

Web Supplementary Materials for: “An Overview of Modern Applications of Negative Binomial Modelling in Ecology and Biodiversity”

Jakub Stoklosa^{*1}, Rachel V. Blakey², and Francis K.C. Hui³

¹School of Mathematics and Statistics and Evolution & Ecology Research
Centre, The University of New South Wales, Sydney, Australia

²UCLA La Kretz Center for California Conservation Science, Institute of
the Environment and Sustainability, University of California, Los Angeles

³Research School of Finance, Actuarial Studies and Statistics, The
Australian National University, Canberra, Australia

^{*}Corresponding author: j.stoklosa@unsw.edu.au; School of Mathematics and Statistics and Evolution & Ecology Research Centre, The University of New South Wales, Sydney, 2052, NSW, Australia

Web Tables

Table S1. Additional notes and examples for log-linear negative binomial (NB) regression models.

Relative abundance models:

Additional notes: The dispersion parameter serves as a measure of the degree of aggregation or clustering for species in question [5], and as $\kappa \rightarrow 0$ this may reflect an increasing degree of spatial aggregation [22] or reflect the stability of host-parasite interactions, for instance [23, 24].

Examples: See Lindén and Mäntyniemi [14] who describe various scenarios how sampling error, flocking, and environmental stochasticity causes overdispersion in daily bird migration count data.

Log-linear NB regression models:

Additional notes: For fixed κ , the NB distribution in equation (2) belongs to the exponential family of distributions [15], allowing us to exploit the fast and stable estimation and inferential methods that have been rigorously established for GLMs and apply them to the NB regression model; see also Section 4.

Examples: A relatively comprehensive example of this is given in Zuur et al. [25], who fitted Poisson, NB and quasi-Poisson regression models to counts of amphibian road killings using various covariates, and through residual analysis concluded that overdispersion was clearly present and thus the NB regression provided a much better fit to the count data compared to the other two models. See also Wilson et al. [23], who compared log-linear NB, Poisson and log-transformed regression models on a long-term study of parasitism on counts of Soay Sheep on St Kilda.

NB generalized additive models:

Additional notes: Various modifications of this include bivariate smoothers to flexibly handle interactions, varying coefficients models where the smoothing function itself depends on a categorical covariate (say), and single index models.

Examples: See Rigby and Stasinopoulos [17] for further details on NB GAMs and Stoklosa and Warton [20] for a recent application of NB semiparametric regression models applied to estimate abundances of multiple species of arthropods using smoothing terms for trait and environment covariates.

Table S2. Additional notes and examples for negative binomial (NB) species distribution models.

NB stacked species distribution models:

Additional notes: Because the SDMs are stacked (and effectively assumed to be independent) then estimation can be done by fitting s separate NB models to each species, see Wang et al. [21] and Naimi and Araujo [16]. For performing inference however, one must be careful to ensure that any potential residual correlation between the s species is accounted for otherwise. With $s > 1$ species, one can extend these methods by forming a SDM for each species separately and stacking to form what is commonly referred to as a stacked species distribution model [e.g., 6].

Examples: A recent example on this type of modelling was given in Caradima et al. [7] who analysed 245 *benthic macroinvertebrate* taxa at multiple sites and times using species richness models. To better describe the variability in these data (*i.e.*, a more realistic representation for the observed variance), they fitted NB stacked SDMs and compared them with joint species distribution models (see Section 3.6).

NB fourth-corner models:

Additional notes: Instead of species-specific coefficients $\beta^{(k)}$, the number of parameters in a NB fourth-corner model is reduced to a common set of parameters γ , which are subsequently used to answer questions related to how traits mediate the species-environmental responses in a community.

Examples: Two examples of NB fourth-corner models can be found in Brown et al. [4], Stoklosa et al. [19], who considered variable selection on a large number of interactions terms consisting of species traits and environmental covariates on arthropod count data with strong overdispersion, and Blakey et al. [2] who used acoustic bat surveys to investigate relationships between forest structure and bat communities across a vegetation structural gradient at community, species and trait levels.

Table S3. Additional notes and examples for negative binomial (NB) generalized linear mixed models.

NB generalized linear mixed models:

Additional notes: The set of random effects included may be much more sophisticated depending on the data and sampling design [18]. Not surprisingly, the NB GLMM can also be combined with other extensions discussed above *e.g.*, NB generalized additive mixed models where $X_{ij}^\top \beta$, say, is replaced by a set of smoothing functions for each covariate, and stacked NB GLMMs for $s > 1$ species. Also, for NB fourth-corner GLMMs, see Ives and Helmus [13].

Examples: An example where NB GLMMs were fitted to count data is given in Blakey et al. [3], who analysed foraging activity of a bat community collected at multiple times at sites. After initially fitting Poisson linear mixed models, they identified patterns of overdispersion from residuals plots. To account for overdispersion and model any heterogeneity between units, both the site and management units were treated as random effects in their NB mixed models. Another example that uses a Bayesian approach is given in Chapter 23.8 of Zuur et al. [25] who considered auto-correlated random effects models for overdispersed count data collected on seals.

NB spatial and spatio-temporal models:

Additional notes: If each observational unit is observed over time, then these models can be combined with the NB GLMMs discussed in the above to even more sophisticated complex random effects structures that incorporate both spatial and temporal information; see Cressie et al. [8], Cressie and Wikle [9] and Irwin et al. [12] for instance. Interpretation of fixed effects should be carefully considered under these models.

Examples: Early examples of NB spatial models in ecology include Gregoire [11] and Diggle and Milne [10], who considered the negative binomial distribution for quadrat counts, and Alexander et al. [1] who used a Bayesian approach to inference on spatial negative binomial models following a similar model structure as above but modelled the dispersion parameter k with a gamma distribution as their prior. They analysed individual-level parasite count data collected across spatial sampling units, and concluded that using “a NB distribution to model extra-Poisson variation gives a natural interpretation when dealing with parasite counts, with the k parameter being a convenient index of overdispersion.”

Web Figures

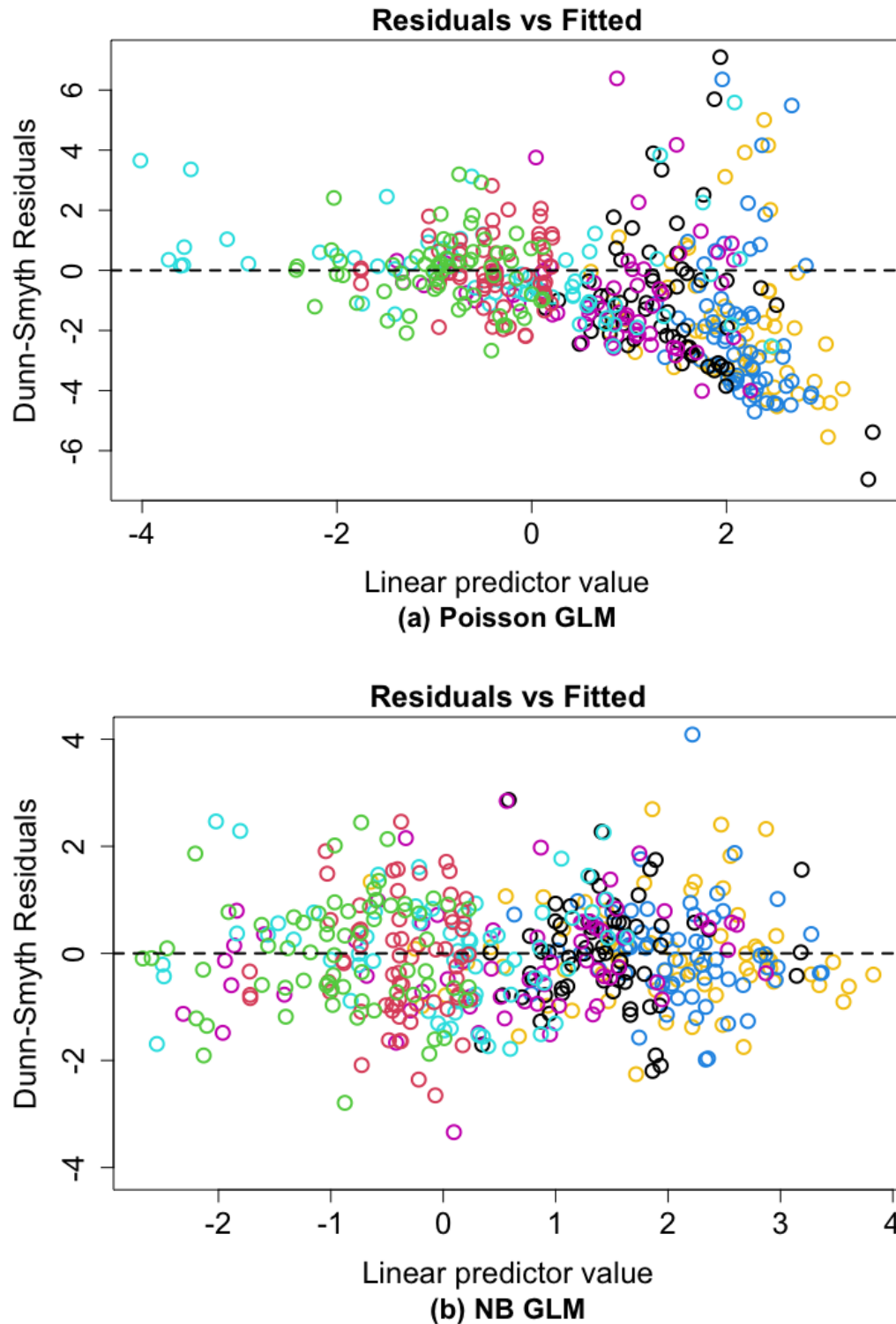


Figure S1. Dunn–Smyth residuals vs the linear predictor values for the (a) Poisson (top) and (b) negative binomial (bottom) GLMs when fitted to the bat acoustic data. Notice the obvious funnelling effect in the top figure but no obvious pattern in the bottom figure, suggesting there is strong overdispersion.

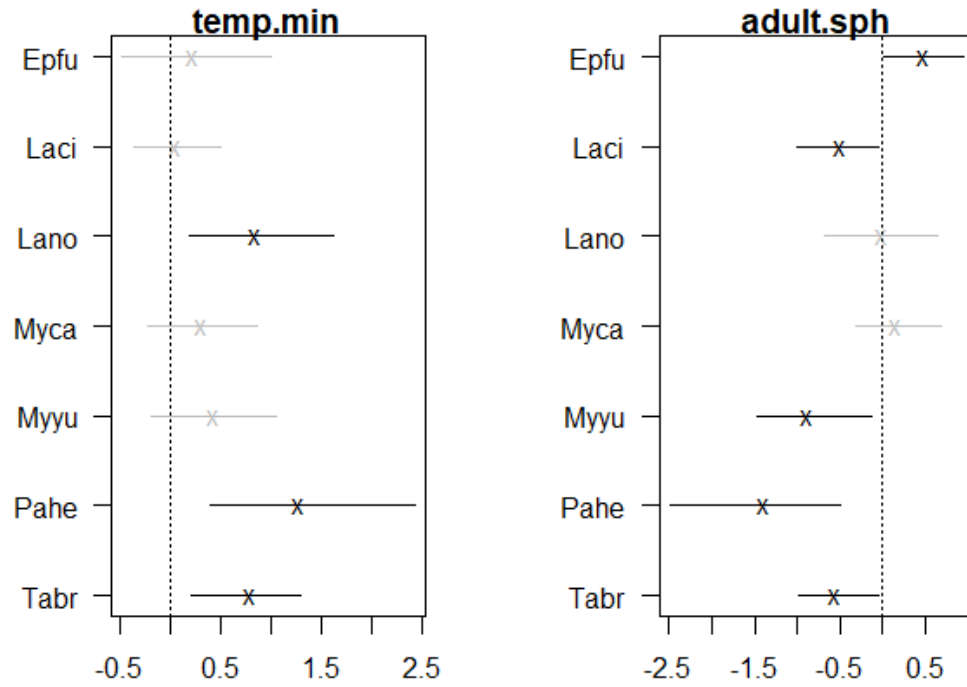


Figure S2. Caterpillar plot of regression coefficients with credible intervals when fitting a NB joint species distribution to the bat acoustic data. The grey lines/crosses indicate that the credible intervals contain the value zero. The two covariates are: `temp.min` minimum temperature and `adult.sph` stem density of adult trees. The seven bat species are labelled by the following abbreviated names: *Tadarida brasiliensis* (Tabr), *Eptesicus fuscus* (Epfu), *Lasionycteris noctivagans* (Lano), *Lasiurus cinereus* (Laci), *Parastrellus hesperus* (Pahe), *Myotis yumanensis* (Myyu), and *Myotis californicus* (Myca).

References

- [1] Alexander, N., Moyeed, R., and Stander, J. (2000). Spatial modelling of individual-level parasite counts using the negative binomial distribution. *Biostatistics*, 1:453–463.
- [2] Blakey, R. V., Law, B. S., Kingsford, R. T., and Stoklosa, J. (2017). Terrestrial laser scanning reveals below-canopy bat trait relationships with forest structure. *Remote Sens. Environ.*, 198:40–51.
- [3] Blakey, R. V., Law, B. S., Kingsford, R. T., Tap, P., Stoklosa, J., and Williamson, K. (2016). Bat and invertebrate communities respond positively to large-scale thinning of forest regrowth. *J. Appl. Ecol.*, 53:1694–1703.
- [4] Brown, A. M., Warton, D. I., Andrew, N. R., Binns, M., Cassis, G., and Gibb, H. (2014). The fourth-corner solution – using predictive models to understand how species traits interact with the environment. *Methods Ecol. Evol.*, 5:344–352.
- [5] Brown, J. H., Mehlman, D. W., and Stevens, G. C. (1995). Spatial variation in abundance. *Ecology*, 76:2028–2043.
- [6] Calabrese, J. M., Certain, G., Kraan, C., and Dormann, C. F. (2013). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecol. Biogeogr.*, 23:99–112.
- [7] Caradima, B., Schuwirth, N., and Reichert, P. (2019). From individual to joint species distribution models: A comparison of model complexity and predictive performance. *J. Biogeogr.*, 46:2260–2274.
- [8] Cressie, N., Calder, C. A., Clark, J. S., Ver Hoef, J. M., and Wikle, C. K. (2009). Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecol. Appl.*, 19:553–570.
- [9] Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-temporal Data*. John Wiley & Sons, New Jersey.
- [10] Diggle, P. J. and Milne, R. K. (1983). Negative binomial quadrat counts and point processes. *Scand. J. Stat.*, 10:257–267.

- [11] Gregoire, G. (1984). Negative binomial distributions for point processes. *Stoch. Proc. Appl.*, 16:179–188.
- [12] Irwin, B., Wagner, T., Bence, J., Kepler, M., Liu, W., and Hayes, D. (2013). Estimating spatial and temporal components of variation for fisheries count data using negative binomial mixed models. *T. Am. Fish. Soc.*, 142:171–183.
- [13] Ives, A. R. and Helmus, M. (2011). Generalized linear mixed models for phylogenetic analyses of community structure. *Ecol. Monogr.*, 81:511–525.
- [14] Lindén, A. and Mäntyniemi, S. (2011). Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92:1414–1421.
- [15] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall/CRC, London, 2nd edition edition.
- [16] Naimi, B. and Araujo, M. B. (2016). sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*, 39:368–375.
- [17] Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *J. R. Stat. Soc. C-Appl.*, 54:507–554.
- [18] Schielzeth, H. and Nakagawa, S. (2013). Nested by design: model fitting and interpretation in a mixed model era. *Methods Ecol. Evol.*, 4:14–24.
- [19] Stoklosa, J., Gibb, H., and Warton, D. I. (2014). Fast forward selection for generalized estimating equations with a large number of predictor variables. *Biometrics*, 70:110–120.
- [20] Stoklosa, J. and Warton, D. I. (2018). A generalized estimating equation approach to multivariate adaptive regression splines. *J. Comput. Graph. Stat.*, 27:245–253.
- [21] Wang, Y., Naumann, U., Wright, S. T., and Warton, D. I. (2012). mvabund – an R package for model-based analysis of multivariate abundance data. *Methods Ecol. Evol.*, 3:471–474.
- [22] White, G. C. and Bennetts, R. E. (1996). Analysis of frequency count data using the negative binomial distribution. *Ecology*, 77:2549–2557.

- [23] Wilson, K., Gerenfell, B., and Shaw, D. (1996). Analysis of aggregated parasite distributions: A comparison of methods. *Funct. Ecol.*, 10:592–601.
- [24] Young, L. and Young, J. (1990). A spatial view of the negative binomial parameter k when describing insect populations. *Conf. App. Stat. Agric.*
- [25] Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. A. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer–Verlag, New York.