



Bayesian Analysis of Nonnegative Data Using Dependency-Extended Two-Part Models

Mariana RODRIGUES-MOTTA and Johannes FORKMAN 

This article is motivated by the challenge of analysing an agricultural field experiment with observations that are positive on a continuous scale or zero. Such data can be analysed using two-part models, where the distribution is a mixture of a positive distribution and a Bernoulli distribution. However, traditional two-part models do not include any dependencies between the two parts of the model. Since the probability of zero is anticipated to be high when the expected value of the positive part is low, and the other way around, this article introduces dependency-extended two-part models. In addition, these extensions allow for modelling the median instead of the mean, which has advantages when distributions are skewed. The motivating example is an incomplete block trial comparing ten treatments against weed. Gamma and lognormal distributions were used for the positive response, although any density on the support of real numbers can be accommodated. In a cross-validation study, the proposed new models were compared with each other and with a baseline model without dependencies. Model performance and sensitivity to choice of priors were investigated through simulation. A dependency-extended two-part model for the median of the lognormal distribution performed best with regard to mean square error in prediction.

Supplementary materials accompanying this paper appear online.

Key Words: Bayesian analysis; Incomplete block design; Mixed-effects models; Hurdle model; Zero-augmented data; Zero-inflated data.

1. INTRODUCTION

In many areas of applied statistics, distributions are positively skewed and variance is increasing with the mean. These phenomena violate the assumptions of normality and homoscedasticity needed for statistical inference in traditional analysis of variance and regression. In addition, the data may include an excess of zeros. For example, in a study of garden bird abundance, average count, which is effectively continuous, spikes at zero for many of the species (Swallow et al. 2016), in a study of precipitation, there are some

M. Rodrigues-Motta, Department of Statistics, State University of Campinas, São Paulo, Brazil.

J. Forkman (✉), Department of Crop Production Ecology, Swedish University of Agricultural Sciences, PO Box 7043, 750 07 Uppsala, Sweden (E-mail: johannes.forkman@slu.se).

© 2021 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics, Volume 27, Number 2, Pages 201–221
<https://doi.org/10.1007/s13253-021-00467-x>

months without rain (Harvey and Van der Merwe 2012; Fuentes et al. 2008; Sun and Stein 2015), and in a study of bycatch data, the endangered hammerhead shark is often missing (Cantoni et al. 2017). Our application is an agricultural plant protection experiment, where the weed of interest does not grow in every plot. In plant protection experiments, treatments effectively controlling the weed either eliminate the weed, giving raise to zeros in the dataset, or result in small positive levels of weed biomass. Non-effective treatments, on the other hand, do not eliminate the weed but typically show large levels of weed biomass. Thus, there is a negative relationship between probability of zero and level of biomass conditioned on the weed being present. We wanted to model probability of zero explicitly as a function of the conditional level of weed biomass, which was not possible using previously proposed methods.

As is well known, the median is more robust against outliers than the mean. Yet, models for means are more common than models for medians. We shall propose dependency-extended two-part models for both the mean and the median.

Using two-part models, the probability of zero and the level of the positive observations are modelled separately. Thus, the distribution is assumed to be a mixture of a Bernoulli distribution and a positive distribution, which cannot take negative values. Duan et al. (1983) used a probit link for modelling the Bernoulli event and fitted a log-linear model for the positive part, while Zhou and Tu (2000) and Hautsch et al. (2013) used a logit link for fitting the Bernoulli piece.

Chen and Qin (2003) and Yang et al. (2016) used an empirical likelihood, thus avoiding exact assumptions on the distribution of the positive part of the data. In its original form, the two parts of the two-part model can be fitted separately, using different or common explanatory variables. This is possible since the likelihood can be written as a product of two factors corresponding to the two parts of the model. However, if the two parts share the same parameters (Moulton et al. 2002), or if the parameters of the two parts are constrained to be related, such factorization is not possible (Mills 2013). In analysis of count data, two-part models are known as hurdle models (Rose et al. 2006). Zero-inflated Poisson or zero-inflated negative binomial models are other popular options for count data. Tang et al. (2018) used such models and assumed a dependency between the probability of zero and the Poisson mean, but only as a result of common predictors.

In this article, the traditional two-part model is extended to accommodate explicit functional dependencies between the level of the positive part and the probability of zero. Four models are explored: a baseline model without correlated parts (Model 0), a model with correlated random effects (Model 1) and two dependency-extended models (Models 2 and 3), which includes functional dependencies between the two parts. Let t , where $t = 1, \dots, T$,

be the treatment index and s , where $s = 1, \dots, S$, be the block index. The models are:

$$\begin{aligned} \text{Models 0 and 1: } & \begin{cases} \log(\mu_{f,st}) = \alpha_t + b_s \\ \text{logit}(p_{st}) = \beta_t + v_s \end{cases} \\ \text{Model 2: } & \begin{cases} \log(\mu_{f,st}) = \gamma_1 + \gamma_2 \text{logit}(p_t) + b_s \\ \text{logit}(p_{st}) = \alpha_t \end{cases} \\ \text{Model 3: } & \begin{cases} \log(\mu_{f,st}) = \alpha_t + b_s \\ \text{logit}(p_{st}) = \gamma_1 + \gamma_2 \log(\mu_{f,st}) \end{cases} \end{aligned}$$

In Model 0, the level of the positive part, $\mu_{f,st}$, and the probability of zero, p_{st} , are modelled by fixed treatment effects: α_t, β_t and random block effects: b_s, v_s , which are uncorrelated. Model 1 is the same as Model 0, but with correlated block effects. In Model 2, the level of the positive part, $\mu_{f,st}$, is dependent, through regression coefficients γ_1 and γ_2 , on the probability of zero, p_{st} , and a random block effect, b_s , whereas the probability of zero, p_{st} , is modelled by fixed treatment effects, α_t . In Model 3, the logarithm of the level of the positive part, $\log(\mu_{f,st})$, is a sum of a fixed treatment effect, α_t , and a random block effect, b_s , whereas the probability of zero, p_{st} , is dependent on the level of the positive part, $\mu_{f,st}$ through regression coefficients γ_1 and γ_2 .

Models similar to Model 1 have been proposed for count data (Min and Agresti 2005; Neelon et al. 2010; Cantoni et al. 2017) and for lognormally distributed repeated measures data (Tooze et al. 2002). Our dependency-extended two-part Models 2 and 3 will be compared with Model 1. It will be shown in Sect. 5.1 that Model 1 does not work well when only zeros are observed for some levels of the random effects, which happens frequently in practice. Note that Models 2 and 3 comprise fewer parameters than Model 1, which can make them more robust and easy to fit.

In Models 0, 1 and 3, $\log(\mu_{f,st}) = \alpha_t + b_s$. However, the logarithm of the conditional level in Model 2 can be written similarly, as with this model, $\log(\mu_{f,st}) = \alpha'_t + b_s$, where $\alpha'_t = \gamma_1 + \gamma_2 \alpha_t$. In terms of the probability of zero, p_{st} , all models use logistic regression. Models 2 and 3 have considerably fewer parameters than Model 1 and are therefore apt to work well for small datasets. Specifically, Model 2 assumes no random effects on the probability of zero. Model 3 uses only two parameters for the logit part, which can be beneficial when the number of zeros is either small or large. However, the great advantage of Models 2 and 3 is that they provide a mathematical description of how the probability of zero and the positive part are related. Similar to a model proposed by Neuhaus et al. (2018) for outcome-dependent visit processes, the parameter γ_2 governs the strength and direction of the association between the conditional mean or median and whether or not data are observed.

For the positive part, we used the lognormal distribution and the gamma distribution. Using the lognormal distribution, attention characteristically shifts from the mean to the median (Goldberger 1968), which is invariant under distributional transformation (Rao and

D’Cunha 2016) and less sensitive to outliers. The lognormal distribution is often a better assumption than the normal, especially in biology, where the lognormal distribution arises asymptotically as a consequence of biological mechanisms (Koch 1969). In several disciplines, e.g. geology, food technology, social sciences and economics, the lognormal distribution is used when effects are multiplicative rather than additive (Limpert et al. 2001). The gamma distribution is another useful distribution when data are positively skewed and heteroscedastic, although rarely used with excess of zeros (Musal and Ekin 2017). Specifically, when the standard deviation is proportional to the mean, this distribution motivates a model with constant coefficient of variation (Lee et al. 2006, p. 71).

Our research was motivated by difficulties encountered in weed science when analysing agricultural field experiments using traditional statistical methods. In order to assume normality and homoscedasticity, data are often log-transformed before analysis. However, when some observations are zero, i.e. when no weed is observed, some small positive constant must first be added to the observed values. This procedure destroys the multiplicative scale and introduces an undesirable arbitrariness, since the choice of the constant may affect the conclusions. Modelling heteroscedasticity is another option (Damesa et al. 2018), but also such modelling, for example, using the power-of-the-mean model (Carroll and Ruppert 1988), would be problematic when only zeros are observed for some treatments, since the variance would be estimated to zero for those treatments.

Two-part models with dependency between the two parts are rare in the literature, perhaps due to the cumbersome maximization of the likelihood (Feuerverger 1979). We suggest instead using Bayesian methodology. Bayesian analysis of two-part models has previously been proposed by Rodrigues-Motta et al. (2015), who assumed the positive distribution to be a member of the bivariate exponential family, and Harvey and Van der Merwe (2012), who recommended Bayesian methods for inference on means and variances in a two-part model with a lognormal distribution. However, these authors did not assume any extension between the two parts. Bayesian models have also been proposed for zero-inflated longitudinal nonnegative data (Swallow et al. 2016; Biswas and Das 2020) and zero-inflated count data (Neelon et al. 2010; King and Song 2019; Bertoli et al. 2020). Specifically, Biswas and Das (2020) and Neelon et al. (2010) proposed models with correlated random effects, which in this regard are similar to our Model 1. Tiao and Draper (1968) and Besag and Higdon (1999) pioneered Bayesian analysis of experiments with incomplete blocks. Several authors, with various focus, have explored Bayesian methods for agricultural field experiments, for example, Donald et al. (2011), who modelled spatial correlation, Forkman and Piepho (2013), who studied prediction of random effects of treatments, Singh et al. (2015), who reported a Bayesian analysis of a crop variety trial with an incomplete block design, and Theobald et al. (2002), who investigated Bayesian analysis of multi-environment crop variety trials. However, none of them considered two-part mixed-effects models.

In analysis of designed experiments, random effects are common, since effects of units involved in randomization should be modelled as random (Piepho et al. 2003). As a means of decreasing residual error variance, agricultural field trials frequently include incomplete blocks, i.e. blocks that do not include all treatments. Effects of incomplete blocks are usually modelled as random, since this allows recovery of inter-block information about treatment differences (Piepho and Edmondson 2018).

Dependence between the zero and the nonzero components of a hierarchical model has also been touched upon in spatio-temporal literature. In this field, Fuentes et al. (2008) considered a zero-inflated log-Gaussian process model and modelled the probability of no rain in terms of the amount of the rainfall, inducing dependency between those parts. In a similar study, Sun and Stein (2015) modelled partly precipitation by means of a space-time Gaussian random field varying along time or space and time, and partly the logit of the probability of precipitation as a function of space and time. However, no relationship was imposed between the parts.

This article has two aims. The first is to present new methodology for analysis of zero-augmented positive observations by introducing dependencies between the probability of zero and the level of the positive part, beyond correlation of random effects. Because of these dependencies, zeros convey information about the mean or the median of the positive part, and the other way around. This assumption makes sense in weed experiments, and presumably also in many other areas of research. The second aim is to propose methods for modelling medians, not just means, since medians are less affected by skewness and extreme values.

Section 2 presents, as a motivating example, the creeping thistle weed dataset. Section 3 details the models, and Sect. 4 describes the Bayesian approach for fitting them. Section 5 evaluates the models and the choice of priors through cross-validation and simulation studies and identifies the best model for the creeping thistle dataset. Section 6 uses this final model for the statistical analysis. Section 7 concludes with a discussion.

2. AN AGRICULTURAL WEED EXPERIMENT

Our motivating example is an agricultural weed trial with ten experimental treatments. The experiment aimed at comparing nine different mixtures of plant protection products (treatments 2–10) with regard to their efficacy on the weed creeping thistle (*Cirsium arvense*) in a field with spring barley (*Hordeum vulgare*). In addition to the active treatments, the experiment included a control treatment with no application of plant protection (treatment 1). The design comprised four replicates with ten plots each, which all received different treatments. Within each replicate, the plots were grouped into two incomplete blocks of five contiguous plots each, following an alpha design (Patterson and Williams 1976). Replicate 1 comprised blocks 1 and 2, replicate 2 comprised blocks 3 and 4, and so on.

In agricultural experiments, it is common to partition the replicates into incomplete blocks of adjacent plots. In this way, homogeneous intra-block variance is achieved. This strategy is successful if combined with an efficient design that ensures that as many pairs of treatments as possible occur together in blocks. The alpha design is such a design (Verdooren 2020).

Table 1 provides an overview of the dataset. Means, medians and standard deviations were computed by treatment and block. These computations were made both with zeros excluded, i.e. when $Y \in (0, \infty)$, and with zeros included, i.e. when $Y \in [0, \infty)$. For treatments 4, 8 and 9, no creeping thistle was observed, i.e. only zeros were recorded. Similarly, no creeping thistle was observed in block 7. Standard deviations vary between treatments, as a result

Table 1. Summary of the creeping thistle dataset

Treatment	% of zeros	Biomass mean (sd) of		Biomass median of	
		$Y \in (0, \infty)$	$Y \in [0, \infty)$	$Y \in (0, \infty)$	$Y \in [0, \infty)$
1	0	125 (76.63)	125 (76.63)	119	119
2	25	68.67 (52.54)	51.50 (54.95)	71	43
3	50	43.00 (36.77)	21.50 (32.66)	43	8.5
4	100	–	0.00 (0.00)	–	0
5	50	44.50 (47.38)	22.75 (37.53)	44.5	5.5
6	0	61.50 (19.12)	61.50 (19.12)	53.5	53.5
7	50	47.50 (45.96)	23.50 (38.16)	47.5	7.5
8	100	–	0.00 (0.00)	–	0
9	100	–	0.00 (0.00)	–	0
10	0	166.25 (186.30)	166.25 (186.30)	107.5	107.5

Block	% of zeros	Biomass mean (sd) of		Biomass median of	
		$Y \in (0, \infty)$	$Y \in [0, \infty)$	$Y \in (0, \infty)$	$Y \in [0, \infty)$
1	20	127.50 (29.86)	102.00 (62.61)	130	120
2	80	80.00 (*)	16.00 (35.78)	80	0
3	60	63.50 (20.51)	25.40 (36.26)	63.5	0
4	20	15.50 (1.00)	12.40 (6.99)	15	15
5	60	256.50 (252.44)	102.60 (188.86)	256.5	0
6	20	102.00 (75.76)	81.60 (79.91)	70	69
7	100	–	0 (0.00)	–	0
8	20	46.75 (26.64)	37.4 (31.13)	50.5	47

Percentages of zeros out of four replicates per treatment and five plots per block. Biomass means, medians and standard deviations (g m^{-2}) were computed with zeros excluded, i.e. when $Y \in (0, \infty)$, and with zeros included, i.e. when $Y \in [0, \infty)$. * There is only one observation larger than zero in block 2

of block effects and residual error heterogeneity. The huge differences between blocks in standard deviation are mainly due to differences between treatments.

Figure A1 in Web Appendix A of the supplementary materials shows all observations by treatment. This figure indicates positive skewness. Table A1 of Web Appendix A includes the dataset. The small size of this dataset (40 observations), combined with a high proportion of zeros (47.5%), is challenging.

3. MODEL DEVELOPMENT

Let $y_{st} \in [0, \infty)$ be the observation in treatment $t = 1, \dots, T$ in block $s = 1, \dots, S$. The blocks are incomplete, since they include subsets of treatments. Given λ and p_{st} , the density function of the response is given by the mixture distribution

$$g(y_{st}|\lambda, p_{st}) = p_{st}^{I_{\{y_{st}=0\}}} \times [(1 - p_{st})f(y_{st}|\lambda)]^{I_{\{y_{st}>0\}}}, \quad y_{st} \geq 0, \quad (1)$$

where $p_{st} \in [0, 1]$ is the probability of observing a zero, and $\lambda \in \Lambda \subset \mathbb{R}^p$. This construction was presented by [Rodrigues-Motta et al. \(2015\)](#) for $p = 2$ with $f(y_{st}|\lambda)$ in (1) being a member of the biparametric exponential family ([Bar-Lev and Reiser 1982](#); [Bose and Boukai](#)

1993) and parametrized by $\lambda = (\mu_{f,st}, \phi)$, where $\mu_{f,st}$ represents the conditional mean or conditional median and ϕ is a dispersion or precision parameter, depending on the choice of $f(\cdot|\lambda)$. When $\mu_{f,st}$ is the conditional mean, the marginal mean is

$$\begin{aligned} \mu_{g,st} &= E[Y_{st}] \\ &= p_{st}E[Y_{st}|Y_{st} = 0] + (1 - p_{st})E[Y_{st}|Y_{st} > 0] \\ &= (1 - p_{st})\mu_{f,st}. \end{aligned} \tag{2}$$

From (1), the cumulative distribution $G(\cdot|\lambda, p_{st})$ of $Y_{st} \in [0, \infty)$ is given by

$$G(y_{st}|\lambda_{st}, p_{st}) = \begin{cases} 0 & \text{if } y_{st} < 0 \\ p_{st} & \text{if } y_{st} = 0 \\ p_{st} + (1 - p_{st})F(y_{st}|\lambda) & \text{if } y_{st} > 0 \end{cases} \tag{3}$$

where $F(\cdot|\lambda)$ is the cumulative distribution corresponding to $f(\cdot|\lambda)$ in (1). Let $\tilde{y}_{g,st}$ be the median of the marginal distribution $g(\cdot|.)$, for treatment t in block s . According to (3), if $p_{st} \geq 0.5$, then $\tilde{y}_{g,st}$ is 0. Otherwise $\tilde{y}_{g,st}$ is given by the quantile $y_{f,st}$ satisfying $F(y_{f,st}|\lambda) = (0.5 - p_{st})/(1 - p_{st})$.

3.1. MODELS

Let $Y_{st} \in (0, \infty)$ have density function $f(\cdot|\mu_{f,st}, \phi)$ parametrized by a location parameter $\mu_{f,st}$ and a dispersion parameter ϕ . Recall that p_{st} is the mixture parameter in $g(\cdot|.)$ given in (1). Models 0–3 specify $\mu_{f,st}$ and p_{st} in four different ways, as follows.

Model 0: The location, $\mu_{f,st}$, and the probability of zero, p_{st} , are modelled as

$$\log(\mu_{f,st}) = \alpha_t + b_s, \tag{4}$$

and

$$\text{logit}(p_{st}) = \beta_t + v_s, \tag{5}$$

where α_t and β_t are fixed effects of treatment t . Furthermore, b_s and v_s are random effects of block s , so that $(b_s, v_s) \sim N(\mathbf{0}, \Sigma)$, where Σ is a diagonal covariance matrix of dimension 2×2 given by

$$\Sigma = \begin{pmatrix} \sigma_b^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}. \tag{6}$$

Model 0 will serve as a baseline model for comparisons in the cross-validation and simulation studies.

Model 1: This model is the same as Model 0, but with correlated random effects:

$$\Sigma = \begin{pmatrix} \sigma_b^2 & \sigma_{b,v} \\ \sigma_{b,v} & \sigma_v^2 \end{pmatrix}. \quad (7)$$

The covariance between $\log(\mu_{st})$ and $\text{logit}(p_{st})$ is induced by $\sigma_{b,v}$, which is the covariance between b_s and v_s .

Model 2: The logarithm of $\mu_{f,st}$ is modelled as a linear function of $\text{logit}(p_{st})$, that is,

$$\log(\mu_{f,st}) = \gamma_1 + \gamma_2 \text{logit}(p_{st}) + b_s, \quad (8)$$

where the block effect, b_s , is $N(0, \sigma_b^2)$ distributed. The probability of observing a zero, p_{st} , is the same across blocks, given by

$$\text{logit}(p_{st}) = \alpha_t, \quad (9)$$

where α_t is a fixed effect of treatment t . Thus, we can write $p_{st} = p_t$. The covariance between $\log(\mu_{f,st})$ and $\text{logit}(p_{st})$ is the variance of $\text{logit}(p_{st})$ multiplied by γ_2 , so that γ_2 determines the sign and the strength of the covariance.

Model 3: The logit of the probability of zero, $\text{logit}(p_{st})$, is modelled as a linear function of $\mu_{f,st}$, which is dependent on a fixed treatment effect, α_t , and a random block effect, b_s . Specifically,

$$\log(\mu_{f,st}) = \alpha_t + b_s, \quad (10)$$

where $b_s \sim N(0, \sigma_b^2)$, and

$$\text{logit}(p_{st}) = \gamma_1 + \gamma_2 \log(\mu_{f,st}). \quad (11)$$

The covariance between $\log(\mu_{f,st})$ and $\text{logit}(p_{st})$ is the variance of $\log(\mu_{f,st})$ multiplied by γ_2 . Thus, also with this model, γ_2 determines the sign and the strength of the covariance.

In Models 2 and 3, the location parameter, $\mu_{f,st}$, and the probability of zero, p_{st} , are functionally related, which they are not in Models 0 and 1. Consequently, Models 2 and 3 require fewer parameters than Models 0 and 1, which can be advantageous for small datasets. According to Model 2, the probabilities of zero depend only on the treatments, whereas according to Model 3, they depend also on the blocks.

3.2. THE DISTRIBUTION OF $Y \in (0, \infty)$

In Models 0–3, $f(y_{st}|\lambda)$ is either a gamma or a lognormal distribution. Parametrized by the mean, $\mu_{f,st}$, and the shape or dispersion parameter, ϕ , the gamma distribution is

$$f(y_{st}|\mu_{f,st}, \phi) = \frac{1}{\Gamma(\phi)} \left(\frac{\phi}{\mu_{f,st}} \right)^\phi y_{st}^{\phi-1} \exp \left\{ -\frac{\phi y_{st}}{\mu_{f,st}} \right\}. \quad (12)$$

Parametrized by the median, $\mu_{f,st}$, and a dispersion parameter, ϕ , the lognormal distribution is

$$f(y_{st}|\mu_{f,st}, \phi) = \frac{1}{y_{st} \phi \sqrt{2\pi}} \exp \left[-\frac{(\log(y_{st}) - \log(\mu_{f,st}))^2}{2 \phi^2} \right]. \tag{13}$$

Note that in (12), $\mu_{f,st}$ is the mean of the distribution, whereas in (13), $\mu_{f,st}$ is the median of the distribution. Under distribution (13), $\log(y_{st})$ is $N(\log(\mu_{f,st}), \phi^2)$, and the mean of y_{st} is $E(y_{st}) = \exp(\phi^2/2)\mu_{f,st}$. Under distribution (12), however, the median has no simple closed form. The lognormal distribution enables modelling the median instead of the mean.

To overcome estimation drawbacks due to a small number of observations per block and treatment, the median can be modelled instead of the mean, as the former is more robust. The median is a natural, always finite and appropriate quantity of centrality for a skewed distribution, which is the case in our application. To assess effects of treatments on the median, rather than the mean, $f(y_{st}|\lambda)$ should be chosen to be a lognormal distribution, as proposed by [Goldberger \(1968\)](#).

The two proposed distributions of $Y > 0$ are both members of the biparametric exponential family. Their density functions are similar in shape, but the lognormal distribution is more skewed than the gamma distribution. The lognormal distribution arises in biological applications when effects are multiplicative rather than additive ([Koch 1969](#); [Limpert et al. 2001](#)). The gamma distribution arises, for example, as a sum of time intervals between events in a Poisson process. However, in this study, the main difference is that the gamma distribution was used for modelling the mean, while the lognormal distribution was used for modelling the median. Therefore, differences in results between using the gamma distribution and using the lognormal distribution are also related to differences between modelling the mean and modelling the median.

4. BAYESIAN INFERENCE

4.1. SPECIFICATION OF PRIORS

Let $IG(\eta_1, \eta_2)$ denote an inverse gamma distribution with shape parameter η_1 and rate parameter η_2 . Let $Cauchy(\eta_1, \eta_2)$ denote a Cauchy distribution with location parameter η_1 and scale parameter η_2 , and $Cauchy(\eta_1, \eta_2)I_{\xi > x}$ the same Cauchy distribution truncated from below at $\xi = x$. Let $N(\eta_1, \eta_2)$ denote a normal distribution with expected value η_1 and variance η_2 , and $N(\eta_1, \eta_2)I_{\xi < x}$ the same normal distribution truncated from above at $\xi = x$. Let $Beta(\eta_1, \eta_2)$ denote a beta distribution with shape parameters η_1 and η_2 . The following independent diffuse priors were assumed.

For Model 0: $\alpha_t \sim N(0.01, 10)$, $\beta_t \sim N(0.01, 10)$, $\phi \sim IG(0.01, 0.01)$, $\sigma_b^2 \sim IG(0.01, 0.01)$, and $\sigma_v^2 \sim IG(0.01, 0.01)$. For Model 1: $\alpha_t \sim N(0.01, 10)$, $\beta_t \sim N(0.01, 10)$, $\phi \sim IG(0.01, 0.01)$. The prior distribution for Σ in (7) was specified as proposed by [Neelon et al. \(2010\)](#), i.e. $b_s \sim N(0, v_1^2)$ and $v_s|b_s \sim N(\psi, v_2^2)$ such that $(b_s, v_s) \sim N(\mathbf{0}, \Sigma_{b,v})$, with diffuse hyperpriors $\psi \sim N(0.01, 100)$, $v_1^2 \sim IG(0.01, 0.01)$,

and $v_2^2 \sim \text{IG}(0.01, 0.01)$. The components of Σ in (7) can then be expressed as

$$\Sigma^* = \begin{pmatrix} v_1^2 & \psi v_1^2 \\ \psi v_1^2 & \psi^2 v_1^2 + v_2^2 \end{pmatrix}, \quad (14)$$

where $\sigma_b^2 = v_1^2$, $\sigma_v^2 = \psi^2 v_1^2 + v_2^2$ and $\rho = (\psi v_1^2) / [\psi^2 v_1^2 + v_2^2]^{1/2}$. For Model 2: $p_t \sim \text{Beta}(1, 1)$, $\gamma_1 \sim \text{N}(0.01, 10)$, $\gamma_2 \sim \text{N}(0.01, 10)I_{\gamma_2 < 0}$, $\sigma_b^2 \sim \text{Cauchy}(0, 0.1)I_{\sigma_b^2 > 0}$, $\phi \sim \text{IG}(0.01, 0.01)$. For Model 3: $\alpha_t \sim \text{N}(0.01, 10)$, $\gamma_1 \sim \text{N}(0.01, 10)$, $\gamma_2 \sim \text{N}(0.01, 10)I_{\gamma_2 < 0}$, $\sigma_b^2 \sim \text{Cauchy}(0, 0.1)I_{\sigma_b^2 > 0}$, $\phi \sim \text{IG}(0.01, 0.01)$.

To ensure proper operation of `OpenBUGS`, expected value 0.01 was used instead of 0 for the normal distributions. The half-Cauchy distribution, $\text{Cauchy}(0, 0.1)I_{\sigma_b^2 > 0}$, was coded as described by [Lunn et al. \(2012\)](#), p. 225. The $\text{Beta}(1, 1)$ distribution is a continuous uniform distribution with minimum value 0 and maximum value 1. For all models, $\text{IG}(0.01, 0.01)$ was used as prior for ϕ , corresponding to a gamma distribution with mean one and variance 100. This prior can be regarded a low-information prior since variance is large. The precision parameter measures the precision of the available information concerning the parameter of interest. If the variance is large, then the precision is low and hence the mean cannot be used as a precise estimate for the random variable of interest ([Ntzoufras 2011](#)).

4.2. HIERARCHICAL STRUCTURE SPECIFICATION

Let $\mathbf{u}_s = (b_s, v_s)$ or $\mathbf{u}_s = b_s$ depending on the imposed model, and let $\Theta = (\lambda, \mathbf{p}, \Sigma)$ be the vector of parameters, where \mathbf{p} is the vector of mixtures probabilities p_{st} , and Σ is the covariance matrix of the random effects. Based on the data $(\mathbf{y}; \mathbf{u})$, the complete likelihood is

$$f(\Theta; \mathbf{y}, \mathbf{u}) = \prod_{s \in S} \prod_{t \in T} \left\{ p_{st}^{I_{\{y_{st}=0\}}} \times [(1 - p_{st}) f(y_{st} | \lambda)]^{I_{\{y_{st}>0\}}} \right. \\ \left. \times \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp \left\{ -\frac{1}{2} \mathbf{u}_s^\top \Sigma \mathbf{u}_s \right\} \right\}.$$

The augmented posterior distribution of (Θ, \mathbf{u}) is recognizable only up to a proportionality constant, and the marginal posterior distributions cannot be obtained analytically. Therefore, the relevant MCMC steps (combination of Gibbs and Metropolis-within-Gibbs sampling) were implemented using the `BRUGS` package ([Thomas et al. 2006](#)), which connects R with the `OpenBUGS` software. Convergence was monitored via MCMC chains, auto-correlation, density plots and the Brooks–Gelman–Rubin potential scale reduction factor \hat{R} , all available in the R `coda` library ([Cowles and Carlin 1996](#)). The R code for the Bayesian analysis is provided as supplementary materials. The full conditional posterior distributions of treatment and block effects and dispersion components are specified in Web Appendix B.

5. ASSESSMENT OF MODEL PERFORMANCE

5.1. POSTERIOR STATISTICS

Posterior statistics were averaged across blocks as follows. The mixture probability was computed as $p_t = \sum_{s \in S_t} p_{st} / \sum_s^S I(s \in S_t)$, where p_{st} is defined in (1). For the gamma distribution, the conditional and marginal means were computed as $\mu_{f,t} = \sum_{s \in S_t} \mu_{f,st} / \sum_s^S I(s \in S_t)$ and $\mu_{g,t} = \sum_{s \in S_t} \mu_{g,st} / \sum_s^S I(s \in S_t)$, respectively, where $\mu_{f,st}$ is defined in (12) and $\mu_{g,st}$ is defined in (2). For the lognormal distribution, the conditional and marginal medians were computed as $\mu_{f,t} = \sum_{s \in S_t} \mu_{f,st} / \sum_s^S I(s \in S_t)$ and $\tilde{y}_{g,t} = \sum_{s \in S_t} \tilde{y}_{g,st} / \sum_s^S I(s \in S_t)$, respectively, where $\tilde{y}_{g,st}$ is defined in Sect. 3 and $\mu_{f,st}$ is defined in (13).

Figure 1 presents, for Models 1–3 and gamma distribution, 95% credible intervals (CIs) for conditional means (displayed on the first row of subfigures), marginal means (on the second row) and probabilities of zero (on the third row). The limits of the 95% CIs are the 2.5th and 97.5th percentiles of the posterior distributions. Similarly, Fig. 2 presents, for Models 1–3 with lognormal distribution, 95% CIs for conditional medians, marginal medians and probabilities of zero. In these figures, CIs for conditional means and medians are much larger using Models 0 and 1 than using Models 2 and 3. Note that the scale of the horizontal axis differs between the models. Using Models 0 and 1, very wide CIs were observed for treatments 4, 8 and 9, for which only zeros were obtained.

Table 2 presents posterior means and 95% CIs for model parameters. The correlation ρ between random effects b_s and v_s in Model 1 is positive, implying a positive correlation between mean or median and probability of zero. On the other hand, the estimated correlation between mean or median and probability of zero using Models 2 and 3 is negative, as the estimate for γ_2 is negative for both these models. A negative correlation seems more reasonable than a positive, as we would expect the probability of zero to decrease with an increasing amount of weed.

5.2. CROSS-VALIDATION OF MODELS FOR THE WEED EXPERIMENT

A leave-one-out cross-validation was performed, using replicates as observations. In this evaluation, Models 1–3 were compared with Model 0 and with each other. Model 0 can be regarded as a baseline model, since the two parts of this model are independent.

Cross-validation method

The means of the posterior distributions of $\tilde{y}_{g,st}$ were used as predictions. With four models and two distributions, altogether eight methods for prediction were compared. One replicate (10 observations) at a time was left out, and the remaining data were used for the analysis. Let $\hat{\boldsymbol{\mu}}_g^{(-r)}$ denote the vector of predictions, sorted by treatment, based on an analysis of the dataset with replicate r removed. Furthermore, let \mathbf{y} denote the vector of observations, sorted by replicate and treatment, and $\hat{\boldsymbol{\mu}}_g$ the concatenated vector $(\hat{\boldsymbol{\mu}}_g^{(-1)}, \hat{\boldsymbol{\mu}}_g^{(-2)}, \hat{\boldsymbol{\mu}}_g^{(-3)}, \hat{\boldsymbol{\mu}}_g^{(-4)})$. Predictive performance was evaluated using the root-mean-square

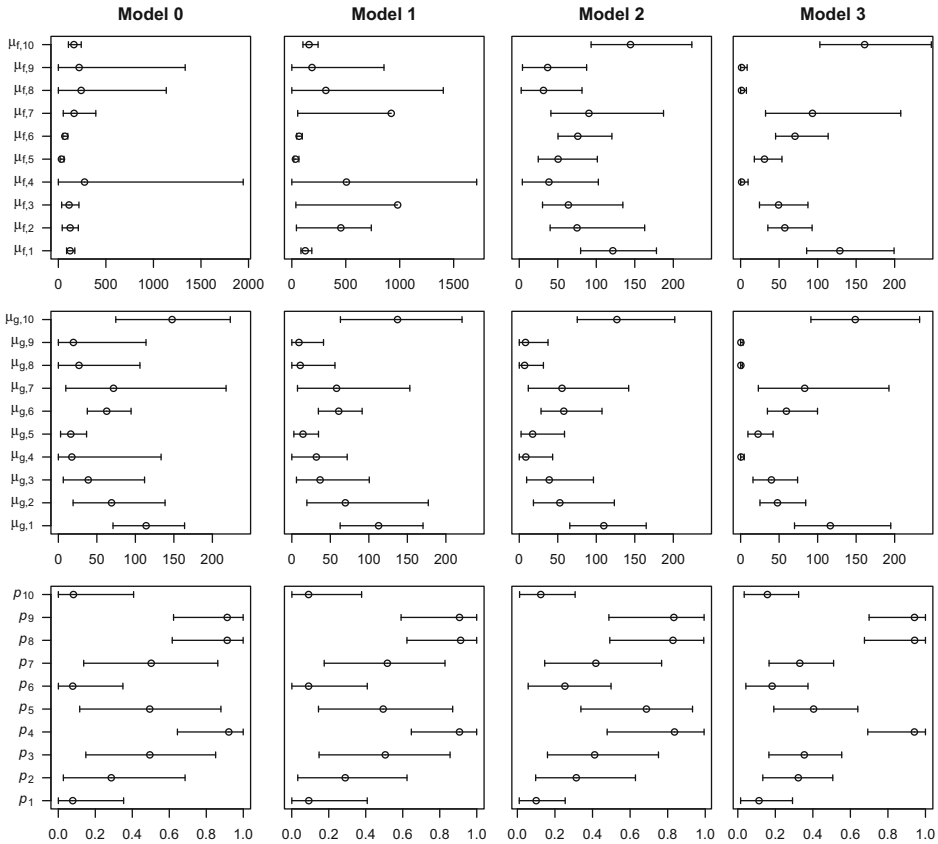


Figure 1. Posterior credible interval (CI) of 95% for conditional mean $\mu_{f,t}$ of $f(\cdot)$, marginal mean $\mu_{g,t}$ of $g(\cdot)$ and probability of zero p_{St} , for treatments $t = 1, \dots, 10$ in Models 0–3 with gamma distribution. Note that the scale differs between models.

error (RMSE) criterion, here defined as the square root of $(\hat{\boldsymbol{\mu}}_g - \mathbf{y})^T(\hat{\boldsymbol{\mu}}_g - \mathbf{y})/N$, where $N = 40$ is the number of observations.

Cross-validation results

Using the gamma distribution, thus modelling the mean, the RMSE was 89.7 for Model 0. The more advanced models performed better. The RMSE was 89.5, 87.3 and 87.3, for Models 1, 2 and 3, respectively. The dependency-extended models, i.e. Models 2 and 3, performed better than Model 1.

Using the lognormal distribution, thus modelling the median, the RMSE was 79.2 for Model 0. Again, the more advanced models performed better. The RMSE was 77.5, 76.4 and 78.9, for Models 1, 2 and 3, respectively. Thus, Model 2 performed best.

The cross-validation clearly showed that for this dataset, it was better to model the median using the lognormal distribution than to model the mean using the gamma distribution.

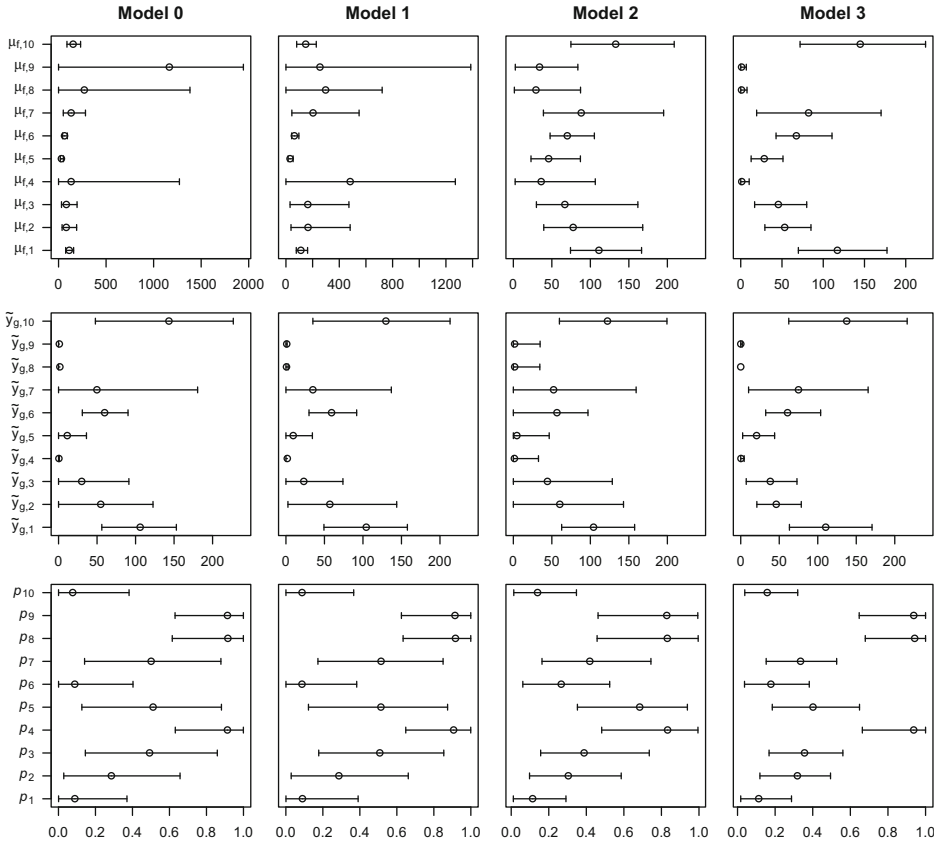


Figure 2. Posterior credible interval (CI) of 95% for conditional median $\tilde{y}_{f,t}$ of $f(\cdot)$, marginal median $\tilde{y}_{g,t}$ of $g(\cdot)$ and probability of zero p_t , for treatments $t = 1, \dots, 10$ in Models 0–3 with lognormal distribution. Note that the scale differs between models .

5.3. SIMULATION STUDIES

To investigate the performance of the models using the median, simulation studies were conducted based on the example of Sect. 2. The objectives of these studies were (1) to compare the models with regard to ability to recover parameter information and (2) to check the sensitivity of the results to the choice of prior distributions. The lognormal distribution was used, since this outperformed the gamma distribution in the cross-validation study.

Data were generated according to the models, where the true parameter values were set to the posterior mean estimates. These were for Model 0: $\hat{\alpha} = (4.12, 3.68, 3.66, -0.06, 2.32, 3.54, 3.68, 0.01, 0.04, 3.97)$, $\hat{\beta} = (-3.38, -1.30, 0.12, 3.56, 0.00, -3.37, -0.17, 3.55, 3.65, -3.44)$, $\hat{\sigma}_b^2 = 1.24^2$, $\hat{\sigma}_v^2 = 1.16^2$, $\hat{\phi} = 0.36$; Model 1: $\hat{\alpha} = (4.29, 3.86, 3.81, -0.01, 2.53, 3.73, 3.80, 0.01, -0.02, 4.16)$, $\hat{\beta} = (-3.51, -1.79, -0.01, 3.77, -0.23, -3.51, -0.54, 3.39, 3.54, -3.74)$, $\hat{\rho} = 0.49$, $\hat{\sigma}_b^2 = 1.19^2$, $\hat{\sigma}_v^2 = 2.07^2$, $\hat{\phi} = 0.37$; Model 2: $\hat{p} = (0.11, 0.30, 0.39, 0.83, 0.68, 0.27, 0.42, 0.83, 0.83, 0.14)$, $\hat{\gamma}_1 = 3.69$, $\hat{\gamma}_2 = -0.44$, $\hat{\sigma}_b^2 = 0.86^2$, $\hat{\phi} = 0.41$; Model 3: $\hat{\alpha} = (4.10, 3.56, 3.37, -2.30, 2.27, 3.54, 3.41, -2.16, -2.43, 3.93)$, $\hat{\gamma}_1 = 2.92$, $\hat{\gamma}_2 = -1.13$, $\hat{\sigma}_b^2 = 1.20^2$, $\hat{\phi} = 0.42$.

Table 2. Posterior means and 95% credible intervals for ϕ , σ_b^2 , σ_v^2 , ρ , γ_1 and γ_2

	Model 0		Model 1		Model 2		Model 3	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
<i>Gamma distribution</i>								
ϕ	10.04	(2.90, 21.05)	9.61	(2.81, 20.65)	7.39	(2.06, 16.53)	8.80	(1.98, 20.43)
σ_b	1.26	(0.57, 2.82)	1.17	(0.54, 2.57)	0.83	(0.37, 1.56)	1.24	(0.45, 2.79)
σ_v	1.19	(0.10, 3.42)	1.93	(0.32, 5.03)				
ρ			0.47	(-0.93, 1.00)				
γ_1					3.76	(2.80, 4.47)	2.99	(0.92, 6.01)
γ_2					-0.41	(-0.88, -0.13)	-1.13	(-2.00, -0.54)
	Model 0		Model 1		Model 2		Model 3	
	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
<i>Lognormal distribution</i>								
ϕ	0.36	(0.22, 0.62)	0.37	(0.23, 0.72)	0.41	(0.24, 0.67)	0.42	(0.23, 0.91)
σ_b	1.24	(0.59, 2.75)	1.19	(0.52, 2.59)	0.86	(0.43, 1.62)	1.20	(0.24, 2.68)
σ_v	1.16	(0.11, 3.74)	2.07	(0.41, 5.29)				
ρ			0.49	(-0.94, 1.00)				
γ_1					3.69	(2.83, 4.57)	2.92	(0.89, 5.91)
γ_2					-0.44	(-1.05, -0.12)	-1.13	(-2.01, -0.85)

The same experimental design was used as in the example. Thus, datasets were generated with $T = 10$ treatments and $S = 8$ incomplete blocks with 5 treatments each. The allocation of treatments to blocks was the same as in the original dataset. For each dataset, $l = 1, \dots, 100$, a total of 40 observations were generated. This was done for each model.

Let \tilde{y}_{0t} denote the averaged marginal median $\tilde{y}_{g,t}$ of treatment t , defined in Sect. 5.1, computed using the original data of the example. Let \tilde{y}_{lt} denote the averaged marginal median $\tilde{y}_{g,t}$ of treatment t , computed in the same way but using the l th simulated dataset. Quality was assessed using the root-mean-square error (RMSE), here defined as

$$RMSE = \left[\frac{1}{L} \sum_{l=1}^L \left(\frac{1}{T} \sum_{t=1}^T (\tilde{y}_{0t} - \tilde{y}_{lt})^2 \right) \right]^{1/2}, \tag{15}$$

where $L = 100$ is the number of generated datasets and $T = 10$ is the number of treatments. The smaller the RMSE, the better the performance of the model.

Impact of the choice of model

To assess simulation objective (1), for each generated dataset l , the models were fitted independently of the model used for sampling. In other words, Model k was fitted to each of the 100 datasets generated using Model r , for $k = 0, 1, 2, 3$, and $r = 0, 1, 2, 3$. Table 3 presents the results. The rows of this table represent four different scenarios. For each scenario, the ranking of the four models is the same. Regardless how the data were generated, Model 2 was consistently the best model for analysing the data. The assumption of correlation between the block effects, which distinguishes Model 1 from Model 0, did not improve

Table 3. Evaluating model performance by means of the root-mean-square error

Data simulated	Model used for the analysis			
	Model 0	Model 1	Model 2	Model 3
From Model 0	43.4	43.7	42.4	46.2
From Model 1	52.0	52.8	50.4	53.2
From Model 2	35.7	36.1	33.0	38.0
From Model 3	52.2	54.0	47.3	54.1

the performance, since the RMSE was always larger for Model 1 than for Model 0. The dependency-extended Model 3, with treatment-by-block-specific probabilities of zero, was less successful than the other models. When the data were generated from Model 2 and this model was also used for the analysis, the smallest RMSE was achieved: RMSE = 33.0. The results of this simulation study indicate that the dependency-extended Model 2 is the best option for analysing experiments of this type.

Impact of the choice of prior distributions

Having identified Model 2 as the best model for the analysis, a study was performed to assess simulation objective (2), i.e. to investigate the sensitivity of the results to the choice of the prior distributions. For this study, the 100 datasets generated using Model 2 were used. Alternative priors to those specified in Sect. 4.1 were considered. Among the alternative priors, continuous uniform distributions, $U(0, 50)$ and $U(0, 100)$, were investigated, as recommended by Gelman (2006). Performance was measured using the RMSE specified in (15). Results are given in Table 4. The first row of this table includes the prior distributions that were used in this article, yielding RMSE = 33.0, as is previously shown in Table 3. Using other diffuse prior distributions does not change the results much. In just two of the investigated cases, the RMSE was smaller than 33.0. When a normal distribution with variance 1 was used for γ_1 and γ_2 , the RMSE was 32.8. Similarly, when a U-shaped prior for p was used instead of a continuous uniform distribution, RMSE was 32.8. In both these cases, more informative priors were substituted for diffuse priors. When there is no pre-knowledge, diffuse priors are preferred. The results of this study point at Model 2 being robust to the choice of diffuse prior distributions.

For supplementary information, Web Appendix C includes tables with posterior means computed using various sets of prior distributions, for all four models and for both distributions.

6. FINAL ANALYSIS OF THE WEED EXPERIMENT

Based on an overall assessment of the results from the cross-validation and the simulation studies of Sect. 5, modelling the median using Model 2 was considered the best option for the weed experiment. Using this model, treatments rank with regard to effectiveness as follows: 4, 8, 9, 5, 7, 3, 2, 6, 10 and 1, with posterior means for the probability of zero equal

Table 4. Evaluating the priors of Model 2 by means of the root-mean-square error

γ_1	γ_2	σ_b^2	ϕ	p	RMSE
N(0.01, 10)	N(0.01, 10) $I_{\gamma_2 < 0}$	Cauchy(0, 0.1) $I_{\sigma_b^2 > 0}$	IG(0.01, 0.01)	Beta(1, 1)	33.0
N(0.01, 1)	N(0.01, 1) $I_{\gamma_2 < 0}$	Cauchy(0, 0.1) $I_{\sigma_b^2 > 0}$	IG(0.01, 0.01)	Beta(1, 1)	32.8
N(0.01, 100)	N(0.01, 100) $I_{\gamma_2 < 0}$	Cauchy(0, 0.1) $I_{\sigma_b^2 > 0}$	IG(0.01, 0.01)	Beta(1, 1)	33.0
N(0.01, 10)	N(0.01, 10)	Cauchy(0, 0.1) $I_{\sigma_b^2 > 0}$	IG(0.01, 0.01)	Beta(1, 1)	33.3
N(0.01, 10)	N(0.01, 10) $I_{\gamma_2 < 0}$	IG(0.01, 0.01)	IG(0.01, 0.01)	Beta(1, 1)	33.0
N(0.01, 10)	N(0.01, 10) $I_{\gamma_2 < 0}$	Cauchy(0, 0.1) $I_{\sigma_b^2 > 0}$	IG(0.1, 0.1)	Beta(1, 1)	33.0
N(0.01, 10)	N(0.01, 10) $I_{\gamma_2 < 0}$	Cauchy(0, 0.1) $I_{\sigma_b^2 > 0}$	IG(0.001, 0.001)	Beta(1, 1)	33.0
N(0.01, 10)	N(0.01, 10) $I_{\gamma_2 < 0}$	Cauchy(0, 0.1) $I_{\sigma_b^2 > 0}$	U(0, 50)	Beta(1, 1)	33.3
N(0.01, 10)	N(0.01, 10) $I_{\gamma_2 < 0}$	Cauchy(0, 0.1) $I_{\sigma_b^2 > 0}$	U(0, 100)	Beta(1, 1)	33.3
N(0.01, 10)	N(0.01, 10) $I_{\gamma_2 < 0}$	Cauchy(0, 0.1) $I_{\sigma_b^2 > 0}$	IG(0.01, 0.01)	Beta(0.5, 0.5)	32.8
N(0.01, 10)	N(0.01, 10) $I_{\gamma_2 < 0}$	Cauchy(0, 0.1) $I_{\sigma_b^2 > 0}$	IG(0.01, 0.01)	Beta(2, 2)	33.6

to 0.83, 0.83, 0.83, 0.68, 0.42, 0.39, 0.30, 0.27, 0.14 and 0.11, respectively. Thus, treatments 4, 8 and 9 are the best treatments for eliminating creeping thistle weed, while treatment 1 is the worst, which was expected as treatment 1 is the control.

Unlike previously published two-part models (Tang et al. 2018; Cantoni et al. 2017; Rose et al. 2006; Tooze et al. 2002), Models 2 and 3 include an explicit functional relationship between the two parts of the model. Under Model 2, conditional on the random effects, the median is $\tilde{y}_{f,st} = \exp(\gamma_1 + \gamma_2 + b_s)$. Its expected value with respect to the random block effect is given by

$$\begin{aligned}
 E(\tilde{Y}_{f,st} | \gamma_0, \gamma_1, \alpha_t) &= E(E(\tilde{Y}_{f,st} | b_s, \gamma_0, \gamma_1, \alpha_t)) \\
 &= \exp(\gamma_1 + \gamma_2 \alpha_t + \sigma_b^2/2),
 \end{aligned}
 \tag{16}$$

because a $N(0, \sigma_b^2)$ variate b_s has moment generating function $E(\exp(t b_s)) = \exp(t^2 \sigma_b^2/2)$. Figure 3 displays the relationship between the probability of zero and the expected value of the conditional median using the posterior means $\hat{\gamma}_1 = 3.69$, $\hat{\gamma}_2 = -0.44$ and $\hat{\sigma}^2 = 0.86$. Since $\hat{\gamma}_2$ is negative, the probability of zero decreases with the conditional median, as illustrated by the curve. As examples, the dashed and dotted lines indicate the probabilities of zero and conditional medians for treatments 1 and 2, respectively. Although the expected value (16) of the conditional median depends on the block variance, the ratio between the expected values of any two treatments is independent of this variance. For example, the ratio between the expected values of the conditional medians for treatments 2 and 1 is, by (16),

$$\frac{E(\tilde{Y}_{f,st} | \gamma_0, \gamma_1, \alpha_2)}{E(\tilde{Y}_{f,st} | \gamma_0, \gamma_1, \alpha_1)} = \exp(\gamma_2(\alpha_2 - \alpha_1)) = \left(\frac{p_2(1 - p_1)}{p_1(1 - p_2)} \right)^{\gamma_2},
 \tag{17}$$

since $\alpha_t = \log(p_t/(1 - p_t))$ by (9). The posterior mean of the slope γ_2 in (8) is $\hat{\gamma}_2 = -0.44$, and the posterior means of the probabilities p_1 and p_2 in (9) are $\hat{p}_1 = 0.11$ and $\hat{p}_2 = 0.30$,

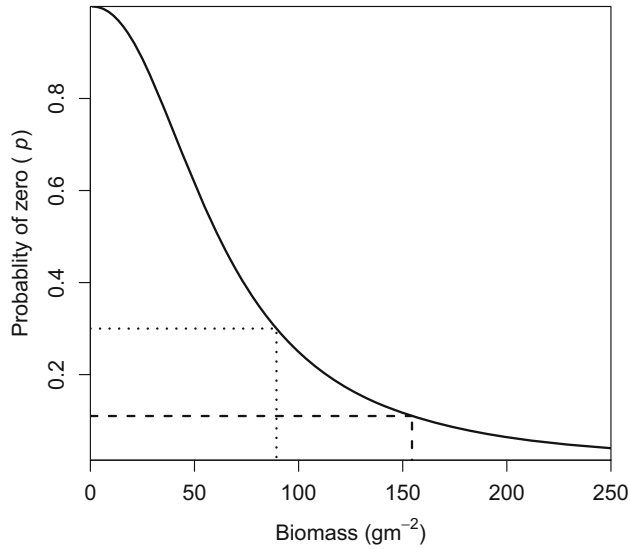


Figure 3. The probability of zero as a function of the conditional median, using Model 2. The dashed and dotted lines indicate treatments 1 and 2, respectively.

respectively. Using these posterior means for computation of (17), the expected conditional median of treatment 2 is 58% of the expected conditional median of treatment 1.

Model 2 provides scientific insights into the performance of the experimental treatments. Specifically, it is not correlation between block effects that causes dependency between the two parts of the model, as would have been assumed using Model 1. According to Model 2, the experimental treatments have varying probabilities of zero and varying conditional means, following the relationship shown in Fig. 3.

7. DISCUSSION

This article introduced dependency-extended two-part models for modelling nonnegative continuous observations in case of zero inflation, heteroscedasticity and skewness. Such data are encountered in many different scientific disciplines; however, we studied specifically an agricultural field experiment with herbicidal treatments. New features were modelling of medians instead of means and modelling of functional dependencies between the parts (Models 2 and 3). The approach of parts correlated through random effects (Cantoni et al. 2017), which was used in Model 1, does not work well when some levels contain only zeros. When this happens, 95% credible intervals for conditional means or medians become very large. The same phenomenon occurs using a traditional two-part model with uncorrelated parts (Model 0). Note that in weed control experiments, it is very common that only zeros, i.e. no weed, are observed for highly effective herbicides. Furthermore, if the seeds of the weed are windblown, the weed can appear in patches on the field and be missing in some incomplete blocks.

Using Model 1, the level of the positive observations, i.e. the mean or the median, and the probability of zero are modelled as functions of fixed and random effects, where the random effects are allowed to be correlated. This is the model that is difficult to fit when only zeros are observed for some levels of the fixed effects or some levels of the random effects. In such cases, using the proposed Bayesian method, the vague prior information completely dominates the likelihood. Using Model 2, the level of the positive observations is modelled as a function of the probability of zero and a random effect, whereas using Model 3, the probability of zero is modelled as a function of the level of the positive observations, which is dependent on a random effect. Models 2 and 3 have the advantage, as compared to Models 0 and 1, that all information is considered in the construction of the joint posterior distribution. These models managed the problem of zero excess better than Model 1. For our crop protection experiment, the cross-validation and the simulation studies indicated that Model 2 performed best when modelling the median using the lognormal distribution.

Dependency-extended two-part models are most useful when there are several predictors. In this event, which includes our case of modelling effects of categorical variables, an assumption of a functional relationship between the two parts enables a more parsimonious model than a model without such an assumption.

All two-part models have the advantage that the positive part and the probability of zero are separated. Thus, two-part models enable assessment of the probability of zero and the conditional means or medians, not just marginal means or medians. In addition, the proposed dependency-extended two-part models make it possible to show graphically how the probability of zero decreases with the level of the positive observations. In weed research, this means that if much weed were observed in the experiment, then the probability is low that the herbicide will kill the weeds completely. This is presumably not because of correlation between block effects (Model 1), but rather because there is some functional relationship between biomass and probability of zero (Models 2 and 3).

Instead of using Bayesian methodology, maximum likelihood estimation can be used. However, if only zeros have been observed for some of the treatments, then the variances in the estimates of the effects of those treatments cannot be computed. For example, the `nlmixed` procedure of the SAS System gives the error message that the Hessian matrix is not positive definite and therefore the estimated covariance matrix may be unreliable.

The underlying problem is the following. The maximum likelihood estimator of a probability, p , is $\hat{p} = y/n$, where y is $\text{Bin}(n, p)$. When $y = 0$ is observed, then $\hat{p} = 0$, which is unrealistic and undesirable. The variance of \hat{p} is usually obtained from the Hessian using the parameter estimate. This variance is $\hat{p}((1 - \hat{p}))/n$, which is also 0 when $\hat{p} = 0$, providing poor information about the precision in \hat{p} . The maximum likelihood estimator does not perform well on the boundary of the parameter space.

Through the use of Bayesian statistics, this problem is avoided. By multiplying the likelihood by a prior distribution for p , a posterior distribution is obtained which does not have all density concentrated at 0. Although the area under the posterior distribution is largest in a vicinity of 0, the posterior distribution does not exclude that p is greater than 0.

Bayesian analysis has the big advantage that statistical inference can be made for any function of the parameters, through sampling from the posterior distribution. In experiments

without pre-knowledge, care should be taken to select vague priors. However, the value of the Bayesian analysis is even larger if prior information is taken into account.

Our models can be extended further by allowing positive distributions that are not members of the exponential family. Other link functions than the log and the logit can be used, e.g. the identity instead of the log or the probit instead of the logit. In larger agricultural field experiments, it would be possible to add effects of complete replicates. For other applications than analysis of incomplete block experiments, the models must be adjusted by including other explanatory variables. In zero-augmented spatial-temporal two-part models, our dependency-extended idea could be used to relate the components of the auto-covariance function, thereby reducing the number of parameters.

In summary, this article introduced two-part models with dependencies between the probability of zero and the level of the positive part and showed how these can be analysed using Bayesian methodology. Particularly Model 2, expressing the median of the lognormal distribution as a function of the probability of zero and a random effect, performed well. We believe the basic construction of this model can be used successfully also in other areas of research.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funding Open access funding provided by Swedish University of Agricultural Sciences. Rodrigues-Motta thanks for the partial support of FAPESP, Brazil, grant #2014/02211-1.

[Received September 2020. Revised June 2021. Accepted July 2021. Published Online August 2021.]

REFERENCES

- Bar-Lev SK, Reiser B (1982) An exponential subfamily which admits UMPU tests based on a single test statistic. *Ann Stat* 10:979–989
- Bertoli W, Conceição KS, Andrade MG, Louzada F (2020) A Bayesian approach for some zero-modified Poisson mixture models. *Stat Model* 20:461–501
- Besag J, Higdon D (1999) Bayesian analysis of agricultural field experiments. *J R Stat Soc B* 61:691–746
- Biswas J, Das K (2020) A Bayesian approach of analysing semi-continuous longitudinal data with monotone missingness. *Stat Model* 20:148–170
- Bose A, Boukai B (1993) Sequential estimation results for a two-parameter exponential family of distributions. *Ann Stat* 21:484–502
- Cantoni E, Flemming JM, Welsh A (2017) A random-effects hurdle model for predicting bycatch of endangered marine species. *Ann Appl Stat* 11:2178–2199
- Carroll RJ, Ruppert D (1988) Transformation and weighting in regression. Chapman and Hall, New York
- Chen SX, Qin J (2003) Empirical likelihood-based confidence intervals for data with possible zero observations. *Stat Probab Lett* 65:29–37

- Cowles MK, Carlin BP (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. *J Am Stat Assoc* 91:883–904
- Damesa TM, Möhring J, Forkman J, Piepho HP (2018) Modeling spatially correlated and heteroscedastic errors in Ethiopian maize trials. *Crop Sci* 58:1575–1586
- Donald M, Alston CL, Young RR, Mengersen KL (2011) A Bayesian analysis of an agricultural field trial with three spatial dimensions. *Comput Stat Data Anal* 55:3320–3332
- Duan N, Manning WG, Morris CN, Newhouse JP (1983) A comparison of alternative models for the demand for medical care. *J Bus Econ Stat* 1:115–126
- Feuerverger A (1979) On some methods of analysis for weather experiments. *Biometrika* 66:655–658
- Forkman J, Piepho HP (2013) Performance of empirical BLUP and Bayesian prediction in small randomized complete block experiments. *J Agric Sci* 151:381–395
- Fuentes M, Reich B, Lee G (2008) Spatial-temporal mesoscale modeling of rainfall intensity using gage and radar data. *Ann Appl Stat* 2:1148–1169
- Gelman A (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal* 1:515–534
- Goldberger AS (1968) The interpretation and estimation of Cobb-Douglas functions. *Econometrica* 36:464–472
- Harvey J, Van der Merwe A (2012) Bayesian confidence intervals for means and variances of lognormal and bivariate lognormal distributions. *J Stat Plan Inference* 142:1294–1309
- Hautsch N, Malec P, Schienle M (2013) Capturing the zero: a new class of zero-augmented distributions and multiplicative error processes. *J Financ Econom* 12:89–121
- King C, Song JJ (2019) A Bayesian two-part quantile regression model for count data with excess zeros. *Stat Model* 19:653–67
- Koch AL (1969) The logarithm in biology: II. Distributions simulating the lognormal. *J Theor Biol* 23:251–268
- Lee Y, Nelder JA, Pawitan Y (2006) Generalized linear models with random effects: unified analysis via H-likelihood. CRC, Boca Raton
- Limpert E, Stahel WA, Abbt M (2001) Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability - normal or log-normal: that is the question. *Bioscience* 51:341–352
- Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D (2012) The BUGS book. CRC, Boca Raton
- Mills ED (2013) Adjusting for covariates in zero-inflated gamma and zero-inflated log-normal models for semi-continuous data. Ph.D. (Doctor of Philosophy) thesis, University of Iowa
- Min Y, Agresti A (2005) Random effect models for repeated measures of zero-inflated count data. *Stat Model* 5:1–19
- Moulton LH, Curriero FC, Barroso PF (2002) Mixture models for quantitative HIV RNA data. *Stat Methods Med Res* 11:317–325
- Musal RM, Ekin T (2017) Medical overpayment estimation: a Bayesian approach. *Stat Model* 17:196–222
- Neelon BH, O'Malley AJ, Normand SLT (2010) A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Stat Model* 10:421–439
- Neuhaus JM, McCulloch CE, Boylan RD (2018) Analysis of longitudinal data from outcome-dependent visit processes: failure of proposed methods in realistic settings and potential improvements. *Stat Med* 37:4457–4471
- Ntzoufras I (2011) Bayesian modeling using WinBUGS, vol 698. Wiley, Hoboken
- Patterson HD, Williams ER (1976) A new class of resolvable incomplete block designs. *Biometrika* 63:83–92
- Piepho HP, Edmondson R (2018) A tutorial on the statistical analysis of factorial experiments with qualitative and quantitative treatment factor levels. *J Agron Crop Sci* 204:429–455
- Piepho HP, Büchse A, Emrich K (2003) A hitchhiker's guide to mixed models for randomized experiments. *J Agron Crop Sci* 189:310–322

- Rao KA, D’Cunha JG (2016) Bayesian inference for median of the lognormal distribution. *J Mod Appl Stat Methods* 15, Article 32
- Rodrigues-Motta M, Galvis Soto DM, Lachos VH, Vilca F, Baltar VT, Junior EV, Fisberg RM, Lobo Marchioni DM (2015) A mixed-effect model for positive responses augmented by zeros. *Stat Med* 34:1761–1778
- Rose CE, Martin SW, Wannemuehler KA, Plikaytis BD (2006) On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *J Biopharm Stat* 16:463–481
- Singh M, Al-Yassin A, Omer SO (2015) Bayesian estimation of genotypes means, precision, and genetic gain due to selection from routinely used barley trials. *Crop Sci* 55:501–513
- Sun Y, Stein ML (2015) A stochastic space-time model for intermittent precipitation occurrences. *Ann Appl Stat* 9:2110–2132
- Swallow B, Buckland ST, King R, Toms MP (2016) Bayesian hierarchical modelling of continuous non-negative longitudinal data with a spike at zero: an application to a study of birds visiting gardens in winter. *Biom J* 58:357–371
- Tang W, He H, Wang W, Chen D (2018) Untangle the structural and random zeros in statistical modelings. *J Appl Stat* 45:1714–1733
- Theobald CM, Talbot M, Nabugoomu F (2002) A Bayesian approach to regional and local-area prediction from crop variety trials. *J Agric Biol Environ Stat* 7:403–419
- Thomas A, O’Hara B, Ligges U, Sturtz S (2006) Making bugs open. *R News* 6:12–17
- Tiao GC, Draper N (1968) Bayesian analysis of liner models with two random components with special reference to the balanced incomplete block design. *Biometrika* 55:101–117
- Tooze JA, Grunwald GK, Jones RH (2002) Analysis of repeated measures data with clumping at zero. *Stat Methods Med Res* 11:341–355
- Verdooren LR (2020) History of the statistical design of agricultural experiments. *J Agric Biol Environ Stat* 25:457–486
- Yang Y, Wang HJ, He X (2016) Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. *Int Stat Rev* 84:327–344
- Zhou X-H, Tu W (2000) Interval estimation for the ratio in means of lognormally distributed medical costs with zero values. *Comput Stat Data Anal* 35:201–210

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.