Statistical Modelling 2024; **24(6)**: 581–598

A combined overdispersed longitudinal model for nominal data

Ricardo K. Sercundes[1], Geert Molenberghs[2,3], Geert Verbeke[3,2], Clarice G.B. Demétrio[1], Sila C. da Silva[4], and Rafael A. Moral[5]

[1]Department of Exact Sciences, ESALQ-University of São Paulo, Piracicaba, Brazil
[2]I-BioStat, Universiteit Hasselt, Hasselt, Belgium
[3]I-BioStat, KU Leuven, Leuven, Belgium
[4]Department of Animal Science, ESALQ-University of São Paulo, Piracicaba, Brazil
[5]Department of Mathematics and Statistics, Maynooth University, Maynooth, Ireland

Abstract: Longitudinal studies involving nominal outcomes are carried out in various scientific areas. These outcomes are frequently modelled using a generalized linear mixed modelling (GLMM) framework. This widely used approach allows for the modelling of the hierarchy in the data to accommodate different degrees of overdispersion. In this article, a combined model (CM) that takes into account overdispersion and clustering through two separate sets of random effects is formulated. Maximum likelihood estimation with analytic-numerical integration is used to estimate the model parameters. To examine the relative performance of the CM and the GLMM, simulation studies were carried out, exploring scenarios with different sample sizes, types of random effects, and overdispersion. Both models were applied to a real dataset obtained from an experiment in agriculture. We also provide an implementation of these models through SAS code.

Key words: Multinomial distribution, beta distribution, hierarchical data

Received March 2023; **revised** July 2023; **accepted** August 2023

1 Introduction

Nominal data may arise from studies in many different subject areas, such as medicine, marketing, education, and agriculture. A nominal outcome has its measurement scale formed by a set of categories that have no intrinsic order, being classified as binary, if only two categories are observed (e.g., dead or alive), or polytomous, if three or more categories are observed (e.g., political party affiliation: democrat, republican, or independent). Although polytomous responses are qualitative, all nominal outcomes may be written as a set of binary variables (Agresti, 2010; Hartzel et al., 2001; Clayton, 1992). For cross-sectional data, a whole collection of modelling approaches can be used, such as the generalized linear modelling (GLM) framework based on the exponential family of distributions (Nelder and Wedderburn, 1972).

Address for correspondence: Rafael A. Moral, Department Mathematics and Statistics, Maynooth University, County Kildare, Ireland.
E-mail: rafael.deandrademoral@mu.ie

10.1177/1471082X231209361

One of the key features of the GLM framework is the mean-variance relationship, where the variance is a deterministic function of the mean. For example, for Bernoulli outcomes with success probability $\mu = \pi$, the variance is $v(\mu) = \pi(1 - \pi)$, which may be overly restrictive depending on the data generating process. Scenarios where the variance is larger or smaller than the mean are reported in the literature as over- or under-dispersion, respectively (Grunwald et al., 2011; Demétrio et al., 2014). For purely binary data, however, hierarchies need to be present for the mean-variance relationship to be violated (Molenberghs et al., 2010, 2012, 2017). Therefore, data from studies where several measurements are taken from the same cluster, subject, or sample unit over time (i.e., longitudinal studies) could violate this assumption.

Some of the main approaches used to analyse longitudinal data with nominal outcomes are generalized estimating equations (GEE) (Liang and Zeger, 1986; Lipsitz et al., 1994; Touloumis et al., 2013), transition models (TM) (Diggle et al., 2002; Molenberghs and Verbeke, 2005; Lara et al., 2017) and generalized linear mixed models (GLMM) (Hartzel et al., 2001; Diggle et al., 2002; Hedeker, 2003; Molenberghs and Verbeke, 2005). These widely used approaches allow for accommodating the correlation between observations induced by the hierarchy of the data collection process, as well as extra-variability. Molenberghs et al. (2007), Molenberghs et al. (2010), Molenberghs et al. (2012), Ivanova et al. (2014), and Molenberghs et al. (2017) showed that accommodating either one of overdispersion or hierarchically induced association may fall short of properly modelling the data. Therefore, they proposed a combined modelling framework encompassing both. Molenberghs et al. (2007) focussed on counts, Molenberghs et al. (2010) laid out a general framework, Molenberghs et al. (2012) worked with binary and binomial outcomes, Ivanova et al. (2014) tackled ordinal outcomes, whereas Molenberghs et al. (2017) contributed with a review of all proposed combined models. Here, we propose a combined model (CM) for nominal outcomes to incorporate the hierarchical data collection process, as well as extra-variability by using two different sets of random effects. Note that Lee and Nelder (1996, 2001, 2003) proposed hierarchical generalized linear models, where random effects can be non-normal, and conjugate, as well. Here, we combine these with normal random effects in the linear predictor.

The remainder of this article is organized as follows. In Section 2, a motivating case study, stemming from an agricultural experiment on elephant grass pasture and dairy cows is introduced. Basic elements for our modelling framework, standard generalized linear models, extensions for overdispersion, the generalized linear mixed model, and the combined modelling framework are the subject of Section 3. The proposed combined model (CM) is described in Section 4, while parameter estimation is the focus of Section 5. A simulation study comparing CM and GLMM is described and results presented in Section 6, while the case study is analysed in Section 7. We offer concluding remarks in Section 8. Finally, we provide the algebraic development in Appendix A and how to implement these models in SAS as Supplementary Materials.

2 Grazing management dataset

This dataset was collected from an experiment on elephant grass pastures (*Pennisetum purpureum* Schum. cv. Napier) grazed by dairy cows (Pereira et al., 2015a,b). It was set up in a randomized complete block design with the treatments allocated according to a 2×2 factorial arrangement,

Table 1 First and last four of the 3, 800 rows in the grazing management dataset.

Seasons	Blocks	Pre-grazing	Post-grazing	Point within paddock	Outcome*
Summer 1	1	maximum	35	1	3
Autumn	1	maximum	35	1	1
Winter	1	maximum	35	1	2
Early Spring	1	maximum	35	1	3
⋮	⋮	⋮	⋮	⋮	⋮
Winter	4	95	45	640	1
Early Spring	4	95	45	640	3
Late Spring	4	95	45	640	3
Summer 2	4	95	45	640	3

*where weed = 1, bare ground = 2 and tussock = 3

where treatments are the combinations of two pre-grazing conditions (95% and maximum canopy light interception during regrowth) and two post-grazing heights (35 and 45 cm). The experiment was carried out from January 2011 until April 2012, encompassing six seasons: 'Summer 1' (Jan–Mar 2011), 'Autumn' (Apr–June 2011), 'Winter' (July–Sept 2011), 'early Spring' (Oct–mid-Nov 2011), 'Late spring' (mid-Nov–Dec 2011) and 'Summer 2' (Jan–Apr 2012).

The response variable is the type of vegetation observed in the field, with three categories: 'weeds', 'bare ground', or 'tussocks'. Forty points were observed in each one of the four paddocks in each block. The data consists of the total number of points where each category was observed, characterising a multinomial outcome with three levels. There are $40 \times 16 = 640$ points per season, but in the early spring, one of the paddocks was affected by frost damage and thus the total number of observations was $640 \times 6 - 40 = 3800$. A sample of the dataset and a sketch of the experiment are show in Table 1 and Figure 1, respectively.

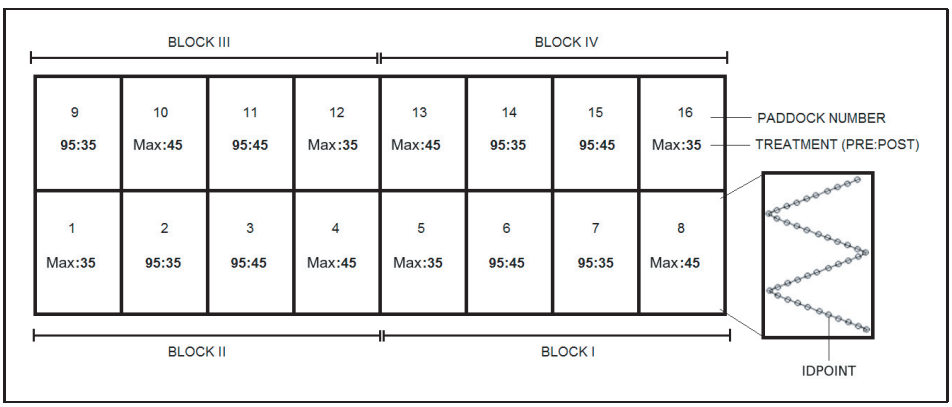

Figure 1 Sketch of the design of the grazing management experiment. Each of the four blocks consists of four paddocks, each one with a combination of the levels of two treatment factors. Forty points were observed within each paddock.

3 Building blocks

Here, we briefly present the main concepts to formulate the combined model for nominal outcomes. In Section 3.1, we introduce the exponential family, generalized linear models and overdispersion. In Section 3.2, we present some properties of generalized linear mixed models and the general framework of combined models.

3.1 Generalized linear models and overdispersion

The class of generalized linear models (GLM) was introduced by Nelder and Wedderburn (1972) as a framework for handling a range of common statistical models for Gaussian and non-Gaussian data. A GLM is defined in terms of three components.

The first component is a set of independent random variables, $Y_1, \ldots, Y_N$, with probability or density function that belongs to the exponential family:

$$f(y_i|\eta_i, \phi) \;=\; \exp\left\{\phi^{-1}[y_i\eta_i - \psi(\eta_i)] + c(y_i, \phi)\right\}, \tag{3.1}$$

where $\psi(\cdot)$ and $c(\cdot)$ are known functions and ϕ and η_i are called dispersion and natural or canonical parameter, respectively. The exponential family includes several distributions, such as the Gaussian, Bernoulli, binomial, Poisson, gamma and multinomial distributions. The first two moments of a distribution that belongs to the exponential family are given by

$$\mathrm{E}(Y_i) = \mu_i = \psi'(\eta_i) \quad \text{and} \quad \mathrm{Var}(Y_i) = \sigma^2 = \phi\psi''(\eta_i).$$

Thus, the mean and variance of these distributions are related through

$$\sigma^2 = \phi\psi''[\psi^{'-1}(\mu)] = \phi v(\mu),$$

with $v(\cdot)$ termed the variance function.

The second component, called linear predictor or natural parameter, η_i, is the quantity that incorporates the information about the independent variables into the model. The third component, called the link function, $h(\cdot)$, provides the relationship between the linear predictor and the mean of the distribution as $\mu_i = h(\eta_i) = h(\boldsymbol{x}_i^T\boldsymbol{\beta})$, where $\boldsymbol{x}_i$ and $\boldsymbol{\beta}$ are covariate vectors and fixed unknown regression coefficients, respectively.

The multinomial distribution is a natural starting point for analysis of polytomous outcomes. This distribution arises as a natural extension of the binomial distribution when each independent trial has more than two possible mutually exclusive outcomes. Consider a series of m independent trials of an experiment, each resulting in one of R mutually exclusive events $\mathrm{E}_1, \ldots, \mathrm{E}_R$. In each replicate within the experiment, the probability of the occurrence of event E_r is equal to π_r with $\sum_{r=1}^{R}\pi_r = 1$. Let $\boldsymbol{Y}^* = (Y_1, \ldots, Y_R)^T$ denote the random vector of the number of occurrences of events $\mathrm{E}_1, \ldots, \mathrm{E}_R$ out of m trials, with $\sum_{r=1}^{R} Y_r = m$. Let $\boldsymbol{y}^* = (y_1, \ldots, y_R)^T$ represent a realization

of $\boldsymbol{Y}^*$, $\sum_{r=1}^{R} y_r = m$. Then, the random vector $\boldsymbol{Y}^*$ is said to have a multinomial distribution with parameters m, $\boldsymbol{\pi}^* = (\pi_1, \ldots, \pi_R)^T$, and joint probability mass function given by

$$P(\boldsymbol{Y}^* = \boldsymbol{y}^*) = \frac{m!}{y_1! y_2! \ldots y_R!} \pi_1^{y_1} \pi_2^{y_2} \ldots \pi_R^{y_R} \quad \pi_i \in [0, 1], \quad i = 1, \ldots, R. \tag{3.2}$$

We can safely reduce the dimensionality of $\boldsymbol{Y}^*$ and $\boldsymbol{\pi}^*$ by removing their respective last elements, since they can be obtained from the remaining $R - 1$ categories. Then, define $\boldsymbol{Y} = (Y_1, \ldots, Y_{R-1})^T$, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{R-1})^T$ and the realization of $\boldsymbol{Y}$ as $\boldsymbol{y} = (y_1, \ldots, y_{R-1})$. Without loss of generality, $\boldsymbol{Y}$ follows a multinomial distribution with parameters m and $\boldsymbol{\pi}$ with joint probability function as in (3.2) with $y_R = m - \sum_{h=1}^{R-1} y_h$ and $\pi_R = 1 - \sum_{h=1}^{R-1} \pi_h$. Hence, the mean and the variance of $\boldsymbol{Y}$ are, respectively,

$$\mathrm{E}(\boldsymbol{Y}) = m\boldsymbol{\pi} \quad \text{and} \quad \mathrm{Var}(\boldsymbol{Y}) = m\Delta(\boldsymbol{\pi}), \tag{3.3}$$

where $\Delta(\boldsymbol{\pi}) = \mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T$. Note that $\Delta(\boldsymbol{\pi})$ is an $R \times R$ variance-covariance matrix of full rank where the diagonal elements are $\pi_r(1 - \pi_r)$ and off-diagonal elements $-\pi_r \pi_{r'}$ for $r \neq r'$.

It is well known that (3.3) implies a restrictive variance function. According to Grunwald et al. (2011) and Demétrio et al. (2014), cases where the variance is greater than the mean are largely reported in the literature as overdispersion, which may occur due to the absence of relevant covariates, heterogeneity of sampling units, correlation induced by hierarchical structures and/or excess of zeros. It should be noted that underdispersion can occur as well (for instance, see Ribeiro Jr et al. (2020) and Morris and Sellers (2022) for extended approaches recently developed that can be used to accommodate underdispersion). Thus, it is important to adapt models to take into account deviations from assumed mean-variance relationships in order to avoid incorrect inferences (Hinde and Demétrio, 1998).

One possibility to extend the multinomial model to handle overdispersion is to multiply the multinomial covariance matrix by a constant scalar parameter. A quasi-likelihood approach using a scale adjustment was presented by McCullagh and Nelder (1983), and later extended by Morel and Koehler (1995) to allow for different levels of overdispersion for each category using a diagonal matrix of overdispersion parameters and a Cholesky decomposition of the multinomial variance-covariance matrix. A mixture of distributions can also be used to allow for overdispersion, such as the random-clumped multinomial distribution proposed by Morel and Nagaraj (1993) and Neerchal and Morel (1998). This model is a finite mixture of multinomial distributions that captures the extra-variability caused by clumped sampling. Another convenient route to take overdispersion into account is through a two-stage approach, which considers a probability distribution for a model parameter, yielding a mixture. For instance, a multinomial model where the parameter $\boldsymbol{\pi}$ follows a Dirichlet distribution is called the Dirichlet-multinomial model (Mosimann, 1962). Although the estimation of a dispersion parameter might provide some flexibility to standard GLMs, this is not always sufficient, especially when hierarchical structures or highly variable data arise.

3.2 Generalized linear mixed models and combined models

When analysing non-Gaussian data that are hierarchically organized (repeated measures or clustering, for example), the generalized linear mixed model (GLMM) is a popular choice (Molen-

berghs and Verbeke, 2005; Diggle et al., 2002). In full generality, one assumes that, conditionally on q-dimensional random effects $\boldsymbol{b}_i$, assumed to be drawn independently from a normal distribution, $N_q(\mathbf{0}, \boldsymbol{D})$, the outcome Y_{ij} measured on the i-th subject or sample unit at the j-th time point ($i = 1, \ldots, N$; $j = 1, \ldots, n_i$) are independent with densities of the form:

$$f_i(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi) = \exp\left\{\phi^{-1}[y_{ij}\eta_{ij} - \psi(\eta_{ij})] + c(y_{ij}, \phi)\right\}, \tag{3.4}$$

where $\eta_{ij} = \eta(\mu_{ij}) = \eta[\mathrm{E}(Y_{ij}|\boldsymbol{b}_i)] = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i$ is the canonical parameter, with $\boldsymbol{x}_{ij}$ ($\boldsymbol{z}_{ij}$) the design vector for the fixed (random) effects. Finally, let $f(\boldsymbol{b}_i|\boldsymbol{D})$ be the density of the Gaussian distribution, $N(\mathbf{0}, \boldsymbol{D})$, for the random effects $\boldsymbol{b}_i$.

For nominal data, it is assumed that the outcome Y_{ij} can take values $r = 1, \ldots, R$. Without loss of generality, we can replace it with a set of R dummy variables where $W_{r,ij}$ is equal to 1 if $Y_{ij} = r$ and 0 otherwise. Evidently, there are redundant dummies, but any subset of $R-1$ components is not, as described in Section 3.1. Thus, $\boldsymbol{W}_{ij} \sim$ multinomial($\boldsymbol{\pi}_{ij}$) with probabilities $\boldsymbol{\pi}_{ij} = (\pi_{1,ij}, \ldots, \pi_{r,ij}, \ldots, \pi_{R,ij})^T$. Assuming that category R is the reference category, a baseline-category logit model (Agresti, 2010; Hartzel et al., 2001) can be written as

$$\ln\left(\frac{\pi_{r,ij}}{\pi_{R,ij}}\right) = \eta_{r,ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_r + \boldsymbol{z}_{ij}^T\boldsymbol{b}_{r,i}, \quad r = 1, \ldots, R-1,$$
$$\boldsymbol{b}_{r,i} \sim N(\mathbf{0}, \boldsymbol{D}),$$

where $\boldsymbol{\beta}_r$ is the fixed-effects coefficient vector of length $p+1$, corresponding to an intercept and p covariates, and $\boldsymbol{b}_{r,i}$ is the random-effects vector, following a multivariate normal distribution. The probabilities of each category for the i-th subject and j-th time can be expressed as

$$\pi_{r,ij} = \begin{cases} \dfrac{\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_r + \boldsymbol{z}_{ij}^T\boldsymbol{b}_{r,i})}{1 + \sum_{h=1}^{R-1}\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_h + \boldsymbol{z}_{ij}^T\boldsymbol{b}_{h,i})} & \text{if } 1 \le r \le R-1, \\ 1 - \sum_{h=1}^{R-1}\pi_{h,ij} & \text{if } r = R. \end{cases}$$

Estimates of $\boldsymbol{\beta}$, $\boldsymbol{D}$ and ϕ for GLMMs are obtained by maximizing the marginal likelihood, computed by integrating out the random effects and commonly written as:

$$\mathrm{L}(\boldsymbol{\beta}, \boldsymbol{D}, \phi) = \prod_{i=1}^{N}\int_{-\infty}^{\infty}\prod_{j=1}^{n_i} f_i(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi) f(\boldsymbol{b}_i|\boldsymbol{D}) d\boldsymbol{b}_i. \tag{3.5}$$

The key problem in maximizing (3.5) is the presence of N integrals over the random effects. Here, numerical methods are needed, such as adaptive Gaussian quadrature (Molenberghs and Verbeke, 2005; Pinheiro and Bates, 1995).

While GLMMs, defined to accommodate within-unit correlation, also capture overdispersion to some extent, a single set of random effects may be insufficient to flexibly capture both. This led Molenberghs et al. (2007) to formulate a flexible and unified modelling framework, which they

termed the combined model. These authors brought together two sets of random effects: the normally distributed subject-specific random effects to capture correlation and a certain amount of overdispersion, and a conjugate measurement-specific random effect on the natural parameter scale to accommodate the remaining overdispersion. Integrating out these two sets of random effects and using the generalized linear model framework, the following general family is introduced:

$$f_i(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \theta_{ij}, \phi) = \exp\left\{\phi^{-1}[y_{ij}\lambda_{ij} - \psi(\lambda_{ij})] + c(y_{ij}, \phi)\right\}, \tag{3.6}$$

with notation similar to the one used in (3.4), but now with conditional mean

$$\mathrm{E}(Y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \theta_{ij}) = \mu^c_{ij} = \theta_{ij}\kappa_{ij}, \tag{3.7}$$

where $\theta_{ij} \sim \mathcal{G}_{ij}(\vartheta_{ij}, \xi_{ij})$ is a conjugate random variable and $\kappa_{ij} = g(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i)$, where g is the inverse of the canonical link function (Molenberghs et al., 2010). Note that we use $g(\cdot)$ rather than $h(\cdot)$, because this transformation applies to only a part of the mean function. Finally, as before, $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$. Unlike in GLMM, we now have two different notations, η_{ij} and λ_{ij}, to refer to the linear predictor and the natural parameter, respectively (i.e., λ_{ij} encompasses θ_{ij}, while η_{ij} refers to the 'GLMM part' only; Molenberghs et al., 2010). Regarding θ_{ij}, three assumptions can be made: (a) they are independent; (b) they are correlated, implying that the collection of univariate distributions $\mathcal{G}_{ij}(\vartheta_{ij}, \xi_{ij})$ needs to be replaced with a multivariate one; and (c) they are shared, in the sense that there is only one realization per cluster, useful in applications with exchangeable outcomes. Assumption (c) is the one adopted in the analysis of the examples. Obviously, parameterization (3.7) allows for random effects θ_{ij} capturing overdispersion, while formulated directly at the mean scale. Opting for a common conjugate random effect for all logits ensures that it can be interpreted as a single overdispersion parameter.

The relationship between the mean and the natural parameter is now given by the function h:

$$\lambda_{ij} = h^{-1}(\mu^c_{ij}) = h^{-1}(\theta_{ij}\kappa_{ij}),$$

We can still apply standard GLM ideas to derive the mean and variance, combined with iterated expectation-based calculations. For the mean, if θ_{ij} and $\boldsymbol{b}_i$ are independent, it follows that

$$\mathrm{E}(Y_{ij}) = \mathrm{E}(\theta_{ij})\mathrm{E}(\kappa_{ij}) = \mathrm{E}[h(\lambda_{ij})].$$

Molenberghs et al. (2010) and Molenberghs et al. (2017) derived explicit expressions for the means, variances, and marginal densities for a number of outcome types, such as Gaussian, Poisson, and time-to-event data. This is not possible for binary data modelled with a logit link and including Gaussian random effects, whether or not other random effects are present.

4 Combined model for nominal outcomes

Analogously to Ivanova et al. (2014), we use a baseline-category logit structure (3.5) and include Gaussian random effects $\boldsymbol{b}_{r,i} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{D})$ in the linear predictor, as well as beta random effects $\theta_{ij} \sim$

Beta(ϑ, ξ) to capture overdispersion (considering θ_{ij} and $\boldsymbol{b}_{r,i}$ independent). We may then write the probabilities of the proposed combined model as:

$$\pi_{r,ij} = \begin{cases} \theta_{ij}\kappa_{r,ij} & \text{if } 1 \leq r \leq R-1, \\ 1 - \sum_{h=1}^{R-1} \theta_{ij}\kappa_{h,ij} & \text{if } r = R, \end{cases}$$

and

$$\kappa_{r,ij} = \frac{\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_{\boldsymbol{r}} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_{r,i})}{1 + \sum_{h=1}^{R-1}\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_{\boldsymbol{h}} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_{h,i})} \text{ if } 1 \leq r \leq R-1, \tag{4.1}$$

where $\boldsymbol{\beta}_{\boldsymbol{r}}$ is the vector of fixed effects (regression coefficients) for each one of the $(R-1)$ categories, and $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are the design vectors for the fixed and random effects, respectively. We considered here the case where θ_{ij} is constant across all categories, resulting in a combined model with constant overdispersion (although one may allow θ_{ij} to depend on covariates through an appropriate link function).

5 Parameter estimation

Molenberghs et al. (2007) and Molenberghs et al. (2010) showed that fitting the combined model is relatively easy, and that standard software tools can be used for maximum likelihood estimation in this case. A priori, fitting a combined model of the type described in Section 4 is done by maximizing the log-likelihood while integrating over the random effects. The joint distribution of the ij-th observation, assuming θ_{ij} and $\boldsymbol{b}_{r,i}$ independent, is given by

$$f(w_{ij}, \boldsymbol{b}_{r,i}, \theta_{ij}) = f(w_{ij}|\boldsymbol{b}_{r,i}, \theta_{ij})\ f(\boldsymbol{b}_{r,i})\ f(\theta_{ij}),$$

and the likelihood function can be written as:

$$L(\boldsymbol{\beta},\ \boldsymbol{D}, \vartheta, \xi) = \prod_{i=1}^{N} \int \int \prod_{j=1}^{n_i} f(w_{ij}|\boldsymbol{\beta},\ \boldsymbol{b}_{r,i}, \theta_{ij})\ f(\boldsymbol{b}_{r,i}|\ \boldsymbol{D})\ f(\theta_{ij}|\vartheta, \xi)\ d\boldsymbol{b}_{r,i}\ d\theta_{ij}.$$

For our proposed model, the three functions in the integrand are, in order, the multinomial, normal and beta distributions probability density or mass functions, which yields:

$$L(\boldsymbol{\beta},\ \boldsymbol{D}, \vartheta, \xi) = \prod_{i=1}^{N} \int \int \prod_{j=1}^{n_i} \prod_{h=1}^{R-1} \left(\theta_{ij}\kappa_{h,ij}\right)^{w_{h,ij}} \left(1 - \sum_{h=1}^{R-1} \theta_{ij}\kappa_{h,ij}\right)^{1-\sum_{h=1}^{R-1} w_{h,ij}} \tag{5.1}$$

$$\frac{1}{\sqrt{(2\pi)^{n_i}}} \frac{1}{\sqrt{|\boldsymbol{D}|}} \exp\left(-\frac{1}{2}\boldsymbol{b}_{r,i}^T\boldsymbol{D}^{-1}\boldsymbol{b}_{r,i}\right) \frac{\theta_{ij}^{\vartheta-1}\left(1-\theta_{ij}\right)^{\xi-1}}{B(\vartheta, \xi)}\ d\boldsymbol{b}_{r,i}\ d\theta_{ij}.$$

The key problem in maximizing (5.1) is the presence of N integrals over the random effects $\boldsymbol{b}_{r,i}$ and θ_{ij}, making this process time consuming and cumbersome to implement if the N integrals are to be calculated using numerical methods. However, we can simplify by integrating analytically over the beta random effects, but not over the normal random effects, leading to a partially marginalized density. In our case, this takes the form (details of calculations are presented in Appendix A):

$$L(\boldsymbol{\beta},\, \boldsymbol{D}, \vartheta, \xi) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} \prod_{h=1}^{R-1} \left(\frac{\vartheta}{\vartheta + \xi}\, \kappa_{h,ij} \right)^{w_{h,ij}} \left(1 - \frac{\vartheta}{\vartheta + \xi} \sum_{h=1}^{R-1} \kappa_{h,ij} \right)^{1-\sum_{h=1}^{R-1} w_{h,ij}}$$
$$\frac{1}{\sqrt{(2\pi)^{n_i}}} \frac{1}{\sqrt{|\boldsymbol{D}|}} \exp\left(-\frac{1}{2} \boldsymbol{b}_{r,i}^T \boldsymbol{D}^{-1} \boldsymbol{b}_{r,i} \right)\, d\boldsymbol{b}_{r,i}.$$

Here, a generic maximum likelihood routine that allows for integration over normal random effects can be used. We follow this route and use the SAS procedure NLMIXED. We opted for the adaptive Gaussian quadrature method (Molenberghs and Verbeke, 2005) and chose the number of quadrature points Q by performing a numerical sensitivity analysis to check whether it was sufficiently large. To ensure identifiability, a constraint needs to be applied. Here, we reparameterise ϑ as $e^{\delta} > 0$ and fix $\xi = 1$. Therefore, larger δ values correspond to weaker overdispersion.

6 Simulation

A simulation study was conducted to compare the performance of a GLMM and the proposed combined model. We simulated longitudinal nominal data with $R = 3$ categories considering a baseline-category logit model, (3.5), and the following linear predictor:

$$\eta_{r,ij} = b_{r,i} + \beta_{r,0} + \beta_{r,1}\text{time}_{ij} + \beta_{r,2}\text{group}_i + \beta_{r,3}\text{time}_{ij} * \text{group}_i,$$

where $\text{time}_{ij} = (j-1)/6$ for $j = 1, \ldots, 6$ and $\text{group}_i = 0$ or 1. The true parameter values were set as:

$$\boldsymbol{\beta}_1 = (\beta_{1,0}; \beta_{1,1}; \beta_{1,2}; \beta_{1,3})^T = (0.1; 0.2; 0.5; 0.7)^T,$$
$$\boldsymbol{\beta}_2 = (\beta_{2,0}; \beta_{2,1}; \beta_{2,2}; \beta_{2,3})^T = (0.2; 0.1; 0.4; 0.5)^T,$$

Finally, the random effects were specified as:

$$\boldsymbol{b}_{r,i} \sim N(\boldsymbol{0},\, \boldsymbol{D}),$$
$$\boldsymbol{D} = \begin{pmatrix} d_1 & c \\ c & d_2 \end{pmatrix},$$

with $c = 0.5$, $d_1 \in \{1, 9\}$ and $d_2 \in \{0.5, 4.5\}$, used to generate 'weak' and 'strong' (nine times higher) correlation values for the simulated datasets.

Table 2 True parameter values used to specify six different scenarios (S1 – S6) used in the simulation study.

Scenario	Size	Variance components	Overdispersion
S1	$N = 300$ or 600	$d_1 = 1.0$, $d_2 = 0.5$	-
S2		$d_1 = 9.0$, $d_2 = 4.5$	-
S3		$d_1 = 1.0$, $d_2 = 0.5$	$\vartheta = 5$, $\xi = 1$
S4		$d_1 = 1.0$, $d_2 = 0.5$	$\vartheta = 20$, $\xi = 1$
S5		$d_1 = 9.0$, $d_2 = 4.5$	$\vartheta = 5$, $\xi = 1$
S6		$d_1 = 9.0$, $d_2 = 4.5$	$\vartheta = 20$, $\xi = 1$

Table 3 Descriptive statistics of the 200 simulated overdispersion values.

Beta parameters	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
$\vartheta = 5$ and $\xi = 1$	0.06	0.76	0.87	0.83	0 94	1
$\vartheta = 20$ and $\xi = 1$	0.48	0.93	0.97	0.95	0.98	1

We simulated 200 datasets with $N = 300$ and $N = 600$, and $n_i = 6, \forall i$, with two groups of equal size each, that is, $N = 150$ and $N = 300$ experimental/observational units per group, respectively. Six scenarios with different magnitudes of random effects and overdispersion were generated to compare the behavior of the GLMM and the CM (Table 2).

To generate data with overdispersion, the simulated probabilities were multiplied by values generated from a beta distribution with the shape parameters specified to 'strongly' ($\vartheta = 5$) or 'weakly' ($\vartheta = 20$) disturb the probabilities generated by the model, while fixing $\xi = 1$ (Table 3).

The GLMM and CM were fitted to the simulated datasets using the estimation method described in Section 5, which was implemented using SAS procedure NLMIXED. We approximated the integrals using adaptive Gaussian quadrature with 10 quadrature points, and optimized the log-likelihood using the quasi-Newton BFGS method. We used, as starting values for the fixed effects, the estimates obtained by fitting a GLM without random effects. For each scenario, we computed the average estimate (AE), bias, and mean squared error (MSE) for each parameter. We set $\vartheta = e^{\delta}$ and $\xi = 1$ for the combined model to ensure identifiability.

In general, the simulation results indicate the estimation procedure produced reliable results, showing that the MSEs of the maximum likelihood estimators of the parameters decay towards zero as the sample size increases, as expected under standard asymptotic theory. For scenarios S1, S2, and S4 (weak overdispersion and/or correlation), the results for both the GLMM and CM are similar (see Table 4, which displays results for scenario S2).

However, if there is a pronounced overdispersion effect (S3) or if it coincides with high correlations (S5 and S6), better performances were observed for the CM, mainly for the variance components (Table 5 displays results for scenario S6). Even with a larger sample size, the predicted random effects for the CM showed smaller bias and MSE when compared to the GLMM.

We present the convergence rates for the two models when 200 datasets were simulated in Table 6. The proportion of datasets for which convergence was achieved is a little bit smaller in the CM than in the GLMM. This can be attributed to sensitivity to the starting values, which points to the need for careful selection. When convergence problems arise, we suggest to start the analysis with the GLM or GLMM estimates and, if necessary, to attempt to fit the CM using different sets of starting values.

Table 4 Average estimate (AE), bias and mean square error (MSE) for the parameters estimated by the GLMM and CM based on 200 simulations for scenario S2.

		GLMM					
		$N = 300$			$N = 600$		
Parameter	True	AE	Bias	MSE	AE	Bias	MSE
$\beta_{1,0}$	0.1	0.123	0.023	0.118	0.112	0.012	0.057
$\beta_{1,1}$	0.2	0.196	−0.004	0.185	0.219	0.019	0.076
$\beta_{1,2}$	0.5	0.524	0.024	0.242	0.463	−0.037	0.127
$\beta_{1,3}$	0.7	0.689	−0.011	0.351	0.692	−0.008	0.153
$\beta_{2,0}$	0.2	0.233	0.033	0.086	0.193	−0.007	0.039
$\beta_{2,1}$	0.1	0.072	−0.028	0.144	0.149	0.049	0.061
$\beta_{2,2}$	0.4	0.402	0.002	0.130	0.428	0.028	0.077
$\beta_{2,3}$	0.5	0.478	−0.022	0.297	0.458	−0.042	0.151
d_1	9.0	9.220	0.220	2.921	9.055	0.055	1.303
d_2	4.5	4.558	0.058	0.682	4.462	−0.038	0.307
c	0.5	0.616	0.116	0.635	0.570	0.070	0.341
		CM					
		$N = 300$			$N = 600$		
Parameter	True	AE	Bias	MSE	AE	Bias	MSE
$\beta_{1,0}$	0.1	0.136	0.036	0.128	0.119	0.019	0.059
$\beta_{1,1}$	0.2	0.195	−0.005	0.187	0.221	0.021	0.077
$\beta_{1,2}$	0.5	0.572	0.027	0.244	0.466	−0.034	0.127
$\beta_{1,3}$	0.7	0.694	−0.006	0.355	0.694	−0.006	0.155
$\beta_{2,0}$	0.2	0.245	0.045	0.090	0.200	0.001	0.039
$\beta_{2,1}$	0.1	0.072	−0.028	0.145	0.150	0.050	0.062
$\beta_{2,2}$	0.4	0.405	0.005	0.132	0.431	0.031	0.078
$\beta_{2,3}$	0.5	0.483	−0.017	0.300	0.459	−0.041	0.153
d_1	9.0	9.288	0.288	2.747	9.110	0.110	1.364
d_2	4.5	4.589	0.089	0.687	4.481	−0.019	0.323
c	0.5	0.654	0.154	0.650	0.600	0.100	0.346
δ	–	9.482	–	–	9.188	–	–

For a particular application that a researcher envisages, it might be useful to conduct a targeted simulation study to assess the convergence rate for a sample size that is envisaged.

7 Analysis of the grazing management data

Here, we present an analysis of the grazing management data, introduced in Section 3.2. This dataset has been previously analysed by Menarin and Lara (2017), through extended generalized estimating equations that use local odds ratios to explain the dependence among the categories (Touloumis et al., 2013). Here, emphasis was placed on a subject-specific interpretation. Let $Y_{ij} = 1, 2, 3$ be the types of vegetation (weed, bare ground, and tussock, respectively) that target

Table 5 Average estimates (AE), bias and mean square errors (MSE) for the parameters estimated by the GLMM and CM based on 200 simulations for scenario S6.

		GLMM					
		$N = 300$			$N = 600$		
Parameter	True	AE	Bias	MSE	AE	Bias	MSE
$\beta_{1,0}$	0.1	−0.251	−0.351	0.215	−0.276	−0.376	0.182
$\beta_{1,1}$	0.2	0.149	−0.051	0.130	0.180	−0.020	0.072
$\beta_{1,2}$	0.5	0.380	−0.120	0.188	0.376	−0.124	0.105
$\beta_{1,3}$	0.7	0.524	−0.176	0.310	0.501	−0.199	0.201
$\beta_{2,0}$	0.2	−0.207	−0.407	0.228	−0.228	−0.428	0.220
$\beta_{2,1}$	0.1	0.068	−0.032	0.128	0.092	−0.008	0.064
$\beta_{2,2}$	0.4	0.293	−0.107	0.145	0.335	−0.065	0.064
$\beta_{2,3}$	0.5	0.353	−0.147	0.282	0.339	−0.161	0.156
d_1	9.0	6.297	−2.703	8.722	6.294	−2.706	7.916
d_2	4.5	3.509	−0.991	1.419	3.517	−0.983	1.153
c	0.5	−0.622	−1.122	1.559	−0.639	−1.139	1.455
		CM					
		$N = 300$			$N = 600$		
Parameter	True	AE	Bias	MSE	AE	Bias	MSE
$\beta_{1,0}$	0.1	0.115	0.015	0.169	0.058	−0.042	0.074
$\beta_{1,1}$	0.2	0.186	0.014	0.147	0.215	0.015	0.070
$\beta_{1,2}$	0.5	0.507	0.007	0.178	0.474	−0.026	0.107
$\beta_{1,3}$	0.7	0.688	−0.012	0.329	0.669	−0.031	0.199
$\beta_{2,0}$	0.2	0.201	0.001	0.117	0.141	−0.059	0.068
$\beta_{2,1}$	0.1	0.091	−0.009	0.122	0.117	0.017	0.054
$\beta_{2,2}$	0.4	0.398	−0.002	0.154	0.419	0.019	0.070
$\beta_{2,3}$	0.5	0.495	−0.005	0.284	0.483	−0.017	0.157
d_1	9.0	8.966	−0.034	4.965	8.636	−0.364	2.158
d_2	4.5	4.435	−0.065	1.056	4.344	−0.156	0.466
c	0.5	0.547	0.047	1.137	0.370	−0.130	0.505
δ	–	3.153	–	–	3.172	–	–

Table 6 Convergence rates for the GLMM and the CM in six simulated scenarios with 200 datasets.

			Scenario					
	Model	Size	S1	S2	S3	S4	S5	S6
Rate (%)	GLMM	300	98.5	100.0	97.3	98.5	100.0	100.0
		600	100.0	100.0	99.5	100.0	100.0	100.0
	CM	300	97.5	100.0	91.0	95.5	100.0	100.0
		600	100.0	100.0	99.0	100.0	100.0	99.5

point i, ($i = 1, \ldots, 640$), reached at season j, ($j = 1, \ldots, 6$). Thus, under a baseline-category logit model, (3.5), the GLMM and the CM can be written as:

$$\text{logit 1: } \ln\left(\frac{\pi_{1,ij}}{\pi_{3,ij}}\right), \quad \text{logit 2: } \ln\left(\frac{\pi_{2,ij}}{\pi_{3,ij}}\right),$$

where $\pi_{r,ij}$ is the probability of the i-th point being classified in the r-th category in season j. The first logit is a log-odds between weeds and tussocks and the second one is the log-odds between bare ground and tussocks. As before, both models were fitted with the SAS procedure NLMIXED using adaptive Gaussian quadrature. We performed a sensitivity analysis by increasing the number of quadrature points up to 5, when the estimates showed stability, and carried out maximization through the quasi-Newton BFGS method. We began with the complete model, including the fixed effects of blocks, pre- and post-grazing conditions, seasons, as well as all two- and three-way interactions between pre- and post-grazing conditions and seasons. We also included a random intercept per point within paddock.

We then performed backwards selection for the fixed effects, by fitting reduced models which did not include higher-order interactions, and carrying out likelihood-ratio tests until the model only included significant interactions and/or main effects. This process yielded the following linear predictor:

$$\begin{aligned}\eta_{r,ij} = b_i + \beta_{r,0} + \overbrace{\beta_{r,1}X_{11i} + \beta_{r,2}X_{12i} + \beta_{r,3}X_{13i}}^{\text{blocks}} + \overbrace{\beta_{r,4}X_{2i}}^{\text{pre-grazing}} \\ + \overbrace{\beta_{r,5}X_{31i} + \beta_{r,6}X_{32i} + \beta_{r,7}X_{33i} + \beta_{r,8}X_{34i} + \beta_{r,9}X_{35i}}^{\text{seasons}} \\ + \overbrace{\beta_{r,10}X_{2i}X_{31i} + \ldots + \beta_{r,14}X_{2i}X_{35i}}^{\text{pre-grazing} \times \text{seasons}},\end{aligned}$$

where $b_i \sim N(0, d)$, and the X variables are dummy covariates for blocks ($X_{11i}, \ldots, X_{13i}$), pre-grazing management (X_{2i}) and seasons ($X_{31i}, \ldots, X_{35i}$). To ensure identifiability, we take the last level of each covariate as reference (Block4 = 0, Pre:maximum = 0 and Summer2 = 0) and for the CM we set $\vartheta = e^{\delta}$ and $\xi = 1$.

The results of both fitted models are presented in Table 7. The estimates are very similar, but there is a reduction in the variance component for the CM, a pronounced value of the overdispersion parameter ($\delta = 1.374$) and also a clear improvement in terms of the log-likelihood. Note that several parameters have shifted a bit in the CM model relative to the GLMM model. This is to be expected to some extent, because the more flexible way in which the CM handles variability and within-unit correlation, combined with the mean-variance link, implies that a change in parameter estimates may occur. Against this background, the shift in the logit 2 intercept is of the same magnitude as that in logit 1, but coincidentally it slides across the zero value.

To compare the models, the likelihood-ratio test was used. The difference between deviances is 11.8; however, since this test is carried out on the boundary of the parametric space, the reference distribution is a mixture of χ^2 distributions (Stram and Lee, 1994; Self and Liang, 1987). To test the hypothesis $H_0 : \theta_{ij} = 1$, the reference distribution is a 50:50 mixture of a χ^2_0 (the degenerate

Table 7 Grazing management data. Parameter estimates (standard errors) from the regression coefficients in the GLMM and CM. Estimation was done by maximum likelihood using numerical integration over the normal and beta random effects, if present.

		GLMM		CM	
Effects	Par.	logit 1	logit 2	logit 1	logit 2
Intercept	$\beta_{r,0}$	−2.826 (0.332)	−0.053 (0.129)	−2.467 (0.392)	0.325 (0.247)
Block 1	$\beta_{r,1}$	0.623 (0.190)	0.117 (0.099)	0.736 (0.198)	0.220 (0.129)
Block 2	$\beta_{r,2}$	0.567 (0.190)	−0.024 (0.098)	0.683 (0.193)	0.032 (0.124)
Block 3	$\beta_{r,3}$	−0.681 (0.252)	0.072 (0.097)	−0.633 (0.248)	0.169 (0.120)
Pre(95%)	$\beta_{r,4}$	0.725 (0.364)	−0.605 (0.168)	0.936 (0.379)	−0.602 (0.209)
Summer1	$\beta_{r,5}$	0.514 (0.370)	−0.810 (0.170)	0.580 (0.390)	−0.677 (0.211)
Autumn	$\beta_{r,6}$	−0.309 (0.444)	−0.345 (0.162)	−0.199 (0.453)	−0.342 (0.205)
Winter	$\beta_{r,7}$	−0.136 (0.426)	−0.364 (0.163)	−0.044 (0.438)	−0.352 (0.206)
Early spring	$\beta_{r,8}$	0.308 (0.403)	−0.286 (0.169)	0.286 (0.432)	−0.217 (0.216)
Late spring	$\beta_{r,9}$	0.969 (0.365)	−0.082 (0.163)	1.029 (0.395)	−0.023 (0.218)
Pre(95%) × summer1	$\beta_{r,10}$	−0.389 (0.464)	0.898 (0.243)	−0.347 (0.487)	0.930 (0.305)
Pre(95%) × autumn	$\beta_{r,11}$	0.346 (0.525)	0.279 (0.237)	0.397 (0.532)	0.274 (0.295)
Pre(95%) × winter	$\beta_{r,12}$	−0.701 (0.551)	0.429 (0.236)	−0.633 (0.547)	0.397 (0.290)
Pre(95%) × early spring	$\beta_{r,13}$	−0.670 (0.504)	0.120 (0.243)	−0.697 (0.530)	0.133 (0.300)
Pre(95%) × late spring	$\beta_{r,14}$	−1.678 (0.496)	0.147 (0.237)	−1.632 (0.513)	0.132 (0.301)
Random effect	d	0.020 (0.042)		0.013 (0.062)	
Overdispersion	δ	–		1.374 (0.437)	
-2loglik		6602.7		**6590.9**	
AIC		6664.7		**6654.9**	
BIC		6803.0		**6797.7**	

chi-squared distribution at 0) and χ^2_1, often denoted as $\chi^2_{0:1}$. Thus, we obtain $p = P(\chi^2_{0:1} \geq 11.8) = 0.5\,P(\chi^2_0 \geq 11.8) + 0.5\,P(\chi^2_1 \geq 11.8) = 0.0003$, showing that the inclusion of the overdispersion parameter was important.

Note that, when using the GEE approach in Menarin and Lara (2017), evidence of significant post-grazing management effect was reported, while here neither the GLMM nor CM confirmed this effect. For this experiment, overdispersion is likely to happen since it is a field experiment that can suffer from several environmental changes, and also because some types of vegetation can occur in an aggregate pattern inside paddocks.

8 Concluding remarks

In this article, we have proposed a model for overdispersed, repeated nominal data. The model combines the baseline-category logit assumption to handle the nominal nature of the outcome, with normal random effects in the linear predictor to deal with correlation across repeated measures, and beta random effects to flexibly account for overdispersion. Similar models were proposed by Molenberghs et al. (2007), Molenberghs et al. (2010), Molenberghs et al. (2012), Ivanova et al. (2014), and Molenberghs et al. (2017) for count data, binary and binomial data, time-to-event and ordinal outcomes. The model is easy to formulate and can be fitted using, for example, the SAS procedure NLMIXED.

A simulation study was conducted to examine the behaviour of the combined model relative to the more conventional GLMM. Both models performed well, but when there is a pronounced effect of overdispersion, or if the overdispersion effect is associated with high correlations between the repeated measurements, better performance was observed for the CM, mainly for the variance components.

We applied the GLMM and the CM to agricultural experimental data to model the probability of occurrence of three types of vegetation. Comparing both models, evidence is found in favour of the CM. It means that besides the parameter to take into account the correlation between measures, an extra overdispersion parameter was useful to accommodate the extra-variability induced by the environmental and biological changes.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

RKS was supported by CAPES and CNPq (proc. no. 233554/2014-9), Brazil. CGBD and SCS were supported by CNPq, Brazil.

Appendix A: Algebraic developments for the CM

The partially marginalized density function of the combined model was obtained by integrating analytically over the beta random effects, leaving the normal random effects untouched. To do this, we need to consider the category to which the outcome belongs in order to proceed with the integration over the beta random effect. To simplify notation, let us consider the case where three categories are analysed. We can rewrite this expression as

$$f(w_{r,ij}|\boldsymbol{b}_{r,i}) = \int \left(\theta_{ij}\kappa_{1,ij}\right)^{w_{1,ij}} \left(\theta_{ij}\kappa_{2,ij}\right)^{w_{2,ij}} \left(1-\theta_{ij}\kappa_{1,ij}-\theta_{ij}\kappa_{2,ij}\right)^{1-w_{1,ij}-w_{2,ij}} \frac{\theta_{ij}^{\vartheta-1}\left(1-\theta_{ij}\right)^{\xi-1}}{B(\vartheta,\xi)}\, d\theta_{ij}.$$

Thus, if the outcome belongs to the first category, that is, $W_{r,ij}$ is equals to 1 if $Y_{ij} = 1$ and 0 otherwise, the following expression is obtained:

$$\begin{aligned} f(w_{1,ij}=1|\boldsymbol{b}_{r,i}) &= \int_0^1 \left(\theta_{ij}\kappa_{1,ij}\right)^1 \frac{\theta_{ij}^{\vartheta-1}\left(1-\theta_{ij}\right)^{\xi-1}}{B(\vartheta,\xi)}\, d\theta_{ij}, \\ &= \frac{\kappa_{1,ij}}{B(\vartheta,\xi)} \int_0^1 \theta_{ij}^{(\vartheta-1)+1}(1-\theta_{ij})^{\xi-1} d\theta_{ij} \end{aligned}$$

$$
\begin{aligned}
&= \kappa_{1,ij} \frac{B(\vartheta+1,\xi)}{B(\vartheta,\xi)} \\
&= \kappa_{1,ij} \frac{\Gamma(\vartheta+1)\Gamma(\xi)}{\Gamma(\vartheta+\xi+1)} \frac{\Gamma(\vartheta+\xi)}{\Gamma(\vartheta)\Gamma(\xi)} \\
&= \kappa_{1,ij} \vartheta \frac{\Gamma(\vartheta)\Gamma(\xi)}{\Gamma(\vartheta+\xi+1)} \frac{\Gamma(\vartheta+\xi)}{\Gamma(\vartheta)\Gamma(\xi)} \\
&= \kappa_{1,ij} \vartheta \frac{\Gamma(\vartheta+\xi)}{(\vartheta+\xi)\Gamma(\vartheta+\xi)} \\
&= \kappa_{1,ij} \frac{\vartheta}{\vartheta+\xi}.
\end{aligned}
$$

Similar results apply to $r = 2$, but multiplied, of course, by their respective κ. For the last category ($r = 3$), the expression is given by:

$$
\begin{aligned}
f(w_{3,ij} = 1|\boldsymbol{b}_{r,i}) &= \int_0^1 \left(1 - \theta_{ij}\kappa_{1,ij} - \theta_{ij}\kappa_{2,ij}\right)^1 \frac{\theta_{ij}^{\vartheta-1}\left(1-\theta_{ij}\right)^{\xi-1}}{B(\vartheta,\xi)} d\theta_{ij} \\
&= \int_0^1 \frac{\theta_{ij}^{\vartheta-1}\left(1-\theta_{ij}\right)^{\xi-1}}{B(\vartheta,\xi)} \\
&\quad - \left(\theta_{ij}\kappa_{1,ij}\right) \frac{\theta_{ij}^{\vartheta-1}\left(1-\theta_{ij}\right)^{\xi-1}}{B(\vartheta,\xi)} - \left(\theta_{ij}\kappa_{2,ij}\right) \frac{\theta_{ij}^{\vartheta-1}\left(1-\theta_{ij}\right)^{\xi-1}}{B(\vartheta,\xi)} d\theta_{ij} \\
&= 1 - \kappa_{1,ij}\frac{\vartheta}{\vartheta+\xi} - \kappa_{2,ij}\frac{\vartheta}{\vartheta+\xi} \\
&= 1 - \frac{\vartheta}{\vartheta+\xi}(\kappa_{1,ij} + \kappa_{2,ij}).
\end{aligned}
$$

Hence, the partially marginalized likelihood function of the combined model considering three categories is given by:

$$
\begin{aligned}
L(\boldsymbol{\beta},\, \boldsymbol{D}, \vartheta, \xi) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} &\left(\frac{\vartheta}{\vartheta+\xi}\,\kappa_{1,ij}\right)^{w_{1,ij}} \left(\frac{\vartheta}{\vartheta+\xi}\,\kappa_{2,ij}\right)^{w_{2,ij}} \left(1 - \frac{\vartheta}{\vartheta+\xi}\sum_{h=1}^{R-1}\kappa_{h,ij}\right)^{1-w_{1,ij}-w_{2,ij}} \\
&\frac{1}{\sqrt{(2\pi)^{n_i}}} \frac{1}{\sqrt{|\boldsymbol{D}|}} \exp\left(-\frac{1}{2}\boldsymbol{b}_{r,i}^T \boldsymbol{D}^{-1}\boldsymbol{b}_{r,i}\right) d\boldsymbol{b}_{r,i}.
\end{aligned}
$$

Supplemental Material

Supplemental material is available for this article online.

References

Agresti A (2010) *Categorical data analysis*. Wiley, New York.

Clayton D (1992) Repeated ordinal measurements: a generalised estimating equation approach. *Medical Research Council Biostatistics Unit Technical Report*, pages 1–11.

Demétrio CGB, Hinde J and Moral RA (2014) Models for overdispersed data in entomology. In *Ecological modeling applied to entomology*. Springer.

Diggle PJ, Heagerty PJ, Liang KY and Zeger SL (2002) *Analysis of longitudinal data*. Oxford University Press, New York.

Grunwald GK, Bruce SL, Jiang L, Strand M and Rabinovitch N (2011) A statistical model for under- or overdispersed clustered and longitudinal count data. *Biometrical Journal*, **53**, 578–594.

Hartzel J, Agresti A and Caffo B (2001) Multinomial logit random effects models. *Statistical Modelling*, **1**, 81–102.

Hedeker D (2003) A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, **1446**, 1433–1446.

Hinde J and Demétrio CGB (1998) Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, **27**, 151–170.

Ivanova A, Molenberghs G and Verbeke G (2014) A model for overdispersed hierarchical ordinal data. *Statistical Modelling*, **14**, 399–415.

Lara IARD, Hinde JP, Castro ACD and da Silva IJO (2017) A proportional odds transition model for ordinal responses with an application to pig behaviour. *Journal of Applied Statistics*, **4763**, 1031–1046.

Lee Y and Nelder JA (1996) Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*, **58**, 619–678.

Lee Y and Nelder JA (2001) Hierarchical generalized linear models: a synthesis of generalized linear models, random-effect models and structured dispersions. *Biometrika*, **88**, 987–1006.

Lee Y and Nelder JA (2003) Extended-reml estimators. *Applied Statistics*, **30**, 845–856.

Liang KY and Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Lipsitz LR, Kim K and Zhao L (1994) Repeated categorical data using generalized estimating equations. *Statistics in Medicine*, **13**, 1149–1163.

McCullagh P and Nelder JA (1983) *Generalized linear models*. Chapman & Hall, London, 1 edition.

Menarin V and Lara IAR (2017) Longitudinal model for categorical data applied in an agriculture experiment about elephant grass. *Scientia Agricola*, **74**, 265–274.

Molenberghs G and Verbeke G (2005) *Models for discrete longitudinal data*. Springer-Verlang, New York.

Molenberghs G, Verbeke G and Demétrio CGB (2017) Hierarchical models with normal and conjugate random effects: a review. *SORT - Statistics and Operations Research Transactions*, **41**, 191–254.

Molenberghs G, Verbeke G and Demétrio CGB (2007) An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, pages 513–531.

Molenberghs G, Verbeke G, Demétrio CGB and Vieira AMC (2010) A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325–347.

Molenberghs G, Verbeke G, Iddi S and Demétrio CGB (2012) A combined beta and normal random-effects model for repeated, overdispersed binary and binomial data. *Journal of Multivariate Analysis*, **111**, 94–109.

Morel JG and Koehler KJ (1995) A one-step Gauss-Newton estimator for modelling categorical data with extraneous variation. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **44**, 187–200.

Morel JG and Nagaraj NK (1993) A finite mixture distribution for modelling multinomial extra variation. *Biometrika*, **80**, 363–371.

Morris D and Sellers K (2022) A flexible mixed model for clustered count data. *Stats*, **5**, 52–69.

Mosimann JE (1962) On the compound multinomial distribution, the multivariate β- distribution, and correlations among proportions. *Biometrika*, **49**, 65–82.

Neerchal NK and Morel JG (1998) Large cluster results for two parametric multinomial extra variation models. *Journal of the American Statistical Association*, **93**, 1078–1087.

Nelder JA and Wedderburn RWM (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A*, **135**, 370–384.

Pereira LET, Paiva AJ, Geremia EV and Silva SC (2015a) Grazing management and tussock distribution in elephant grass. *Grass and Forage Science*, **70**, 406–417.

Pereira LET, Paiva AJ, Geremia EV and Silva SC (2015b) Regrowth patterns of elephant grass (Pennisetum purpureum Schum) subjected to strategies of intermittent stocking management. *Grass and Forage Science*, **70**, 195–204.

Pinheiro JC and Bates DM (1995) Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, **4**, 12–35.

Ribeiro Jr E, Zeviani W, Bonat W, Demétrio C and Hinde J (2020) Reparametrization of compoisson regression models with applications in the analysis of experimental data. *Statistical Modelling*, **20**, 443–466.

Self SG and Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605–610.

Stram DO and Lee JW (1994) Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics*, **50**, 1171–1177.

Touloumis A, Agresti A and Kateri M (2013) GEE for Multinomial Responses Using a Local Odds Ratios Parameterization. *Biometrics*, **69**, 633–640.