Wiley Interdisciplinary Reviews: Computational Statistics

WILEY

WIREs COMPUTATIONAL STATISTICS

OVERVIEW OPEN ACCESS

A Review of Generalized Linear Latent Variable Models and Related Computational Approaches

Pekka Korhonen | Klaus Nordhausen | Sara Taskinen

Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland

Correspondence: Sara Taskinen (sara.l.taskinen@jyu.fi)

Received: 31 May 2024 | **Revised:** 25 August 2024 | **Accepted:** 21 September 2024

Edited by: Emily Frieben, Commissioning Editor and David W. Scott, Editor-in-Chief

Funding: This study was supported by European Cooperation in Science and Technology (Grant CA21163); Luonnontieteiden ja Tekniikan Tutkimuksen Toimikunta (Grants 453691 and 363261); and Koneen Säätiö.

Keywords: factor analysis | Gauss–Hermite | Laplace approximation | likelihood function | MCMC | quasi-likelihood

ABSTRACT

Generalized linear latent variable models (GLLVMs) have become mainstream models in this analysis of correlated, m-dimensional data. GLLVMs can be seen as a reduced-rank version of generalized linear mixed models (GLMMs) as the latent variables which are of dimension $p \ll m$ induce a reduced-rank covariance structure for the model. Models are flexible and can be used for various purposes, including exploratory analysis, that is, ordination analysis, estimating patterns of residual correlation, multivariate inference about measured predictors, and prediction. Recent advances in computational tools allow the development of efficient, scalable algorithms for fitting GLLMVs for any response distribution. In this article, we discuss the basics of GLLVMs and review some options for model fitting. We focus on methods that are based on likelihood inference. The implementations available in R are compared via simulation studies and an example illustrates how GLLVMs can be applied as an exploratory tool in the analysis of data from community ecology.

1 | Introduction

Latent variable (LV) modeling is nowadays one of the mainstream methods in the analysis of complex, multivariate data. Although the origins of LV modeling are mainly in social sciences and psychology, models are nowadays increasingly important also in other fields of science. Some recent examples include applications to finance (Huber, Scaillet, and Victoria-Feser 2009), neuroscience (Turner, Wang, and Merkle 2017), medicine (Bianconcini and Cagnone 2021), microbiology (Zeng, Zhao, and Wang 2021), and ecology (Dombrovski, Zhurauliou, and Ashton-Butt 2022).

By LV models we mean statistical models that relate latent, that is, unobserved variables to observed variables in some way. The core idea is that a small amount of LVs drive the dependency structure of the observed data. Traditionally, LV models have been divided into four categories depending on the assumption made on observed and LVs. In latent profile models (for continuous observed variables) and latent class models (for discrete observed variables) the LVs are assumed to be discrete and the methods aim to recover hidden groups from observed data (Gibson 1959; Lazarsfeld and Henry 1968). These methods can be seen as model-based clustering methods and they belong to a larger family of LV techniques called finite mixture models. See Oberski (2016) for a recent review on the two models. In this review, we focus on models in which observed variables can be of any type and LVs are assumed to be continuous. A general framework for such models is generalized linear latent variable modeling framework (GLLVM, Skrondal and Rabe-Hesketh 2004; Bartholomew, Knott, and Moustaki 2011) which includes models for continuous and

discrete responses (or mixed responses) as special cases. GLLVMs that assume continuous responses, that is, factor analysis (FA) models, have been classically used to estimate causes of correlation in multivariate data (confirmatory FA) or as an explanatory tool to reduce the dimension of multivariate data (explanatory FA). However, they also serve as a basis for more general methods such as structural equation models (SEM; Jöreskog 1969, 1970). As will be discussed later, generalized linear mixed models (GLMMs) share certain similarities with GLLVMs, as in GLMMs so-called random effects play similar roles as the LVs in GLLVMs. Both GLMMs, GLLVMs, and SEMs model complex dependencies in data using components that are not directly measurable. In GLMMs, random effects typically account for variability because of clustering effects, such as repeated measures on the same subject or students nested within school classes. In contrast, in GLLVMs, the LVs are of direct interest, particularly in how they affect the outcome variables. SEMs further expand on GLLVMs by specifying causal relationships among LVs and between latent and observed variables. For a more detailed discussion of the similarities and differences among these three models, see, for example Skrondal and Rabe-Hesketh (2004).

Nowadays, when recent developments in computational tools have allowed fast and efficient LV model fitting for any response type, GLLVMs have become a standard joint modeling tool for multivariate analysis that can address various data exploration and analysis questions (Warton et al. 2015; Ovaskainen and Abrego 2020).

Consider for a moment a continuous response case and write $\mathbf{Y} = \left(\mathbf{y}_1 \cdots \mathbf{y}_n\right)^\top$ for a $n \times m$ response matrix, where $\mathbf{y}_i = \left(y_{i1}, \ldots, y_{im}\right)^\top$ is the m-vector of responses recorded at observational unit $i = 1, \ldots, n$. The most widely known dimension reduction method is principal component analysis (PCA, Pearson 1901; Hotelling 1933) which transforms multivariate data into uncorrelated components, that is, $\mathbf{y}_i$ is decomposed into

$$\mathbf{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\mathbf{u}_i, \tag{1}$$

where $\boldsymbol{\nu}$ is a m-dimensional location vector, $\boldsymbol{\Lambda}$ is an orthogonal $m \times m$ matrix specifying the principal axes, and the components of m-vector $\mathbf{u}_i = \left(u_{i1}, \ldots, u_{im}\right)^\top$, that is, principal component scores, are uncorrelated and ordered according to their variances. PCA is often used as a pre-processing step and only p components explaining most of the variation in data are retained for further analyses. PCA is a non-probabilistic method as no distributional assumptions are made. If we make some moment conditions on $\mathbf{u}_i$, methods such as independent component analysis (ICA, Comon 1994) can be used to decompose data into components that are independent and ordered according to some measure of non-Gaussianity. Notice that $\boldsymbol{\Lambda}$ and $\mathbf{u}_i$ in model (1) are confounded and the model is not uniquely defined. We return to a similar indeterminacy issue in connection with FA models later in this article. For a comprehensive review of ICA methods and their variants, see Comon and Jutten (2010) and Nordhausen and Oja (2018). For the extensions of model (1) to dependent data settings, see Ensor (2013), Bachoc et al. (2020), Pan et al. (2021), Virta et al. (2020), and Muehlmann, De Iaco, and Nordhausen (2023), for example. PCA has been extended to discrete data settings especially in the matrix factorization literature, see, for example, Cao and Xie (2016), Lee and Seung (2000), Collins, Dasgupta, and Schapire (2001), and Smallman, Artemiou, and Morgan (2018).

The popularity of PCA and ICA arises from their computational simplicity, but a notable drawback is the absence of a probabilistic model for the observed data. By assuming a probability distribution for LVs we can better account for key statistical properties of the data at hand. It also allows easy comparison with other probabilistic techniques, and the availability of inferential, prediction, model selection, and diagnostic tools in addition to dimension reduction. A probabilistic version of PCA (probabilistic PCA) assumes that the observations are generated via

$$\mathbf{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\mathbf{u}_i + \boldsymbol{\epsilon}_i, \tag{2}$$

where $m \times p$ matrix $\boldsymbol{\Lambda}$ now relates the p-dimensional LVs $\mathbf{u}_i = \left(u_{i1}, \ldots, u_{ip}\right)^\top$ to observed data (Lawley 1953; Anderson and Rubin 1956). The noise components in m-vector $\boldsymbol{\epsilon}_i = \left(\epsilon_{i1}, \ldots, \epsilon_{im}\right)^\top$ are assumed to be uncorrelated and normally distributed, $\boldsymbol{\epsilon}_i \sim N_m\left(\mathbf{0}, \sigma^2\mathbf{I}_p\right)$. Tipping and Bishop (1999) give a detailed discussion of model (2) and show how the estimation of $\boldsymbol{\Lambda}$ and σ^2 can be conducted iteratively using an EM algorithm. Recently, the extensions of probabilistic PCA to exponential family case and related computational approaches have been considered in series of articles, see, for example, Chiquet, Mariadassou, and Robin (2018), Wang and Carvalho (2024), and references therein.

In this article, we review GLLVMs, which extend (2) to any response distributions, and related computational approaches focusing on maximum likelihood estimation (MLE)-based methods. The structure of this article is as follows. Section 2 reviews the considered models in detail. Additionally, we discuss also the tools available for selecting the optimal number of factors and address the identifiability issues inherent in these models. Related to this, we focus in this review on exploratory methods and only refer to confirmatory methods where the interpretability of factors is of main interest. The most recent computational tools are reviewed in Section 3, and some of these methods are compared using simulation studies in Section 4. The article is concluded with some discussion in Section 5. For a list of abbreviations used in this article, see Table 1.

2 | Generalized Linear Latent Variable Models

2.1 | Factor Analysis

The main idea in factor analysis (FA) is to explain the correlation across continuous, multivariate responses using a small number of common factors. It is thus assumed that there are some important unmeasured predictors that introduce correlation across responses, and the main idea is to find those predictors. FA models are the most popular LV models used in psychology, and its origins date back to studies of intelligence by Spearman (1904). The name FA was introduced in Thurstone (1931), and a statistical treatment of the model was considered in Anderson and Rubin (1956) and more comprehensively in Lawley and Maxwell (1962).

TABLE 1 | Commonly occurring abbreviations in this article.

Term	Abbreviation
(Adaptive) Gauss–Hermite quadrature	(A)GHQ
Akaike Information Criterion (corrected)	AIC(c)
Automatic differentiation	AD
Bayesian Information Criterion	BIC
Expectation propagation	EP
Expectation–maximization	EM
Extended variational approximation	EVA
Factor analysis	FA
Generalized linear latent variable model	GLLVM
Generalized linear mixed model	GLMM
Generalized linear model	GLM
Independent component analysis	ICA
Integrated nested Laplace approximations	INLA
Item response theory	IRT
Laplace's approximation	LA
Latent Gaussian modeling	LGM
Latent variable	LV
Markov Chain Monte Carlo	MCMC
Maximum likelihood estimation	MLE
Penalized quasi-likelihood	PQL
Principal component analysis	PCA
Probability density/mass function	PDF/PMF
Quasi-maximum likelihood	QML
Root mean square error	RMSE
Simulated maximum likelihood	SML
Stochastic partial differential equation	SPDE
Structural equation modeling	SEM
Variational approximation/inference	VA/VI

2.1.1 | Model Formulation

Write now $\mathbf{Y} = \left(\mathbf{y}_1 \cdots \mathbf{y}_n\right)^\top$ for a $n \times m$ response matrix, where $\mathbf{y}_i = \left(y_{i1}, \ldots, y_{im}\right)^\top$ is the m-vector of responses recorded at observational unit $i = 1, \ldots, n$. The FA model assumes that the responses are generated via

$$\mathbf{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\mathbf{u}_i + \boldsymbol{\epsilon}_i, \tag{3}$$

where the m-vector $\boldsymbol{\nu} = \left(\nu_1, \ldots, \nu_m\right)^\top$ specifies the location of the model. The $m \times p$ matrix $\boldsymbol{\Lambda} = \left[\lambda_1 \cdots \lambda_m\right]^\top$, where $\lambda_j = \left(\lambda_{j1}, \ldots, \lambda_{jp}\right)^\top$, is a matrix of factor loadings, and the p-dimensional LVs (common factors) $\mathbf{u}_i = \left(u_{i1}, \ldots, u_{ip}\right)^\top$ are usually assumed to be independent vectors from a standard multivariate normal distribution, $\mathbf{u}_i \sim N_p\left(\mathbf{0}, \mathbf{I}_p\right)$, where $\mathbf{I}_p$ is a $p \times p$ identity matrix. This fixes the scales and locations of factors, but they still remain rotation invariant. We return to this topic in Section 2.1.2. If the additional noise vectors $\boldsymbol{\epsilon}_i = \left(\epsilon_{i1}, \ldots, \epsilon_{im}\right)^\top$ are also assumed to be normally distributed, $\boldsymbol{\epsilon}_i \sim N_m(\mathbf{0}, \boldsymbol{\Psi})$, where $\boldsymbol{\Psi}$ is a $m \times m$ diagonal matrix with variances $\sigma_1^2, \ldots, \sigma_m^2$ as diagonal elements, and $\mathbf{u}_i$ is independent on $\boldsymbol{\epsilon}_i$, then model (3) implies that

$$\mathbf{y}_i \,|\, \mathbf{u}_i \sim N_m\left(\boldsymbol{\nu} + \boldsymbol{\Lambda}\mathbf{u}_i, \boldsymbol{\Psi}\right), \tag{4}$$

that is, responses are independent after conditioning on LVs. By integrating out the LVs, the marginal distribution of responses becomes

$$\mathbf{y}_i \sim N_m\left(\boldsymbol{\nu}, \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}\right), \tag{5}$$

that is, LVs induce a (reduced-rank) covariance structure for the model. As $\boldsymbol{\Psi}$ is assumed to be diagonal, information about correlations across responses is given by a matrix of factor loadings $\boldsymbol{\Lambda}$, and $\boldsymbol{\epsilon}_i$ only contribute to variances. When $\sigma_1^2 = \ldots = \sigma_m^2 =: \sigma^2$, that is, $\boldsymbol{\Psi} = \sigma^2\mathbf{I}_m$, the FA model (3) is known as the isotropic error model and it is equivalent to the probabilistic PCA model (2). For connections between the two models and algorithms for estimating $\boldsymbol{\Lambda}$ and σ^2, see Anderson (1963); Tipping and Bishop (1999).

With special choices of p and $\boldsymbol{\Lambda}$, model (3) reduces to a linear mixed effects model (Pinheiro and Bates 1995; Searle, Casella, and McCulloch 2009). A simple variance component model is obtained by selecting $p = 1$ and $\lambda_1 = \ldots = \lambda_m =: \lambda$, then λ^2 corresponds to the variance of random intercept, that is, we assume that $u_i \sim N\left(0, \lambda^2\right)$. The presence of univariate random intercept induces a constant positive correlation between each pair of responses y_{ij} and y_{ik}, where $j \neq k$. However, in most applications such a model is too far too simple. If we select $p = m$, and allow m-dimensional random effects to be distributed as $\mathbf{u}_i \sim N_m(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is an unstructured $m \times m$ covariance matrix, this corresponds to the choice $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$, that is, we can for example formulate the model using $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{1/2}$, where $\boldsymbol{\Sigma}^{1/2}$ is the unique symmetric root of $\boldsymbol{\Sigma}$. Models that include m-dimensional random effects offer a flexible framework for accounting for any correlation structure in data. With large m, the model fitting becomes however computationally demanding as the number of parameters in the covariance matrix increases quadratically with m. Hence, as seen in (5), by letting $p \ll m$, the factor loadings in $m \times p$ matrix $\boldsymbol{\Lambda}$ describe the correlation across responses, but use a lot less parameters than the mixed effects model with general correlation structure. We return to the methods for choosing the number of factors in Section 2.1.2.

Finally, notice that the FA model (3) can be extended to account for measured predictors. If we write $\mathbf{x}_i = \left(x_{i1}, \ldots, x_{iq}\right)^\top$, for a q-vector of observed predictors for the observational unit i, then the model (3) can be written for example as

$$\mathbf{y}_i = \boldsymbol{\nu} + \mathbf{B}\mathbf{x}_i + \boldsymbol{\Lambda}\mathbf{u}_i + \boldsymbol{\epsilon}_i, \tag{6}$$

where now $m \times q$ matrix $\mathbf{B} = \left[\boldsymbol{\beta}_1 \cdots \boldsymbol{\beta}_m\right]^\top$ with $\boldsymbol{\beta}_j = \left(\beta_{j1}, \ldots, \beta_{jq}\right)^\top$ is a matrix of regression coefficients. Now $\boldsymbol{\Lambda}\mathbf{u}_i$ accounts for correlation in responses not accounted for by the observed predictors.

Such correlation can be driven by some missing predictors or other unmeasured characteristics of the observational units.

2.1.2 | Rotation of Factors and Choosing Their Number

The FA model (3) is not uniquely defined; if we write $\mathbf{\Lambda}^\star = \mathbf{L}\mathbf{\Lambda}$ and $\mathbf{u}_i^\star = \mathbf{L}^\top \mathbf{u}_i$, where $\mathbf{L}$ is an orthogonal $p \times p$ matrix (e.g., a rotation), then the model (3) is equivalent to

$$\mathbf{y}_i = \boldsymbol{\nu} + \mathbf{\Lambda}^\star \mathbf{u}_i^\star + \boldsymbol{\epsilon}_i,$$

that is, the model is defined only up to a rotation. This indeterminacy is both a challenge and an opportunity. Assuming that the number of latent factors p is known, this indeterminacy must be taken into account when considering the identifiability of the model and its parameter estimation. Identifiability is considered in great detail in Anderson and Rubin (1956), and many estimation approaches proposed in the literature often fix the form of $\mathbf{\Lambda}$ in some way; see, for example, Mulaik (2009), Govindasamy et al. (2024), and references therein. The most popular estimation method by far is the MLE, which assumes that both LVs and noise are normally distributed. The MLE can essentially be based solely on the covariance matrix (or correlation matrix) but has no closed-form solution, and iterative approaches such as the EM algorithm, which will be discussed in detail in Section 3, have to be used. Once estimates have been obtained, however, the matrix $\mathbf{L}$ can be chosen in such a way that it best reflects the purpose of the analysis.

If the aim is to give the loading matrix $\mathbf{\Lambda}$ a simple structure and the factors $\mathbf{u}_i$ meaningful interpretations, then one might choose $\mathbf{L}$ to be a Varimax rotation that pushes the loadings toward zero or one (Kaiser 1958). However, there are many other objective functions to choose $\mathbf{L}$ with quite similar goals; see, for example, Mulaik (2009, chap. 10 and 11). Note that in some cases $\mathbf{L}$ is not restricted to being an orthogonal matrix to achieve better interpretability. In such case the rotation is denoted as an oblique rotation where again many options exist (see, e.g., Mulaik 2009, chap. 12). However, if the aim of the LV modeling is to use it as a data exploration method, one may, for example, be mainly interested in visualizing the data in a lower-dimensional space so that the axes are defined by factors (ordination axes). In such an exploratory analysis (i.e., ordination analysis), rotation does not matter as only the relative position of points in the ordination plot is relevant. In such cases, the rotation is fixed in some convenient way, for example, by constraining the upper triangular components of $\mathbf{\Lambda}$ to be zero and the diagonal elements to be positive (Huber, Ronchetti, and Victoria-Feser 2004; Niku et al. 2017). Notice that when FA is used to explore the correlation structure of the data, the method is called as the exploratory FA. The counterpart of the method is the confirmatory FA in which case the goal is to decide if a given dataset follows some hypothetical factor model (see Mulaik 2009, for more details).

Estimation in exploratory FA is done assuming that the number of LVs is known, which is often not the case in practical scenarios. Hence, extensive research has been conducted on the optimal choice for p, as reviewed, for example, in Iantovics, Rotar, and Morar (2019). Heuristic approaches for choosing p are commonly based on the eigenvalues of the sample correlation matrix, and are still widely used. Kaiser's rule (Kaiser 1960, 1961) suggests selecting p as the number of eigenvalues greater than 1, as this indicates that each factor explains more of the variation than a single variable. Cattell (1966) recommends plotting the eigenvalues against their ordinal number and visually identifying an "elbow"—the point where the plot bends significantly, which then suggests the number of components to be used. This method is known as the scree plot method.

These heuristic approaches, while not depending on strong assumptions or a specific estimation method, are considered quite crude. Nonetheless, the popularity of using the MLE for FA also stems from its immediate provision of tools for model selection, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), and for inference, such as likelihood ratio tests that compare the unrestricted covariance matrix with the model-based one (Bartlett 1950). More details on FA can be found in Anderson (2003), Mulaik (2009), Bartholomew, Knott, and Moustaki (2011), and discussions on how to perform FA with various packages in R are available in Govindasamy et al. (2024).

2.2 | Generalized Linear Latent Variable Models

Quite often the multivariate data are discrete and extensions of classical FA are needed. One example of such data is questionnaire data in psychology and sociology, where responses to test items are often binary, ordinal or categorical (or of mixed type). The aim of the analysis may then be to describe the data using a set of LVs and interpret the variables as psychological or sociological traits such as intelligence (Andersen 1982; Moustaki and Knott 2000; Skrondal and Rabe-Hesketh 2004). In educational testing data consist of students' answers to a set of questions and a single LV corresponding to students' performances is used to locate students on some chosen scale. Such a method is known as the item response theory (IRT, Andersen 1973; Andersen and Madsen 1977). In community ecology, GLLVMs are nowadays widely used to build joint models for abundance data. Such data occur when observations (e.g., presence–absences, counts, biomass, coverage, etc.) of multiple interacting species are recorded from a set of sites or samples. When GLLVM is used as an explanatory tool, predicted LVs (often with $p = 2$) can be plotted to illustrate how different samples or sites differ in terms of their species composition. Such a method is known as the model-based ordination analysis (Hui et al. 2015; van der Veen et al. 2021). If covariates related to study sites and/or traits are recorded, GLLVMs can be used to make valid inferences about their effects while accounting for any residual correlation between taxa not accounted for by the measured covariates. For more applications of GLLVMs on ecology, see Warton et al. (2015), Ovaskainen and Abrego (2020), and references therein.

2.2.1 | Model Formulation

The principles of FA can be easily extended to discrete or semi-continuous data setting using the GLLVM framework. A general treatment of GLLVMs for any response types stems from

Bartholomew (1980, 1984) who proposed a model for binary data. Later, Moustaki (1996) and Moustaki and Knott (2000) allowed the response variables to be of mixed type. For a general framework for GLLVMs, see for example, Skrondal and Rabe-Hesketh (2004) and Bartholomew, Knott, and Moustaki (2011). In GLLVMs we assume that responses $\mathbf{y}_i$ come from some distribution with known mean–variance relationship. Often it is assumed that the distribution belongs to the exponential family of distributions, but there is no reason to restrict to exponential family. The p-dimensional LVs are again assumed to follow a standard multivariate normal distribution, $\mathbf{u}_i \sim N_p(\mathbf{0}, \mathbf{I}_p)$, and the conditional mean of $\mathbf{y}_i$ given LVs $\mathbf{u}_i$ is linked to the linear predictor via some known link function. To be more specific, write the linear predictor $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{im})^\top$ as $\boldsymbol{\eta}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\mathbf{u}_i$, where $\boldsymbol{\nu}$ is again a m-vector specifying the location and $m \times p$ matrix $\boldsymbol{\Lambda}$ is a matrix of factor loading as in (3). In GLLVMs, the normality assumption in (4) can be replaced by

$$y_{ij} \mid \mathbf{u}_i \sim F(\mu_{ij}, \boldsymbol{\phi})$$

where $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)^\top$ denotes possible additional response-specific parameters for modeling dispersion, depending on the distributional families of the responses; present in, for example, negative binomial models for modeling overdispersion in count data (Niku et al. 2017), or in beta models for modeling vegetation cover data (Korhonen et al. 2024). The conditional mean $\mu_{ij} = \mathbb{E}(y_{ij} \mid \mathbf{u}_i)$ is then linked to the linear predictor via

$$g(\mu_{ij}) = \eta_{ij} = \nu_j + \boldsymbol{\lambda}_j^\top \mathbf{u}_i, \tag{7}$$

where $g(\cdot)$ is a known link function. As in FA, the model is not uniquely defined. Assumption $\mathbf{u}_i \sim N_p(\mathbf{0}, \mathbf{I}_p)$ fixes the mean and scale of LVs, but not the rotation as the vectors $\boldsymbol{\lambda}_j$ and $\mathbf{u}_i$ can be multiplied by an orthogonal matrix without changing η_{ij}. To fix the rotation, one can force the upper triangular components of $\boldsymbol{\Lambda}$ to be zero and the diagonal elements to be positive as in Huber, Ronchetti, and Victoria-Feser (2004).

As in (6), the model (7) can be extended to account for measured predictors. Then the model can be specified, for example, by

$$g(\mu_{ij}) = \nu_j + \boldsymbol{\beta}_j^\top \mathbf{x}_i + \boldsymbol{\lambda}_j^\top \mathbf{u}_i. \tag{8}$$

The above model is widely used, for example, in ecology community studies, where the aim may be in conducting multivariate inference about the effect of environmental covariates on community composition while accounting for correlation between species via LVs. Note that (8) now bears a lot of similarities to a multivariate GLMM (Breslow and Clayton 1993), which models the mean μ_{ij} by, for example,

$$g(\mu_{ij}) = \nu_j + \boldsymbol{\beta}_j^\top \mathbf{x}_i + \gamma_{ij}, \tag{9}$$

where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{im})^\top$ are elements of a multivariate random effect, which is typically assumed to be distributed according to $N_m(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is used to model the covariance structure among the responses. As already discussed in Section 2.1, although the mixed modeling approach is flexible, if $\boldsymbol{\Sigma}$ is left unstructured, the number of parameters needed to be estimated in $\boldsymbol{\Sigma}$ grow quadratically with the number of responses m, making GLMMs impractical for modeling large multivariate datasets. The GLLVM (7) can then be seen as a rank-reduced version of (9), lowering the number of parameters needed through the factor-analytic decomposition that is obtained by $\gamma_{ij} = \boldsymbol{\lambda}_j^\top \mathbf{u}_i$.

The model as defined in (8) can be extended further to more complex settings. For models specific to studies in ecology and various examples of case studies, we refer to Warton et al. (2015), Niku, Hui, et al. (2019), and Ovaskainen and Abrego (2020).

3 | Estimation and Computational Tools

In this section we review some most common computational tools for fitting GLLVMs. For simplicity, we consider here models without measured predictors, but their inclusion would not change the parameter estimation. Let $f_j(y_{ij} \mid \mu_{ij}, \phi_j)$ denote the pdf. (for continuous) or pmf. (for discrete) of the ith observation of the jth response y_{ij}, conditional on the LVs and the parameters of the model as in (7). Note the subscript j in $f_j(\cdot)$, indicating that the conditional distributions are allowed to be *mixed*, that is, they can vary among the responses $j = 1, \dots, m$. To simplify notation, from now on we assume one common distributional family $f_j(\cdot) = f(\cdot)$ for all j.

Because of the relatedness of GLMMs and GLLVMs as shown in the previous section, likelihood-based estimation methods for the latter mirror those developed earlier for the former. For this reason, we will be referring to GLMMs several times during the rest of this section. Note however that no specific additional knowledge on GLMMs is required on part of the reader.

3.1 | Complete Likelihood Function

Denote by $\boldsymbol{\theta} = (\boldsymbol{\nu}^\top, \mathrm{vec}(\boldsymbol{\Lambda})^\top, \boldsymbol{\phi}^\top)^\top$ and $\mathbf{u} = (\mathbf{u}_1^\top, \dots, \mathbf{u}_n^\top)^\top$ the vectors collecting all of the parameters and the LVs in the model, respectively. Under the assumption, that conditional on the LVs $\mathbf{u}_i$ the responses are distributed independently, the complete likelihood function takes the form

$$\mathcal{L}(\boldsymbol{\theta}, \mathbf{u}) = \prod_{i=1}^{n} \left[\prod_{j=1}^{m} f(y_{ij}, \mu_{ij} \mid \phi_j) \right] f(\mathbf{u}_i) = f(\mathbf{y}, \boldsymbol{\theta} \mid \mathbf{u}) f(\mathbf{u}),$$

where $f(\mathbf{u}_i)$ are densities of the standard multivariate normal distribution $N_p(0, \mathbf{I}_p)$, $f(\mathbf{y} \mid \theta, \mathbf{u}) = \prod_{i=1}^{n} \prod_{j=1}^{m} f(y_{ij} \mid \mu_{ij}, \phi_j)$, and $f(\mathbf{u}) = \prod_{i=1}^{n} f(\mathbf{u}_i)$. As follows, the complete log-likelihood function is then given by

$$\begin{aligned} \ell(\boldsymbol{\theta}, \mathbf{u}) &= \log \mathcal{L}(\boldsymbol{\theta}, \mathbf{u}) = \log f(\mathbf{y}, \boldsymbol{\theta} \mid \mathbf{u}) + \log f(\mathbf{u}) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} \log f(y_{ij}, \mu_{ij} \mid \phi_j) + \sum_{i=1}^{n} \log f(\mathbf{u}_i). \end{aligned} \tag{10}$$

3.1.1 | Expectation–Maximization Algorithm

As the complete log-likelihood (10) depends on the unobserved quantities in $\mathbf{u}$, we cannot maximize $\ell(\boldsymbol{\theta}, \mathbf{u})$ as is using the standard routines of maximum likelihood estimation (MLE).

Instead, it is popular to maximize (10) using algorithms such as the *expectation–maximization* (EM algorithm, Dempster, Laird, and Rubin 1977)—a general estimation tool for models with missing observations and/or LVs. The EM algorithm works by iterating through the so-called *E* (expectation) and *M* (maximization) steps:

1. *E* step: Let $\boldsymbol{\theta}^{(k)}$ stand for the current estimate of the parameters $\boldsymbol{\theta}$ after k iterations. Determine then the expected value of the complete log-likelihood (10) w.r.t. the LV conditional density $f\left(\mathbf{u}|\,\mathbf{y},\boldsymbol{\theta}^{(k)}\right)$:

$$Q\left(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(k)}\right) := \mathbb{E}_{\mathbf{u}\sim f\left(\cdot|\mathbf{y},\boldsymbol{\theta}^{(k)}\right)}\left[\ell(\boldsymbol{\theta},\mathbf{u})\right] = \int_{\mathbb{R}^{np}} f\left(\mathbf{u}|\,\mathbf{y},\boldsymbol{\theta}^{(k)}\right)\left[\log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u}) + \log f(\mathbf{u})\right] d\mathbf{u}.$$

2. *M* step: Update the parameter estimates by maximizing $Q\left(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(k)}\right)$ w.r.t. $\boldsymbol{\theta}$:

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, Q\left(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(k)}\right).$$

These steps are repeated in sequence until convergence, say the Kth iteration. Then, the EM estimates for the parameters of the model (7) can be read from $\boldsymbol{\theta}^{(K)}$, and the conditional distribution $f\left(\mathbf{u}|\,\mathbf{y},\boldsymbol{\theta}^{(K)}\right)$ (or posterior in Bayesian terms) may be used for predicting the LV scores. The EM algorithm is often easy to implement, and $Q\left(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(k)}\right)$ may admit closed-form expression even when the model's marginal log-likelihood—more on which below—does not. Furthermore, based on Jensen's inequality, the EM algorithm is guaranteed to improve the marginal log-likelihood at each iteration (see, e.g., the convergence discussions in the seminal work Dempster, Laird, and Rubin 1977).

In the modeling frameworks related to this review, variants of the EM algorithm have been employed in a plethora of settings, for example, in Sammel, Ryan, and Legler (1997) EM was used for LV modeling of mixed discrete and continuous responses, and in Hui et al. (2015) coupled with Monte Carlo integration for model-based ordination using GLLVMs. In Daolin Pang and Wang (2023), a variational EM algorithm was employed for the analysis of microbiome data using a multinomial response GLLVM. Wang and Carvalho (2024) proposed a general algorithm for the estimation of GLLVMs based on the EM algorithm and numerical integration.

The R package ltm (Rizopoulos 2006), aimed for applications in IRT, can be used to fit logistic GLLVMs on multivariate Bernoulli (i.e., binary) responses, with the LV dimension $p \leq 2$. It utilizes the EM algorithm together with the Gauss–Hermite quadrature (GHQ, Bock and Lieberman 1970; Bock and Aitkin 1981) to fit estimate the model.

3.2 | Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC, e.g., Metropolis et al. 1953; Hastings 1970) sampling based on the complete likelihood (10) presents another popular approach for estimation of models with unobserved parameters. In general, MCMC algorithms aim to draw samples from a given target distribution (typically the marginal posterior densities of the model parameters and LVs). Specifically, the samples are drawn by constructing a Markov chain, that is, sequence where the probability of a given member depends only on the one preceding it. The Markov chain is constructed in a way which ensures that after a sufficient amount of draws, the distribution of the chain closely approximates the target. For GLLVM estimation specifically, MCMC has been utilized, for example, in Ovaskainen, Abrego, et al. (2016) and Hui et al. (2017). Big advantage of the MCMC sampling based approach is that, compared with alternatives, it caters naturally to estimation of very complex, hierarchical models; Tikhonov et al. (2020) fitted GLLVMs with spatially correlated LVs, replacing the assumption that $\mathbf{u}_i \sim N_p\left(\mathbf{0},\mathbf{I}_p\right)$ by a more complex correlation structure. Lammel et al. (2018) utilized MCMC to estimate a GLLVM augmented with a phylogenetic correlation matrix. The latter two studies used the popular R package Hmsc (Tikhonov et al. 2019, 2024) to conduct the analyses. Motivated by ecological applications, Hmsc provides a user-friendly interface for Bayesian analysis of multivariate community data. As an alternative, one might consider boral (Hui 2016, 2024), another ecologically focused MCMC-based R package, with a larger number of supported response distributions while being more limited when it comes to including phylogenetic information or more complicated hierarchical structure. HMSC also employs a clever way of finding the optimal number for the LV dimension p; by allowing p to be essentially unlimited, HMSC controls the effective number of LVs by employing the multiplicative gamma process shrinking prior of Bhattacharya and Dunson (2011) for the loadings λ_j, leading to sparse loading matrices $\boldsymbol{\Lambda}$ regardless of p.

As an alternative to software specific to GLLVMs—such as the aforementioned HMSC or boral—one can choose to build their own modeling tools using some general purpose MCMC modeling software, including, for example, Stan (for the R interface, see Stan Development Team 2024), greta (Golding 2019), JAGS (Plummer 2003), and many more. This approach can offer great flexibility in the kinds of models one can fit, but may in turn require considerably more effort on part of the user.

3.3 | Marginal Likelihood

In the case of estimating GLLVMs based on (10), the two aforementioned methods—the EM algorithm and MCMC samplers—are able to provide accurate inferences and predictions, but tend to lead to very long computation times (Warton et al. 2015), rendering them impractical in many scenarios involving high or even moderate-dimensional data. Furthermore, for those not familiar with Bayesian statistics, MCMC convergence might be hard to assess (Gelman and Rubin 1996), particularly when the amount of parameters in the model is high—as is typical in scenarios where GLLVMs are employed. In this review, we are mainly focusing on the alternative and generally much more computationally efficient approach of approximating the marginal log-likelihood function

$$\ell(\boldsymbol{\theta}) = \log \mathcal{L}(\boldsymbol{\theta}) = \log\left(\int_{\mathbb{R}^{np}} f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u}) f(\mathbf{u})\, d\mathbf{u}\right). \tag{11}$$

where the LVs $\mathbf{u}_i$ have been integrated out, leading into a proper likelihood function for just the parameters $\boldsymbol{\theta}$. To facilitate efficient MLE requires the integral in (11) to be available as closed-form expression, a condition fulfilled only by the normal-response identity-link (i.e., factor analytic) model. For other response types, some approximation technique is needed; popular "classical" approaches include those that approximate the integrand, for example, Laplace's approximation (LA) (Tierney and Kadane 1986) and penalized quasi-likelihood (Green 1987; Breslow and Clayton 1993), and those that approximate the integration itself, for example, numerical integration using Monte Carlo methods (Danielsson 1994; Cappellari and Jenkins 2003) or (adaptive) Gaussian quadrature rules such as the Gauss–Hermite (Bock and Lieberman 1970; Bock and Aitkin 1981; Naylor and Smith 1982). More recently, the method of variational approximations has also gained significant amount of traction in likelihood based statistics (Ormerod and Wand 2010, 2012). On the Bayesian side, the method of integrated nested Laplace approximations (Rue, Martino, and Chopin 2009) has seen a lot of adaptation in general latent variable modeling. While these approximative methods are usually a lot faster to run than EM algorithm or MCMC, there comes a trade-off in terms of estimation accuracy. However, often the loss in accuracy is relatively small compared with the gain in computational efficiency, making these approximation methods attractive from a practical point of view (Warton et al. 2015).

3.3.1 | Approximations to the Integrand

The methods belonging to this class seek to bypass the integration in (11) by approximating the integrand, that is, the function $f(\mathbf{y}|\boldsymbol{\theta},\mathbf{u})f(\mathbf{u}) = f(\mathbf{y},\mathbf{u}|\boldsymbol{\theta})$, by a Gaussian density to leverage the fact that the integral then has a closed form. Popular methods of this class include the laplace's approximation (LA) (Tierney and Kadane 1986) and the penalized quasi-likelihood estimation (Green 1987; Breslow and Clayton 1993). Common characteristic of these kind of methods is that their performance generally improves with larger number of responses m. Conversely, for small m or very "discretely behaving" responses—for example, binary 0 / 1 or low-count Poisson or negative-binomial—the accuracy of the Gaussian approximation might be very poor.

3.3.1.1 | Laplace's Approximation. The *Laplace's method* or *Laplace's approximation* (LA, Tierney and Kadane 1986) is one of the best-known and most widely employed tools of approximate inference in statistics, including both Bayesian and frequentist methodologies. LA bypasses the need for high-dimensional integration in (11) by instead considering an optimization problem, namely that of finding the mode of the joint density w.r.t. $\mathbf{u}$, that is, $\hat{\mathbf{u}} = \text{argmax}_{\mathbf{u}} \log(\mathbf{y},\mathbf{u}\,|\,\boldsymbol{\theta})$—in Bayesian terminology, this corresponds to the *maximum* a posteriori—or MAP—estimate of LVs in $\mathbf{u}$. The joint density is then approximated by a multivariate normal distribution (un-normalized), $\log f(\mathbf{y},\mathbf{u}\,|\,\boldsymbol{\theta}) \approx N\left(\mathbf{u}|\,\hat{\mathbf{u}},\hat{\mathbf{S}}\right)$, with the mean $\hat{\mathbf{u}}$ and the precision (i.e., inverse covariance) matrix

$$\hat{\mathbf{S}}^{-1} = -\nabla^2_{\mathbf{u}}\left[\log f(\mathbf{y},\mathbf{u}|\boldsymbol{\theta})\right]_{\mathbf{u}=\hat{\mathbf{u}}}.$$

As mentioned earlier, (11) has a well-known closed-form solution for Gaussian integrands, meaning that LA is able to deliver closed-form approximations of (11) for GLLVMs with any type of response distribution. Asymptotic properties of LA are also well established; Tierney and Kadane (1986) have shown the asymptotic error of the approximation to be of order $\mathcal{O}\left(m^{-1}\right)$, and for GLLVMs specifically, Huber, Ronchetti, and Victoria-Feser (2004) have shown the estimates based on LA to be asymptotically normal and consistent.

Popular software packages that implement the Laplace's method for fitting GLLVMs or closely related models include glmmTMB (Brooks et al. 2017); gllvm (Niku et al. 2023) by setting method = "LA" in gllvm(); and lme4 (Bates et al. 2015) by setting nAGQ = 1 in the function glmer(). Of the three packages, gllvm focuses on GLLVMs, while glmmTMB and lme4 are oriented toward GLMMs, with glmmTMB also having the option of fitting GLMMs with rank-reduced correlation structures, that is, essentially GLLVMs. Notably, in their implementations, both glmmTMB and gllvm leverage the R/C++ TMB library (Kristensen et al. 2016); a state-of-the-art model-building software combining fast LA with automatic/algorithmic differentiation (AD, e.g., Rall 1981) for efficient and relatively easy implementation of many types of models with random effects.

The basic TMB workflow consists of writing and compiling a C++ function, the model *template* for the objective, that is, usually the negative log-likelihood. On the R side, the user then specifies the data and parameter starting values and denotes the random effects using MakeADFun() on the compiled template, producing the objective function together with its gradient, which is calculated using AD. These can then be fed into any gradient-based optimization function in R, for example, optim() or nlminb(), to obtain estimates for the parameters $\boldsymbol{\theta}$ and predictions for $\mathbf{u}$. Through the R_inla namespace TMB can be combined with the stochastic partial differential equation approach from the INLA package (Rue, Martino, and Chopin 2009; Lindgren, Rue, and Lindström 2011; Martins et al. 2013) to create models with very complex spatial or spatio-temporal correlation structures. By using the package tmbstan in addition, users can do MCMC sampling based on a TMB model object using the No-U-Turn Hamiltonian Monte Carlo sampler (Hoffman and Gelman 2014) implemented in the Bayesian modeling software platform Stan (e.g., Stan Development Team 2024). Prospective users familiar with R but not with C++, can instead opt for using the RTMB package (Kristensen 2023), a R-only interface to TMB, with most of the same functionalities in place.

3.3.1.2 | Penalized Quasi-Likelihood. *Quasi-maximum likelihood* (QML) estimation refers to a broad category of ways to relaxing some of the distributional assumptions inherent in popular statistical methodologies, particularly generalized linear models—the context in which it was originally introduced in Wedderburn (1974). Unlike the typical likelihood function, which is always related to a given distribution assumed for the response variable y, the quasi-likelihood function is formed by specifying only the assumed mean–variance relationship

$$\mathbb{E}(y) = \mu,$$
$$\text{var}(y) = \phi\mathbb{V}(\mu),$$

where ϕ is an (unknown) dispersion parameter and $\mathbb{V}(\cdot)$ is some known variance function depending on the mean μ—which itself is connected to some linear predictor through a known link function $g(\mu) = \eta$. The quasi-likelihood is then given by integrating the so-called quasi-score function defined as

$$ql(\mu, \phi) = \int_y^{\mu} \frac{y - t}{\phi \mathbb{V}(t)}\, dt, \tag{12}$$

and it behaves similarly to ordinary log-likelihood, and actually coincides with it when y is in the exponential family (Wedderburn 1974). QML estimates for the regression parameters related to the mean μ are then achieved by maximizing (12) w.r.t. them; interestingly, ϕ has no effect in the estimation of the mean parameters. Under certain fairly broad conditions, the estimates based on (12) enjoy similar properties as ML estimates, for example, asymptotic normality and consistency, but suffer in terms of efficiency (Wedderburn 1974; McCullagh 1983; Firth 1987).

Following Green (1987), the method of *penalized quasi-likelihood* (PQL) has been developed as a fast estimation method for GLMMs in Breslow and Clayton (1993), and more recently extended for GLLVMs in Kidzinski et al. (2022), particularly in the context of very high-dimensional datasets. Their line of work derives the PQL for LV models by considering first a LA to a quasi-likelihood formulation of GLMM/GLLVM, which is followed by then discarding an expensive to compute log-determinant term from the resulting approximation. Justification for this exclusion is, that based on theory (GLMMs, Breslow and Clayton 1993) and empirical evidence (GLLVMs, Kidzinski et al. 2022), the magnitude of the particular log-determinant diminishes compared with other terms, when n and m grows large.

Assuming that y_{ij} is in the exponential family with canonical link, that is, $y_{ij} \sim \exp\big(\big(y_{ij}\eta_{ij} - b\big(\eta_{ij}\big)\big)/\phi_j\big)h\big(y_{ij}, \phi\big)$ with known $b(\cdot)$ and $h(\cdot)$, following Kidzinski et al. (2022), the likelihood contribution of the ith observation is in the first step (Laplace) approximated by

$$\ell_i(\boldsymbol{\theta}) \approx -\frac{1}{2}\log\left|\frac{\partial^2 L_i\big(\mathbf{u}_i\big)}{\partial \mathbf{u}_i \partial \mathbf{u}_i^{\top}}\right|_{\mathbf{u}_i = \hat{\mathbf{u}}_i} - L_i\big(\hat{\mathbf{u}}_i\big),$$

where $L_i(\mathbf{u}) = \sum_{j=1}^{m}\big(y_{ij}\eta_{ij} - b\big(\eta_{ij}\big)\big)/\phi_j - \mathbf{u}^{\top}\mathbf{u}/2$, and $\hat{\mathbf{u}}_i$ is such that $L_i(\cdot)$ is minimal. After calculating the necessary derivatives (see eqs. 4 and 5 in Kidzinski et al. 2022) the approximation for the ith likelihood contribution becomes

$$\ell_i(\boldsymbol{\theta}) \approx -\frac{1}{2}\log\left|\boldsymbol{\Lambda}^{\top}\mathbf{W}_i\boldsymbol{\Lambda} + \mathbf{I}_p\right| + \sum_{j=1}^{m}\frac{1}{\phi_j}\big(y_{ij}\hat{\eta}_{ij} - b\big(\hat{\eta}_{ij}\big)\big) - \frac{1}{2}\hat{\mathbf{u}}_i^{\top}\hat{\mathbf{u}}_i, \tag{13}$$

where $\mathbf{W}_i$ is a $m \times m$ diagonal matrix with diagonal elements $\mathbb{V}\big(\mu_{i1}\big)/\phi_1, \ldots, \mathbb{V}\big(\mu_{im}\big)/\phi_m$ and $\hat{\eta}_{ij} = \nu_i + \lambda_j^{\top}\hat{\mathbf{u}}_i$. In the context of GLMMs, where the loading matrix $\boldsymbol{\Lambda}$ would be replaced by a known random-effect design matrix, Breslow and Clayton (1993) argue that as the matrix $\mathbf{W}_i$ tends to vary slowly w.r.t. $\boldsymbol{\theta}$, the log-determinant term could be ignored. In GLLVMs, where $\boldsymbol{\Lambda}$ need to be estimated, one cannot categorically discard the first term in (13). However, backed by empirical evidence, Kidzinski et al. (2022) argue that in high-dimensional settings with large numbers of observations n and responses m, the log-determinant terms become negligible lending to the following PQL approximation of the marginal log-likelihood

$$\ell(\boldsymbol{\theta}) \approx -\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{\phi_j}\big(y_{ij}\eta_{ij} - b\big(\eta_{ij}\big)\big) + \frac{1}{2}\sum_{i=1}^{n}\mathbf{u}_i^{\top}\mathbf{u}_i, \tag{14}$$

The term $\frac{1}{2}\sum_{i=1}^{n}\mathbf{u}_i^{\top}\mathbf{u}_i$ in (14) resembles a ridge-like penalty term (e.g., McDonald 2009)—hence the name *penalized* quasi-likelihood. By essentially treating $\mathbf{u}_i$ as fixed effects instead of random, the PQL estimation can be done very efficiently leveraging well-established Newton and quasi-Newton methods, as shown by the two algorithms provided in Kidzinski et al. (2022). Both algorithms are implemented in the R package gmf (Kidzinski et al. 2020). The package also implements an option of using a regularization scheme to control the effective number of LVs p. This procedure, as described in Kidzinski et al. (2022), is done by setting a bound on (e.g., $p < \sqrt{m}$), and then including in (14) L_2 penalties on the LV score and loading matrices with some scaling factor γ. During the model estimation, these penalty terms then push some of the columns toward zero, thus limiting the effective number of LVs.

3.3.2 | Approximations to the Integral

The other approach to approximating (11) is to consider approximations to the operation of integration itself. This includes, for example, numerical integration techniques such as the GHQ (Bock and Lieberman 1970; Bock and Aitkin 1981) and simulated maximum likelihood (SML) (Danielsson 1994; Cappellari and Jenkins 2003).

3.3.2.1 | Gauss–Hermite Quadrature. Finding its way into the statistical literature through uses in item response theory (Bock and Lieberman 1970; Bock and Aitkin 1981), the GHQ is a popular numerical integration technique for approximating univariate integrals of the following form

$$\int_{-\infty}^{\infty}\exp\big(-u^2\big)h(u)\,du \approx \sum_{q=1}^{n_Q} w_q h\big(u_q\big), \tag{15}$$

by a weighted sum of $h(\cdot)$ evaluated on a set of n_Q quadrature points or *nodes* $u_q \in Q \subset \mathbb{R}$. The set of quadrature points Q corresponds to the roots of the physicist's Hermite polynomial of order n_Q. The roots and the corresponding weights have long been tabulated for moderate values of n_Q, see, for example, Stroud and Secrest (1966). Using R, for example, the function gaussHermiteData() from the package fastGHQuad (Blocker 2022) can be used to create tables for a desired number of nodes and weights $\big(u_q, w_q\big)$. Accuracy of the method increases by increasing the number of quadrature points n_Q, but so does the computational load. The univariate quadrature (15) can be extended to the multivariate case $\mathbf{u} \in \mathbb{R}^p$ by considering p nested Gauss–Hermite integrals together with the note that $\exp\big(-\mathbf{u}^{\top}\mathbf{u}\big) = \prod_{k=1}^{p}\exp\big(-u_k^2\big)$ factorizes, thus making GHQ suitable for marginal MLE of GLLVMs (7). However, these nested integrals get computationally cumbersome very

quickly as the LV dimension p grows, hampering the practicality of the method for direct marginal MLE of GLLVMs (11). Nowadays, GHQ is more often seen as an intermediary step in other algorithms; particularly with intractable integrals appearing within the "*E*-step" in the EM algorithm, as in Moustaki (1996), Moustaki and Knott (2000), and the R package ltm (Rizopoulos 2006), for fitting GLLVMs with mixed exponential family responses; or within variational approximation as in Korhonen et al. (2023).

3.3.2.2 | Adaptive Gauss–Hermite Quadrature. In practice, the usability of GHQ in the estimation of multivariate random effect models is hindered by the fact that the integrands $f(\mathbf{y}_i|\boldsymbol{\theta},\mathbf{u}_i) = \prod_{j=1}^{m} f(y_{ij}|\boldsymbol{\theta},\mathbf{u}_i)$ often exhibit high peakedness (Lesaffre and Spiessens 2001), causing the need for high number of quadrature points to accurately assess the integral. This issue is further compounded in scenarios with high number of responses m, and by the nested integration needed for multivariate random effects $p > 1$. Pruning can be employed to remove the points in the grid with the smallest weights w_q, but this is often small relief. As a better and more comprehensive solution, Naylor and Smith (1982) and Pinheiro and Bates (1995) considered an extension of GHQ (15) called the *adaptive Gauss–Hermite quadrature* (AGHQ). AGHQ differs from regular GH by using a technique akin to importance sampling to shift the nodes u_q and to transform the weights w_q to better target the high density regions of the integrand which are of interest. As a result, a much smaller n_Q suffices for accurate approximation of the integral, easing the computational burden. Still, the method is often considered impractical even for relatively low numbers of LVs, $p > 2$.

Since its introduction, the AGHQ has become one of the standard methods for integration in random effect models, particularly for GLMMs. For example, it is available as a fitting method used by glmer() in the mixed modeling R package lme4, though only for univariate random effects; for $p \geq 2$ the function switches to LA (Tierney and Kadane 1986)—which is often seen as a special case of AGHQ with just one quadrature point. AGHQ was first applied and implemented for GLLVMs by Rabe-Hesketh, Skrondal, and Pickles (2002) in the associated gllamm package for the proprietary statistical software STATA. Bianconcini (2014) showed that for GLLVM estimators based on AGHQ, the asymptotic normality, and consistency holds.

3.3.2.3 | Simulated Maximum Likelihood. SML (e.g., Danielsson 1994; Cappellari and Jenkins 2003) estimation is a Monte Carlo method for evaluating the marginal log-likelihood function (11). Study on SML estimation of GLLVM/FA models was first conducted in Wedel and Kamakura (2001) and recently revisited in Wang and Carvalho (2024). SML can be an attractive choice for estimation method particularly because of its simplicity; the Monte Carlo estimate of (11) is constructed by using the following procedure:

1. Store $k = 1, \dots, K$ draws of $\tilde{\mathbf{u}}^{(k)} = \left(\tilde{\mathbf{u}}_{k1}^{\top}, \dots, \tilde{\mathbf{u}}_{kn}^{\top}\right)^{\top}$ with each $\tilde{\mathbf{u}}_{ki} \sim N_p(\mathbf{0}, \mathbf{I}_p)$, as assumed in (7).
2. Choose starting values $\boldsymbol{\theta}_0$ for the model parameters $\boldsymbol{\theta}$.
3. Based on the values $\tilde{\mathbf{u}}$, evaluate the simulated log-likelihood function:

$$\tilde{\ell}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \prod_{j=1}^{m} f(y_{ij}|\boldsymbol{\theta},\tilde{\mathbf{u}}_{ki}).$$

4. For gradient-based maximization (e.g., Newton–Raphson) of $\tilde{\ell}(\boldsymbol{\theta})$, evaluate the simulated gradient:

$$\frac{\partial \tilde{\ell}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\sum_{k=1}^{K} \partial f(y_{ij}|\boldsymbol{\theta},\tilde{\mathbf{u}}_{ki})/\partial \boldsymbol{\theta}}{\sum_{k=1}^{K} f(y_{ij}|\boldsymbol{\theta},\tilde{\mathbf{u}}_{ki})}.$$

The gradients $\partial f(y_{ij}|\boldsymbol{\theta},\tilde{\mathbf{u}}_{ki})/\partial\boldsymbol{\theta}$ are typically easy to find, ensuing SML as a simple method to implement for most types of LV models. Based on strong law of large numbers, $\tilde{\ell}(\boldsymbol{\theta}) \to \ell(\boldsymbol{\theta})$ when $K \to \infty$, making $\tilde{\ell}(\boldsymbol{\theta})$ a consistent simulator of the marginal log-likelihood (11), resulting in the SML estimator $\tilde{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} \tilde{\ell}(\boldsymbol{\theta})$ being consistent and asymptotically equivalent to the ML estimator—but only if $K \to \infty$ when $n \to \infty$. However, the SML estimator has been shown in Lee (1995) to be asymptotically biased when the amount of Monte Carlo draws K does not grow at a rate of at least $\sqrt{n}$. Furthermore, Wang and Carvalho (2024) record that empirically, to ensure reasonably low variances for the Monte Carlo gradients $\partial f(y_{ij}|\boldsymbol{\theta},\tilde{\mathbf{u}}_{ki})/\partial\boldsymbol{\theta}$, a much larger amount of draws $K \gg \sqrt{n}$ might be needed, making SML impractical for modern big data settings. As a further practical hindrance, Wang and Carvalho (2024) also note that the denominator in the simulated gradient is easily prone to issues of numerical underflow and requires good choices for starting values $\boldsymbol{\theta}_0$, which can be challenging especially when the amount of responses m is large.

3.3.3 | Recent Developments

This section reviews some of the more recently developed approximations to the marginal log-likelihood (11), namely the methods of (mean-field) variational inference (Ormerod and Wand 2010, 2012), expectation–propagation (Minka 2001; Hall et al. 2020) and integrated nested Laplace approximations (Rue, Martino, and Chopin 2009).

3.3.3.1 | Variational Approximations. Already popular in Bayesian statistics and machine learning literature, the use of *variational inference* (VI) was introduced to frequentist setting by Ormerod and Wand (2010, 2012), for the purpose of fitting GLMMs. As a rank-reduced form of mixed effect models, GLLVMs fit into their approach naturally, as evidenced by Hui et al. (2017), Niku, Brooks, et al. (2019), and Daolin Pang and Wang (2023). Furthermore, for GLLVMs, as demonstrated by Warton et al. (2015), VI is able to exhibit a desirable balance between accuracy and speed, compared with popular alternatives such as (adaptive) numerical quadratures, MCMC, or LA reviewed above.

Instead of the marginal log-likelihood (11), VI seeks to maximize the so-called *evidence lower bound* (ELBO), derived using the Jensen's inequality

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= \log\left(\int_{\mathbb{R}^{np}} f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u})f(\mathbf{u})\,d\mathbf{u}\right) = \log\left(\int_{\mathbb{R}^{np}} f(\mathbf{y},\mathbf{u}|\boldsymbol{\theta})\frac{q(\mathbf{u}|\,\boldsymbol{\xi})}{q(\mathbf{u}|\,\boldsymbol{\xi})}\,d\mathbf{u}\right)\\ &= \log \mathbb{E}_{\mathbf{u}\sim q}\left[\frac{f(\mathbf{y},\mathbf{u}|\boldsymbol{\theta})}{q(\mathbf{u}|\,\boldsymbol{\xi})}\right] \geq \mathbb{E}_{\mathbf{u}\sim q}\left[\log f(\mathbf{y},\mathbf{u}|\boldsymbol{\theta}) - \log q(\mathbf{u}|\,\boldsymbol{\xi})\right]\\ &=: \mathrm{ELBO}(\boldsymbol{\theta},\boldsymbol{\xi}),\end{aligned} \tag{16}$$

(Needham 1993), where $q(\mathbf{u}|\,\boldsymbol{\xi})$ is referred to as the *variational distribution* for the LVs $\mathbf{u}$, with $\boldsymbol{\xi}$ denoting the *variational parameters*. Maximizing (16) involves minimizing the Kullback–Leibler divergence—or *distance*—(KL divergence, Kullback and Leibler 1951) from the variational distribution $q(\mathbf{u}|\,\boldsymbol{\xi})$ to the actual LV distribution $f(\mathbf{u})$, more evident from the fact that

$$\begin{aligned}\mathrm{ELBO}(\boldsymbol{\theta},\boldsymbol{\xi}) &= \mathbb{E}_{\mathbf{u}\sim q}\left[\log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u})\right] - \mathbb{E}_{\mathbf{u}\sim q}\left[\log\left(\frac{q(\mathbf{u}|\,\boldsymbol{\xi})}{f(\mathbf{u})}\right)\right]\\ &= \mathbb{E}_{\mathbf{u}\sim q}\left[\log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u})\right] - D_{\mathrm{KL}}(q(\mathbf{u})\,\|\,f(\mathbf{u})).\end{aligned} \tag{17}$$

Undeniably, the most popular assumption for q is the *mean-field approximation*, under which the variational distribution factorizes into independent parts as follows

$$q(\mathbf{u}|\,\boldsymbol{\xi}) = \prod_{i=1}^{n} q_i(\mathbf{u}_i|\,\boldsymbol{\xi}_i). \tag{18}$$

Mean-field VI is often preferred because of computational simplicity it provides, even though independence might be a too strong assumption for some applications. Furthermore, the individual q_i are typically chosen to be multivariate Gaussians, that is, $q_i(\mathbf{u}_i|\,\boldsymbol{\xi}_i) = N_p(\mathbf{u}_i|\,\mathbf{a}_i,\mathbf{A}_i)$. In particular, this is the approach employed when specifying method = "VA" when calling gllvm() in the R package gllvm (Niku, Brooks, et al. 2019; Niku et al. 2023). Mean-field approximation was also used in Chiquet, Mariadassou, and Robin (2018) for probabilistic Poisson PCA, and in Daolin Pang and Wang (2023) to fit a multinomial GLLVMs to microbiome data. As GLLVMs (7) and FA (3) by default already assume independent normal LVs, the mean-field assumptions might not be as big of a concern. Noteworthily, in the GLLVM setting, the number of variational mean and covariance parameters in $\boldsymbol{\xi}_i = \{\mathbf{a}_i,\mathbf{A}_i\}$ depends on the number of rows n and LVs p. Particularly, with higher LV dimension, if the matrices $\mathbf{A}_i$ are assumed to be unstructured (default in gllvm), then the optimization process can get hindered by the swiftly increasing additional parameter count. This can be alleviated to a degree by specifying a simpler structure for $\mathbf{A}_i$, for example,, diagonal—achieved through the control.va argument, when calling gllvm().

When both $q(\mathbf{u})$ and $f(\mathbf{u})$ are Gaussian densities, the term $D_{\mathrm{KL}}(q(\mathbf{u})\,\|\,f(\mathbf{u}))$ in (17) has a well-known closed-form expression. Meanwhile, the term $\mathbb{E}_{\mathbf{u}\sim q}\left[\log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u})\right]$ is not guaranteed to have closed-form for all response-link combinations; for example presence-absence responses require specifically the use of the probit link function, that is, $g(\cdot) = \Phi(\cdot)$ in (7), to arrive in closed expression. To use logit (or cloglog) link on binary responses, one needs to apply numerical integration (such as GHQ) or additional approximations on $\mathbb{E}_{\mathbf{u}\sim q}\left[\log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u})\right]$, possibly losing the property of being a lower bound to (11) in the process.

Readers interested in variational inference are recommended to refer to Blei, Kucukelbir, and McAuliffe (2017) for a general review on VI methodology intended specifically for statistics-oriented audience, and Zhang et al. (2019) for a review on recent advanced techniques of VI-driven mainly by machine learning and AI research, touching, for example, on the potential scalability issues mentioned above.

3.3.3.2 | Extended Variational Approximations. To circumvent the inherent restrictions in VA regarding the choices of response family and link function, Korhonen et al. (2023) introduced the method of *extended variational approximations* (EVA) in the GLLVM context. In EVA, $\log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u})$ in (17) is replaced by its second-order Taylor expansion around the variational mean $\mathbf{a}$, that is,

$$\begin{aligned}\log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u}) \approx{}& \log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{a}) + \nabla_{\mathbf{u}}\left[\log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u})\right]_{\mathbf{u}=\mathbf{a}}(\mathbf{u}-\mathbf{a})\\ &+ \frac{1}{2}(\mathbf{u}-\mathbf{a})^{\top}\nabla^2_{\mathbf{u}}\left[\log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u})\right]_{\mathbf{u}=\mathbf{a}}(\mathbf{u}-\mathbf{a}),\end{aligned}$$

Now, calculating the mean of the expression above w.r.t. the variational density $q(\cdot\,|\mathbf{a},\mathbf{A})$ results in a closed-form approximation for any combination of response distribution $f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u})$ and link function $g(\cdot)$. On the other hand, the resulting objective function is no longer guaranteed to be a lower bound for the marginal log-likelihood, as is the case with the ELBO (16). This means that improving the objective could in fact weaken the actual likelihood. Regardless, EVA has shown to be a competitive estimation method both in terms of scalability/computational speed and estimation accuracy when fitting GLLVMs (Korhonen et al. 2023). This is further validated by similar developments and findings in machine learning and variational Bayes literature; Wang and Blei (2013) proposed two estimation methods, *Laplace VI* and *delta method VI*, for nonconjugate models in the contexts of Bayesian logistic regression and correlated topic models. Similarly to EVA, delta method VI also uses the variational mean as a center of Taylor expansion, for updating q, while Laplace VI builds the Taylor expansion around the MAP estimate $\hat{\mathbf{u}}$ of $\log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u})$. Braun and McAuliffe (2010) proposed two *surrogate* ELBO functions for closed-form variational approximations for Bayesian discrete choice models, one—like EVA—based on multivariate delta method, and another based on the application of Jensen's inequality onto the complex log-sum-exp term present in $\mathbb{E}_{\mathbf{u}\sim q}\left[\log f(\mathbf{y}|\,\boldsymbol{\theta},\mathbf{u})\right]$. Even though the latter surrogate ELBO preserves the lower-bound property of the original, the authors noted the former to be empirically superior.

EVA—accessible through the argument method = "EVA" when calling gllvm()—has been implemented for select response distributions and link functions in the gllvm package (Niku et al. 2023).

3.3.3.3 | Expectation Propagation. Closely related to variational inference, *expectation propagation* (EP, Minka 2001) is an algorithm or framework for estimating models containing random effects. Rooted deeply within machine learning literature and adopted first and foremost for the estimation of Bayesian graphical models, EP was introduced to frequentist methodology by Hall et al. (2020), again as a way to fit GLMMs with multivariate random effects

efficiently. The method is implemented in the glmmEP package for binary response probit GLMM with one level of nesting—a model more specific, but with lots of similarities to GLLVM as formulated in (7).

The key difference between VI and EP is in the application of the KL divergence; in VI (16), the objective is in fact to minimize the "forward" KL divergence from $q(\mathbf{u})$ to $f(\mathbf{y},\mathbf{u})$, while EP seeks to minimize the "reverse" KL divergence $D_{\mathrm{KL}}(f(\mathbf{y},\mathbf{u}) \,\|\, q(\mathbf{u}))$. Because of the asymmetry of the KL divergence, there are notable differences in the resulting algorithms; in minimization of the forward KL-divergence *zero-forcing* is rewarded, that is, $q(\mathbf{u})$ needs to be close to zero whenever $f(\mathbf{u})$ is close to zero, while for reverse KL to be minimal zeros need to be *avoided*. Particularly, in situations where $f(\mathbf{u})$ is possibly multimodal, VI tends to concentrate around *a* mode, while EP seeks to cover the whole support. Unlike VI, EP does not possess the lower bound property and has no guarantees of convergence (generally). Nonetheless, it has still displayed comparable or even improved performance to those of MCMC or LA in some settings (Kuss, Rasmussen, and Herbrich 2005; Hall et al. 2020).

As a final note, both EP and VI fall under a more broader family of estimation algorithms known as *Power EP*, in which the KL-divergence is replaced by a more general class of divergence measures, namely the α-divergences (e.g., Minka 2004; Hernandez-Lobato et al. 2016). For a recent overview of the EP framework/philosophy—particularly from a distributed computing point of view—see Vehtari et al. (2020).

3.3.3.4 | Integrated Nested Laplace Approximations. Introduced in Rue, Martino, and Chopin (2009), the method of *integrated nested Laplace approximations*, or INLA, is an approximate estimation tool for general (Bayesian) latent Gaussian models (LGM), that is, models containing Gaussian random effects. As such, it has successfully been applied for estimation of GLMMs in, for example, Fong, Rue, and Wakefield (2010), and a R package implementing the INLA approach for fitting GLLVMs is in development, under the name LatentINLA (O'Hara and van der Veen 2024). The core INLA software library—that the LatentINLA also utilizes—is called R-INLA (Martins et al. 2013).

Compared with MCMC, INLA is regarded as a fast and accurate alternative for Bayesian inference in latent Gaussian models (e.g., Held, Schrödle, and Rue 2010; De Smedt et al. 2015). Many of the frequently used LGMs admit properties of conditional independence—that is, the inverse covariance matrix $\boldsymbol{\Sigma}^{-1}$ of the Gaussian random effects is a sparse matrix—or are reasonably well approximated by such a simplified model (see, e.g., Rue and Held 2005; Lindgren, Rue, and Lindström 2011). The computational efficiency of the INLA approach is largely driven by its ability to take advantage of such sparsity-inducing assumptions, as long as the amount of model hyperparameters also stays relatively small.

As a predominately Bayesian method, INLA is concerned with providing accurate approximations to marginal posteriors of the model parameters and the LVs using (roughly) the following steps:

1. Explore the parameter space $\boldsymbol{\theta}$ through Laplace approximation to the marginal posterior

$$f(\boldsymbol{\theta}|\,\mathbf{y}) \propto \frac{f(\mathbf{y}|\,\mathbf{u},\boldsymbol{\theta})f(\mathbf{u}|\,\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathbf{u}|\,\boldsymbol{\theta},\mathbf{y})} \approx \left.\frac{f(\mathbf{y}|\,\mathbf{u},\boldsymbol{\theta})f(\mathbf{u}|\,\boldsymbol{\theta})f(\boldsymbol{\theta})}{\widetilde{f}_G(\mathbf{u}|\,\boldsymbol{\theta},\mathbf{y})}\right|_{\mathbf{u}=\mathbf{u}^*},$$

   where $\widetilde{f}_G(\mathbf{u}|\,\boldsymbol{\theta},\mathbf{y})$ is the Gaussian approximation built by matching the mode $\mathbf{u}^*$ and the curvature at the mode. Find a set of K high-density points $\{\boldsymbol{\theta}^{(1)},\ldots,\boldsymbol{\theta}^{(K)}\}$, which can be formed by, for example, taking a grid of equidistant points or using a strategy inspired by central composite design (Box and Draper 2007).

2. For each $\boldsymbol{\theta}^{(k)}$, approximate the conditional marginal posteriors

$$f(\mathbf{u}_i|\,\boldsymbol{\theta}^{(k)},\mathbf{y}) \propto \frac{f(\mathbf{y}|\,\mathbf{u},\boldsymbol{\theta}^{(k)})f(\mathbf{u}|\,\boldsymbol{\theta}^{(k)})f(\boldsymbol{\theta}^{(k)})}{f(\mathbf{u}_{-i}|\,\mathbf{u}_i,\boldsymbol{\theta}^{(k)},\mathbf{y})}$$

   by $\widetilde{f}(\mathbf{u}_i|\,\boldsymbol{\theta}^{(k)},\mathbf{y})$, using one of three options:

   a. The marginal Gaussian approximation $\widetilde{f}_G(\mathbf{u}_{-i}|\,\mathbf{u}_i,\boldsymbol{\theta}^{(k)},\mathbf{y})$, computed fast from the full conditional Gaussian approximation $\widetilde{f}_G(\mathbf{u}|\,\boldsymbol{\theta},\mathbf{y})$ from step 1 using simple recursive formulas. This is the fastest option, but often inaccurate.

   b. The Laplace approximation

$$\widetilde{f}_{LA}(\mathbf{u}_i|\,\boldsymbol{\theta}^{(k)},\mathbf{y}) = \left.\frac{f(\mathbf{y}|\,\mathbf{u},\boldsymbol{\theta}^{(k)})f(\mathbf{u}|\,\boldsymbol{\theta}^{(k)})f(\boldsymbol{\theta}^{(k)})}{\widetilde{f}_G(\mathbf{u}_{-i}|\,\mathbf{u}_i,\boldsymbol{\theta}^{(k)},\mathbf{y})}\right|_{\mathbf{u}_{-i}=\mathbf{u}^*_{-i}}$$

      This is considered the most accurate option, but its performance in practice is hindered by the need to factorize large sparse matrices for each i separately.

   c. The simplified Laplace approximation $\widetilde{f}_{SLA}(\mathbf{u}_i|\,\boldsymbol{\theta}^{(k)},\mathbf{y})$, which uses a third-order Taylor expansion of $\widetilde{f}_{LA}(\mathbf{u}_i|\,\boldsymbol{\theta}^{(k)},\mathbf{y})$ around the mean $\mathbb{E}[\mathbf{u}_i]$ to essentially correct the Gaussian approximation $\widetilde{f}_G(\mathbf{u}_{-i}|\,\mathbf{u}_i,\boldsymbol{\theta}^{(k)},\mathbf{y})$ for both location and skewness. Offers a balance between accuracy and speed, and is indeed the default choice in the R-INLA package.

3. Using numerical integration, approximate the LV marginal posteriors as $f(\mathbf{u}_i|\,\mathbf{y}) \approx \sum_{k=1}^{K}\widetilde{f}(\mathbf{u}_i|\,\boldsymbol{\theta}^{(k)},\mathbf{y})\widetilde{f}(\boldsymbol{\theta}^{(k)}|\,\mathbf{y})w_k$, using the approximations calculated in steps 1 and 2. The integration weights w_k depend on how the points $\{\boldsymbol{\theta}^{(1)},\ldots,\boldsymbol{\theta}^{(K)}\}$ are picked, for example, for regular grids the weights can be set to be equal $w_1 = \cdots = w_K$.

   The approximate marginal posteriors for each component of $\boldsymbol{\theta}$ can be attained in a similar fashion.

Note that the above outline of the algorithm is very scarce on details. For an in-depth view of each of the different steps, we refer to Rue, Martino, and Chopin (2009). The "nested Laplace approximations" in the name of the method refer to the approximations taken in steps 1 and 2, while "integrated" comes from the numerical integration employed in the last

step (Table 2). We provide a summary of all of the software referenced in this review with some details in Table 2.

4 | Comparative Studies

To compare some of the computational approaches from Section 3 with freely available R software implementations in their ability to estimate GLLVMs, we used two simple simulation studies intending to also highlight two kinds of datasets typically encountered in ecological studies. Afterwards, a short example study based on a real dataset about abundances of Australian ant species from Gibb and Cunningham (2011) was conducted, involving two of the approaches included in the simulations. These comparisons are by no means intended as conclusive, as some of the methods are designed for widely differing use cases and feature slightly different parametrizations or constraints. Nevertheless, the intention of these studies is to be able to provide the reader some guidance on how to approach the choice computational method, or how to determine a suitable amount of LVs in a given scenario.

TABLE 2 | A summary of all of the software referenced in this review with the capabilities of being used to estimate GLLVMs.

	Software	Method(s)	Response families	Short description	Citation
Specialized packages	boral(R)	MCMC	(Zero-truncated) Poisson and negative binomial, binomial, normal, Tweedie; beta, ordinal, and more.	Highly accurate, at the cost of very slow computation. Provides samples drawn from the posterior distribution. Model fit may be harder to assess without knowledge in Bayesian methods.	Hui (2016, 2024)
	gllamm (Stata)	AGHQ	Poisson, binomial, normal, gamma, multinomial, ordinal.	Accurate estimation. Model fitting can slow down considerably when LV count exceeds two.	Rabe-Hesketh, Skrondal, and Pickles (2005)
	gllvm (R)	(E)VA, LA	(Zero-inflated) Poisson and negative binomial, binomial, normal, Tweedie, beta, ordinal, and more.	Aims to offer a balance between accuracy and computational efficiency. Focuses solely on GLLVMs. Implements several schemes for model-based ordination.	Niku, Brooks, et al. (2019); Niku et al. (2023)
	glmmTMB (R)	LA	(Zero-inflated) Poisson and negative binomial, binomial, normal, Tweedie, beta, ordinal, and more.	Aims to offer a good balance between accuracy and speed. Is focused on mixed effect models, with the option to specify a reduced-rank covariance structure.	Brooks et al. (2017)
	gmf (R)	PQL	Poisson, binomial, normal, gamma, inverse Gaussian, quasi-Poisson, quasi-binomial.	Best suited for very high-dimensional datasets and multiple LVs. Implements regularization to help control the effective number of parameters and LVs.	Kidzinski et al. (2020, 2022)
	Hmsc (R)	MCMC	Poisson, Bernoulli, normal, log-normal, log-normal Poisson.	Similar strengths and weaknesses to boral. More limited in featured response types, but offers more tools for incorporating spatiotemporal dependencies, phylogeny trees, etc.	Tikhonov et al. (2019, 2024)
	LatentINLA (R)	INLA	Everything that comes with R-INLA.	Under development. Relative strengths of INLA for GLLVMs remain untested.	O'Hara and van der Veen (2024)
	ltm (R)	EM, GHQ	Bernoulli	Limited to two LVs at most. Comes with many tools for researchers working in IRT.	Rizopoulos (2006)

(Continues)

TABLE 2 | (Continued)

	Software	Method(s)	Response families	Short description	Citation
General model building libraries	R-INLA (R)	INLA	Long list of readily available distributions, including majority of the ones listed above.	Offers the INLA method for approximate Bayesian inference of models with LVs. Conveniently integrates the stochastic partial differential equation, or SPDE, approach for models with spatially correlated effects.	Martins et al. (2013)
	Stan (R, Stata, MATLAB, Python, Julia, shell)	MCMC, VA, LA	Long list of readily available distributions. Allows the user to specify custom likelihood functions.	Provides full Bayesian inference of complex high-dimensional hierarchical LV models through a Hamiltonian Monte Carlo sampler. Approximate inference is also available with implementations of Laplace's and variational methods.	Stan Development Team (2024)
	TMB (R, C++)	LA	Long list of readily available distributions. Allows the user to specify custom likelihood functions.	Efficient marginal MLE by combination of fast matrix operations with LA and automatic differentiation. Pairs with R-INLA for SPDE approximation of spatial models, and with Stan for fully Bayesian analysis by the package tmbstan.	Kristensen et al. (2016)

Code for reproducing both simulation studies and the example can be found at https://github.com/pekolako/lvm-review.

4.1 | Simulation Studies

Both simulation setups consisted of four cases with varying sample sizes: (a) $n = 50$, $m = 50$; (b) $n = 50$, $m = 150$; (c) $n = 200$, $m = 50$, and (d) $n = 200$, $m = 150$, where n and m denote the amounts of rows and columns in the desired response matrices. Then, each setup/case used similar approach to generating the parameters used in simulation. The values for column effects and LV loadings were drawn from the uniform distribution, $\nu_j \sim U(-1, 1)$ and $\lambda_{j1}, \dots, \lambda_{jp} \sim U(-2, 2)$, respectively, with the addition that the upper triangle of the loading matrix was fixed to zero and the diagonal entries set to be positive—as per the usual GLLVM identifiability assumptions (Huber, Ronchetti, and Victoria-Feser 2004). The number of LVs was set to $p = 2$ in the first setup and $p = 5$ for the second setup. Then, in each of the cases (eight in total), the fixed values of ν_j and λ_j were used to generate 1000 datasets of simulated responses, according to (7), after first drawing the LV scores $\mathbf{u}_i = (u_{i1}, \dots, u_{ip})^\top \sim N_p(\mathbf{0}, \mathbf{I}_p)$. In the first setup, the responses were generated from Bernoulli distribution with probit link, while the second setup instead used Poisson distribution and log link, thus representing both presence–absence and count data, that is, some of the most common types of multivariate abundance data. Figure 1 shows the densities of values for the linear predictor η_{ij} (left) and the Bernoulli probabilities $p_{ij} = \Phi(\eta_{ij})$ (right) that were used to generate the presence–absence data during the simulations.

The sample sizes chosen for n and m were partly inspired by some real world ecological datasets. Namely, the size 50×150 roughly corresponds to the Bornean bird dataset of Cleary et al. (2005), which had 37 observational sites and 177 species, or the SBC LTER coastal kelp forest dataset (Reed and Miller 2023) with 44 observational units and about 150 species. This type of "wide" structure (i.e., $n \ll m$) is also very commonly encountered when dealing with sequencing data, for example, in studies on microbial species such as in Mach et al. (2015), Edwards et al. (2018), and Jernfors et al. (2024). To contrast this, the "long" datasets of size 200×50 are similar in size and form to, for example, the testate amoebae data in Daza Secco et al. (2016). The size 50×50 resembles that of the Australian ant data (Gibb and Cunningham 2011) used in the example in Section 4.2.

The methods compared were the VA, LA, and EVA from gllvm (Niku et al. 2023), MCMC from boral (Hui 2024), PQL from gmf (Kidzinski et al. 2020), LA from glmmTMB, and a combination of EM algorithm and GHQ from ltm (Rizopoulos 2006). It is important to note that excepting the last, all of the methods use very similar parametrizations and allow estimation of probit models on Bernoulli data; the ltm package instead only allows the use of logistic link, putting it at a disadvantage in these comparisons. The ltm package also cannot fit Poisson GLLVMs, and is limited to at most two LVs, thus excluding it from the simulation setup 2. Moreover, ltm does not allow the user to enforce positivity constraints on the diagonal elements of the loading matrix $\mathbf{\Lambda}$ without fixing them at some specific values. The package has been developed with focus in item response theory, while the rest of the compared packages have-been developed with ecological applications in mind, explaining the differing approaches.

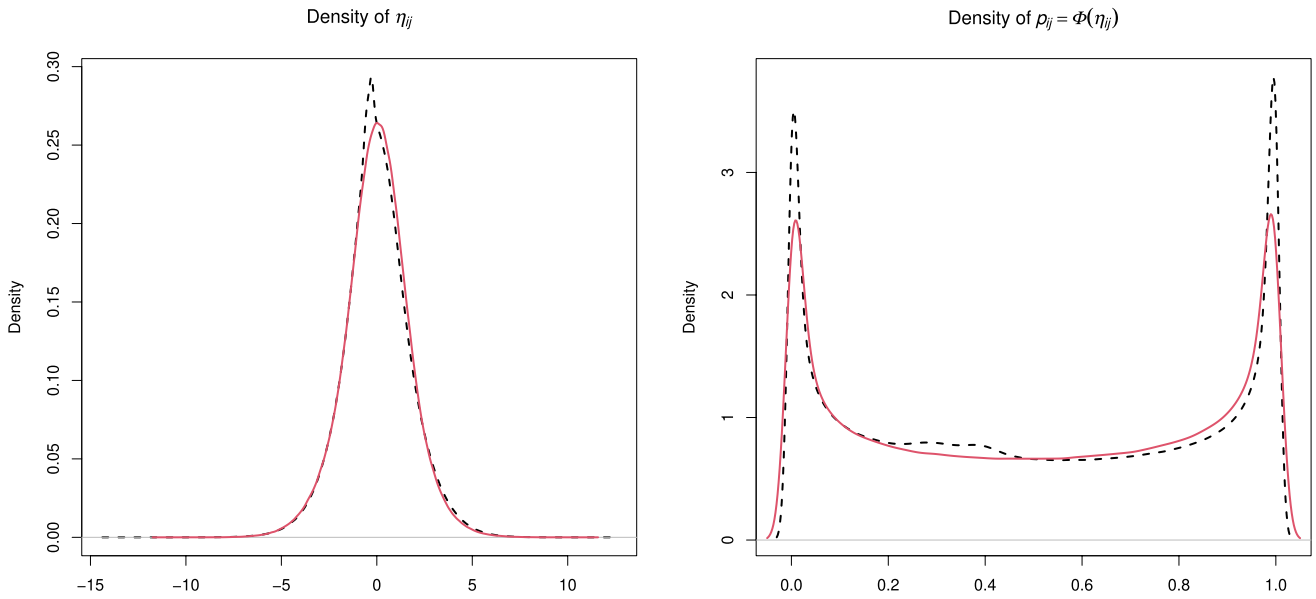

FIGURE 1 | Density plots for the values of the linear predictor $\eta_{ij} = \nu + \mathbf{\Lambda u}_i$ (left), and the resulting probabilities $p_{ij} = \Phi(\eta_{ij})$ (right), which were used to generate the Bernoulli data during the simulation studies. The solid (red) lines correspond to the case with $n = 50$ rows and $m = 50$ columns, and the dashed (black) lines correspond to the case with $n = 200$ rows and $m = 150$ columns.

Default settings for starting values and stopping criteria were used for all of the methods. Metrics of comparison included the bias and root mean square error (RMSE) of the column effect estimates $\widehat{\nu}_j$, as well as Procrustes errors (e.g., Bartholomew, Knott, and Moustaki 2011) of both the predicted LV scores $\widehat{\mathbf{u}}_i$ and the estimated loadings $\widehat{\boldsymbol{\lambda}}_j$. Procrustes error can be thought of as the Frobenius distance of two matrices, after accounting for differences in location, scale, and rotation. It is a measure often used for comparing the quality of ordinations w.r.t. some reference, such as the true $\mathbf{u}_i$ and $\boldsymbol{\lambda}_j$ here. Additionally, computation times were recorded for each method.

Tables 3 and 4 contain the summaries of the results from the first and second simulation setups, respectively. EVA proved to be the method with the lowest median computation time across all cases in setup 1, followed closely by VA. In the setup 2, PQL was the fastest method by several orders of magnitude, with both of the variational methods slowing down drastically (cases 2c, d), as the amount of rows n—and thus the amount of variational mean and covariance parameters—increased, compounded by the higher LV dimension of $p = 5$. Interestingly, the computation times were quite varied between the two LA implementations, even though both gllvm and glmmTMB leverage the TMB library. Generally, glmmTMB was the faster of the two, while gllvm produced better ordinations, the differences being most probably because of differing choices in regard to starting values or optimization algorithms used. In terms of bias and RMSE, the LA methods were on equal footing, and fairly close to VA/EVA—which were the two best methods overall, performing well in every metric across all cases.

LA made big improvements when the column count m increased, that is, cases b and d in Tables 3 and 4, reflecting similar findings in the supplementary simulation studies done in Korhonen et al. (2023). The smaller number of columns proved especially problematic for the PQL method, best evidenced by cases 1a and (to a lesser degree) 1c, where it failed to produce finite estimates for the row effects ν_j. The relatively subpar performance of the PQL in these comparisons is not surprising, as the method was developed with emphasis (and in a sense, dependence) on very large datasets—setting at which it excels at according to the numerical studies conducted in Kidzinski et al. (2022). The EM and GHQ hybrid method from ltm performed the worst in general, producing the highest and most spread-out biases and RMSEs, and high Procrustes errors—unsurprising when taking into account the vast differences in implementation compared with the others. In terms of estimation accuracy, the MCMC was firmly on par with VA/EVA/LA, but its computational burden was enormous in comparison.

4.2 | Real Data Example

Fitting GLLVMs is a complex process and a typical workflow is presented in Figure 2. To illustrate the process and to showcase that the algorithm choice matters we conduct a small case study using the best (and the worst) performer from the simulation studies done in Section 4.1. We fit GLLVMs using the packages gllvm and ltm on a dataset containing abundances of $m = 41$ species of ants collected at $n = 30$ observation sites in March–April 2008 near Canberra, Australia (Gibb and Cunningham 2011). The original data come as counts, so to facilitate the use of ltm, we transformed the data into binary presence/absence responses. With gllvm, we use EVA as the estimation method so that we can use logit link, similar to ltm. The dataset is freely available within the gmf package (Kidzinski et al. 2020), and in Kidzinski et al. (2022) the authors used it to compare their proposed two PQL estimation algorithms to the VA method from gllvm.

Figure 3 shows (residual) ordination plots resulting from the two methods; the upper plot is based on ltm fit and the lower is based on gllvm. Ordination analysis is an important application of GLLVMs in community ecology. To assess similarities and dissimilarities between the observational units, ordination

TABLE 3 | Results from the first simulation setup, in which 1000 datasets with (a) $n = 50$, $m = 50$; (b) $n = 50$, $m = 150$; (c) $n = 200$, $m = 50$, and (d) $n = 200$, $m = 150$ rows and columns, respectively, were generated according to a Bernoulli GLLVM with probit link and $p = 2$ LVs.

	Method	Package	Bias (SD)	RMSE (SD)	Proc. (LV)	Proc. (Λ)	Time (s)
(a)	MCMC	boral	−0.003 (0.123)	0.341 (0.064)	0.088	0.109	393.404
	VA	gllvm	0.003 (0.052)	0.300 (0.044)	**0.081**	**0.085**	1.528
	LA-1	gllvm	**−0.001** (0.543)	0.671 (0.717)	0.086	0.625	9.809
	EVA	gllvm	−0.002 (0.072)	**0.278** (0.042)	0.146	0.105	**1.162**
	LA-2	glmmTMB	−0.001 (0.606)	0.749 (0.882)	0.086	0.618	34.394
	PQL	gmf	—	—	0.120	0.257	2.529
	EM	ltm	−0.031 (2.154)	2.131 (2.337)	0.948	0.757	8.274
(b)	MCMC	boral	−0.003 (0.076)	0.274 (0.041)	0.040	0.088	1278.034
	VA	gllvm	**0.000** (0.076)	0.342 (0.081)	**0.034**	0.112	5.698
	LA-1	gllvm	−0.009 (0.339)	0.434 (0.440)	0.037	0.718	74.106
	EVA	gllvm	0.001 (0.065)	**0.237** (0.026)	0.041	**0.073**	**2.308**
	LA-2	glmmTMB	−0.002 (0.213)	0.414 (0.231)	0.037	0.664	195.586
	PQL	gmf	0.001 (0.122)	0.325 (0.101)	0.041	0.315	7.129
	EM	ltm	−0.027 (1.039)	1.069 (0.916)	0.948	0.802	32.387
(c)	MCMC	boral	−0.005 (0.036)	0.161 (0.039)	0.074	0.036	2954.051
	VA	gllvm	0.001 (0.046)	**0.138** (0.025)	**0.073**	**0.026**	4.931
	LA-1	gllvm	−0.001 (0.024)	0.156 (0.038)	**0.073**	0.034	21.077
	EVA	gllvm	−0.001 (0.036)	0.148 (0.025)	0.077	0.028	**3.112**
	LA-2	glmmTMB	−0.001 (0.024)	0.156 (0.038)	**0.073**	0.034	107.141
	PQL	gmf	0.001 (0.115)	0.235 (0.079)	0.101	0.272	68.320
	EM	ltm	**0.000** (0.543)	0.600 (0.217)	0.987	0.051	21.422
(d)	MCMC	boral	0.002 (0.027)	0.143 (0.033)	0.029	0.028	22048.171
	VA	gllvm	**0.000** (0.012)	0.145 (0.031)	**0.027**	**0.023**	24.639
	LA-1	gllvm	0.001 (0.021)	0.152 (0.038)	**0.027**	0.030	100.963
	EVA	gllvm	−0.002 (0.049)	**0.135** (0.028)	0.040	0.024	**7.776**
	LA-2	glmmTMB	0.001 (0.021)	0.152 (0.038)	**0.027**	0.030	687.591
	PQL	gmf	0.004 (0.079)	0.159 (0.043)	0.029	0.027	11.430
	EM	ltm	0.005 (0.484)	0.562 (0.196)	0.987	0.040	94.358

Note: Bolded values indicate the best performer on each of the metrics used; mean bias and RMSE of the column effect estimates $\hat{v}_j$, mean Procrustes errors of LVs and loadings $\mathbf{\Lambda}$, and median computation time in seconds. The means were calculated using trimming factor of 0.05 to remove effects of the most extreme values. The Procrustes errors were scaled by the number of LV scores or loading parameters. Note, that in the smallest setting (a) the PQL method failed to produce finite biases and RMSEs for 27 of the 50 column effects involved. PQL also failed on two column effects in the case c, but these were simply left out when calculating the mean.

methods seek a low-dimensional representation of the original high-dimensional dataset—two-dimensional here, with $p = 2$. In the plots in Figure 3, sites that are closer to each other are deemed more similar (because of species composition, unobserved environmental factors, et cetera.). Interestingly, the arrangement of the sites here are quite different between the two estimation methods. Most notably, the site #9 is very distant from the rest according to ltm, while belonging to the "main" cluster of points as deemed by gllvm. On the other hand, some formations are similar, as for example, sites #22 and #30 lie close to each other on either plots, as are sites #8 and #24. Note that we used the same starting values for the model parameters with both methods.

Although $p = 2$ LVs is the most common choice for the purpose of ordination, a model with higher p can be more suitable to the data at hand—or other applications. The decision can be guided by considering information criteria, such as AIC or BIC. These values, together with the corrected AIC—or AICc—are printed, for example, when calling summary() upon a gllvm object. Table 5 lists these quantities for one to six LVs from GLLVMs fitted to the ant data with the EVA approach, along

TABLE 4 | Results from the second simulation setup, in which 1000 datasets with (a) $n = 50$, $m = 50$; (b) $n = 50$, $m = 150$; (c) $n = 200$, $m = 50$, and (d) $n = 200$, $m = 150$ rows and columns, respectively, were generated according to a Poisson GLLVM with log link and $p = 5$ LVs.

	Method	Package	Bias (SD)	RMSE (SD)	Proc. (LV)	Proc. (Λ)	Time (s)
(a)	MCMC	boral	−0.073 (0.192)	0.452 (0.097)	0.034	0.033	728.192
	VA	gllvm	−0.023 (0.025)	0.341 (0.064)	**0.026**	**0.029**	81.637
	LA-1	gllvm	−0.024 (0.023)	0.341 (0.065)	**0.026**	**0.029**	103.360
	EVA	gllvm	−0.023 (0.025)	0.341 (0.064)	**0.026**	**0.029**	81.359
	LA-2	glmmTMB	**−0.016** (0.110)	**0.291** (0.050)	0.069	0.107	44.847
	PQL	gmf	−0.192 (0.357)	2.071 (0.600)	0.072	0.115	**0.255**
(b)	MCMC	boral	**0.003** (0.480)	0.831 (0.201)	0.080	0.072	2136.711
	VA	gllvm	−0.021 (0.021)	**0.348** (0.066)	**0.021**	**0.028**	287.885
	LA-1	gllvm	−0.021 (0.020)	**0.348** (0.066)	**0.021**	**0.028**	502.301
	EVA	gllvm	−0.021 (0.021)	0.350 (0.066)	**0.021**	**0.028**	286.059
	LA-2	glmmTMB	0.036 (0.243)	0.356 (0.110)	0.157	0.345	349.671
	PQL	gmf	0.038 (0.302)	2.395 (0.610)	0.101	0.153	**0.529**
(c)	MCMC	boral	0.007 (0.031)	0.170 (0.039)	0.026	0.022	2938.741
	VA	gllvm	−0.004 (0.006)	0.162 (0.034)	**0.009**	**0.006**	1610.825
	LA-1	gllvm	**−0.003** (0.009)	0.163 (0.032)	**0.009**	0.007	486.576
	EVA	gllvm	−0.004 (0.006)	**0.162** (0.033)	**0.009**	**0.006**	1623.731
	LA-2	glmmTMB	0.004 (0.149)	0.191 (0.058)	0.068	0.105	168.984
	PQL	gmf	−0.622 (1.389)	2.701 (1.744)	0.126	0.209	**0.695**
(d)	MCMC	boral	0.035 (0.175)	0.614 (0.143)	0.066	0.067	8515.737
	VA	gllvm	**−0.002** (0.011)	0.176 (0.029)	0.008	0.009	3106.656
	LA-1	gllvm	0.002 (0.032)	**0.168** (0.042)	**0.006**	**0.007**	4136.681
	EVA	gllvm	**−0.002** (0.011)	0.175 (0.029)	0.008	0.008	3117.498
	LA-2	glmmTMB	0.031 (0.312)	0.391 (0.127)	0.179	0.410	1451.624
	PQL	gmf	0.431 (0.354)	1.967 (0.516)	0.182	0.382	**1.210**

Note: Bolded values indicate the best performer on each of the metrics used; mean bias and RMSE of the column effect estimates $\hat{v}_j$, mean Procrustes errors of LVs and loadings $\boldsymbol{\Lambda}$, and median computation time in seconds. The means were calculated using trimming factor of 0.05 to remove effects of the most extreme values. The Procrustes errors were scaled by the number of LV scores or loading parameters.

with the corresponding log-likelihood approximations and degrees of freedom. Lowest—that is, the best—AIC is achieved with the model with two LVs, while the model with only one LV is the best in terms of AICc and BIC. AICc differs from AIC by including an additional penalty $\left(2k^2 + 2k\right) / (nm - k - 1)$ on the number of model parameters k, as AIC might encourage overfitting with small nm (Burnham and Anderson 2002). Based on Table 5, going beyond $p = 2$ brings no considerable benefits in this case.

In addition to similarities/dissimilarities between the observation sites, the dependence patterns among the different species in the data are also often of interest. These can be assessed, for example, using biplots, which simultaneously display both LV scores and loadings in the same figure, or through inspecting the residual covariation matrix—attained from the estimated loading matrix fit by $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}^{\top}$ (with a possible additional correction term depending on the response family and link). In gllvm, this can be done easily by calling the function getResidualCov(). By transforming covariances into correlations, and with help of the packages corrplot (Wei and Simko 2021) and gclus (Hurley 2019), one can visualize the inter-species correlations in the manner of Figure 4. In the correlation plot, sizes and colors of the squares signal the magnitude and direction of dependence. Pairs of species with positive correlation (blue) tend to exhibit co-occurrence among the environments present in the study, while pairs of species with negative correlation (red) tend to compete for space or resources, et cetera. Figure 4 suggests, that among the ants, for example, the species *Monomorium rothsteini* and *Cardiocondyla atalanta* have a high positive correlation, meaning that they commonly share environments. Meanwhile, the species *Camponotus cinereus amperei* and *Monomorium sydneyense* have high negative correlation, that is, they do not often appear in the same environment.

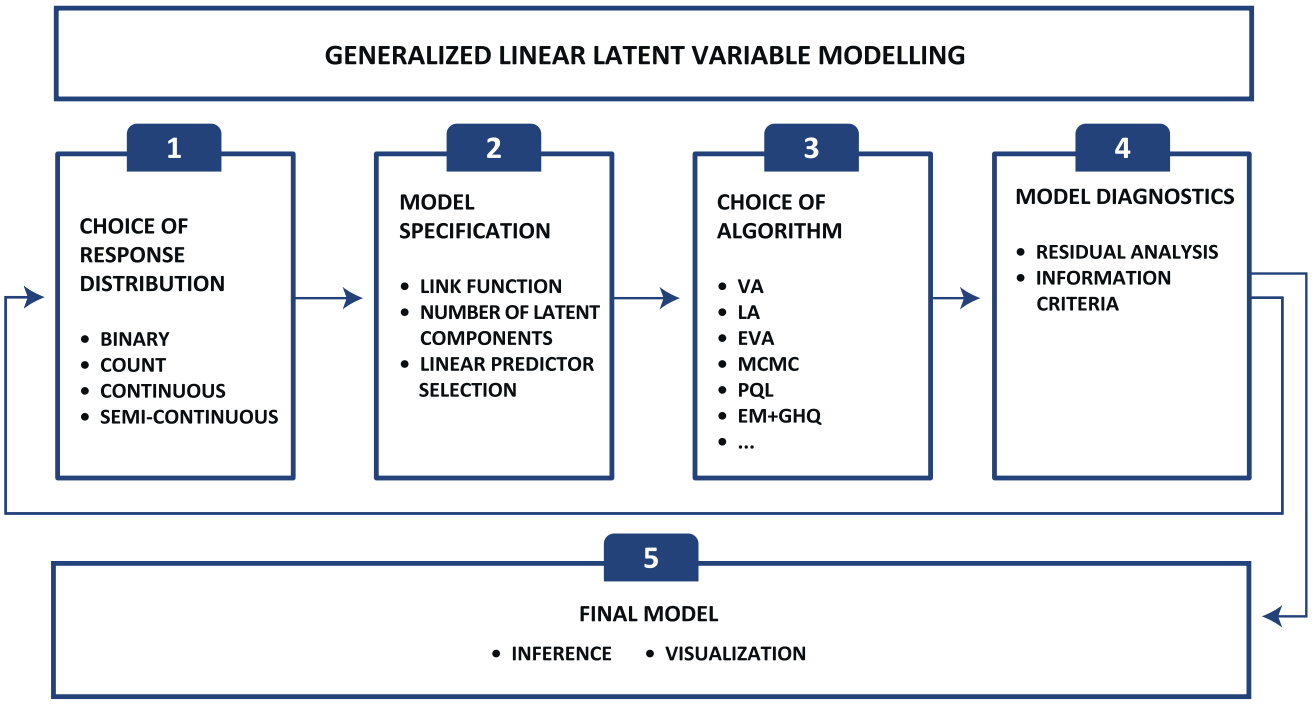

FIGURE 2 | Workflow of a typical GLLVM modeling process.

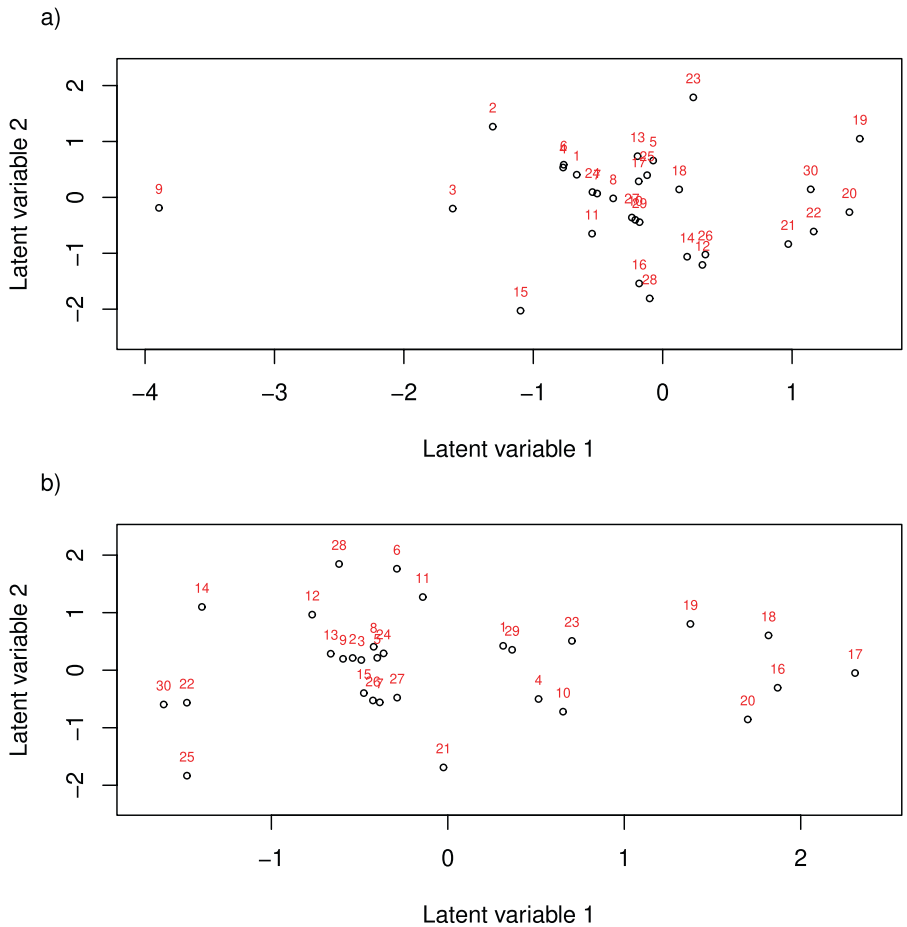

FIGURE 3 | Residual ordination of the ant dataset (when converted to presence–absence responses), formed based on the predicted LV scores using both (a) ltm and (b) gllvm. With the latter, EVA was chosen as the method of estimation. Ordination plots can be used for determining observation sites/units (numbered here), that are similar based on their species composition, environment, et cetera. Here, the arrangements of the sites differ notably; for example, site #9 is deemed very distant from the rest by ltm, but grouped firmly among the main mass by gllvm.

Finally, the goodness of a GLLVM fit on data can be assessed in a manner shared by many popular statistical modeling frameworks—by inspecting the residuals. With discrete responses specifically, the *Dunn–Smyth residuals*, also known as *randomized quantile residuals* (Dunn and Smyth 1996), are commonly employed, for they yield quantities on a continuous scale through clever application of sampling. Calling plot() on a gllvm object produces plots displaying the Dunn–Smyth residuals against the linear predictor, normal quantiles or row/column index. Figure 5 shows these plots for the $p = 2$ GLLVM fitted on the ant dataset. Here, no visible patterns deviating from the norm are revealed, suggesting that the model fits the data well.

TABLE 5 | Akaike, corrected Akaike, and Bayesian information criteria resulting from GLLVMs fitted on the ant dataset using gllvm(), with differing number of assumed LVs p.

p	AIC	AICc	BIC	$\log \mathcal{L}$	df
1	1330.29	**1342.16**	**1749.70**	−583.14	82
2	**1324.91**	1352.02	1948.91	−540.45	122
3	1325.68	1374.52	2149.16	−501.84	161
4	1335.17	1412.45	2353.01	−468.59	199
5	1374.86	1487.51	2581.95	−451.43	236
6	1373.13	1528.32	2764.35	−414.57	272

Note: Corresponding values for the approximate log-likelihood and the degrees of freedom are also shown. The values of the criteria indicate, that either $p = 1$ or $p = 2$ are the most likely to give the best model candidates, for this scenario. The lowest values for information criteria are marked in bold.

5 | Conclusions

In this article, we reviewed GLLVMs and some widely used computational approaches for model fitting focusing on methods based on likelihood inference. Six methods that are easily available in R were compared using two simulation studies and a real data example from community ecology. The simulation studies revealed that two recently implemented methods available in R package gllvm (Niku et al. 2019, 2023), that is, the method based on variational approximations (VA, Hui 2017)

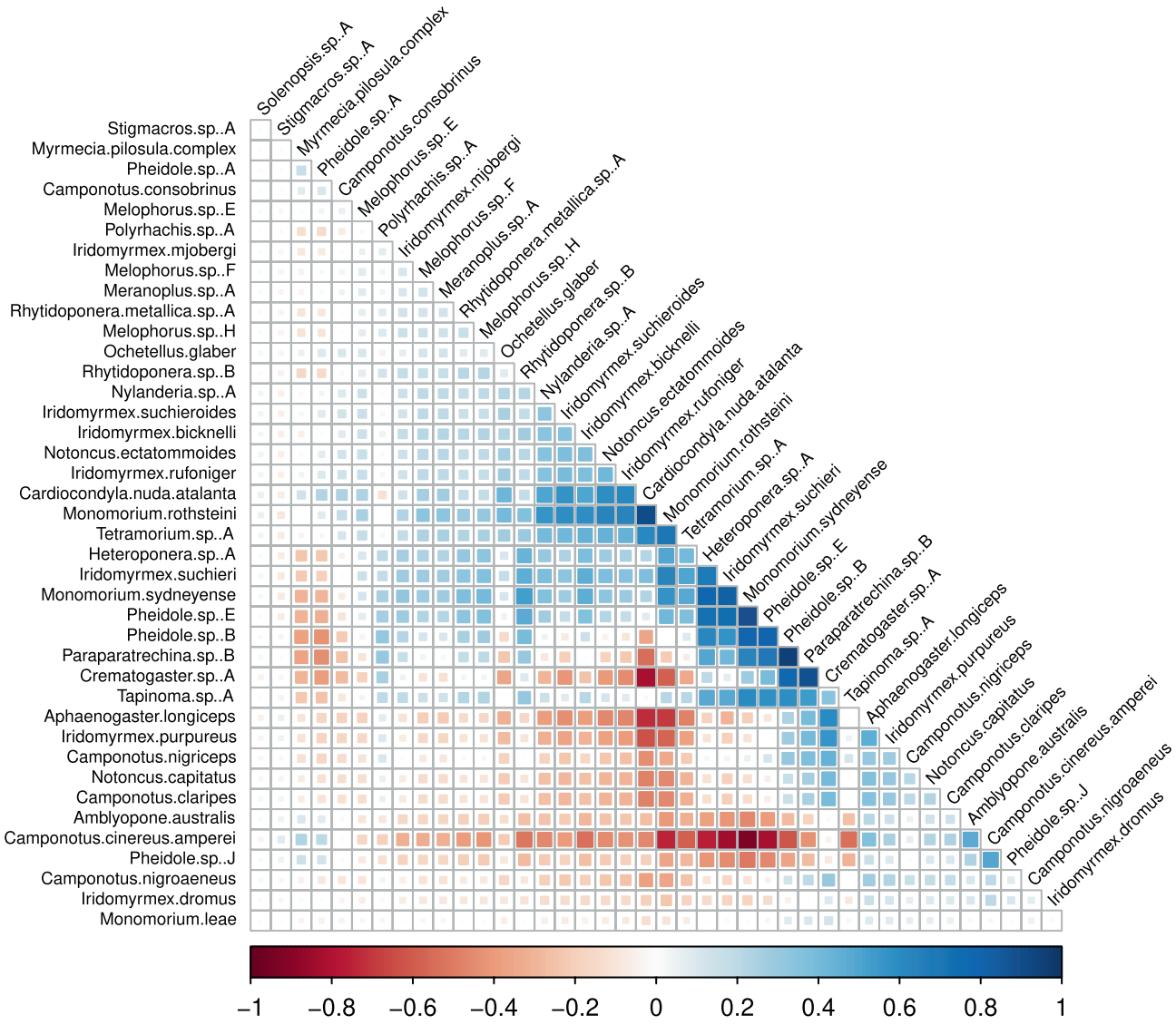

FIGURE 4 | Residual correlation plot resulting from a GLLVM fit. Blue and red squares indicate the pairs of ant species that exhibit co-occurrent or competitive relationships, respectively, along the environments or observational units present in the study.

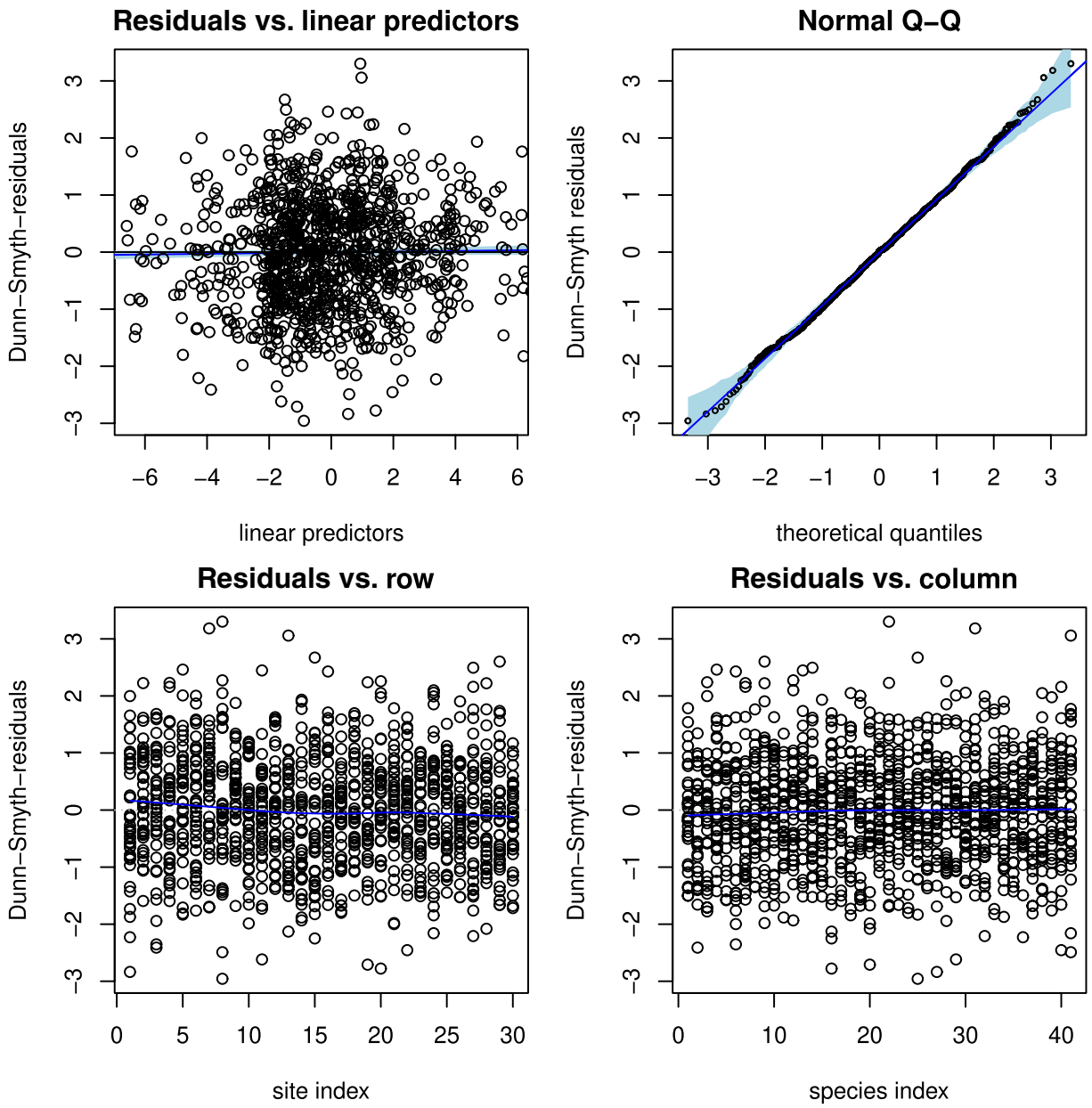

FIGURE 5 | Four kinds of residual plots from calling plot() on gllvm object, after fitting GLLVM with two LVs on the Australian ant dataset. Deviant patterns in the Dunn–Smyth residuals would be a sign of weak fit or violation of model assumptions.

and the method based on EVA (Korhonen et al. 2023), had the best overall performances when compared with the other methods. The method utilizing EVAs had the lowest computation times across all sample sizes in the first setting, followed closely by the other variational approach. As discussed in Korhonen et al. (2023), EVA seems to be a promising method for LV model fitting as one can obtain a closed-form approximation to the marginal likelihood for any response type and link function combination. However, as the second simulation setup shows, with a higher assumed LV count, the methods based on variational approximations start to exhibit scalability issues because of the rapidly increasing amount of variational parameters needed—prompting further development of scalable variational methods for estimation of GLLVMs. Related to this, see, for example, Zhang et al. (2019) for a recent review from machine learning standpoint.

In this review, we focused only on exploratory methods. In confirmatory FA one builds a hypothetical factor model based on the prior knowledge on the phenomenon and then tries to confirm that base on the data; see, for example, Mulaik (2009). Confirmatory FA is also used in structural equation models (SEMs, Jöreskog 1969, 1970) to estimate LVs. SEM can be seen as an extension of the FA model as described in Section 2.1. The method stems from the path analysis developed in Wright (1921) and is widely used in psychology and sociology to capture the relationship among a set of variables (Bollen 2002). SEMs consist of two types of LVs often referred to as dependent (endogenous) variables and independent (exogenous) variable. Models allow causal relationships between LVs by relating variables by the LV model known as the structural model. For more details on SEMs, we refer to Bollen (1989) and Skrondal and Rabe-Hesketh (2004). For the

review of SEMs applied to the analysis of ecological data, see Fan et al. (2016). For R software packages for confirmatory FA and SEMs, see lavaan (Rosseel 2012) and sem (Fox, Nie, and Byrnes 2022), for example.

Standard GLLVMs as reviewed in this article have been extended to more general settings in various articles. Here we assumed that LVs are independent vectors from a standard multivariate normal distribution. When modeling spatial and/or temporal data, one has to choose a more general covariance structure for LVs. FA for continuous responses has been extended to longitudinal data in Raffalovich and Bohrnstedt (1987) and Marsh and Grayson (1994), for example. The ordinal response case is covered in Cagnone, Moustaki, and Vasdekis (2009). GLLVMs in spatial and spatio-temporal settings and related computational approaches have been discussed in Wang and Wall (2003), Zhu, Eickhoff, and Yan (2005), Lopes, Gamerman, and Salazar (2011), Hui, Hill, and Welsh (2022), and Hui et al. (2023) among others. For models specific for community ecology, see Thorson et al. (2015, 2016), Ovaskainen et al. (Ovaskainen, Abrego, et al. 2016 and Ovaskainen, Roy, et al. 2016), and Tikhonov et al. (2020). For robust and semiparametric approaches for GLLVMs, we refer to Ma and Genton (2010), Irincheeva, Cantoni, and Genton (2012), and Moustaki and Victoria-Feser (2006), for example.

Author Contributions

Pekka Korhonen: formal analysis (lead), investigation (lead), methodology (equal), writing – original draft (equal), writing – review and editing (equal). **Klaus Nordhausen:** methodology (equal), writing – original draft (equal), writing – review and editing (equal). **Sara Taskinen:** methodology (equal), writing – original draft (equal), writing – review and editing (equal).

Acknowledgments

The work of P.K. and S.T. was supported by the Research Council of Finland (453691) and the Kone Foundation. The work of K.N. was supported by the Research Council of Finland (363261). The work of K.N. and S.T. was supported by the HiTEc COST Action (CA21163). The authors wish to acknowledge CSC—IT Center for Science, Finland, for computational resources, and the two anonymous reviewers for their insightful comments on the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Related Wires Articles

Time series factor models

Latent variable, matrix decomposition, and constrained uniqueness formulations

A review of second-order blind identification methods

A systematic review of exploratory factor analysis packages in R software

Independent component analysis: A statistical perspective

References

Andersen, E. B. 1973. "Conditional Inference for Multiple-Choice Questionnaires." *British Journal of Mathematical and Statistical Psychology* 26: 31–44.

Andersen, E. B. 1982. "Latent Structure Analysis: A Survey." *Scandinavian Journal of Statistics* 9, no. 1: 1–12.

Andersen, E. B., and M. Madsen. 1977. "Estimating the Parameters of the Latent Population Distribution." *Psychometrika* 42: 357–374.

Anderson, T. W. 1963. "The Use of Factor Analysis in the Statistical Analysis of Multiple Time Series." *Psychometrika* 28: 1–25.

Anderson, T. W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd ed. New York: Wiley.

Anderson, T. W., and H. Rubin. 1956. "Statistical Inference in Factor Analysis." In *Third Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman, 111–150. Berkeley, CA: University of California Press.

Bachoc, F., M. G. Genton, K. Nordhausen, A. Ruiz-Gazen, and J. Virta. 2020. "Spatial Blind Source Separation." *Biometrika* 107: 627–646.

Bartholomew, D. J. 1980. "Factor Analysis for Categorical Data." *Journal of the Royal Statistical Society: Series B* 42, no. 3: 293–321.

Bartholomew, D. J. 1984. "The Foundations of Factor Analysis." *Biometrika* 71, no. 2: 221–232.

Bartholomew, D. J., M. Knott, and I. Moustaki. 2011. *Latent Variable Models and Factor Analysis: A Unified Approach*. New York: Wiley.

Bartlett, M. S. 1950. "Tests of Significance in Factor Analysis." *British Journal of Statistical Psychology* 3, no. 2: 77–85.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67, no. 1: 1–48.

Bhattacharya, A., and D. B. Dunson. 2011. "Sparse Bayesian Infinite Factor Models." *Biometrika* 98, no. 2: 291–306.

Bianconcini, S. 2014. "Asymptotic Properties of Adaptive Maximum Likelihood Estimators in Latent Variable Models." *Bernoulli* 20, no. 3: 1507–1531.

Bianconcini, S., and S. Cagnone. 2021. "Dynamic Latent Variable Models for the Analysis of Cognitive Abilities in the Elderly Population." *Statistics in Medicine* 40, no. 20: 4410–4429.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe. 2017. "Variational Inference: A Review for Statisticians." *Journal of the American Statistical Association* 112, no. 518: 859–877.

Blocker, A. W. 2022. "fastGHQuad: Fast 'Rcpp' Implementation of Gauss-Hermite Quadrature." R Package Version 1.0.1. https://CRAN.R-project.org/package=fastGHQuad.

Bock, R. D., and M. Aitkin. 1981. "Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm." *Psychometrika* 46: 443–459.

Bock, R. D., and M. Lieberman. 1970. "Fitting a Response Model Forn Dichotomously Scored Items." *Psychometrika* 35, no. 2: 179–197.

Bollen, K. 1989. *Structural Equations With Latent Variables*. New York: John Wiley & Sons.

Bollen, K. 2002. "Latent Variables in Psychology and the Social Sciences." *Annual Review of Psychology* 53: 605–634.

Box, G. E., and N. R. Draper. 2007. *Response Surfaces, Mixtures, and Ridge Analyses*. New York: John Wiley & Sons.

Braun, M., and J. McAuliffe. 2010. "Variational Inference for Large-Scale Models of Discrete Choice." *Journal of the American Statistical Association* 105, no. 489: 324–335.

Breslow, N. E., and D. G. Clayton. 1993. "Approximate Inference in Generalized Linear Mixed Models." *Journal of the American Statistical Association* 88: 9–25.

Brooks, M. E., K. Kristensen, K. J. van Benthem, et al. 2017. "glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling." *R Journal* 9, no. 2: 378–400.

Burnham, K., and D. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer.

Cagnone, S., I. Moustaki, and V. Vasdekis. 2009. "Latent Variable Models for Multivariate Longitudinal Ordinal Responses." *British Journal of Mathematical and Statistical Psychology* 62, no. 2: 401–415.

Cao, Y., and Y. Xie. 2016. "Poisson Matrix Recovery and Completion." *IEEE Transactions on Signal Processing* 64, no. 6: 1609–1620.

Cappellari, L., and S. P. Jenkins. 2003. "Multivariate Probit Regression Using Simulated Maximum Likelihood." *Stata Journal* 3, no. 3: 278–294.

Cattell, R. B. 1966. "The Scree Test for the Number of Factors." *Multivariate Behavioral Research* 1, no. 2: 245–276.

Chiquet, J., M. Mariadassou, and S. Robin. 2018. "Variational Inference for Probabilistic Poisson PCA." *Annals of Applied Statistics* 12, no. 4: 2674–2698.

Cleary, D. F. R., M. J. Genner, T. J. B. Boyle, T. Setyawati, C. D. Angraeti, and S. B. J. Menken. 2005. "Associations of Bird Species Richness and Community Composition With Local and Landscape-Scale Environmental Factors in Borneo." *Landscape Ecology* 20: 989–1001.

Collins, M., S. Dasgupta, and R. E. Schapire. 2001. "A Generalization of Principal Components Analysis to the Exponential Family." In *Advances in Neural Information Processing Systems 14: Proceedings of the 2001 Conference*, edited by T. Dietterich, S. Becker, and Z. Ghahramani. Cambridge, UK: MIT Press.

Comon, P. 1994. "Independent Component Analysis, a New Concept?" *Signal Processing* 36, no. 3: 287–314.

Comon, P., and C. Jutten. 2010. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Oxford: Academic Press.

Danielsson, J. 1994. "Stochastic Volatility in Asset Prices Estimation With Simulated Maximum Likelihood." *Journal of Econometrics* 64, no. 1: 375–400.

Daolin Pang, H. Z., and T. Wang. 2023. "Factor Augmented Inverse Regression and Its Application to Microbiome Data Analysis." *Journal of the American Statistical Association* 119, no. 547: 1957–1967.

Daza Secco, E., T. Haapalehto, J. Haimi, K. Meissner, and T. Tahvanainen. 2016. "Do Testate Amoebae Communities Recover in Concordance With Vegetation After Restoration of Drained Peatlands?" *Mires and Peat* 18: 1–14.

De Smedt, T., K. Simons, A. Van Nieuwenhuyse, and G. Molenberghs. 2015. "Comparing MCMC and INLA for Disease Mapping With Bayesian Hierarchical Models." *Archives of Public Health* 73: 1.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood From Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society: Series B* 39: 1–38.

Dombrovski, V. C., D. V. Zhurauliou, and A. Ashton-Butt. 2022. "Long-Term Effects of Rewilding on Species Composition: 22 Years of Raptor Monitoring in the Chernobyl Exclusion Zone." *Restoration Ecology* 30, no. 8: e13633.

Dunn, P. K., and G. K. Smyth. 1996. "Randomized Quantile Residuals." *Journal of Computational and Graphical Statistics* 5, no. 3: 236–244.

Edwards, J. A., C. M. Santos-Medellín, Z. S. Liechty, et al. 2018. "Compositional Shifts in Root-Associated Bacterial and Archaeal Microbiota Track the Plant Life Cycle in Field-Grown Rice." *PLoS Biology* 16, no. 2: e2003862.

Ensor, K. B. 2013. "Time Series Factor Models." *WIREs Computational Statistics* 5, no. 2: 97–104.

Fan, Y., J. Chen, G. Shirkey, et al. 2016. "Applications of Structural Equation Modeling (SEM) in Ecological Studies: An Updated Review." *Ecological Processes* 5, no. 11: 1–12.

Firth, D. 1987. "On the Efficiency of Quasi-Likelihood Estimation." *Biometrika* 74, no. 2: 233–245.

Fong, Y., H. Rue, and J. Wakefield. 2010. "Bayesian Inference for Generalized Linear Mixed Models." *Biostatistics* 11, no. 3: 397–412.

Fox, J., Z. Nie, and J. Byrnes. 2022. "sem: Structural Equation Models." R Package Version 3.1-15. https://CRAN.R-project.org/package=sem.

Gelman, A., and D. B. Rubin. 1996. "Markov Chain Monte Carlo Methods in Biostatistics." *Statistical Methods in Medical Research* 5, no. 4: 339–355.

Gibb, H., and S. A. Cunningham. 2011. "Habitat Contrasts Reveal a Shift in the Trophic Position of Ant Assemblages." *Journal of Animal Ecology* 80, no. 1: 119–127.

Gibson, W. A. 1959. "Three Multivariate Models: Factor Analysis, Latent Structure Analysis, and Latent Profile Analysis." *Psychometrika* 24: 229–252.

Golding, N. 2019. "GRETA: Simple and Scalable Statistical Modelling in R." *Journal of Open Source Software* 4, no. 40: 1601.

Govindasamy, P., N. J. M. Isa, N. F. Mohamed, et al. 2024. "A Systematic Review of Exploratory Factor Analysis Packages in R Software." *WIREs Computational Statistics* 16, no. 1: e1630.

Green, P. J. 1987. "Penalized Likelihood for General Semi-Parametric Regression Models." *International Statistical Review* 55: 245–259.

Hall, P., I. Johnstone, J. Ormerod, M. Wand, and J. Yu. 2020. "Fast and Accurate Binary Response Mixed Model Analysis via Expectation Propagation." *Journal of the American Statistical Association* 115, no. 532: 1902–1916.

Hastings, W. K. 1970. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika* 57, no. 1: 97–109.

Held, L., B. Schrödle, and H. Rue. 2010. "Posterior and Cross-Validatory Predictive Checks: A Comparison of MCMC and INLA." In *Statistical Modelling and Regression Structures: Festschrift in Honour of Ludwig Fahrmeir*, edited by T. Kneib and G. Tutz, 91–110. Heidelberg: Physica-Verlag HD.

Hernandez-Lobato, J., Y. Li, M. Rowland, T. Bui, D. Hernández-Lobato, and R. Turner. 2016. "Black-Box Alpha Divergence Minimization." In *International Conference on Machine Learning*, PMLR (1511–1520).

Hoffman, M. D., and A. Gelman. 2014. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research* 15, no. 1: 1593–1623.

Hotelling, H. 1933. "Analysis of a Complex of Statistical Variables Into Principal Components." *Journal of Educational Psychology* 24, no. 6: 417–441.

Huber, P., E. Ronchetti, and M. Victoria-Feser. 2004. "Estimation of Generalized Linear Latent Variable Models." *Journal of the Royal Statistical Society: Series B* 66: 893–908.

Huber, P., O. Scaillet, and M.-P. Victoria-Feser. 2009. "Assessing Multivariate Predictors of Financial Market Movements: A Latent Factor Framework for Ordinal Data." *Annals of Applied Statistics* 3, no. 1: 249–271.

Hui, F. K. 2017. "Model-Based Simultaneous Clustering and Ordination of Multivariate Abundance Data in Ecology." *Computational Statistics & Data Analysis* 105: 1–10.

Hui, F. K. 2024. "Boral: Bayesian Ordination and Regression Analysis." R Package Version 2.0.2. https://CRAN.R-project.org/package=boral.

Hui, F. K., N. A. Hill, and A. Welsh. 2022. "Assuming Independence in Spatial Latent Variable Models: Consequences and Implications of Misspecification." *Biometrics* 78, no. 1: 85–99.

Hui, F. K. C. 2016. "Boral – Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in R." *Methods in Ecology and Evolution* 7: 744–750.

Hui, F. K. C., S. Taskinen, S. Pledger, S. D. Foster, and D. I. Warton. 2015. "Model-Based Approaches to Unconstrained Ordination." *Methods in Ecology and Evolution* 6: 399–411.

Hui, F. K. C., D. I. Warton, S. D. Foster, and C. R. Haak. 2023. "Spatiotemporal Joint Species Distribution Modelling: A Basis Function Approach." *Methods in Ecology and Evolution* 14, no. 8: 2150–2164.

Hui, F. K. C., D. I. Warton, J. T. Ormerod, V. Haapaniemi, and S. Taskinen. 2017. "Variational Approximations for Generalized Linear Latent Variable Models." *Journal of Computational and Graphical Statistics* 26, no. 1: 35–43.

Hurley, C. 2019. "gclus: Clustering Graphics." R Package Version 1.3.2. https://CRAN.R-project.org/package=gclus.

Iantovics, L. B., C. Rotar, and F. Morar. 2019. "Survey on Establishing the Optimal Number of Factors in Exploratory Factor Analysis Applied to Data Mining." *WIREs Data Mining and Knowledge Discovery* 9, no. 2: e1294.

Irincheeva, I., E. Cantoni, and M. G. Genton. 2012. "Generalized Linear Latent Variable Models With Flexible Distribution of Latent Variables." *Scandinavian Journal of Statistics* 39, no. 4: 663–680.

Jernfors, T., A. Lavrinienko, I. Vareniuk, et al. 2024. "Association Between Gut Health and Gut Microbiota in a Polluted Environment." *Science of the Total Environment* 914: 169804.

Jöreskog, K. G. 1969. "A General Approach to Confirmatory Maximum Likelihood Factor Analysis." *Psychometrika* 34, no. 2: 183–202.

Jöreskog, K. G. 1970. "A General Method for Analysis of Covariance Structures." *Biometrika* 57, no. 2: 239–251.

Kaiser, H. F. 1958. "The Varimax Criterion for Analytic Rotation in Factor Analysis." *Psychometrika* 23: 187–200.

Kaiser, H. F. 1960. "The Application of Electronic Computers to Factor Analysis." *Educational and Psychological Measurement* 20, no. 1: 141–151.

Kaiser, H. F. 1961. "A Note on Guttman's Lower Bound for the Number of Common Factors." *British Journal of Statistical Psychology* 14, no. 1: 1–2.

Kidzinski, L., F. K. Hui, D. I. Warton, and T. J. Hastie. 2020. "gmf: Generalized Matrix Factorization." R Package Version 0.2.1, Commit 32500abb38d6fd114187bf646ded9d32d451c926. https://github.com/kidzik/gmf.

Kidzinski, L., F. K. Hui, D. I. Warton, and T. J. Hastie. 2022. "Generalized Matrix Factorization: Efficient Algorithms for Fitting Generalized Linear Latent Variable Models to Large Data Arrays." *Journal of Machine Learning Research* 23, no. 291: 1–29.

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, no. 1: 26.

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." arXiv Preprint. https://arxiv.org/abs/2403.11562v1.

Kristensen, K. 2023. "RStan: The R Interface to Stan." R Package Version 1.4. https://cran.r-project.org/web/packages/RTMB/index.html.

Kristensen, K., A. Nielsen, C. Berg, H. Skaug, and B. Bell. 2016. "TMB: Automatic Differentiation and Laplace Approximation." *Journal of Statistical Software* 70, no. 5: 1–21.

Kullback, S., and R. A. Leibler. 1951. "On Information and Sufficiency." *Annals of Mathematical Statistics* 22, no. 1: 79–86.

Kuss, M., C. E. Rasmussen, and R. Herbrich. 2005. "Assessing Approximate Inference for Binary Gaussian Process Classification." *Journal of Machine Learning Research* 6, no. 57: 1679–1704.

Lammel, D. R., G. Barth, O. Ovaskainen, et al. 2018. "Direct and Indirect Effects of a pH Gradient Bring Insights Into the Mechanisms Driving Prokaryotic Community Structures." *Microbiome* 6, no. 1: 6–106.

Lawley, D. N. 1953. "A Modified Method of Estimation in Factor Analysis and Some Large Sample Results." In *Uppsala Symposium on Psychological Factor Analysis*, 35–42.

Lawley, D. N., and A. E. Maxwell. 1962. "Factor Analysis as a Statistical Method." *Journal of the Royal Statistical Society: Series D* 12, no. 3: 209–229.

Lazarsfeld, P., and N. Henry. 1968. *Latent Structure Analysis*. Mifflin: Houghton.

Lee, D., and H. S. Seung. 2000. "Algorithms for Non-Negative Matrix Factorization." In *Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference*, edited by T. Leen, T. Dietterich, and V. Tresp. Cambridge, UK: MIT Press.

Lee, L.-F. 1995. "Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models." *Econometric Theory* 11, no. 3: 437–483.

Lesaffre, E., and B. Spiessens. 2001. "On the Effect of the Number of Quadrature Points in a Logistic Random Effects Model: An Example." *Journal of the Royal Statistical Society, Series C* 50, no. 3: 325–335.

Lindgren, F., H. Rue, and J. Lindström. 2011. "An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach." *Journal of the Royal Statistical Society: Series B* 73, no. 4: 423–498.

Lopes, H. F., D. Gamerman, and E. Salazar. 2011. "Generalized Spatial Dynamic Factor Models." *Computational Statistics & Data Analysis* 55, no. 3: 1319–1330.

Ma, Y., and M. G. Genton. 2010. "Explicit Estimating Equations for Semiparametric Generalized Linear Latent Variable Models." *Journal of the Royal Statistical Society: Series B* 72, no. 4: 475–495.

Mach, N., M. Berri, J. Estellé, et al. 2015. "Early-Life Establishment of the Swine Gut Microbiome and Impact on Host Phenotypes." *Environmental Microbiology Reports* 7, no. 3: 554–569.

Marsh, H. W., and D. Grayson. 1994. "Longitudinal Confirmatory Factor Analysis: Common, Time-Specific, Item-Specific, and Residual-Error Components of Variance." *Structural Equation Modeling: A Multidisciplinary Journal* 1, no. 2: 116–145.

Martins, T. G., D. Simpson, F. Lindgren, and H. Rue. 2013. "Bayesian Computing With INLA: New Features." *Computational Statistics & Data Analysis* 67: 68–83.

McCullagh, P. 1983. "Quasi-Likelihood Functions." *Annals of Statistics* 11, no. 1: 59–67.

McDonald, G. C. 2009. "Ridge Regression." *Wiley Interdisciplinary Reviews: Computational Statistics* 1, no. 1: 93–100.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. "Equation of State Calculations by Fast Computing Machines." *Journal of Chemical Physics* 21, no. 6: 1087–1092.

Minka, T. 2004. *Power EP. Technical Report*. Cambridge: Microsoft Research.

Minka, T. P. 2001. "Expectation Propagation for Approximate Bayesian Inference." In *Proceedings of the Seventeenth Conference on Uncertainty*

in Artificial Intelligence, edited by J. Breese and D. Koller, 362–369. San Francisco: Morgan Kaufmann Publishers Inc.

Moustaki, I. 1996. "A Latent Trait and a Latent Class Model for Mixed Observed Variables." *British Journal of Mathematical and Statistical Psychology* 49: 313–334.

Moustaki, I., and M. Knott. 2000. "Generalized Latent Trait Models." *Psychometrika* 65: 391–411.

Moustaki, I., and M.-P. Victoria-Feser. 2006. "Bounded-Influence Robust Estimation in Generalized Linear Latent Variable Models." *Journal of the American Statistical Association* 101, no. 474: 644–653.

Muehlmann, C., S. De Iaco, and K. Nordhausen. 2023. "Blind Recovery of Sources for Multivariate Space-Time Random Fields." *Stochastic Environmental Research and Risk Assessment* 37: 1593–1613.

Mulaik, S. 2009. *Foundations of Factor Analysis*. 2nd ed. Boca Raton: Chapman and Hall/CRC.

Naylor, J. C., and A. F. M. Smith. 1982. "Applications of a Method for the Efficient Computation of Posterior Distributions." *Journal of the Royal Statistical Society, Series C* 31, no. 3: 214–225.

Needham, T. 1993. "A Visual Explanation of Jensen's Inequality." *American Mathematical Monthly* 100, no. 8: 768–771.

Niku, J., W. Brooks, R. Herliansyah, et al. 2023. "gllvm: Generalized Linear Latent Variable Models." R Package Version 1.4.3. https://CRAN.R-project.org/package=gllvm.

Niku, J., W. Brooks, R. Herliansyah, F. K. C. Hui, S. Taskinen, and D. I. Warton. 2019. "Efficient Estimation of Generalized Linear Latent Variable Models." *PLoS One* 14, no. 5: e0216129.

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, no. 12: 2173–2182.

Niku, J., D. I. Warton, F. K. C. Hui, and S. Taskinen. 2017. "Generalized Linear Latent Variable Models for Multivariate Count and Biomass Data in Ecology." *Journal of Agricultural, Biological, and Environmental Statistics* 22, no. 4: 498–522.

Nordhausen, K., and H. Oja. 2018. "Independent Component Analysis: A Statistical Perspective." *WIREs Computational Statistics* 10: e1440.

Oberski, D. 2016. "Mixture Models: Latent Profile and Latent Class Analysis." In *Modern Statistical Methods for HCI*, edited by J. Robertson and M. Kaptein, 275–287. Germany: Springer.

O'Hara, B., and B. van der Veen. 2024. "LatentINLA: Package to Write GLLVMs in INLA." R Package Version 0.1.1, Commit 23bfe5e78c5ca84f4996ab66ce540a2799b218e1. https://github.com/oharar/LatentINLA.

Ormerod, J., and M. Wand. 2012. "Gaussian Variational Approximate Inference for Generalized Linear Mixed Models." *Journal of Computational and Graphical Statistics* 21: 2–17.

Ormerod, J. T., and M. P. Wand. 2010. "Explaining Variational Approximations." *American Statistician* 64, no. 2: 140–153.

Ovaskainen, O., and N. Abrego. 2020. *Joint Species Distribution Modelling: With Applications in R*. Cambridge: Cambridge University Press.

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: 549–555.

Ovaskainen, O., D. B. Roy, R. Fox, and B. J. Anderson. 2016. "Uncovering Hidden Spatial Structure in Species Communities With Spatially Explicit Joint Species Distribution Models." *Methods in Ecology and Evolution* 7, no. 4: 428–436.

Pan, Y., M. Matilainen, S. Taskinen, and K. Nordhausen. 2021. "A Review of Second-Order Blind Identification Methods." *WIREs Computational Statistics* 14, no. 4: e1550.

Pearson, K. 1901. "LIII. On Lines and Planes of Closest Fit to Systems of Points in Space." *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2, no. 11: 559–572.

Pinheiro, J. C., and D. M. Bates. 1995. "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model." *Journal of Computational and Graphical Statistics* 4, no. 1: 12–35.

Plummer, M. 2003. "Jags: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Austria, 1–10.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2002. "Reliable Estimation of Generalized Linear Mixed Models Using Adaptive Quadrature." *Stata Journal* 2: 1–21.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2005. "Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models With Nested Random Effects." *Journal of Econometrics* 128, no. 2: 301–323.

Raffalovich, L. E., and G. W. Bohrnstedt. 1987. "Common, Specific, and Error Variance Components of Factor Models: Estimation With Longitudinal Data." *Sociological Methods & Research* 15, no. 4: 385–405.

Rall, L. B. 1981. *Automatic Differentiation: Techniques and Applications. Lecture Notes in Computer Science*. Berlin Heidelberg: Springer.

Reed, D. C., and R. J. Miller. 2023. *SBC LTER: Reef: Kelp Forest Community Dynamics: Cover of Sessile Organisms, Uniform Point Contact*. LTER Network Member Node. https://doi.org/10.6073/pasta/0af1a5b0d9dde5b4e5915c0012ccf99c.

Rizopoulos, D. 2006. "ltm: An R Package for Latent Variable Modelling and Item Response Theory Analyses." *Journal of Statistical Software* 17: 1–25.

Rosseel, Y. 2012. "Lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* 48, no. 2: 1–36.

Rue, H., and L. Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. New York: Chapman and Hall/CRC.

Rue, H., S. Martino, and N. Chopin. 2009. "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations." *Journal of the Royal Statistical Society: Series B* 71, no. 2: 319–392.

Sammel, M. D., L. M. Ryan, and J. M. Legler. 1997. "Latent Variable Models for Mixed Discrete and Continuous Outcomes." *Journal of the Royal Statistical Society: Series B* 59: 667–678.

Searle, S., G. Casella, and C. McCulloch. 2009. *Variance Components*. New York: Wiley.

Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton: Chapman & Hall.

Smallman, L., A. Artemiou, and J. Morgan. 2018. "Sparse Generalised Principal Component Analysis." *Pattern Recognition* 83: 443–455.

Spearman, C. 1904. "'General Intelligence,' Objectively Determined and Measured." *American Journal of Psychology* 15, no. 2: 201–292.

Stan Development Team. 2024. "RStan: The R Interface to Stan." R Package Version 2.32.6. https://mc-stan.org/.

Stroud, A., and D. Secrest. 1966. *Gaussian Quadrature Formulas*. Englewood Cliffs, NJ: Prentice-Hall.

Thorson, J. T., J. N. Ianelli, E. A. Larsen, et al. 2016. "Joint Dynamic Species Distribution Models: A Tool for Community Ordination and Spatio-Temporal Monitoring." *Global Ecology and Biogeography* 25: 1144–1158.

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: 627–637.

Thurstone, L. L. 1931. "Multiple Factor Analysis." *Psychological Review* 38, no. 5: 406–427.

Tierney, L., and J. B. Kadane. 1986. "Accurate Approximations for Posterior Moments and Marginal Densities." *Journal of the American Statistical Association* 81, no. 393: 82–86.

Tikhonov, G., L. Duan, N. Abrego, et al. 2020. "Computationally Efficient Joint Species Distribution Modeling of Big Spatial Data." *Ecology* 101: e02929.

Tikhonov, G., Ø. Opedal, N. Abrego, A. Lehikoinen, and O. Ovaskainen. 2019. "Joint Species Distribution Modelling With HMSC-R." *Methods in Ecology and Evolution* 11, no. 3: 442–447.

Tikhonov, G., O. Ovaskainen, J. Oksanen, M. de Jonge, O. Opedal, and T. Dallas. 2024. "Hmsc: Hierarchical Model of Species Communities." R Package Version 3.1-2, Commit 932c09cdb59cb1ceacb6467d-d2ce5e4126870218. https://github.com/hmsc-r/HMSC.

Tipping, M. E., and C. M. Bishop. 1999. "Probabilistic Principal Component Analysis." *Journal of the Royal Statistical Society: Series B* 61, no. 3: 611–622.

Turner, B. M., T. Wang, and E. C. Merkle. 2017. "Factor Analysis Linking Functions for Simultaneously Modeling Neural and Behavioral Data." *NeuroImage* 153: 28–48.

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, no. 7: 1288–1300.

Vehtari, A., A. Gelman, T. Sivula, et al. 2020. "Expectation Propagation as a Way of Life: A Framework for Bayesian Inference on Partitioned Data." *Journal of Machine Learning Research* 21, no. 17: 1–53.

Virta, J., B. Li, K. Nordhausen, and H. Oja. 2020. "Independent Component Analysis for Multivariate Functional Data." *Journal of Multivariate Analysis* 176: 104568.

Wang, C., and D. M. Blei. 2013. "Variational Inference in Nonconjugate Models." *Journal of Machine Learning Research* 14, no. 1: 1005–1031.

Wang, F., and M. M. Wall. 2003. "Generalized Common Spatial Factor Model." *Biostatistics* 4: 569–582.

Wang, L., and L. Carvalho. 2024. "Computational Approaches for Exponential-Family Factor Analysis." arXiv Preprint. https://arxiv.org/abs/2403.14925v2.

Warton, D. I., F. G. Blanchet, R. O'Hara, et al. 2015. "So Many Variables: Joint Modeling in Community Ecology." *Trends in Ecology & Evolution* 30: 766–779.

Wedderburn, R. W. M. 1974. "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method." *Biometrika* 61, no. 3: 439–447.

Wedel, M., and W. Kamakura. 2001. "Factor Analysis With (Mixed) Observed and Latent Variables." *Psychometrika* 66: 515–530.

Wei, T., and V. Simko. 2021. "R Package 'Corrplot': Visualization of a Correlation Matrix." Version 0.92. https://github.com/taiyun/corrplot.

Wright, S. 1921. "Correlation and Causation." *Journal of Agricultural Research* 20, no. 7: 557.

Zeng, Y., H. Zhao, and T. Wang. 2021. "Model-Based Microbiome Data Ordination: A Variational Approximation Approach." *Journal of Computational and Graphical Statistics* 30, no. 4: 1036–1048.

Zhang, C., J. Bütepage, H. Kjellström, and S. Mandt. 2019. "Advances in Variational Inference." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, no. 8: 2008–2026.

Zhu, J., J. Eickhoff, and P. Yan. 2005. "Generalized Linear Latent Variable Models for Repeated Measures of Spatially Correlated Multivariate Data." *Biometrics* 61, no. 3: 674–683.