ecosystems in the riparian zones of the Three Gorges Reservoir, China. *Global Ecology and Conservation* 54: e03174. https://doi.org/10.1016/j.gecco.2024.e03174 .

Niku, J. 2024. Correlation structures for latent variables and row effects. https://jenniniku.github.io/gllvm/articles/vignette9.html.

Niku, J. 2026. Assessing Convergence in gllvm models. https://jenniniku.github.io/gllvm/articles/vignette10.html.

Niku, J., W. Brooks, R. Herliansyah, F.K.C. Hui, P. Korhonen, S. Taskinen, B. van der Veen, and D.I. Warton 2025. *gllvm: Generalized Linear Latent Variable Models*. R package version 2.0.10. https://doi.org/10.32614/CRAN.package.gllvm.

Niku, J., F.K.C. Hui, S. Taskinen, and D.I. Warton. 2019. Gllvm: Fast analysis of multivariate abundance data with generalized linear latent variable models in r. *Methods in Ecology and Evolution 10*(12): 2173–2182. https://doi.org/10.1111/2041-210X.13303 .

Niku, J., F.K.C. Hui, S. Taskinen, and D.I. Warton. 2021. Analyzing environmental-trait interactions in ecological communities with fourth-corner latent variable models. *Environmetrics 32*(6): e2683. https://doi.org/10.1002/env.2683 .

Økland, R.H. 1996. Are ordination and constrained ordination alternative or complementary strategies in general ecological studies? *Journal of Vegetation Science 7*(2): 289–292. https://doi.org/10.2307/3236330 .

Ovaskainen, O., N. Abrego, P. Halme, and D. Dunson. 2016. Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution 7*(5): 549–555. https://doi.org/10.1111/2041-210X.12501 .

Ovaskainen, O., G. Tikhonov, D. Dunson, V. Grøtan, S. Engen, B.E. Sæther, and N. Abrego. 2017. How are species interactions structured in species-rich communities? A new method for analysing time-series data. *Proceedings of the Royal Society B: Biological Sciences 284*(1855): 20170768. https://doi.org/10.1098/rspb.2017.0768 .

Ovaskainen, O., G. Tikhonov, A. Norberg, F. Guillaume Blanchet, L. Duan, D. Dunson, T. Roslin, and N. Abrego. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters 20*(5): 561–576. https://doi.org/10.1111/ele.12757 .

O'Hara, R.B. and B.v.d. Veen. 2024, January. Hierarchical Ordination, A unifying framework for drivers of community processes. Pages: 2024.01.08.574701 Section: New Results.

Pedley, S.M., P. Wolstenholme, and P.M. Dolman. 2023. Plantation clear-fell patches benefit heathland arthropods. *Ecological Solutions and Evidence 4*(3): e12281. https://doi.org/10.1002/2688-8319.12281 .

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. Coda: Convergence diagnosis and output analysis for mcmc. *R News 6*(1): 7–11 .

Podani, J. and I. Miklós. 2002. Resemblance Coefficients and the Horseshoe Effect in Principal Coordinates Analysis. *Ecology 83*(12): 3331–3343. https://doi.org/10.1890/0012-9658(2002)083[3331:RCATHE]2.0.CO;2 .

Popovic, G.C., D.I. Warton, F.J. Thomson, and A.T. Moles. 2019. Untangling direct species associations from indirect mediator species effects with graphical models. *Methods in Ecology and Evolution 10*(9): 1571–1583. https://doi.org/10.1111/2041-210X.13247 .

Reeve, S., D.C. Deane, C. McGrannachan, G. Horner, C. Hui, and M. McGeoch. 2022. Rare, common, alien and native species follow different rules in an understory plant community. *Ecology and Evolution 12*(3): e8734. https://doi.org/10.1002/ece3.8734 .

Reinsel, G.C., R.P. Velu, and K. Chen. 2022. Reduced-Rank Regression Model, In *Multivariate Reduced-Rank Regression: Theory, Methods and Applications*, eds. Reinsel, G.C., R.P. Velu, and K. Chen, 19–73. New York, NY: Springer. https://doi.org/10.1007/978-1-0716-2793-8_2.

Reis, B.P., K. Szitár, A. Kövendi-Jakó, K. Török, N. Sáradi, E. Csákvári, and M. Halassy. 2022. The long-term effect of initial restoration intervention, landscape composition, and time on the progress of Pannonic sand grassland restoration. *Landscape and Ecological Engineering 18*(4): 429–440. https://doi.org/10.1007/s11355-022-00512-y .

Ribeiro, L.G., H.H. Puerari, A.O. Silva, K.A. Vaz, J.V. dos Santos, C.A. Nunes, M.V. Barbosa, M.R. da Rocha, J.O. Siqueira, and M.A.C. Carneiro. 2023. Structure and composition of the nematode community in a restoration area affected by iron tailings. *Pedobiologia* 97–98: 150864. https://doi.org/10.1016/j.pedobi.2023.150864 .

Roy, V. 2020, March. Convergence Diagnostics for Markov Chain Monte Carlo. *Annual Review of Statistics and Its Application 7*(Volume 7, 2020): 387–412. https://doi.org/10.1146/annurev-statistics-031219-041300 .

Russell, L.K., S.J. Evans, L. Smith, C.M. Bevers, A.P. Luxford, W.J. Stubbs, and J.B. Wilson. 2005. Distribution/abundance relations in a New Zealand grassland landscape. *New Zealand Journal of Ecology 29*(1): 61–68 .

Sahade, R., C. Lagger, L. Torre, F. Momo, P. Monien, I. Schloss, D.K.A. Barnes, N. Servetto, S. Tarantelli, M. Tatián, N. Zamboni, and D. Abele. 2015. Climate change and glacier retreat drive shifts in an Antarctic benthic ecosystem. *Science Advances 1*(10): e1500050. https://doi.org/10.1126/sciadv.1500050 .

Sainani, K.L. 2014. Explanatory Versus Predictive Modeling. *PM&R 6*(9): 841–844. https://doi.org/10.1016/j.pmrj.2014.08.941 .

Sanchez, K.A., L. Benedict, and E.A. Holt. 2023. Landscape composition is a stronger determinant than noise and light of avian community structure in an urbanizing county. *Frontiers in Ecology and Evolution* 11. https://doi.org/10.3389/fevo.2023.1254280 .

Shembo, A.K., S.S. Ayichew, I. Stiers, A. Geremew, and L. Carson. 2024. Classification and ordination analysis of wild medicinal plants in Ada'a district, Ethiopia: Implication for sustainable conservation and utilization. *Ecological Frontiers 44*(4): 809–819. https://doi.org/10.1016/j.ecofro.2024.04.002 .

Shmueli, G. 2010. To Explain or to Predict? *Statistical Science 25*(3): 289–310. https://doi.org/10.1214/10-STS330 .

Souza-Alonso, P., Y. Lechuga-Lago, A. Guisande-Collazo, and L. González. 2022. Evidence of functional and structural changes in the microbial community beneath a succulent invasive plant in coastal dunes. *Journal of Plant Ecology 15*(6): 1154–1167. https://doi.org/10.1093/jpe/rtac026 .

Stan Development Team. 2026a. How to Diagnose and Resolve Convergence Problems.

Stan Development Team. 2026b. Stan Reference Manual. Version 2.39. https://mc-stan.org.

Suárez-Tangil, B.D. and A. Rodríguez. 2023. Environmental filtering drives the assembly of mammal communities in a heterogeneous Mediterranean region. *Ecological Applications 33*(2): e2801. https://doi.org/10.1002/eap.2801 .

ter Braak, C. 1987. *Unimodal models to relate species to environment*. Ph. D. thesis, Ter Braak, [S.l.]. Proefschrift Wageningen.

Ter Braak, C.J.F. and L.G. Barendregt. 1986, March. Weighted averaging of species indicator values: Its efficiency in environmental calibration. *Mathematical Biosciences 78*(1): 57–72. https://doi.org/10.1016/0025-5564(86)90031-3 .

ter Braak, C.J.F. and I.C. Prentice. 1988. A Theory of Gradient Analysis, In *Advances in Ecological Research*, eds. Begon, M., A.H. Fitter, E.D. Ford, and A. Macfadyen, Volume 18, 271–317. Academic Press. https://doi.org/10.1016/S0065-2504(08)60183-X.

ter Braak, C.J.F. and I.C. Prentice. 2004. A Theory of Gradient Analysis, *Advances in Ecological Research*, Volume 34 of *Advances in Ecological Research: Classic Papers*, 235–282. Academic Press. https://doi.org/10.1016/S0065-2504(03)34003-6.

ter Braak, C.J.F. and P. Šmilauer. 2015. Topics in constrained and unconstrained ordination. *Plant Ecology 216*(5): 683–696. https://doi.org/10.1007/s11258-014-0356-5 .

ter Braak, C.J.F., P. Šmilauer, and S. Dray. 2018. Algorithms and biplots for double constrained correspondence analysis. *Environmental and Ecological Statistics 25*(2): 171–197. https://doi.org/10.1007/s10651-017-0395-x .

Thorson, J.T., M.D. Scheuerell, A.O. Shelton, K.E. See, H.J. Skaug, and K. Kristensen. 2015. Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution 6*(6): 627–637. https://doi.org/10.1111/2041-210X.12359 .

Tikhonov, G., L. Duan, N. Abrego, G. Newell, M. White, D. Dunson, and O. Ovaskainen. 2019. Computationally efficient joint species distribution modeling of big spatial data. https://doi.org/10.1002/ecy.2929 .

Tikhonov, G., O. Ovaskainen, J. Oksanen, M. de Jonge, O. Opedal, and T. Dallas 2025. *Hmsc: Hierarchical Model of Species Communities*. R package version 3.3-7. https://doi.org/10.32614/CRAN.package.Hmsc.

van der Veen, B. 2022. Next generation ordination with Generalized Linear Latent Variable Models, *Statistical advances in multispecies models for community ecology*. Norges teknisk-naturvitenskapelige universitet.

van der Veen, B., F.K.C. Hui, K.A. Hovstad, and R.B. O'Hara. 2023. Concurrent ordination: Simultaneous unconstrained and constrained latent variable modelling. *Methods in Ecology and Evolution 14*(2): 683–695. https://doi.org/10.1111/2041-210X.14035 .

van der Veen, B., F.K.C. Hui, K.A. Hovstad, E.B. Solbu, and R.B. O'Hara. 2021. Model-based ordination for species with unequal niche widths. *Methods in Ecology and Evolution 12*(7): 1288–1300. https://doi.org/10.1111/2041-210X.13595 .

van der Veen, B. and R.B. O'Hara. 2025. Fast fitting of phylogenetic mixed-effects models. Preprint, arXiv, https://doi.org/10.48550/arXiv.2408.05333.

van der Veen, B., R.B. O'Hara, F.K. Hui, and K.A. Hovstad. 2024. Predicting niche overlap with model-based ordination. *Ecography 2024*(4): e06938. https://doi.org/10.1111/ecog.06938 .

Vehtari, A., A. Gelman, D. Simpson, B. Carpenter, and P.C. Bürkner. 2021, June. Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis 16*(2): 667–718. https://doi.org/10.1214/20-BA1221 .

Wang, D., Y. Zhu, Z. Li, X. Yang, S. Kwon, Z. Shi, and T. Indree. 2025. Vegetation community reassembly changes in Eastern Eurasian degraded steppe: Roles of environmental filtering and biotic interaction. *Ecological Frontiers* 45. https://doi.org/10.1016/j.ecofro.2025.01.015 .

Warton, D.I., S.D. Foster, G. De'ath, J. Stoklosa, and P.K. Dunstan. 2015. Model-based thinking for community ecology. *Plant Ecology 216*(5): 669–682. https://doi.org/10.1007/s11258-014-0366-3 .

Warton, D.I. and F.K.C. Hui. 2011. The arcsine is asinine: the analysis of proportions in ecology. *Ecology 92*(1): 3–10. https://doi.org/10.1890/10-0340.1 .

Warton, D.I. and F.K.C. Hui. 2017. The central role of mean-variance relationships in the analysis of multivariate abundance data: A response to Roberts (2017). *Methods in Ecology and Evolution 8*(11): 1408–1414. https://doi.org/10.1111/2041-210X.12843 .

Warton, D.I., S.T. Wright, and Y. Wang. 2012. Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution 3*(1): 89–101. https://doi.org/10.1111/j.2041-210X.2011.00127.x .

Wong, R., N. Perkins, J. Monk, M. Prall, A. Lauermann, and N. Barrett. 2026. Decadal changes in California's temperate mesophotic reef invertebrate community through the 2014–2016 northeast Pacific marine heatwave. *Marine Environmental Research* 213: 107644. https://doi.org/10.1016/j.marenvres.2025.107644 .

Wood, Simon N., a. 2017. *Generalized additive models : an introduction with R* (Second edition. ed.). Boca Raton, FL :: CRC Press, Taylor & Francis Group,.

Yee, T.W. 2004. A New Technique for Maximum-Likelihood Canonical Gaussian Ordination. *Ecological Monographs 74*(4): 685–701. https://doi.org/10.1890/03-0078 .

Yee, T.W. 2025. *VGAM: Vector Generalized Linear and Additive Models*. R package version 1.1-14. https://doi.org/10.32614/CRAN.package.VGAM.

Young, E.L., K.M. Halanych, D.J. Amon, I. Altamira, J.R. Voight, N.D. Higgs, and C.R. Smith. 2022. Depth and substrate type influence community structure and diversity of wood and whale-bone habitats on the deep NE Pacific margin. *Marine Ecology Progress Series* 687: 23–42. https://doi.org/10.3354/meps14005 .

Zuur, A.F. and E.N. Ieno. 2016. A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution* *7*(6): 636–645. https://doi.org/10.1111/2041-210X.12577 .

Zuur, A.F., E.N. Ieno, and C.S. Elphick. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* *1*(1): 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x .

Zuur, A.F., E.N. Ieno, N. Walker, A.A. Saveliev, and G.M. Smith. 2009. *Mixed Effects Models and Extensions in Ecology with R* (1 ed.). Statistics for Biology and Health. New York, NY: Springer. https://doi.org/10.1007/978-0-387-87458-6.