# Devil in the details: how can we avoid potential pitfalls of CATS regression when our data do not follow a Poisson distribution?

**Zoltán Botta-Dukát** [Corresp. 1]

[1] Centre for Ecological Research, Vácrátót, Humgary

Corresponding Author: Zoltán Botta-Dukát
Email address: botta-dukat.zoltan@okologia.mta.hu

**Background:** Community assembly by trait selection (CATS) allows for the detection of environmental filtering and estimation of the relative role of local and regional (meta-community-level) effects on community composition from trait and abundance data without using environmental data. CATS has been shown to be equivalent to fitting a Poisson regression. Abundance data do not necessarily follow a Poisson distribution, and in these cases, other generalized linear models should be fitted to obtain unbiased parameter estimates.

**Aims:** This paper discusses how the original algorithm for calculating the relative role of local and regional effects has to be modified if Poisson model is not appropriate.

**Results:** It can be shown that the use of the logarithm of regional relative abundances as an offset is appropriate only if a log-link function is applied. Otherwise, the link function should be applied to the product of local total abundance and regional relative abundances. Since this product may be outside the domain of the link function, the use of log-link is recommended, even if it is not the canonical link. An algorithm is also suggested for calculating the offset when data are zero-inflated. The relative role of local and regional effects is measured by Kullback-Leibler $R^2$. The formula for this measure presented in Shipley's paper is valid only if the abundances follow a Poisson distribution. Otherwise, slightly different formulas have to be applied. Beyond theoretical considerations, the proposed refinements are illustrated by numerical examples. CATS regression could be a useful tool for community ecologists, but it has to be slightly modified when abundance data do not follow a Poisson distribution. This paper gives detailed instructions on the necessary refinement.

1 **Devil in the details: how can we avoid potential pitfalls**
2 **of CATS regression when our data do not follow a**
3 **Poisson distribution?**
4
5
6 Zoltán Botta-Dukát[1]
7
8 [1] Centre for Ecological Research, Vácrátót, Hungary
9
10
11 Corresponding Author:
12 Zoltán Botta-Dukát [1]
13 Alkotmány út 2-4, Vácrátót, 2163, Hungary
14 Email address: botta-dukat.zoltan@ecolres.hu
15

## Abstract

16
17 **Background:** Community assembly by trait selection (CATS) allows for the detection of

18 environmental filtering and estimation of the relative role of local and regional (meta-

19 community-level) effects on community composition from trait and abundance data without

20 using environmental data. It has been shown Poisson regression of abundances against trait data

21 results in the same parameter estimates. Abundance data do not necessarily follow a Poisson

22 distribution, and in these cases, other generalized linear models should be fitted to obtain

23 unbiased parameter estimates.

24 **Aims:** This paper discusses how the original algorithm for calculating the relative role of local

25 and regional effects has to be modified if Poisson model is not appropriate.

26 **Results:** It can be shown that the use of the logarithm of regional relative abundances as an

27 offset is appropriate only if a log-link function is applied. Otherwise, the link function should be

28 applied to the product of local total abundance and regional relative abundances. Since this

29 product may be outside the domain of the link function, the use of log-link is recommended,

30 even if it is not the canonical link. An algorithm is also suggested for calculating the offset when

31 data are zero-inflated. The relative role of local and regional effects is measured by Kullback-

32 Leibler $R^2$. The formula for this measure presented by Shipley (2014) is valid only if the

33 abundances follow a Poisson distribution. Otherwise, slightly different formulas have to be

34 applied. Beyond theoretical considerations, the proposed refinements are illustrated by numerical

35 examples.

36 CATS regression could be a useful tool for community ecologists, but it has to be slightly

37 modified when abundance data do not follow a Poisson distribution. This paper gives detailed

38 instructions on the necessary refinement.

39

## Introduction

40
41 The community assembly by trait selection (CATS) method developed by Shipley et al. allows

42 for the detection of environmental filtering of traits without using data on acting environmental

43 variables (Shipley, Vile & Garnier, 2006; Shipley, 2010). Moreover, its extension (Shipley, 2014)

44 can estimate the relative role of environmental filtering, meta-community effects (mass effect or

45 dispersal limitation), and demographic stochasticity. In its original form, CATS minimizes the

46 Kullback-Leibler divergence between the relative abundance expected a priori and the predicted

47  relative abundances under the constraint that weighted trait means (CWMs) have to be equal in

48  the predicted and observed communities. This approach uses observed trait means as the input

49  and does not require the observed abundances (see its implementation in FD package; Laliberté,

50  Legendre & Shipley, 2014). This fact suggests that the method could be applied to any type of

51  abundance data if relative abundances can be calculated (i.e., relative abundances are the weights

52  in the CWM calculation).

53  Warton et al. (2015) have shown that a Poisson-regression for abundances using species' traits as

54  predictors and the logarithm of relative abundances expected a priori as an offset results in the

55  same parameter estimates. The Poisson regression has three assumptions: (1) the abundance

56  values follow a Poisson distribution; (2) the logarithm of expected abundances is a linear

57  function of trait values; and (3) observations are independent (conditional on trait values).

58  Violation of the third assumption has implications for interference (see eg. Warton, Shipley &

59  Hastie, 2015; ter Braak, Peres-Neto & Dray, 2017 for discussion on interference, when

60  independence assumtion is violated due to species interactions). However, it does not lead to

61  biased estimates of model parameters. Violation of the first and second assumptions may result

62  in biased parameter estimates.

63  The "examples" section will illustrate that violation of the Poisson assumption may lead to false

64  conclusions.

65  A Poisson distribution of abundances can be assumed when they are measured by a number of

66  individuals. Even in this case, the abundances may be over-dispersed (i.e., the variance is higher

67  than the mean), while a Poisson distribution implies equal mean and variance. Abundance is

68  often measured in other units: estimated cover, biomass, and frequency of presence, and

69  sometimes, only presence/absence data are available. Replacing the maximum entropy formalism

70  with maximum likelihood (ML) fitting of generalized linear models (GLMs) allows for a

71  generalization of CATS that relaxes the distribution assumption. Following Warton et al. (2015),

72  we refer to GLMs for abundances with trait values as predictors as in CATS regression. These

73  models are not mathematically equivalent to CATS, but they have the same goal as CATS using

74  distributional assumptions that better fit the abundance data at hand.

75  Depending on the unit of abundances, different distributions can be assumed. Presence/absence

76  data can be modeled by logit-regression and assuming a binomial (Bernoulli) distribution

77  (Warton & Hui, 2010). Assuming independent sampling points, the frequency of occurrence in

78    sampling points (e.g. pin points; Goodall, 1952) can also be modeled by a binomial distribution

79    (see Damgaard (2008, 2009) for relaxing of the independence assumption). If the number of

80    individuals counted in a sampling with fixed intensity (e.g., fixed sampling area or trapping

81    time), the simplest assumption is that abundances follow a Poisson distribution. However,

82    abundance data may be over-dispersed (i.e., the variance is higher than mean), zero-inflated (the

83    number of zeros is higher than expected from the fitted distribution), or both.

84    Over-dispersed counts can be modeled by a negative binomial distribution (O'Hara & Kotze,

85    2010) or Conway-Maxwell-Poisson distribution (Lynch, Thorson & Shelton, 2014). For count

86    data with excessive zeros, zero-inflated or two-part (hurdle) models can be fitted (Zuur et al.,

87    2009; Blasco-Moreno et al., 2019). If the total number of individuals is fixed in sampling instead

88    of the sampling intensity, the number of individuals in each species follows a multinomial

89    distribution (Chong & Spencer, 2018). The abundance of plant species is often described by their

90    cover (which is often visually estimated). Cover data can be analyzed by (zero-inflated) beta-

91    regression (Damgaard & Irvine, 2019). When individuals considerably differ in size, biomass

92    may be a better abundance measure than the number of individuals. For modeling biomass data,

93    a Tweedie distribution could be applied. It assumes that the mean-power relationship follows the

94    Taylor law. If the power parameter p is in the range of $1<p<2$, a Tweedie distribution is

95    mathematically equivalent to a compound Poisson-gamma distribution (i.e., the sums of the

96    Poisson-distributed number of individuals each have a gamma distributed mass). This

97    distribution has a point mass at zero (i.e., an absence of species) (Dunstan et al., 2013).

98    All of these distributions can be applied in generalized linear models. The interpretation of fitted

99    parameters is similar for all distributions: a positive parameter value means that a higher trait

100   value results in higher expected abundance. However, the relationship is nonlinear (except when

101   applying an identity link) and depends on the applied link function. Therefore, plotting the

102   expected abundances against traits gives a more detailed picture. For interference, the same

103   procedures can be applied irrespective of the distribution (Warton, Shipley & Hastie, 2015; ter

104   Braak, Peres-Neto & Dray, 2017). Thus, at first glance, generalization of the CATS regression

105   seems to be straightforward. However, there are two points that need more consideration:

106   choosing/interpreting offset terms and calculating the explained variation. The aims of this paper

107   are (1) to show that recommendations for the original CATS model should be reconsidered when

108    a Poisson distribution is replaced by another distribution and (2) to give a general solution for

109    this replacement and detailed recommendations for the most often used distributions.

110

## Modeling meta-community effect via offset

111

112    Beyond local trait selection, larger-scale effects can also influence species' local abundance.

113    High propagule pressure can increase the local abundance of species that are abundant in the

114    surroundings. On the other hand, locally well-adapted species may be missing from the local

115    community due to propagule limitation. The unique property of CATS is that it can measure the

116    relative importance of local and meta-community-scale (dispersal) processes (Shipley, 2014).

117    If species survival, growth, and reproduction were independent of their traits, local abundances

118    would differ from the meta-community-level means due to demographic stochasticity only

119    (including stochasticity of dispersal). In this case, local abundances could be predicted well from

120    mean abundances at the meta-community level, while using traits as predictors would not

121    improve the fit. At the other extreme, when species abundances are independent of dispersal

122    processes (i.e., no mass effect or propagule limitation) and fully determined by local processes,

123    knowledge on the meta-community-level abundance would not be able to improve our ability to

124    predict local abundances (Shipley, 2014). Therefore, the heart of CATS is the fitting of models

125    with and without information on abundances in the meta-community level. Shipley (2014) called

126    this information the "neutral prior." However, Warton et al. (2015) called attention to the term

127    "prior" as being associated with Bayesian statistics, where it has a specific meaning. To avoid

128    confusion, we will refer to it as "abundances expected *a priori*," where "*a priori*" means "before

129    knowing local conditions."

130    The abundances expected a priori can be included into CATS regression models via offset terms.

131    Warton et al. (2015) suggested using the logarithm of relative abundances at the meta-

132    community level in a Poisson regression to reproduce the original CATS model. This study

133    discusses whether this suggestion is generally valid irrespective of the link function.

134    The aim of CATS models is to predict relative abundances of species, not to explain differences

135    in total abundances among sites, which may be caused by differences in sampling intensity.

136    Therefore, they always contain an intercept. The relative abundance predicted by a model

137    containing only an intercept and offset should be equal to the relative abundances expected a

138    priori ($\pi$):

$$\frac{\hat{y}_i}{\sum\limits_{i=1}^{S} \hat{y}_i} = \pi_i \tag{1}$$

139 where $\hat{y}_i$ is the predicted abundance of species $i$, $S$ is the number of species, and $\pi_i$ is relative

140 abundances expected a priori of species $i$. For most of the fitted models (but not for zero-inflated

141 and two-stage models), $\hat{y}_i = \mu_i$, where $\mu_i$ is the location parameter of the fitted distribution.

142 Therefore, $\mu_i$ will be used instead of $\hat{y}_i$ where appropriate. If a canonical link is applied and the

143 model contains an intercept term, the sum of predicted values is equal to the sum of observed

144 values ($y_{tot}$). Thus, requirement (1) could be written in the following form:

$$\mu_i = \pi_i y_{tot} \tag{2}$$

145 The GLM with an intercept and offset but no predictors can be written in the following general

146 form:

$$h(\mu_i) = \beta_0 + O_i \tag{3}$$

147 where h() is the link function, $\beta_0$ is the intercept, and $O_i$ is the offset for species $i$. Substituting

148 (2) into (3), we obtain the following system of linear equations (note that the left side of

149 equations can be replaced by numbers calculated from data on hand):

$$h(\pi_i y_{tot}) = \beta_0 + O_i \tag{4}$$

150 This system contains $S+1$ variables ($O_1$, $O_2$, …, $O_S$ and $\beta_0$), but only $S$ equations, so it has no

151 unique solution. It can be solved by choosing an arbitrary value for $O_1$. Then, the other offsets

152 can be calculated with the following formula:

$$O_i = h(\pi_i y_{tot}) - h(\pi_1 y_{tot}) + O_1 \tag{5}$$

153 In Poisson and negative binomial regression, the canonical link is the natural logarithm (Dobson,

154 2002). Thus,

$$O_i = \ln(\pi_i y_{tot}) - \ln(\pi_1 y_{tot}) + O_1 = \ln(\pi_i) - \ln(\pi_1) + O_1 \tag{6}$$

155 In this case, it is appropriate to choose $O_1 = \ln(\pi_1)$, which leads to the offset recommended by

156 Warton et al. (2015). For other link functions, the simplest choice is $O_1 = 0$. However, this

157 approach has two limitations. First, $\pi_i y_{tot}$ has to be within the domain of the link function. For

158 example, for binomial and beta distributions, where the canonical link is logit($x$), offsets can be

159 calculated only if $\pi_i y_{tot} < 1$; otherwise, $\ln\left(\pi_i y_{tot} / (1 - \pi_i y_{tot})\right)$ cannot be calculated. The second

160 limitation is that a canonical link is not always the most appropriate link function, and

161 sometimes, another link function has to be chosen. For example, for a Tweedie distribution with

162 power parameter 1<p<2, the canonical link would be (Ohlsson & Johansson, 2006):

$$h(\mu) = -\frac{1}{p-1}\mu^{-(p-1)} \tag{7}$$

163 When the exact value of power parameter $p$ is unknown, it can be estimated from data during

164 model fitting, but in this case, the log-link is applied in R packages mgcv (Wood, 2017) and

165 glmmTMB (Brooks et al., 2017). If a canonical link is not used, the sum of expected values may

166 differ from the sum of observed values. But equation (1) remains true irrespective of the link

167 function, and it can be converted to:

$$\frac{\mu_i}{\mu_1} = \frac{\pi_i}{\pi_1} \tag{8}$$

168 Combining equations (3) and (8) results in:**Error! Reference source not found.**

$$\frac{h^{-1}(\beta + O_i)}{h^{-1}(\beta + O_1)} = \frac{\pi_i}{\pi_1} \tag{9}$$

169 where $h^{-1}(x)$ is the inverse of link function.

170 Setting $O_1$ to an arbitrary value leads to a nonlinear equation system with $S$ equations and $S$

171 variables. Solving such a system is often a hard task. The situation would be much simpler if the

172 following were true:

$$h^{-1}(\beta + O_i) = h^{-1}(\beta)h^{-1}(O_i) \tag{10}$$

173 In this case, choosing $O_i = h(\pi_i)$ satisfies condition (9). Since $h(x) = \ln(x)$ and $h^{-1}(x) = e^x$ satisfy

174 condition (10), it is reasonable to use log-link even if it is not the canonical link (for example,

175 instead of logit in a binomial model). In medical statistics, binomial GLM with log-link is called

176 relative risk regression, which is often recommended due to the easier interpretation of

177 proportions than odds ratios (Marschner, 2015). If log-link is used with a binomial distribution,

178 iteratively reweighted least squares (the standard method for fitting GLMs) may fail to converge

179 to the maximum likelihood estimate (Marschner & Gillett, 2012). Therefore, alternative

180 estimation procedures were developed and are implemented in the *logbin* R package (Donoghoe

181 & Marschner, 2018). Log-link for beta regression is also available in the *betareg* R package

182 (Cribari-Neto & Zeileis, 2010).

183    Using the mentioned distributions, we suppose that all data come from the same distribution, and

184    only their parameters depend on species. In this case, the relative abundances at the meta-

185    community level can be estimated by:

$$\pi_i = \frac{\overline{m_i}}{\sum \overline{m_i}} \tag{11}$$

186    where $\overline{m_i}$ is the mean abundance of species $i$ in plots representing the meta-community. Note that

187    since only the ratio of relative abundances expected a priori are used, the offset could be simply

188    $O_i = \ln(\overline{m_i})$ instead of $O_i = ln(\pi_i)$.

189    If there are excessive zeros, it could be supposed that some of the zeros do not come from the

190    distribution of "normal" abundances. Zero-inflated and two-part (hurdle) models are based on

191    this assumption. In these models, there are two equations for two location-type parameters.

192    A zero-inflated model supposes that positive counts and some of the zeros comes from a Poisson

193    or negative binomial process, while some zeros are "structural zeros" (i.e., species cannot occur

194    there). The two parameters in this case are the probability of structural zeros ($p$) and the expected

195    value of the Poisson or negative binomial process ($\mu$) (Zuur et al., 2009). For estimating offsets,

196    we must know the probability of structural zeros at the meta-community. Therefore, it seems that

197    zero-inflated models have low practical relevance when a meta-community effect has to be

198    modeled.

199    A two-part (hurdle) model fits two separate models: a binomial model for presence/absence data

200    and a truncated Poisson or negative binomial model for positive abundances. In this case, the two

201    parameters are the probability of presence ($p$) and mean of the Poisson or negative binomial

202    distribution ($\mu$), from which the fitted zero-truncated distribution is deduced (not the mean of the

203    truncated distribution itself) (Zuur et al., 2009). For simplicity, let us imagine that we really fit a

204    two-part model as two separate GLMs. The first GLM is a binomial model for binary data.

205    Therefore, we estimate offsets using the standard procedure (but mean abundances at the meta-

206    community level have to be calculated from binary data). In the second GLM, we fit a truncated

207    Poisson or truncated negative binomial distribution for the non-zero abundances. Similar to

208    Poisson regression, it is assumed that $ln(\mu_i)$ is a linear combination of trait values and the offset,

209    but the expected abundance is the following for a Poisson distribution:

$$\hat{y}_i = \frac{\mu_i}{1 - \exp(-\mu_i)} \tag{12}$$

210 For a negative binomial distribution, the expected abundance is:

$$\hat{y}_i = \frac{\mu_i}{1 - \left(\frac{\mu_i + \theta}{\theta}\right)^{-\theta}} \tag{13}$$

211 $\hat{y}_i$ is the expected or mean abundance of species $i$ when it present, while $\mu_i$ is the expected or

212 mean abundance when only structural zeros are excluded. For setting the offset, we need a priori

213 expectation for the latter.

214 The meta-community level mean of species' abundance when present $(\overline{m}_i^+)$ can be easily

215 estimated. Assuming a Poisson process, the mean abundance excluding structural zeros $(\tilde{m}_i)$ can

216 be estimated by solving the following nonlinear equation:

$$\overline{m}_i^+ = \frac{\tilde{m}_i}{1 - \exp(-\tilde{m}_i)} \tag{14}$$

217 If a negative binomial distribution is assumed, a similar approach can be applied if $\theta$ is known.

218 Then, $\ln(\tilde{m}_i)$ could be used as an offset.

219

## Relative importance of environmental selection and dispersal
## processes

222 The relative importance of local and meta-community-level processes can be calculated from

223 variation explained by models containing only traits (as independent variables), only offset

224 (calculated from meta-community-level abundances; see above), or both traits and offset

225 (Shipley, 2014). Using the classic $R^2$ as a measure of explained variance is suitable only in OLS

226 regression. Different generalizations of $R^2$ are suggested for GLMs (Cameron & Windmeijer,

227 1996, 1997; Menard, 2000; Nakagawa & Schielzeth, 2013). Shipley (2014) proposed using a

228 generalization based on Kullback-Leibler divergence (Cameron & Windmeijer, 1997).

229 I will show below that formula (4) from Shipley (2014) is valid for only a Poisson-model, and

230 different formulas have to be used for other distributions. A definition of Kullback-Leibler $R^2$ is

231 not available for models with an offset; therefore, solution for this case by Shipley (2014) and its

232 alternatives will be discussed. Increasing the number of predictors (traits) always improves the fit

233    of the model (i.e., increases the Kullback-Leibler $R^2$). Therefore, $R^2$ values of models with

234    different numbers of predictors (traits) cannot be compared. Shipley (2014) proposed an

235    "adjustment" procedure based on randomization of traits. Although this procedure is correct, it is

236    time consuming for large datasets . Thus, an alternative deterministic adjustment is proposed.

237

## $R^2$ for models without offset

239    Kullback-Leibler $R^2$ is a generalization of the classic $R^2$ used in ordinary least squares

240    regression:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = \frac{\sum(y_i - \bar{y})^2 - \sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{15}$$

241    where $\sum(y_i - \hat{y}_i)^2$ and $\sum(y_i - \bar{y})^2$ are the squared Euclidean distances between observed values

242    and predictions of models with and without predictors, respectively. Therefore, $R^2$ is a

243    proportional decrease of distance between model prediction and observed values due to the

244    inclusion of predictors in the model. For other distributions, the squared Euclidean distance can

245    be replaced with Kullback-Leibler divergence with the same interpretation:

$$R_{KL}^2 = 1 - \frac{K(\mathbf{y};\boldsymbol{\mu})}{K(\mathbf{y};\boldsymbol{\mu^0})} = \frac{K(\mathbf{y};\boldsymbol{\mu^0}) - K(\mathbf{y};\boldsymbol{\mu})}{K(\mathbf{y};\boldsymbol{\mu^0})} \tag{16}$$

246    where $\mathbf{y}$ is the vector of observed values, and $\boldsymbol{\mu}$ and $\boldsymbol{\mu^0}$ are vectors of values predicted by the

247    evaluated and intercept-only models, respectively.

248    Table 1 shows that distributions that could be used in CATS regression belong to the exponential

249    family, so their density functions can be written in the following general form (McCullagh &

250    Nelder, 1999):

$$f(y;\psi,\phi) = \exp\left\{\frac{y\psi - b(\psi)}{a(\phi)} - c(y,\phi)\right\} \tag{17}$$

251    where $\psi$ is the natural or canonical parameter, $\phi$ is the dispersal parameter, and a, b, and c are

252    specific functions. The mean and variance of y are:

$$\begin{aligned} E(y) &= \mu = b'(\psi) \\ Var(y) &= a(\phi)b''(\psi) \end{aligned} \tag{18}$$

253 where $b'$ and $b''$ are the first and second derivatives of function $b$. Note that negative binomial

254 and Tweedie distributions belong to this family only if parameters $\theta$ and $p$ are known constants.

255 Function $h(\mu) = \psi$ is called a canonical link function.

256 For members of the exponential family, Kullback-Leibler divergence can be calculated as the

257 difference between the likelihood of a full model (i.e., a model where predicted and observed

258 values are equal) and a fitted model (Cameron & Windmeijer, 1997):

$$K(\mathbf{y};\boldsymbol{\mu}) = 2\left[l(\boldsymbol{\mu}^{\mathbf{full}};\mathbf{y}) - l(\boldsymbol{\mu};\mathbf{y})\right] \tag{19}$$

259 Thus, for members of the exponential family, $R_{KL}^2$ could be deduced as a corrected version of

260 likelihood ratio $R^2$ or McFadden $R^2$:

$$R_L^2 = 1 - \frac{l(\boldsymbol{\mu};\mathbf{y})}{l(\boldsymbol{\mu}^0;\mathbf{y})} \tag{20}$$

261 A drawback of $R_L^2$ is that its maximum is not 1, but $1 - \left[l(\boldsymbol{\mu}^{\mathbf{full}};\mathbf{y})/l(\boldsymbol{\mu}^0;\mathbf{y})\right]$. Since its minimum is

262 zero, $R_L^2$ can be rescaled to the interval of 0–1 by dividing it by its maximum, which results in

263 $R_{KL}^2$:

$$\frac{R_L^2}{\left(1 - \dfrac{l(\boldsymbol{\mu}^{\mathbf{full}};\mathbf{y})}{l(\boldsymbol{\mu}^0;\mathbf{y})}\right)} = \frac{l(\boldsymbol{\mu}^0;\mathbf{y}) - l(\boldsymbol{\mu};\mathbf{y})}{l(\boldsymbol{\mu}^0;\mathbf{y})} \bigg/ \frac{l(\boldsymbol{\mu}^0;\mathbf{y}) - l(\boldsymbol{\mu}^{\mathbf{full}};\mathbf{y})}{l(\boldsymbol{\mu}^0;\mathbf{y})} = \frac{l(\boldsymbol{\mu}^0;\mathbf{y}) - l(\boldsymbol{\mu};\mathbf{y})}{l(\boldsymbol{\mu}^0;\mathbf{y}) - l(\boldsymbol{\mu}^{\mathbf{full}};\mathbf{y})} = R_{KL}^2 \tag{21}$$

264 Substituting (19) into (16) results in:

$$R_{KL}^2 = 1 - \frac{l(\boldsymbol{\mu}^{\mathbf{full}};\mathbf{y}) - l(\boldsymbol{\mu};\mathbf{y})}{l(\boldsymbol{\mu}^{\mathbf{full}};\mathbf{y}) - l(\boldsymbol{\mu}^0;\mathbf{y})} = \frac{l(\boldsymbol{\mu};\mathbf{y}) - l(\boldsymbol{\mu}^0;\mathbf{y})}{l(\boldsymbol{\mu}^{\mathbf{full}};\mathbf{y}) - l(\boldsymbol{\mu}^0;\mathbf{y})} \tag{22}$$

265 McCullagh and Nelder (1999) called $2\left[l(\boldsymbol{\mu}^{\mathbf{full}};\mathbf{y}) - l(\boldsymbol{\mu};\mathbf{y})\right]$ the scaled deviance ($D^*$), so $R_{KL}^2$ can

266 also be calculated from scaled deviances ($D^*$) or deviances ($D$) of fitted and intercept-only

267 models:

$$R_{KL}^2 = 1 - \frac{D^*(\boldsymbol{\mu};\mathbf{y})}{D^*(\boldsymbol{\mu}^0;\mathbf{y})} = 1 - \frac{D(\boldsymbol{\mu};\mathbf{y})/a(\phi)}{D(\boldsymbol{\mu}^0;\mathbf{y})/a(\phi)} = 1 - \frac{D(\boldsymbol{\mu};\mathbf{y})}{D(\boldsymbol{\mu}^0;\mathbf{y})} \tag{23}$$

268

269 If there is no offset, in a generalized linear model fitted by ML estimation with a canonical link,

270 the expectations in an intercept-only model is equal to the mean of observed values: $\boldsymbol{\mu}^0 = \overline{y}$.

271    Formulas for this case are listed in Appendix 1. Appendix 2 shows that formula (4) from Shipley

272    (2014) is equivalent to the formula given for a Poisson regression in Appendix 1.

273    These formulas assume that the likelihood is a function of $\mu$ only, and if there are other

274    parameters, their values are constants known a priori (i.e., not estimated during regression). If

275    these parameters are estimated in regression, we can obtain different estimates for the evaluated

276    and the intercept-only models.  For a negative binomial distribution with unknown dispersion ($\theta$

277    ), Cameron and Windmeijer  (1996) suggested using a parameter estimated for an evaluated

278    model when the likelihood of full and intercept-only models is calculated. Applying this

279    approach, $R^2_{KL}$ may decrease when a regressor is added to the model due to changes in estimated

280    $\theta$. The same approach can be applied for the power parameter of a Tweedie distribution. GLM

281    fitting programs usually give the log-likelihood of the fitted model and the deviance of fitted and

282    intercept-only models.

283

## 284  $R^2$ for models with offset

285    In the previous section, $\mu$ was the prediction of model containing an intercept and predictors

286    (traits), while $\mu^0$ was the prediction of an intercept-only model. A possible solution for how we

287    should include the offset is to define $\mu^0$ as the prediction of a model without predictors (i.e., a

288    model containing only an intercept and offset). At first glance, it seems to be a natural

289    generalization of $R^2$ shown in the previous section. However, there is a drawback in this

290    approach: the effect of meta-community-level processes cannot be directly measured since $R^2$ for

291    models containing only an intercept and offset (but no traits) would always be zero. Shipley

292    (2014) suggested an indirect measure of a pure meta-community effect:

$$\frac{R^2(\text{traits; offset}) - R^2(\text{traits})}{1 - \overline{R^2(\text{random traits})}} \qquad (24)$$

293    The nominator is an adjustment for removing bias (see next section on adjustment), so now, we

294    should focus on the denominator. This subtraction is based on the assumption that $R^2(\text{traits})$ is

295    the variation explained by traits, while $R^2(\text{traits; offset})$ is the variation explained by traits and

296    offset (i.e., meta-community effect) together. However, this assumption is not satisfied when in

297    calculation of R-squared $\mu^0$ is the prediction of a model containing only an intercept and offset.

298  To understand why, we should recall the geometric interpretation of Kullback-Leibler $R^2$ shortly

299  mentioned above: $R^2$ is the proportional decrease of distance between observed and predicted

300  values (or proportional improvement of fit) due to the inclusion of predictors in the model.

301  Therefore, in (24), both $R^2$(traits) and $R^2$(traits; offset) are proportional improvements of fit due

302  to the inclusion of traits, but they are proportional to different original distances of observed and

303  predicted values. Therefore, their difference has no simple interpretation and does not measure

304  the pure meta-community effect.

305  Instead of defining $\boldsymbol{\mu^0}$ as a prediction of a model without a predictor, it could be defined as a

306  prediction of an intercept-only model, even if an offset is applied. This definition allows us to

307  calculate a meaningful $R^2$ for models with offset but no predictors as a direct measure of the

308  meta-community effect. This definition may result in negative $R^2$ when including an offset

309  increases the distance between observed and predicted values instead of decreasing it. A negative

310  value is nonsense if $R^2$ is interpreted as explained variation, but it is meaningful if $R^2$ is

311  interpreted as a proportional change in the distance between observed and predicted values. This

312  geometric interpretation seems more useful in CATS regression, where it has a meaning that

313  includes meta-community-level relative abundances decreases the goodness-of-fit.

314  Note that in R environment, to avoid negative $R^2$ values, the following formula is applied instead

315  of (15):

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \hat{y}_i)^2 + \sum(\hat{y}_i - \overline{y})^2} \tag{25}$$

316  If there is no offset, (15) and (25) result in the same value, but they differ if offset is applied.  A

317  generalization of (25) could be:

$$R^2 = 1 - \frac{K(\mathbf{y};\boldsymbol{\mu})}{K(\mathbf{y};\boldsymbol{\mu}) + K(\boldsymbol{\mu};\boldsymbol{\mu^0})} \tag{26}$$

318

# Adjusted $R^2$ and partitioning of explained variation

320  As a goodness-of-fit measure, a drawback of $R^2$ (both in the classic form and its generalization)

321  is that including an additional predictor in the model always increases $R^2$, even if the predictor is

322  independent of the dependent variable. Therefore, it has a positive expected value instead of zero

323   when a dependent variable is not related to the predictors. To remove this bias, Fisher (1925)

324   suggested using the following adjustment:

$$R_{adj}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1} \tag{27}$$

325   where $n$ is the number of data points, and $k$ is the number of predictors.

326   Unfortunately, this adjustment is valid for only for "classic" $R^2$ of ordinary least squares

327   regression. Since Shipley (2014) has not found a similar solution for Kullback-Leibler $R^2$, he

328   proposed a procedure based on reshuffling trait values to remove the bias. Although the

329   suggested procedure is correct, it has no unique result, in contrast to the correction using a closed

330   form. Ricci (2010) has shown that there is a simple general adjustment for $R_{KL}^2$ when the

331   dependent variable follows a distribution belonging to the exponential family. Let us write the

332   formula for $R_{KL}^2$ using scaled deviances:

$$R_{KL}^2 = \frac{D^*(\mathbf{\mu^0};\mathbf{y}) - D^*(\mathbf{\mu};\mathbf{y})}{D^*(\mathbf{\mu^0};\mathbf{y})} \tag{28}$$

333   Recall that $R_{KL}^2$ is the proportional improvement of model fit. Let us focus on the denominator,

334   which is the absolute improvement. In a bias-corrected version, the absolute improvement should

335   be zero when predictors have no effect on the dependent variable. Under this condition, for

336   members of the exponential family, $D^*(\mathbf{\mu^0};\mathbf{y}) - D^*(\mathbf{\mu};\mathbf{y})$ approximately follows a Chi-square

337   distribution with degree of freedom equals to the number of predictors ($k$). Since the expected

338   value of the Chi-square distribution is its degree of freedom, the following is an approximately

339   bias-free goodness-of-fit measure:

$$R_{KL,adj}^2 = \frac{D^*(\mathbf{\mu^0};\mathbf{y}) - D^*(\mathbf{\mu};\mathbf{y}) - k}{D^*(\mathbf{\mu^0};\mathbf{y})} \tag{29}$$

340   Appendix 3  shows that (27) is a special case of (29) for a Gaussian distribution with dispersion

341   estimated from the data. Note that Ricci (2010) applied an alternative derivation of adjustment by

342   generalization of the shrinkage factor and used the deviance instead of the scaled deviance in the

343   formulas.

344   For partitioning variation, we should fit models containing both traits and offset and models with

345   only traits and only offset. Let us denote the corresponding adjusted $R^2$ values by $R^2$(trait;

346   offset), $R^2$(trait), and $R^2$(offset), respectively. $R^2$(trait; offset) measures the whole variation

347    explained by studied traits and relative abundances at the meta-community level. The pure trait

348    effect (i.e., variation explained only by traits) is $R^2$(trait; offset)- $R^2$(offset), while the pure meta-

349    community effect is $R^2$(trait; offset)- $R^2$(trait). Variation that can be explained by both traits and

350    the meta-community effect is $R^2$(trait)+$R^2$(offset)- $R^2$(trait; offset). These formulas are analogous

351    to partitioning of the variation of community composition into environmental and spatial

352    components (Borcard, Legendre & Drapeau, 1992; Peres-Neto et al., 2006).

353

## Examples

355    Examples are presented to illustrate the main messages of the paper. First, example 1 shows how

356    setting an inappropriate distribution leads to biased parameter estimates. Next, Example 2 shows

357    that it is important to choose an appropriate offset, and finally Example 3 illustrates why

358    variation components should be estimated in a new way.

359    Examples uses new R package *CATSregression* publicly available on GitHub

360    (https://github.com/BottaDZ/CATSregression/). The package's vignette shows more examples

361    using field data.

362

# Example 1: fitting Poisson model to over-dispersed counts

364    The first example illustrates the problems that arise when an inappropriate model is fitted. The

365    type of abundance often clearly determines the type of model to be fitted. However, when

366    abundance is measured by a number of individuals, a Poisson or negative binomial model should

367    be fitted depending on whether there is a random or aggregated spatial pattern of individuals. If

368    the spatial pattern is random, the number of individuals will follow a Poisson distribution, where

369    the variance is equal to the mean. An aggregated spatial pattern leads to over-dispersed counts

370    (i.e., the variance is higher than mean), which can be modeled by a negative binomial

371    distribution.

372    The presence of over-dispersion can be checked by comparing the Akaike Information Criteria

373    of Poisson and negative binomial models or using diagnostic plots of residuals (Figure 1). Dunn-

374    Smyth (or randomized quantile) residuals (Dunn & Smyth, 1996) are especially useful for this

375    purpose because if the model's assumptions (specified distribution and log-linear relationship)

376  are satisfied, they follow a standard normal distribution (Warton, Shipley & Hastie, 2015; Feng, Li

377  & Sadeghpour, 2020).

378  This example uses simulated data. Abundances (y) of 20 species in a plot were simulated.

379  Abundances follow a negative binomial distribution with a mean that has a log-linear relation to

380  values of a trait. The dispersion parameter is constant, and trait values follow a normal

381  distribution:

$$y_i \sim NegBin(\mu_i = exp(0.5 * x_i), \theta = 1)$$
$$x_i \sim N(m = 10, \sigma = 3)$$

(30)

382  The simulation was repeated 50 times, and Poisson and negative binomial models were fitted to

383  each simulated plots separately. Figure 1 shows a typical diagnostic plot, which has a strong

384  nonlinearity of the QQ plot indicating that the distributional assumption of a Poisson regression

385  is not satisfied. The diagnostic plots of a negative binomial model do not indicate any problem.

386  The estimated slopes are dispersed around the real value (0.5) in both Poisson and negative

387  binomial models (Figure 2). The variation among estimates was higher in the Poisson

388  distribution, while the confidence intervals were narrower due to the neglect of over-dispersion.

389  These two facts together may result in over-interpretation of differences in the strength of

390  selection among plots.

391

392  ## Example 2: choosing appropriate offset

393  The second example illustrates the importance of choosing an appropriate offset, without which

394  fitted relative abundances may differ considerably from the relative abundances in the meta-

395  community, even if no traits are included in the model. The example uses the dataset of Raevel et

396  al. (2012), which contains the abundance (number of individuals) of 97 species at 52 sites. Data

397  were transformed into a presence/absence scale, and then meta-community-level abundances

398  were measured as the number of occurrences at the 52 sites.

399  For modeling of the presence/absence data, a binomial distribution has to be applied. The

400  canonical link for this distribution is the logit link. The proposed way of calculating the offset for

401  the logit link in this paper cannot be applied because the product of the number of species in the

402  plot and relative abundance at the meta-community scale was larger than one in 39 species-site

403    combinations (and it excludes 22 of 52 sites). Therefore, the logarithm of relative abundances in

404    the meta-community was used.

405    Two link functions were tested: the canonical (logit) link and log link, as suggested in this paper.

406    The latter was fitted using the *logbin* package (Donoghoe & Marschner, 2018).

407    A model containing only intercept and offset terms was fitted, so the predicted relative

408    abundances in plots should equal to the relative abundance in the meta-community. This

409    requirement was satisfied in the model using log link (not shown). However, when the canonical

410    link was used, there is a non-linear relationship between two vectors of relative abundances

411    (Figure 3).

412

## Example 3: comparing formulas for estimation of variation components

415    The aim of the next example is to compare variation components estimated by the method of

416    Shipley (2014) and the new method proposed in this paper. Simulated data were used, where

417    trends of variation components were predictable. Data were generated using the following

418    model:

$$y_i \sim Poisson(\lambda_i)$$
$$\log\lambda_i = a + \log\pi_i + s * t_i \tag{31}$$

419    where $\pi_i$ is the meta-community-level relative abundances, $|s|$ is the strength of selection, and $t_i$

420    is the trait value. To remove changes due to total community size, the intercept (a) was set to:

$$a = \log A - \log\sum_i \pi_i e^{st_i} \tag{32}$$

421    Thus, $\Sigma\lambda_i = A$ for any value of $s$.

422    The species pool consists of 50 species, and their traits follow a standard normal distribution.

423    The expected community size (A) was set to 2500, and the strength of selection (s) changes from

424    0 to 3. Pure selection and pure meta-community effects were calculated for each simulated

425    community separately using formulas from Shipley (2014) and the method proposed in this

426    paper.

427    There is a good agreement between variation components calculated by the two ways (Figure 4).

428    As expected, the pure meta-community effect decreases, while the pure selection effect increases

429     with increasing strength of selection, and the later was about zero at s=0 (Figure 5). The joint

430     effect and unexplained variation do not change considerably with changing strength of selection.

431     c

## Conclusions

433     CATS regression is a useful part of community ecologists' toolbox to understand how the

434     environment selects species through trait-environment relationships, as well as to estimate the

435     relative role of local environmental selection and meta-community-level processes in the

436     assembly of communities. In its original version using maximum entropy formalism (Shipley,

437     Vile & Garnier, 2006; Shipley, 2010), the assumptions of the methods remain hidden. Converting

438     the maximum entropy formalism into a GLM (Warton, Shipley & Hastie, 2015) made the

439     assumptions explicit. Warton et al. (2015) focused on the most important assumption (the

440     distribution of abundance values) and the central part of the method (the estimation of

441     parameters). The original version assumes a Poisson distribution, and not only parameter

442     estimates, but also the additional parts of the method (defining offset terms and calculating R-

443     squared values) may change when data follow an other distribution.

444     Parameter estimation for different distributions is a well-known statistical problem, and a user

445     can easily choose the appropriate function (or option of the applied function). This paper focused

446     on additional parts of the method, which are more specific and have thus received little attention

447     so far. Theoretical considerations and examples illustrated that naively using algorithms

448     developed for a Poisson distribution may be misleading when data follow other distributions.

449     The recommendations formulated in this paper could help to avoid these potential pitfalls.

450

## Acknowledgements

453

454

## References

456     Blasco-Moreno A, Pérez-Casany M, Puig P, Morante M, Castells E. 2019. What does a zero

457           mean? Understanding false, random and structural zeros in ecology. *Methods in*

458           *Ecology and Evolution* 10:949–959. DOI: 10.1111/2041-210X.13185.

459 Borcard D, Legendre P, Drapeau P. 1992. Partialling out the Spatial Component of

460       Ecological Variation. *Ecology* 73:1045–1055.

461 ter Braak CJF, Peres-Neto P, Dray S. 2017. A critical issue in model-based inference for

462       studying trait-based community assembly and a solution. *PeerJ* 5:e2885. DOI:

463       10.7717/peerj.2885.

464 Brooks ME, Kristensen K, Van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ,

465       Machler M, Bolker BM. 2017. glmmTMB balances speed and flexibility among

466       packages for zero-inflated generalized linear mixed modeling. DOI: 10.3929/ETHZ-

467       B-000240890.

468 Cameron AC, Windmeijer FAG. 1996. R-Squared Measures for Count Data Regression

469       Models with Applications to Health-Care Utilization. *Journal of Business & Economic*

470       *Statistics* 14:209. DOI: 10.2307/1392433.

471 Cameron CA, Windmeijer FAG. 1997. An R-squared measure of goodness of fit for some

472       common nonlinear regression models. *Journal of Econometrics* 77:329–342. DOI:

473       10.1016/S0304-4076(96)01818-0.

474 Chong F, Spencer M. 2018. Analysis of relative abundances with zeros on environmental

475       gradients: a multinomial regression model. *PeerJ* 6:e5643. DOI: 10.7717/peerj.5643.

476 Cribari-Neto F, Zeileis A. 2010. Beta Regression in *R. Journal of Statistical Software* 34. DOI:

477       10.18637/jss.v034.i02.

478 Damgaard C. 2008. Modelling pin-point plant cover data along an environmental gradient.

479       *Ecological Modelling* 214:404–410. DOI: 10.1016/j.ecolmodel.2008.03.012.

480 Damgaard C. 2009. On the distribution of plant abundance data. *Ecological Informatics*

481       4:76–82. DOI: 10.1016/j.ecoinf.2009.02.002.

482 Damgaard CF, Irvine KM. 2019. Using the beta distribution to analyse plant cover data.

483 *Journal of Ecology* 107:2747–2759. DOI: 10.1111/1365-2745.13200.

484 Dobson AJ. 2002. *An introduction to generalized linear models*. Chapman&Hall/CRC.

485 Donoghoe MW, Marschner IC. 2018. **logbin** : An *R* Package for Relative Risk Regression

486 Using the Log-Binomial Model. *Journal of Statistical Software* 86. DOI:

487 10.18637/jss.v086.i09.

488 Dunn PK, Smyth GK. 1996. Randomized Quantile Residuals. *Journal of Computational and*

489 *Graphical Statistics* 5:236–244. DOI: 10.1080/10618600.1996.10474708.

490 Dunstan PK, Foster SD, Hui FKC, Warton DI. 2013. Finite Mixture of Regression Modeling

491 for High-Dimensional Count and Biomass Data in Ecology. *Journal of Agricultural,*

492 *Biological, and Environmental Statistics* 18:357–375. DOI: 10.1007/s13253-013-

493 0146-x.

494 Feng C, Li L, Sadeghpour A. 2020. A comparison of residual diagnosis tools for diagnosing

495 regression models for count data. *BMC Medical Research Methodology* 20:175. DOI:

496 10.1186/s12874-020-01055-2.

497 Fisher RA. 1925. The influence of rainfall on the yield of wheat at Rothamsted. *Philosophical*

498 *Transactions of the Royal Society of London. Series B, Containing Papers of a*

499 *Biological Character* 213:89–142. DOI: 10.1098/rstb.1925.0003.

500 Goodall D. 1952. Some Considerations in the Use of Point Quadrats for the Analysis of

501 Vegetation. *Australian Journal of Biological Sciences* 5:1. DOI: 10.1071/BI9520001.

502 Laliberté E, Legendre P, Shipley B. 2014. *FD: measuring   functional diversity from multiple*

503 *traits, and other tools for   functional ecology. R package version 1.0-12.*

504 Lynch HJ, Thorson JT, Shelton AO. 2014. Dealing with under- and over-dispersed count data

505    in life history, spatial, and community ecology. *Ecology* 95:3173–3180. DOI:

506    10.1890/13-1912.1.

507 Marschner IC. 2015. Relative Risk Regression for Binary Outcomes: Methods and

508    Recommendations. *Australian & New Zealand Journal of Statistics* 57:437–462. DOI:

509    10.1111/anzs.12131.

510 Marschner IC, Gillett AC. 2012. Relative risk regression: reliable and flexible methods for

511    log-binomial models. *Biostatistics* 13:179–192. DOI: 10.1093/biostatistics/kxr030.

512 McCullagh P, Nelder JA. 1999. *Generalized linear models*. London: Chapman & Hall.

513 Menard S. 2000. Coefficients of Determination for Multiple Logistic Regression Analysis.

514    *The American Statistician* 54:17–24. DOI: 10.1080/00031305.2000.10474502.

515 Nakagawa S, Schielzeth H. 2013. A general and simple method for obtaining $R^2$ from

516    generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4:133–

517    142. DOI: 10.1111/j.2041-210x.2012.00261.x.

518 O'Hara RB, Kotze DJ. 2010. Do not log-transform count data. *Methods in Ecology and*

519    *Evolution* 1:118–122. DOI: 10.1111/j.2041-210X.2010.00021.x.

520 Ohlsson E, Johansson B. 2006. Exact Credibility and Tweedie Models. *ASTIN Bulletin*

521    36:121–133. DOI: 10.1017/S0515036100014422.

522 Peres-Neto PR, Legendre P, Dray S, Borcard D. 2006. Variation Partitioning of Species Data

523    Matrices: Estimation and Comparison of Fractions. *Ecology* 87:2614–2625.

524 Raevel V, Violle C, Munoz F. 2012. Mechanisms of ecological succession: insights from plant

525    functional strategies. *Oikos* 121:1761–1770. DOI: 10.1111/j.1600-

526    0706.2012.20261.x.

527    Ricci L. 2010. Adjusted -squared type measure for exponential dispersion models. *Statistics*

528         *& Probability Letters* 80:1365–1368. DOI: 10.1016/j.spl.2010.04.019.

529    Shipley B. 2010. *From Plant Traits to Vegetation Structure. Chance and Selection in the*

530         *Assembly of Ecological Communities*. Cambridge, UK: Cambridge University Press.

531    Shipley B. 2014. Measuring and interpreting trait-based selection versus meta-community

532         effects during local community assembly. *Journal of Vegetation Science* 25:55–65.

533         DOI: 10.1111/jvs.12077.

534    Shipley B, Vile D, Garnier É. 2006. From Plant Traits to Plant Communities: A Statistical

535         Mechanistic Approach to Biodiversity. *Science* 314:812–814. DOI:

536         10.1126/science.1131344.

537    Warton DI, Hui FKC. 2010. The arcsine is asinine: the analysis of proportions in ecology.

538         *Ecology* 92:3–10. DOI: 10.1890/10-0340.1.

539    Warton DI, Shipley B, Hastie T. 2015. CATS regression – a model-based approach to

540         studying trait-based community assembly. *Methods in Ecology and Evolution* 6:389–

541         398. DOI: 10.1111/2041-210X.12280.

542    Wood SN. 2017. *Generalized additive models: an introduction with R*. Boca Raton London

543         New York: CRC Press/Taylor & Francis Group.

544    Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. 2009. Zero-Truncated and Zero-

545         Inflated Models for Count Data. In: *Mixed effects models and extensions in ecology*

546         *with R*. Statistics for Biology and Health. New York, NY: Springer New York, 261–

547         293. DOI: 10.1007/978-0-387-87458-6_11.

548

549

550

# Table 1(on next page)

Defining distributions widely used for modeling abundances using notations of exponential family

See equation (17) for explanation of notations

1

| Distribution | $\psi$ | $b(\psi)$ | $a(\phi)$ | $c(y,\phi)$ |
|---|---|---|---|---|
| Gaussian (Normal) | $\mu$ | $\dfrac{\psi^2}{2}$ | $\sigma^2$ | $\dfrac{\sigma^2 ln(2\pi\sigma^2) + y^2}{2\sigma^2}$ |
| Poisson | $\ln\mu$ | $e^\psi$ | 1 | $-\ln y!$ |
| Binomial | $\ln\dfrac{\mu}{n-\mu}$ | $n\ln(1+e^\psi)$ | 1 | $\ln\binom{n}{y}$ |
| Negative binomial | $\ln\dfrac{\mu}{\mu+\theta}$ | $-\theta ln(1-e^\psi)$ | 1 | $\ln\dfrac{\Gamma(\theta)y!}{\Gamma(y+\theta)}$ |
| Tweedie (1<p<2) | $\dfrac{\mu^{1-p}}{1-p}$ | $\dfrac{[\psi(1-p)]^{(2-p)/(1-p)}}{2-p}$ | $\phi$ | $\begin{array}{ll} 0 & \text{if } y = 0 \\ \ln W(y,0) - \ln y & \text{if } y > 0 \end{array}$ |
| Zero-truncated Poisson | $\ln\mu$ | $e^\psi + \ln\left[1 - exp(-e^\psi)\right]$ | 1 | $-\ln y!$ |
| Zero-truncated negative binomial | $\ln\dfrac{\mu}{\mu+\theta}$ | $-\theta ln(1-e^\psi) + ln\left[1 - (1-e^\psi)^{-\theta}\right]$ | 1 | $\ln\dfrac{\Gamma(\theta)y!}{\Gamma(y+\theta)}$ |

2    Notation: $\Gamma(x)$ is the gamma-function.

3

# Figure 1

Diagnostic plots of models fitted to one community of Example 1

The fan shape of points in residuals vs. fitted values plot (upper row) and departure from the expected line in QQ-plot (bottom row) indicate that Poisson model is inappropriate due to over-dispersion.
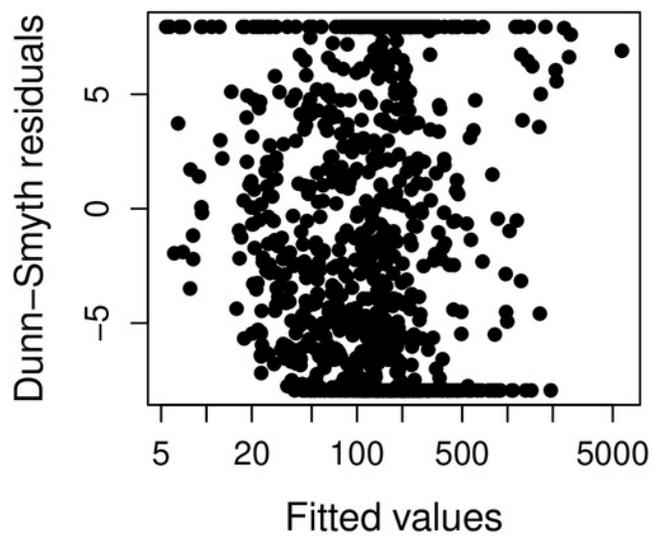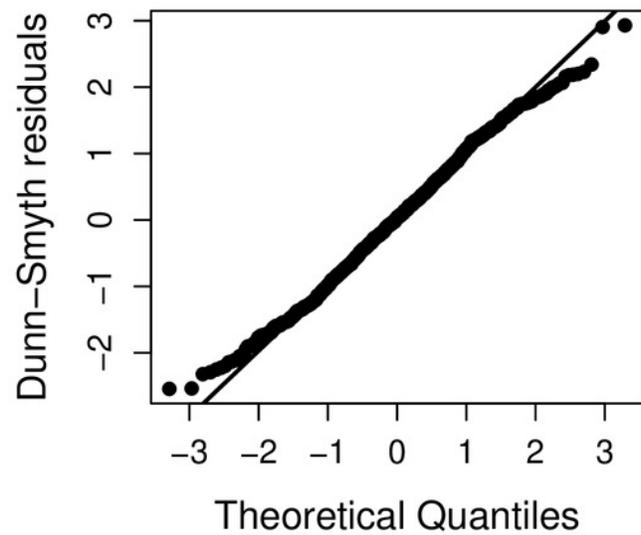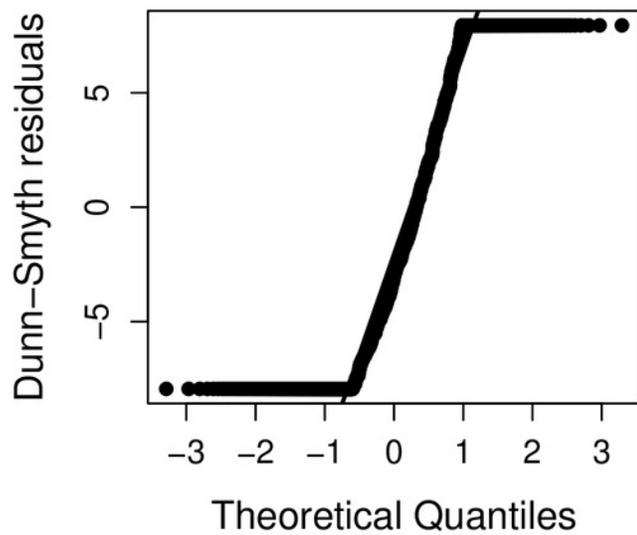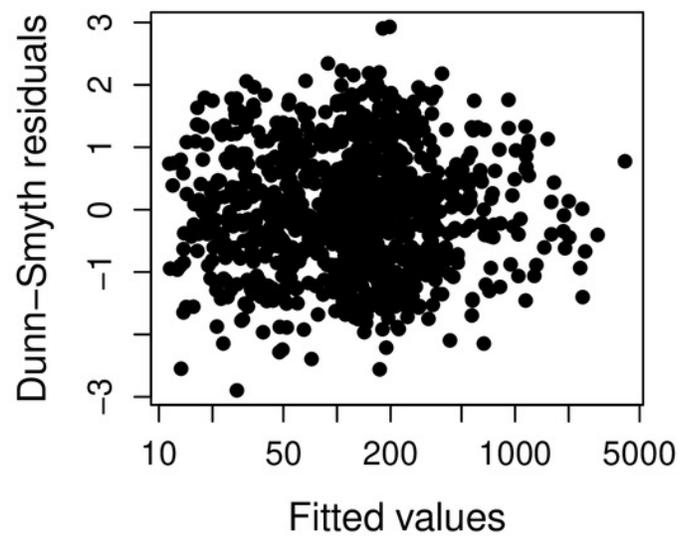
# Figure 2

Estimated slopes with their 95% confidence intervals in 50 simulated plots of Example 1.

Red horizontal line indicates the real slope used in the simulation.
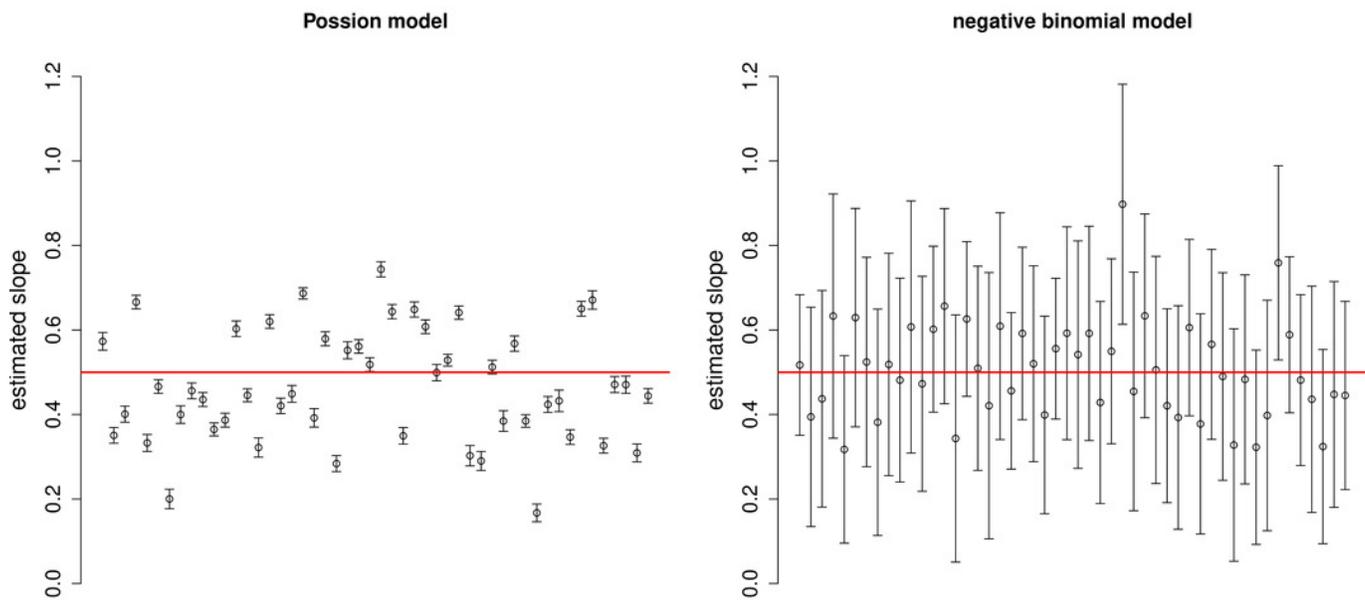
# Figure 3

Relationship between meta-community level and predicted relative abundances in model without traits using logit link

Since local selection is not modelled, points should lie the red 1:1 line.
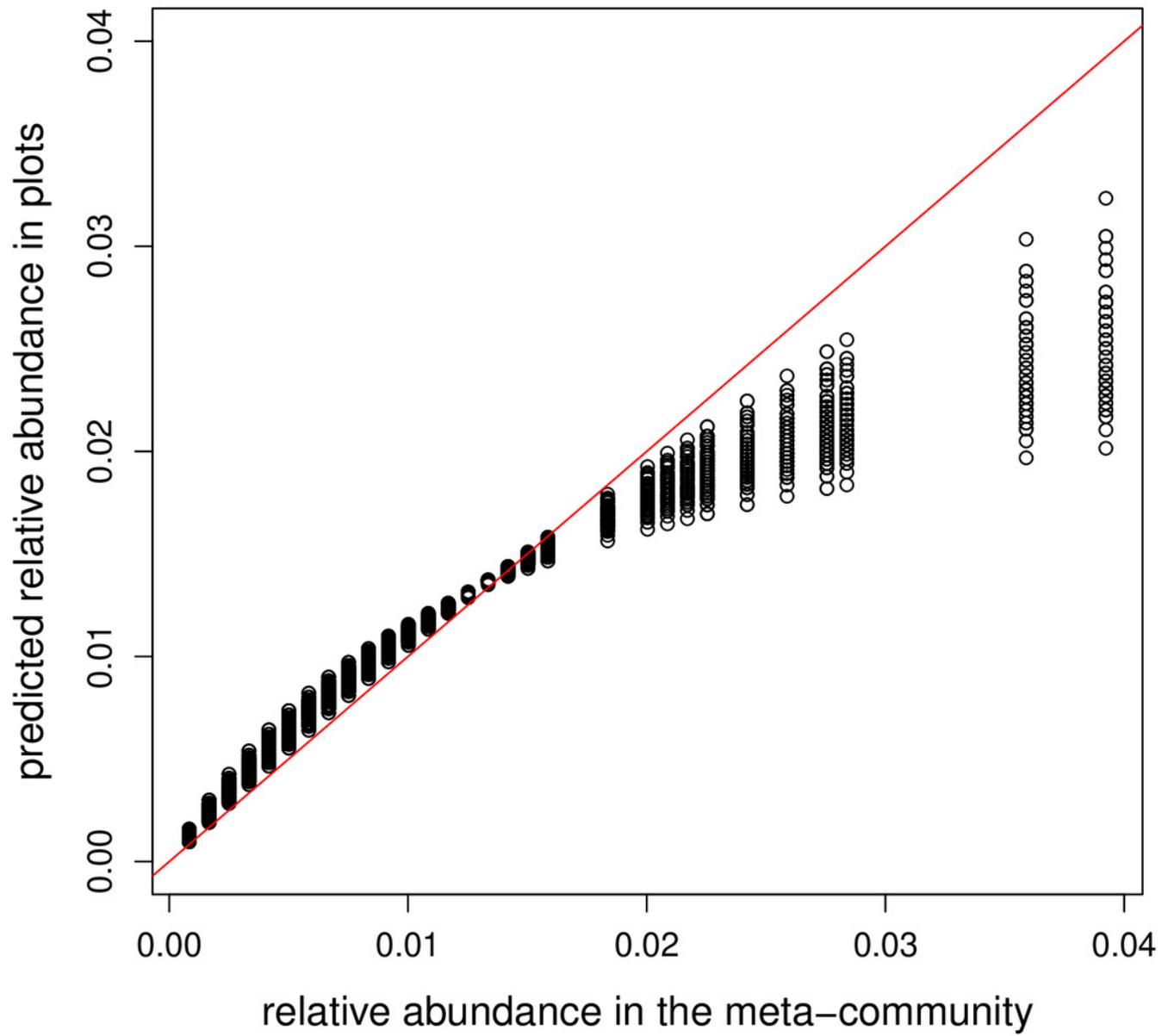
# Figure 4

Comparing variation components calculated by Shipley's formulas and new formulas proposed in this paper.

Components calculated by two ways show good agreement.
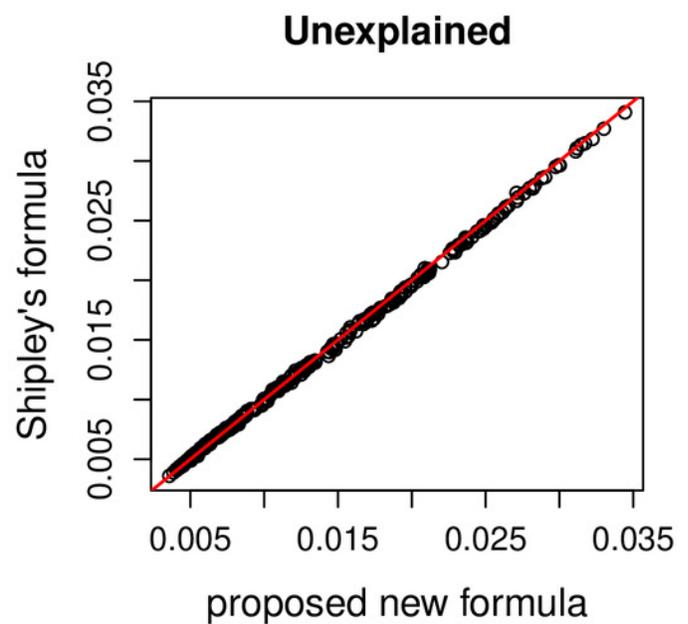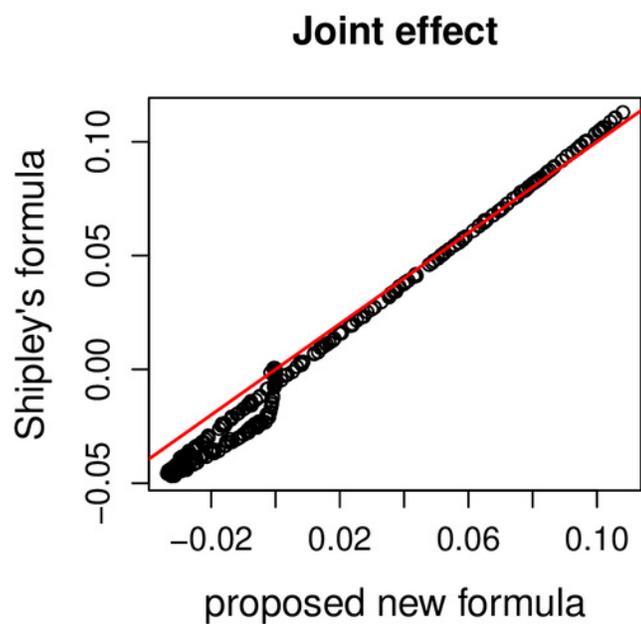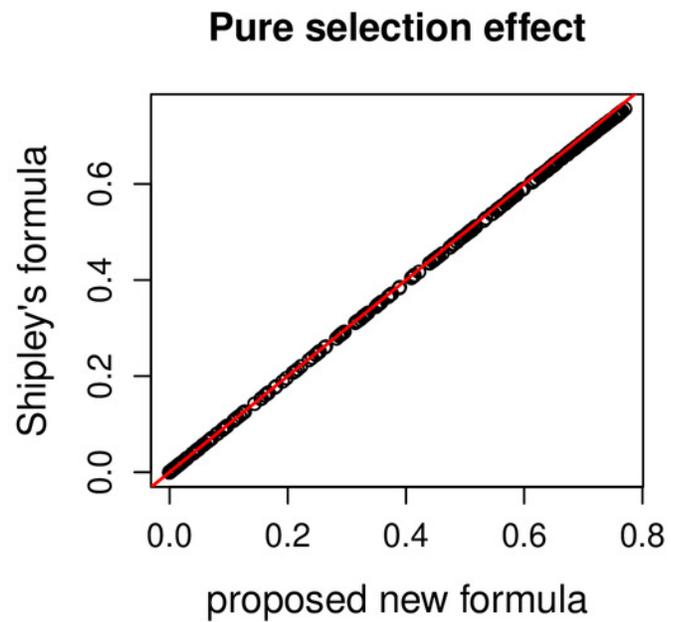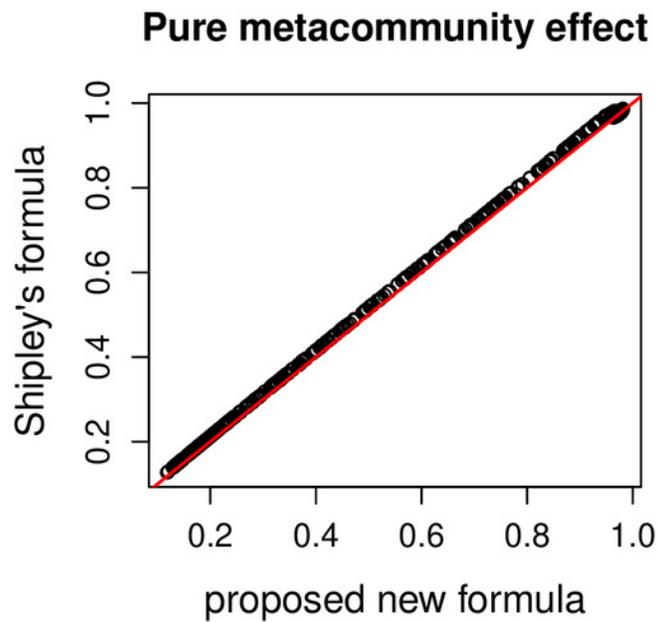
# Figure 5

Variation components in simulated communities that differ in strength of selection calculated by formulas proposed in this paper.

As expected, meta-community effect decreases, while selection effect increases with increasing strength of selection, and the former is near zero at s=0 when there is no selection in the simulation.