# Generalized Linear Mixed Models can detect unimodal species-environment relationships

Niche theory predicts that species occurrence and abundance show non-linear, unimodal relationships with respect to environmental gradients. Unimodal models, such as the Gaussian (logistic) model, are however more difficult to fit to data than linear ones, particularly in a multi-species context in ordination, with trait modulated response and when species phylogeny and species traits must be taken into account. Adding squared terms to a linear model is a possibility but gives uninterpretable parameters.

This paper explains why and when generalized linear mixed models, even without squared terms, can effectively analyse unimodal data and also presents a graphical tool and statistical test to test for unimodal response while fitting just the generalized linear mixed model. R-code for this is supplied in supplementary information.

2 Tahira Jamil[1,2] and Cajo J. F. ter Braak[1,*]

4 [1]Biometris, Wageningen University and Research Centre, Box 100, 6700 AC Wageningen,

5 the Netherlands.

6 [2] Department of Mathematics, COMSATS Institute of Information Technology, Park Road,

7 Islamabad, Pakistan

8 [*] corresponding author. Tel.:+31 317 480803 Fax +31 317 483554 E-mail address:

9 cajo.terbraak@wur.nl

## Introduction

15 Niche theory predicts that species occurrence and abundance show non-linear, unimodal

16 relationships with respect to environmental gradients (Austin 1987; Palmer & Dixon 1990;

17 Whittaker 1967). Many studies fail to test for unimodal response (Austin 2007). Thus

18 straight-line relationships are often fitted without justification (e.g. Gibson et al.(2004) ).

19 Pollock et al. (2012) propose a generalized linear mixed model for investigating trait

20 modulation of the environmental response of a number of species. In their data unimodal

21 response was said to be precluded, presumably as they examined relatively short

22 environmental gradients. But might their method or a small modification thereof  have

23 worked for unimodal response? Ives & Helmus (2011) recently proposed phylogenetic

24 generalized linear mixed models. Can these models usefully analyze unimodal response?

25 A similar question arises in community ecological ordination, a class of multivariate methods

26 to analyze the occurrence and/or abundance of  a set of species in a set of sites and resulting

27 in a configuration of the sites in a factorial plane, the directions of which can be interpreted as

28 latent environmental variables (Jongman et al. 1995; ter Braak & Prentice 2004; Walker &

29 Jackson 2011). Principal component analysis and redundancy analysis are linear ordination

30 methods whereas (detrended) correspondence analysis and canonical correspondence analysis

31 are claimed to be able to analyze unimodal response (ter Braak 1985; ter Braak 1986).

32 Nevertheless, (canonical) correspondence analysis is an eigen vector method and therefore

33 inherently linear. This is most apparent in the reconstitution formula of (canonical)

34 correspondence (Greenacre 1984; ter Braak & Verdonschot 1995). How can it be understood

35 that these methods are able to model unimodal data but are inherently linear?

36  Some insight in this question is given by Ihm & van Groenewoud (1984) and further worked

37  out by ter Braak (1987) and de Rooij (2007) who showed the relation between the unimodal

38  model and a generalized (bi)linear model, also known as Goodman's RC model. The relation

39  can be used both ways. Ihm & van Groenewoud (1984) use the relationship to justify their

40  model B which is a (bi)linear model), for ecological ordination and de Rooij (2007) uses it to

41  transform the linear predictor of the RC model into a quadratic form, with the graphical

42  purpose to transform a vector representation or biplot to a distance representation that is

43  supposed to be easier to interpret for naïve users of multivariate methods. Additional insight

44  is given by Zhu et al.(2005) who introduced a weighted sample model to show the

45  equivalence of canonical correspondence analysis and linear discriminant analysis.

46  In this paper we propose a graphical tool and statistical test to test for unimodal response

47  while fitting just a generalized linear mixed model (GLMM) without squared terms. GLMMs

48  are model-based, inferential statistical tools. GLMMs are very useful for describing the

49  community patterns and are becoming popular in ecological and evolutionary studies (Bolker

50  et al. 2009; Ives & Helmus 2011; Zuur et al. 2009). Even when unimodality is detected using

51  the proposed tool and test, we claim that GLMMs can effectively analyze unimodal data

52  when the niche width is not very different among species and illustrate this claim by

53  comparing the GLMM approach with an explicit unimodal model approach on data that show

54  unimodal response. We focus on explicit environmental variables. In this case of direct

55  gradient analysis we can of course add the squares of environmental variables to the model

56  (ter Braak & Looman 1986) and test the statistical significance of the addition, but the

57  argument extends to the trait modulated response (Jamil et al. 2012) and to latent variable

58  models (Walker & Jackson 2011) for which models without squared terms are easier to

59  handle with software that is more widely available.

60

61    Material  & Methods

62    Generalized linear mixed models and unimodal response

63

64    For ease of exposition we start with a logistic linear mixed model for presence-absence data

65    as example GLMM. The same approach can be followed for count data and loglinear models,

66    which would relate to the RC model (de Rooij 2007). Consider the logistic linear mixed

67    model that relates the probability of occurrence $p_{ij}$ of species $j$ in site $i$ to a quantitative

68    environmental variable $x_i$ by the formula

69    $$logit\left(p_{ij}\right)=\alpha_j+\beta_j x_i+\gamma_i \qquad (i=1,\ldots, n; j = 1,\ldots, m).\qquad\qquad (1)$$

70    with $\alpha_j$ an intercept, $\beta_j$ a slope and $\gamma_i$ a site effect, which we all take as random

71    parameters with normal distributions with zero mean and variances $\sigma_\alpha^2$ , $\sigma_\beta^2$ and $\sigma_\gamma^2$ .

72    In this random intercept, random slope model (Gelman & Hill 2007; Zuur et al. 2009) it is

73    prudent to have an additional parameter $\rho$ for the correlation between the intercepts $\alpha_j$

74    and slopes $\beta_j$ ; otherwise the model would change by just centering the environmental

75    variable. Inclusion of the random site effects $\gamma_i$ are a means to avoid pseudoreplication

76    (Hurlbert 1984) as they introduce correlation among species. These correlations were not

77    modelled by Pollock et al. (2012) which makes their statistical tests liberal. The site effects

78    may account for the size of the site, the fertility of the site or any other unknown factors that

79    influence the probability of occurrence of all species in the site. The site effect $\gamma_i$ will thus

80  be expected to be related to the expected number of species in a site, that is to $\sum_j p_{ij}$ and,

81  in terms of the data, to the number of species that is observed in a site, for short the site total,

82  defined as $S_i = \sum_j y_{ij}$ . The site total and the site effect are expected to have a monotonic

83  positive relationship.

84  We now turn to one of the simplest unimodal curves for presence-absence, the Gaussian

85  logistic curve (Oksanen et al. 2001; ter Braak & Looman 1986)

86  
$$logit\left(p_{ij}\right) = a_j - \frac{\left(x_i - u_j\right)^2}{2t_j^2} \tag{2}$$

87  with $a_j$ a coefficient related to maximum probability of occurrence, $u_j$ the species

88  optimum and $t_j$ the tolerance of species $j$. This model has thus a logistic form but is

89  nonlinear in this parameterization.  By expanding the quadratic term,

90  
$$a_j - \frac{\left(x_i - u_j\right)^2}{2t_j^2} = a_j - \frac{1}{2t_j^2}x_i^2 - \frac{1}{2t_j^2}u_j^2 + \frac{1}{t_j^2}x_i u_j = \left(a_j - \frac{1}{2t_j^2}u_j^2\right) + \left(\frac{u_j}{t_j^2}\right)x_i - \frac{1}{2t_j^2}x_i^2 \tag{3}$$

91  it can be fitted to data of each individual species by a generalized linear model (GLM) by

92  using $x$ and $x^2$ as predictors (Jongman et al. 1995; ter Braak & Looman 1986). The fit is

93  unimodal if the regression coefficient of $x^2$ is negative.

94  Might we be able to model unimodal response even without the squared term? On assuming

95  $t_j = t$ and setting

96  
$$\alpha_j = a_j - \frac{1}{2t^2}u_j^2 \quad , \quad \beta_j = \frac{u_j}{t^2} \quad \text{and} \quad \gamma_i = \frac{-1}{2t^2}x_i^2 \quad , \tag{4}$$

97   we obtain the equation (1) again. If $t$   would vary among species then equation (1) does

98   not exactly hold because $x_i^2/t_j^2$   then also depends on *j*. With equal tolerances, unimodal

99   response can thus be represented by a simple linear model with site effects and, as we

100  propose, be fitted by a GLMM based on equation (1). The GLMM has additional normality

101  assumptions. In case of unimodal response, the assumption that the site effects { $\gamma_i$  } are

102  independent normal is false, as the site effects then depend on $x_i$   through $(x_i^2)/t^2$   . This

103  will be the basis of our test on unimodal response in the next section. The site effects in

104  equation (4) also have a nonzero mean, but that is not a problem, as the mean can be taken

105  out and transferred to the intercepts $\alpha_j$   .

106  The unimodal model with two or more environmental variables (ter Braak & Prentice 1988)

107  can similarly be rewritten as a simple linear model without squared terms if the tolerances are

108  equal (see Appendix). In conclusion, up to distributional assumptions, the GLMM (e.g.

109  equation (1)) can be interpreted as a Gaussian logistic model with equal tolerances for the

110  species.

111  **A graphical tool and statistical test for unimodal response**

112  Equation (4)  suggests a graphical tool for detecting unimodal response and also a statistical

113  test. The idea is to fit a GLMM to the binary data {$y_{ij}$} with respect to the environmental

114  variable with values  {$x_i$} (*i* = 1,…, *n*). In the R package lme4  (Bates et al. 2011) ,  the model

115  can be fitted by

116  ```
lmer (y ~ 1+ x + (1+ x | sp) + (1| site),
```

117  ```
family=binomial(link="logit") ,data),
```

118 where y represents the vectorized response data while `sp` and `site` are factors indicating

119 species and sites. The site effects $\gamma_i$ obtained from the fit are then plotted against the

120 environmental variable $\{x_i\}$. There is an indication of unimodal response in terms of the

121 species response with respect to the environmental variable $x$ if this graph shows a n-shaped

122 (as opposed to u-shaped) quadratic relationship. If the shape is not quadratic but curved, a

123 transformation of x may improve it.

124 In the statistical test for unimodal response, the null model is the GLMM of equation (1)

125 with, specifically, independent and normally distributed site effects. The alternative model is

126 that the site effects depend quadratically on the environmental variable. As the site effects

127 typically also depend on the site total $S$, a sensitive test on unimodal response is obtained by

128 regressing the site effects on $x$, $x^2$ and $S$, according to the model formula

129 $$\gamma \sim x + x^2 + S \ .$$ (5)

130 There is evidence of unimodal response if the squared term is significant as judged by a z-test

131 or, equivalently, an anova test on its regression coefficient, the null model being $\gamma \sim$

132 $x + S$ . R-code for making the graph and performing the test on unimodality is supplied in

133 supplementary information.

## Simulated data

135

136 In the first example series the procedure to simulate data is the following:

137     1)  Generate $n=50$ values of an environmental variable $x$ as a random sample from the

138         uniform distribution, $x \sim U(-2,2)$ .

139    2) Generate a vector $u$ of length *m* from a uniform distribution $U(-\tau,\tau)$, where

140        $\tau=2+t$, for a fixed value of *t,* to ensure that optima are also placed outside the

141        sample range of *x*.

142    3) Generate a vector $a$ of length *m* drawn at random from the standard normal

143        distribution.

144    4) Generate binomial probabilities    from the unimodal response curve

$$p_{ij}=logit^{-1}\left(a_j-\frac{(x_i-u_j)^2}{2t_j^2}\right)$$    (6)

146    where $y_{ij}$ presence–absence data were generated at random from a Bernoulli distribution

147    with probability $p_{ij}$ and $t_j=t$. We simulate data with constant tolerance    in each

148    data set for *m*=100 species and vary $t$ between data sets (*t* = 0.5, 1 and 4). Fig. 1 indicates

149    how the simulated species response curves look like in each data set. We repeated the

150    example with *x* and *u* having normal distributions instead (*x* ~ $N(0,1)$ and *u* ~ $N(0,t^2)$

151    ), but do not show the results as they were very similar to the uniformly distributed case.

152    In the second example series, we vary the tolerance lognormally among species, *t* ~

153    $LogN(0,\sigma)$ with $\sigma$ = 0.25, 0.5 and 1 so that the median tolerance is 1. For $\sigma=1$,

154    the coefficient of variation is larger than 100%. In the third example series we vary the

155    number of species (*m* =10, 50 and 100) with $t$ =1. For the rest the simulation is set up as

156 in the first series. We also simulated data according equation (1) with the strict assumptions

157 of the GLMM of independent normal site effects.

158 Each dataset was characterized by beta diversity and length of gradient. The index of beta

159 diversity is $\beta_w = T/\acute{S} - 1$ , where $T$ is the total number of species, and $\acute{S}$ is the average

160 number of species per site (Whittaker 1960). Length of gradient, obtained by analyzing the

161 data with detrended correspondence analysis (DCA), is expressed in standard deviation (S.D.)

162 units (Hill & Gauch 1980). Values greater than 4 are commonly taken to indicate unimodal

163 response. Beta diversity was calculated using the asbio package (Aho 2011) and DCA was

164 performed in the *vegan* package (Oksanen et al. 2011), both in R.

165 **Real data**
166

167 We illustrate our method with three real data sets. The first is the Dune Meadow data

168 (Jongman et al. 1995). This is a small data set of 28 higher plants in 20 sites in a dune area in

169 the Netherlands. Environmental variables, related to soil and management, were measured at

170 each site; we use the variable Moisture.

171 The second data set is about the vegetation of the rising seashore on the island Skabbholmen

172 in the Stockholm archipelago, eastern central Sweden (Cramer & Hytteborn 1987) and is part

173 of the Canoco package (ter Braak & Smilauer 1998). The data set consists of 63 sites

174 sampled in both 1978 and 1984 and contains 68 species. The environmental variable is

175 Elevation.

176 The third data set involves phytoplankton communities of 203 lakes located within four

177 climate zones and associated measurements on various environmental variables and

178 morphological species traits of 60 species (Kruk et al. 2011). We consider the environmental

179 variables Temperature.

180   For each data set we fit a GLMM according to equations (1) with $x$ the noted environmental

181   variable, plot the resulting site effects against $x$ and test for unimodal response at the 5%

182   significance level explained by and below equation (5). For the sea-shore data, we analyze

183   years 1978 and 1984 separately. We also compare the regression coefficient $\beta_j$ as

184   estimated by GLMM with the optimum $u_j$ as obtained by explicitly fitting equation (2)

185   using GLM (ter Braak & Looman 1986) for species with a well-defined optimum, that is, the

186   squared term of which has a z-ratio smaller than -1 in the GLM model. In the small Dune

187   Meadow data set we use z-ratio <0.

188

189   Results

190

191   **Simulated data**

192   Fig. 1 shows the simulated response curves in example series 1; the sampled range of the

193   environmental variable is the range of $x$ shown. With increasing tolerance the part of the

194   curves that is sampled shows less unimodal response. This is expressed quantitatively in the

195   length of gradient SD units which varies between about 1 SD (not so unimodal) to 6 SD (very

196   unimodal); the beta diversity varies correspondingly between 1 and 5.

197   In Fig. 2, the site effects estimated by the GLMM analysis of each of the simulated data sets

198   are plotted against the environmental variable. In all three series, site effects shows a clear

199   quadratic relationship with the environmental variable except for large tolerance ( $t=4$ ) in

200   the first series. Note the decreasing range of sites effects as the tolerance increases in the first

201   row of Fig. 2; for large tolerance, the site effects are close to zero. Nevertheless, the squared

202   term in equation (5) was significant in all examples (P<0.001) so that the method detects

203  unimodal response even if it is moderate ($t = 4$). When the data are simulated according to

204  equation (1) with the strict normality assumptions of the GLMM, the squared term was not

205  judged significant more often than expected on the basis of  Type I error of the test.

206  Fig. 3 shows the relationship between the random slopes ( $\beta_j$ ) estimated by GLMM and

207  the true optima ( $u_j$ ) in the three simulation series. The relation is positive as predicted by

208  equation (4). The relationship is weaker the larger the tolerance (Fig. 3, first row). With

209  tolerance varying across species, the relationship continues to hold true surprisingly well (Fig

210  3, second row), except perhaps when the coefficient of variation of the tolerance is large

211  (>100%). The larger the number of species the clearer the predicted relationships (last row of

212  Figs 2 and 3).

213  Real data example

214  The site effects estimated by GLMM show a quadratic relationship with the noted

215  environmental variable in each of the three data sets (Fig. 4 top row), with the least unimodal

216  response in the Dune Meadow data. Unimodal response is significant in all cases, as judged

217  by our proposed significance test (P<0.001) and the relationship between the random slopes (

218  $\beta_j$ ) estimated by GLMM and the optima ( $u_j$ ) obtained from a fit of the unimodal

219  model of equation (1) is close to linear (Fig. 4, bottom row). In the Dune Meadow data there

220  is one outlier for a species with a z-ratio close to 0. In the phytoplankton data, the species

221  with similar low values for the optimum received differential values for the slope, but

222  otherwise there is a good agreement.

223

## Discussion

<sub>224</sub>

<sub>225</sub>

<sub>226</sub> Our analysis of multi-species data sets showed that a GLMM without squared terms but with

<sub>227</sub> site effects is able to detect unimodal response. The theory required equal tolerances among

<sub>228</sub> species, but the simulations showed remarkable robustness to this assumption. With

<sub>229</sub> significant unimodal response, as judged by our test, the assumption of independence and

<sub>230</sub> normality of the site effects underlying the GLMM is clearly violated.  This result can be

<sub>231</sub> interpreted in two ways. The first way is to try and adapt the model so that the assumptions

<sub>232</sub> are no longer (grossly) violated, for example, by extending the GLMM with an explicit

<sub>233</sub> square of the environmental variable as fixed effect, yielding

<sub>234</sub>
$$logit\left(p_{ij}\right)=\alpha_j+\beta_{1j}x_i+\beta_2 x_i^2+\gamma_i \tag{7}$$

<sub>235</sub> and then testing the normality assumption on the new site effects, for example by making a

<sub>236</sub> Q-Q plot (Gelman & Hill 2007). The site effects $\gamma_i$ obtained from the fit can then plotted

<sub>237</sub> against the environmental variable to assess whether or not the unimodality is adequately

<sub>238</sub> modeled using quadratic terms. The extended model still assumes constant tolerance (as does

<sub>239</sub> equation (1)) as it can be rewritten as

<sub>240</sub>
$$logit\left(p_{ij}\right)=a_j-\frac{\left(x_i-u_j\right)^2}{2t^2}+\tilde{\gamma}_i \tag{8}$$

<sub>241</sub> with $t=1/\sqrt{-2\beta_2}$ , $u_j=t^2\beta_j$ , $a_j=\alpha_j+\frac{1}{2t^2}u_j^2$ , and $\tilde{\gamma}_i=\gamma_i+\frac{1}{2t^2}x_i^2$ . $\qquad$ (9)

<sub>242</sub> We can go one step further and test the assumption of equi-tolerance by adding the squared

<sub>243</sub> term $x^2$ also as a random (species-dependent) component to equation (7) and testing the

244 significance of this extra variance component. In the R package lme4  (Bates et al. 2011),  the

245 two models to compare are (with xx = x2)

```
246 lmer (y ~ 1+ x + xx + (1+ x | sp) + (1| site)
247          family=binomial(link="logit") ,data)
248 and
249 lmer (y ~ 1+ x + xx + (1+ x + xx| sp) + (1| site)
250          family=binomial(link="logit") ,data).
251
```

252 We are currently investigating robust ways of testing the equi-tolerance assumption on the

253 basis of these two models, as the default ANOVA test has inflated type I error for small *t*.

254 The second way of interpretation is to conclude that the GLMM based on equation (1) is

255 remarkably robust to the normality assumption on the site effects and the equi-tolerance

256 assumption and that it can be used as a basis of more complicated models, such as the trait

257 modulated response model (Jamil et al. 2012; Pollock et al. 2012) and latent variable models

258 (Walker & Jackson 2011), even for unimodal response. In this way, the site effects are treated

259 as nuisance parameters and the independence and normal distributions, needed for efficient

260 computation, as a prior distribution that a posteriori may turn out to be false. Cormont et al

261 (2011) used this type of rationale to claim that their linear trait-environment method is well

262 suited to analyse unimodal response.

263

264 Conclusions

265 Site effects in multi-species GLMM serve three purposes in ecological data, first to avoid

266 pseudoreplication (Hurlbert 1984), second to account for differences in species richness

267 among sites (Ives & Helmus 2011) and third (this paper) to allow for common nonlinear,

268 unimodal response. The results of this paper imply that the phylogenetic generalized linear

269 mixed models of Ives & Helmus (2011) are not in conflict with niche theory as they include

270 random site effects and can thus deal with unimodal response.

271  Walker & Jackson (2011)  used a latent variable approach to test for unimodal response. We

272  tried their approach with the phytoplankton data, but failed to get an answer because the

273  program for fitting the quadratic model crashed. We focussed this paper on the easier task of

274  detection unimodal response to measured environmental variables by using a GLMM without

275  squared terms. To scale up to a latent variable approach we need factor analytic structure

276  within a GLMM. This already exists for linear mixed models (Thompson et al. 2003; Verbyla

277  et al. 2003) and it is a matter of time that it becomes standardly available for GLMM.  This

278  paper shows the utility of such factor analytic models in ecology (Walker & Jackson 2011) if

279  they allow additional random site effects.

280  A GLMM with terms that are linear in quantitative predictors is, of course, linear. But with

281  random site effects included, GLMMs can detect and fit unimodal response, with the

282  provision that the differences in niche widths among species is not too large.  The application

283  scope of GLMM in ecology is thus much broader than one might think at first glance.

284

285  Acknowledgements

286
287  This paper benefited from input from Mike Palmer, Steve Walker and Mu Zhu.

288
289
290  References

291  Aho K. 2011. asbio: A collection of statistical tools for biologists. R package version 0.3-36.

292      http://CRANR-projectorg/package=asbio.

293  Austin M. 2007. Species distribution models and ecological theory: A critical assessment and

294      some possible new approaches. *Ecological Modelling* 200:1-19.

295  Austin MP. 1987. Models for the analysis of species' response to environmental gradients.

296      *Plant Ecology* 69:35-45.

297     Bates D, Maechler M, and Bolker B. 2011. lme4: Linear mixed-effects models using S4

298          classes. R package version 0.999375-39,.

299          *http://cranr-projectorg/web/packages/lme4/indexhtml*.

300     Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, and White JSS.

301          2009. Generalized linear mixed models: a practical guide for ecology and evolution.

302          *Trends in Ecology and Evolution* 24:127-135.

303     Cormont A, Vos CC, van Turnhout CAM, Foppen RPB, and ter Braak CJF. 2011. Using

304          life-history traits to explain bird population responses to increasing weather variability

305          *Climate Research* 49:59-71.

306     Cramer W, and Hytteborn H. 1987. The separation of fluctuation and long-term change in

307          vegetation dynamics of a rising seashore. *Plant Ecology* 69:157-167.

308     de Rooij M. 2007. The distance perspective of generalized biadditive models: scalings and

309          transformations *Journal of Computational and Graphical Statistics* 16 210-227.

310     Gelman A, and Hill J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical*

311          *Models*: Cambridge University Press.

312     Gibson LA, Wilson BA, Cahill DM, and Hill J. 2004. Spatial prediction of rufous bristlebird

313          habitat in a coastal heathland: a GIS-based approach. *Journal of Applied Ecology*

314          41:213-223.

315     Greenacre MJ. 1984. *Theory and applications of correspondence analysis.* London:

316          Academic Press.

317     Hill MO, and Gauch HG. 1980. Detrended correspondence analysis: An improved ordination

318          technique. *Plant Ecology* 42:47-58.

319     Hurlbert SH. 1984. Pseudoreplication and the Design of Ecological Field Experiments.

320          *Ecological Monographs* 54:187-211.

321    Ihm P, and Van Groenewoud H. 1984. Correspondence analysis and Gaussian ordination.

322        *COMPSTAT Lectures* 3:5–60.

323    Ives AR, and Helmus MR. 2011. Generalized linear mixed models for phylogenetic analyses

324        of community structure. *Ecological Monographs* 81:511-525.

325    Jamil T, Ozinga WA, Kleyer M, and ter Braak CJF. 2012. Selecting traits that explain

326        species–environment relationships: a generalized linear mixed model approach.

327        *Journal of Vegetation Science*:10.1111/j.1654-1103.2012.12036.x.

328    Jongman RHG, ter Braak CJF, and van Tongeren OFR. 1995. *Data analysis in community*

329        *and landscape ecology*. Cambridge: Cambridge University Press.

330    Kruk C, Peeters E, Van Nes EH, Huszar VLM, Costa LS, and Scheffer M. 2011.

331        Phytoplankton community composition can be predicted best in terms of

332        morphological groups. *Limnology and Oceanography* 56:110-118.

333    Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL, Solymos P, Stevens

334        MHH, and Wagner H. 2011. vegan: Community Ecology Package. R package version

335        1.17-12. *http://CRANR-projectorg/package=vegan*.

336    Oksanen J, Läärä E, Tolonen K, and Warner BG. 2001. Confidence Intervals for the Optimum

337        in the Gaussian Response Function. *Ecology* 82:1191-1197.

338    Palmer MW, and Dixon PM. 1990. Small-Scale Environmental Heterogeneity and the

339        Analysis of Species Distributions along Gradients. *Journal of Vegetation Science*

340        1:57-65.

341    Pollock LJ, Morris WK, and Vesk PA. 2012. The role of functional traits in species

342        distributions revealed through a hierarchical model. *Ecography* 35:716-725.

343    Rue H, and Held L. 2005 *Gaussian Markov Random Fields: Theory and Applications*:

344        London: Chapman and Hall–CRC Press.

345  ter Braak CJF. 1985. Correspondence analysis of incidence and abundance data: properties in

346      terms of a unimodal response model. *Biometrics* 41:859-873.

347  ter Braak CJF. 1986. Canonical correspondence analysis: a new eigenvector technique for

348      multivariate direct gradient analysis. *Ecology* 67:1167-1179.

349  ter Braak CJF. 1987. *Unimodal models to relate species to environment*: DLO-Agricultural

350      Mathematics Group, 1996: Wageningen, the Netherlands.

351  ter Braak CJF, and Looman CWN. 1986. Weighted averaging, logistic regression and the

352      Gaussian response model. *Vegetatio* 65:3-11.

353  ter Braak CJF, and Prentice IC. 1988. A theory of gradient analysis. *Advances in Ecological*

354      *Research* 18:271-317.

355  ter Braak CJF, and Prentice IC. 2004. A theory of gradient analysis. *Advances in Ecological*

356      *Research* 34:235-282.

357  ter Braak CJF, and Smilauer P. 1998. *CANOCO Reference Manual and User's Guide to*

358      *Canoco for Windows: Software for Canonical Community Ordination (version 4)*.

359      Ithaca, NY, USA: Microcomputer Power.

360  ter Braak CJF, and Verdonschot PFM. 1995. Canonical correspondence analysis and related

361      multivariate methods in aquatic ecology. *Aquatic Sciences* 57:255-289.

362  Thompson R, Cullis B, Smith A, and Gilmour A. 2003. A sparse implementation of the

363      average information algorithm for factor analytic and reduced rank variance models.

364      *Australian & New Zealand Journal of Statistics* 45:445-459.

365  Verbyla AP, Eckermann PJ, Thompson R, and Cullis BR. 2003. The analysis of quantitative

366      trait loci in multi-environment trials using a multiplicative mixed model. *Australian*

367      *Journal of Agricultural Research* 54:1395-1408.

368  Walker SC, and Jackson DA. 2011. Random-effects ordination: describing and predicting

369      multivariate correlations and co-occurrences. *Ecological Monographs* 81:635-663.

370  Whittaker RH. 1960. Vegetation of the Siskiyou Mountains, Oregon and California.

371      *Ecological Monographs* 30:279-338.

372  Whittaker RH. 1967. Gradient analysis of vegetation. *Biological Reviews of the Cambridge*

373      *Philosophical Society* 49:207-264.

374  Zhu M, Hastie TJ, and Walther G. 2005. Constrained ordination analysis with flexible

375      response functions. *Ecological Modelling* 187:524-536.

376  Zuur AF, Ieno EN, Walker N, Saveliev AA, and Smith GM. 2009. *Mixed Effects Models and*

377      *Extensions in Ecology with R*: Springer.

378

379  # Appendix

380

381  This Appendix shows that similarity between the Gaussian logit models and GLMMs

382  extends beyond the Gaussian logistic curve. The Gaussian logistic model can be extended to

383  two environmental variables as follows (ter Braak & Prentice 1988)

384
$$logit(p_{ij}) = a_j - \frac{1}{2}\left(d_1(x_{i1} - u_{1j})^2 + d_2(x_{i2} - u_{2j})^2 - 2d_{12j}(x_{i1} - u_{1j})(x_{i2} - u_{2j})\right) .$$

385  where $d$'s are precision parameters, in the context of the bivariate normal distribution (Rue &

386  Held 2005 ). By setting

387
$$\alpha_j = a_j - \frac{1}{2}\left(d_1 u_{1j}^2 + d_2 u_{2j}^2 - 2d_{12j} u_{1j} u_{2j}\right) ,$$

388
$$\beta_{1j} = d_1 u_{1j} - d_{12j} u_{2j} ,$$

389
$$\beta_{2j} = d_2 u_{2j} - d_{12j} u_{1j} ,$$

390     $\beta_{3j} = d_{12j}$ ,

391     and

392     $\gamma_i = \dfrac{-1}{2}\left(d_1 x_{i1}^2 + d_2 x_{i2}^2\right)$

393     we can write

394     $logit\left(p_{ij}\right) = \alpha_j + \beta_{1j} x_{i1} + \beta_{2j} x_{i2} + \beta_{3j} x_{i1} x_{i2} + \gamma_i$ .

395     Here   $\beta_{3j}$   are random effects for interactions. If the 'co-precisions' are equal (   $d_{12j} = d_{12}$   )

396     the term   $\beta_{3j} x_{i1} x_{i2}$   can be subsumed in to the site effects   $\gamma_i$   and the model can do

397     without interactions. The   $\gamma_i$   account for the quadratic term arising from the Gaussian

398     (logit) model. Extension to more than two environmental variables is immediate. In

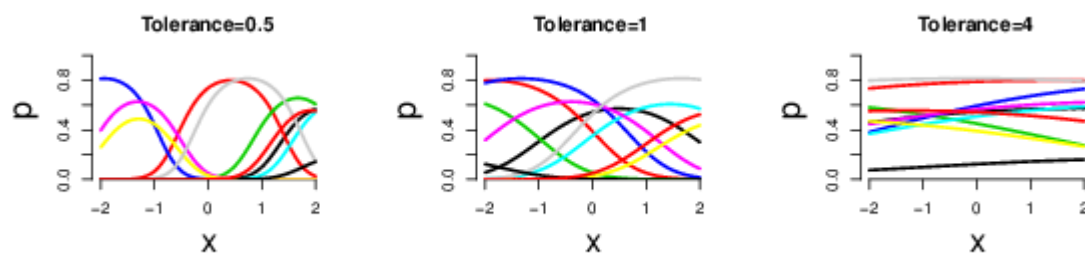
399     conclusion, up to distributional assumptions, the GLMM can be interpreted as a Gaussian

400     logistic model with equal tolerances for the species.

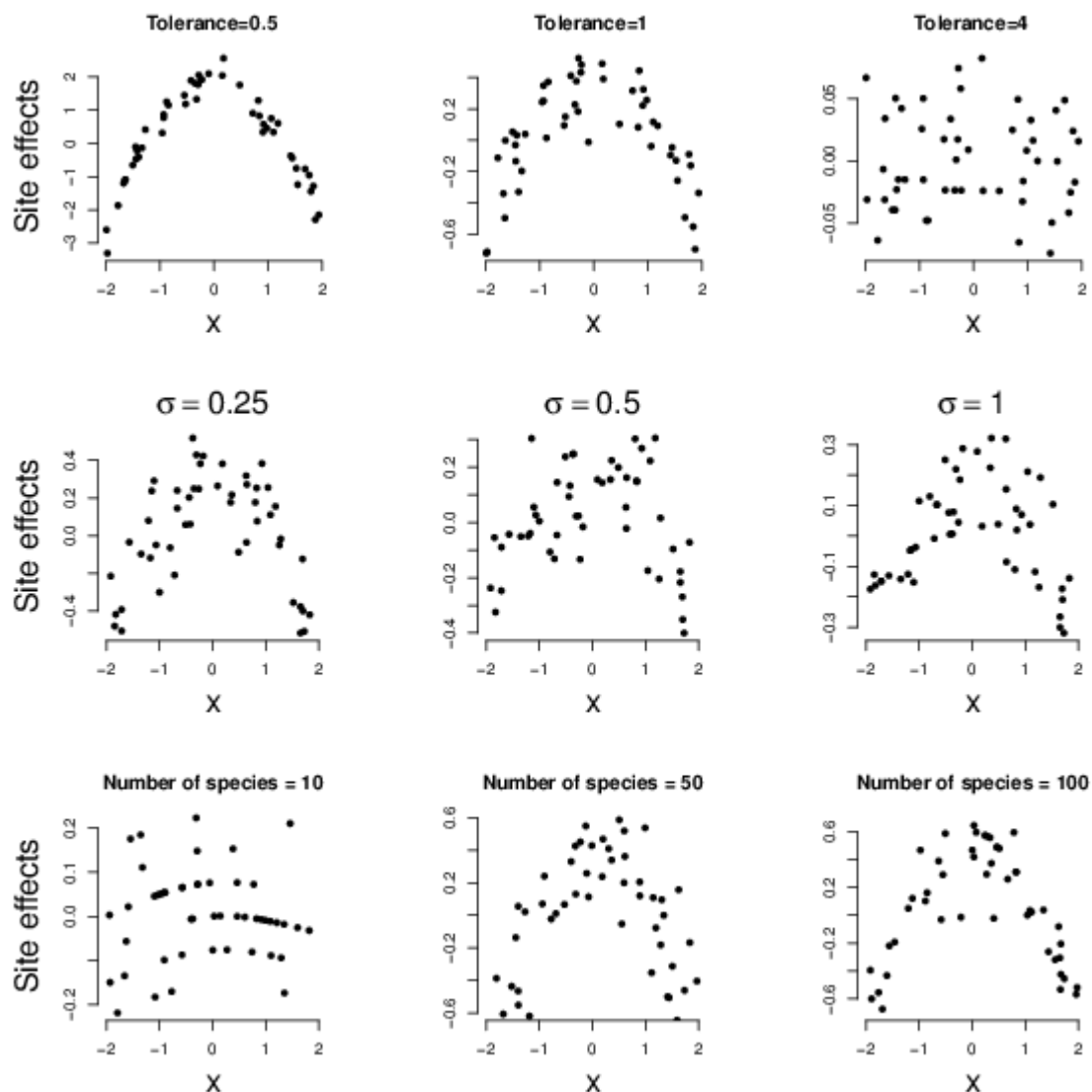
401

# Figure 1

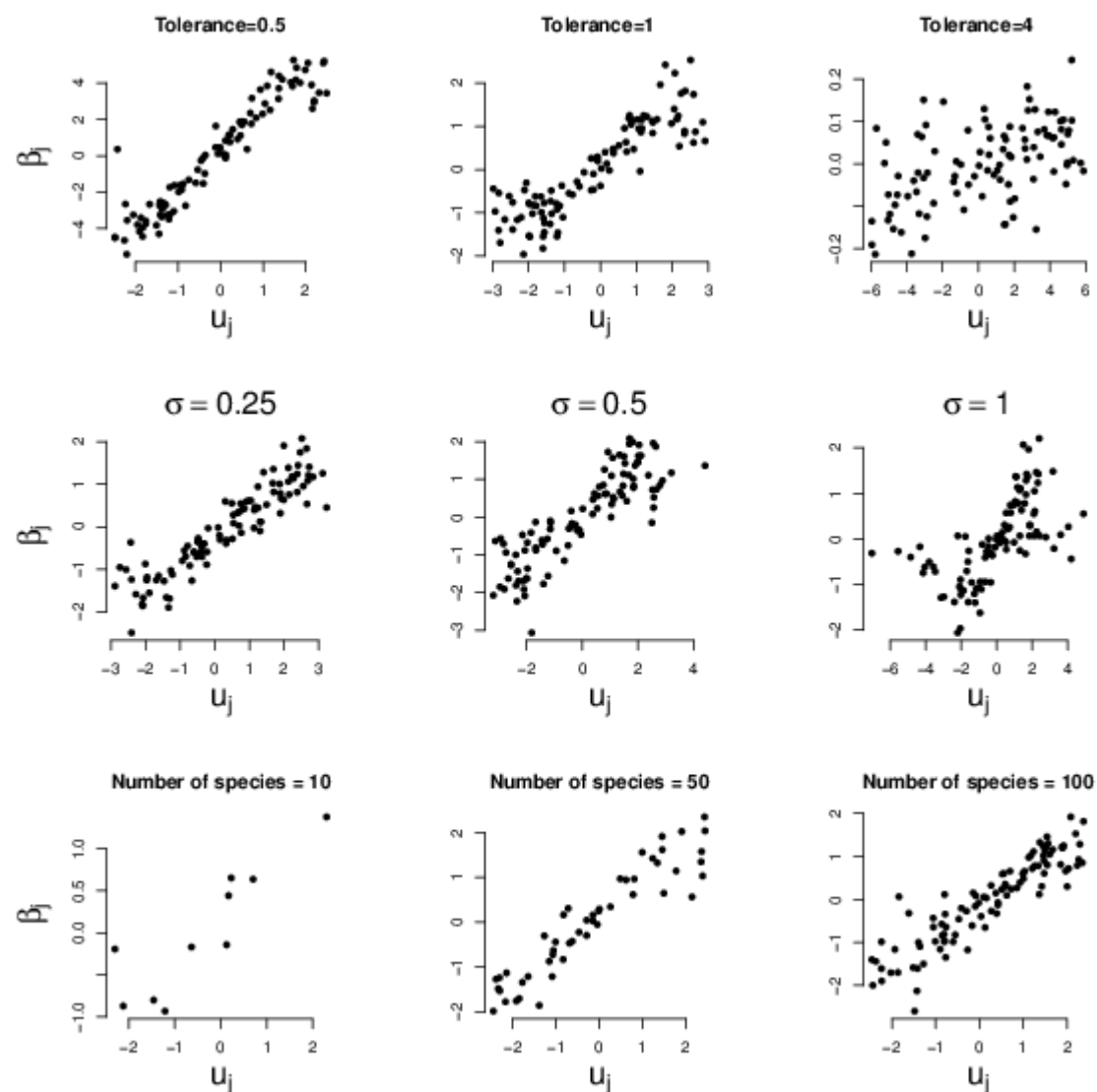
Simulated unimodal response (occurrence probability $p_{ij}$ of a species at a site) against environmental variable x for a selection of species.

For left to right: three tolerances ($t = 0.5$, 1 and 4).

# Figure 2

Simulated data: site effects as estimated by GLMM based on equation (1) plotted against environmental variable x. A quadratic relationship indicates a unimodal response.

A quadratic relationship indicates a unimodal response. Rows: example series 1-3, columns: parameter varied.

# Figure 3

Simulated data: GLMM random slopes of equation (1) against the true optima of the Gaussian logistic curve of equation (2).

Rows: example series 1-3, columns: parameter varied. In the two bottom rows the (median) tolerance is 1.

# Figure 4

Real data: site effects (top) and the GLMM random slopes (bottom) plotted against the environmental variable *x* and the optima of the Gaussian logistic curve as estimated by GLM, respectively.