

Generalized linear mixed models

Ben Bolker

McMaster University, Mathematics & Statistics and Biology

28 April 2016

Acknowledgments

- lme4: Doug Bates, Martin Mächler, Steve Walker
- Data: Josh Banta, Adrian Stier, Sea McKeon, David Julian, Jada-Simone White
- NSERC (Discovery)
- SHARCnet

Outline

- 1 Examples and definitions
- 2 Estimation
 - Overview
 - Methods
- 3 Inference
- 4 Challenges & open questions

Outline

- 1 Examples and definitions
- 2 Estimation
 - Overview
 - Methods
- 3 Inference
- 4 Challenges & open questions

(Generalized) linear mixed models

(G)LMMs: a statistical modeling framework incorporating:

- combinations of categorical and continuous predictors, and interactions
- (some) non-Normal responses
(e.g. binomial, Poisson, and extensions)
- (some) nonlinearity
(e.g. logistic, exponential, hyperbolic)
- non-independent (grouped) data

(Generalized) linear mixed models

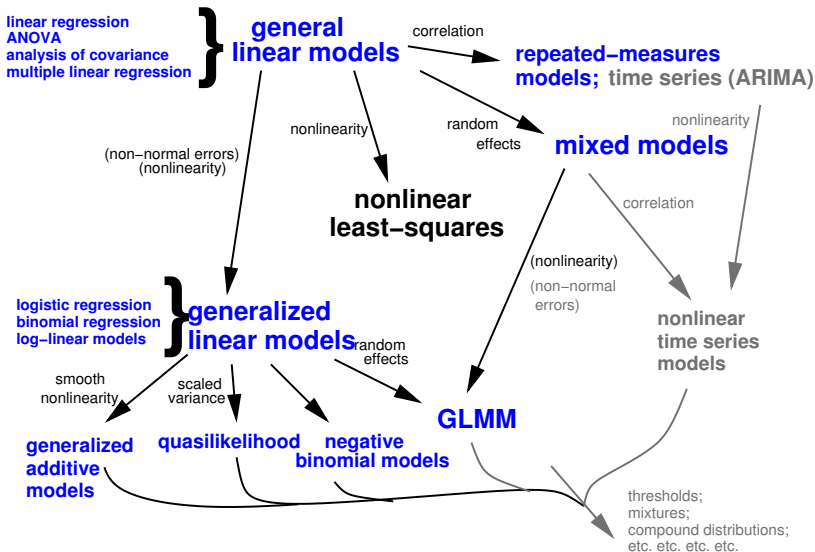
(G)LMMs: a statistical modeling framework incorporating:

- combinations of categorical and continuous predictors, and interactions
- (some) non-Normal responses
(e.g. binomial, Poisson, and extensions)
- (some) nonlinearity
(e.g. logistic, exponential, hyperbolic)
- non-independent (grouped) data

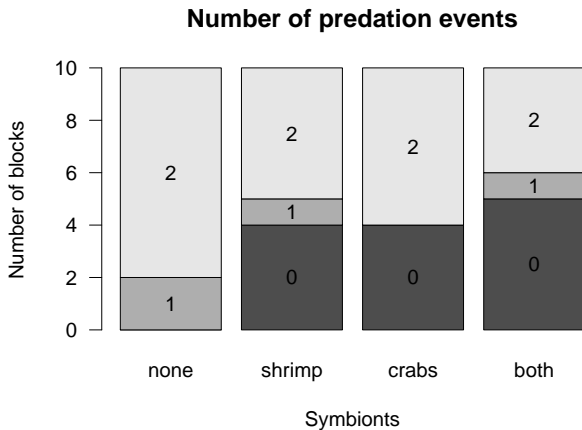
(Generalized) linear mixed models

(G)LMMs: a statistical modeling framework incorporating:

- combinations of categorical and continuous predictors, and interactions
- (some) non-Normal responses
(e.g. binomial, Poisson, and extensions)
- (some) nonlinearity
(e.g. logistic, exponential, hyperbolic)
- non-independent (grouped) data

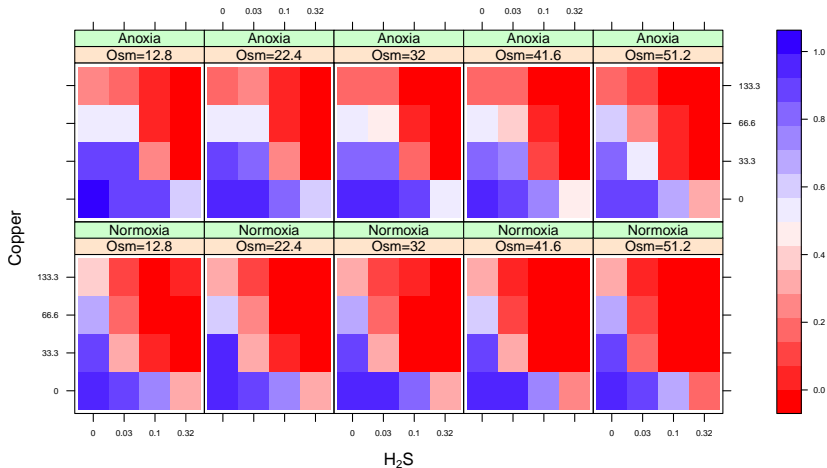


Coral protection from seastars (*Culcita*) by symbionts (McKeon et al., 2012)

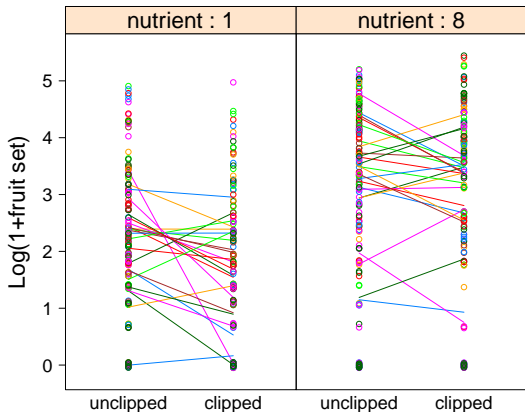


Environmental stress: *Glycera* cell survival

(D. Julian unpubl.)

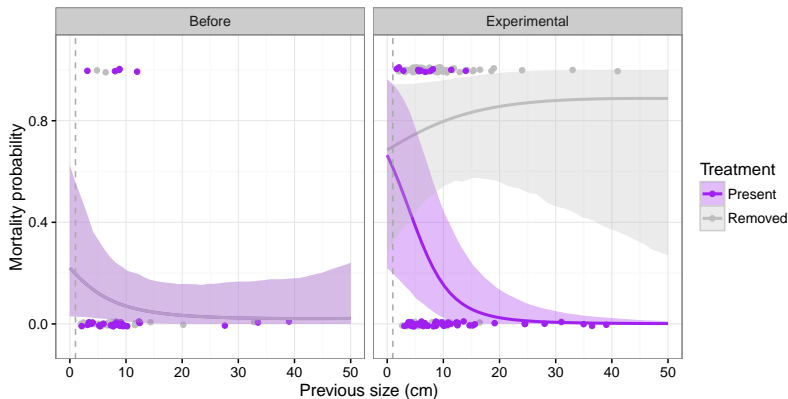


Arabidopsis response to fertilization & herbivory (Banta et al., 2010)



Coral demography

(J.-S. White unpubl.)



Technical definition

$$\underbrace{Y_i}_{\text{response}} \sim \underbrace{\text{Distr}}_{\text{conditional distribution}} \left(\underbrace{g^{-1}(\eta_i)}_{\text{inverse link function}}, \underbrace{\phi}_{\text{scale parameter}} \right)$$

Technical definition

$$\underbrace{Y_i}_{\text{response}} \sim \overbrace{\text{Distr}}^{\text{conditional distribution}} \left(\underbrace{g^{-1}(\eta_i)}_{\text{inverse link function}}, \underbrace{\phi}_{\text{scale parameter}} \right)$$
$$\underbrace{\eta}_{\text{linear predictor}} = \underbrace{X\beta}_{\text{fixed effects}} + \underbrace{Zb}_{\text{random effects}}$$

Technical definition

$$\underbrace{Y_i}_{\text{response}} \sim \overbrace{\text{Distr}}^{\text{conditional distribution}} \left(\underbrace{g^{-1}(\eta_i)}_{\text{inverse link function}}, \underbrace{\phi}_{\text{scale parameter}} \right)$$
$$\underbrace{\eta}_{\text{linear predictor}} = \underbrace{X\beta}_{\text{fixed effects}} + \underbrace{Zb}_{\text{random effects}}$$
$$\underbrace{b}_{\text{conditional modes}} \sim \text{MVN}(\mathbf{0}, \underbrace{\Sigma(\theta)}_{\text{variance-covariance matrix}})$$

What are random effects?

A method for . . .

- accounting for among-individual, within-block correlation
- compromising between
 - complete pooling** (no among-block variance)
 - and **fixed effects** (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

What are random effects?

A method for . . .

- accounting for among-individual, within-block correlation
- compromising between
 - complete pooling** (no among-block variance)
 - and **fixed effects** (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

What are random effects?

A method for . . .

- accounting for among-individual, within-block correlation
- compromising between
 - complete pooling** (no among-block variance)
 - and **fixed effects** (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

What are random effects?

A method for . . .

- accounting for among-individual, within-block correlation
- compromising between
 - complete pooling** (no among-block variance)
 - and **fixed effects** (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

What are random effects?

A method for . . .

- accounting for among-individual, within-block correlation
- compromising between
 - complete pooling** (no among-block variance)
 - and **fixed effects** (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

What are random effects?

A method for . . .

- accounting for among-individual, within-block correlation
- compromising between
complete pooling (no among-block variance)
and **fixed effects** (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

Random-effect myths

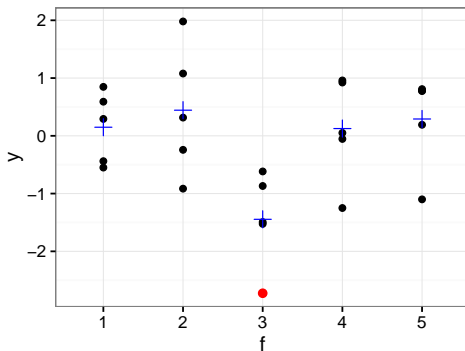
- levels of random effects must always be sampled at random
- a complete sample cannot be treated as a random effect
- random effects are always a **nuisance variable**
- nothing can be said about the predictions of a random effect
- you should always use a random effect no matter how few levels you have

Outline

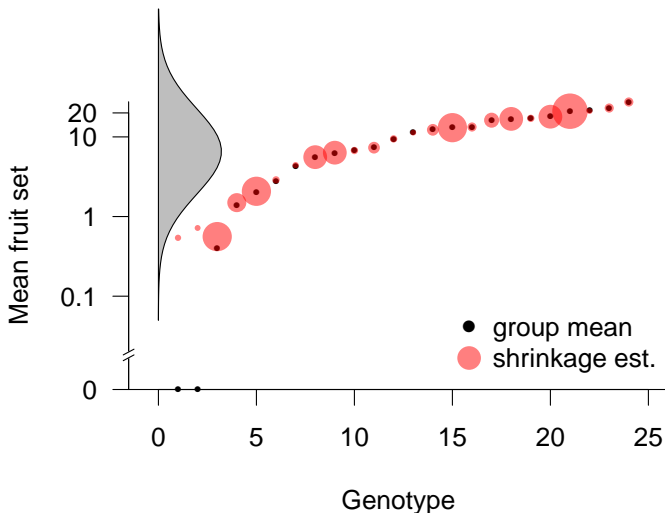
- 1 Examples and definitions
- 2 Estimation
 - Overview
 - Methods
- 3 Inference
- 4 Challenges & open questions

Maximum likelihood estimation

- Best fit is a compromise between two components
(consistency of data with fixed effects and conditional modes;
consistency of random effect with RE distribution)
- Goodness-of-fit *integrates* over conditional modes



Shrinkage: *Arabidopsis* conditional modes



Estimation methods

deterministic : various approximate integrals (Breslow, 2004)

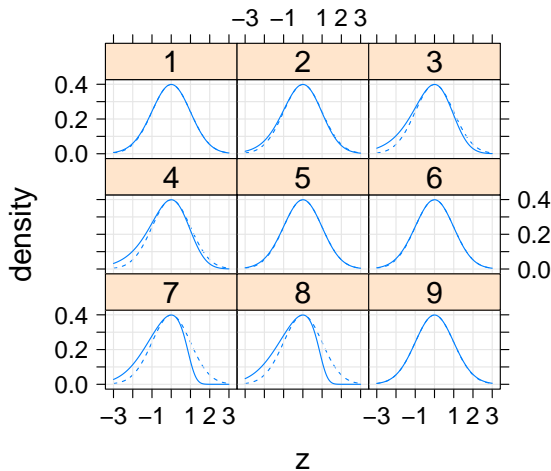
- Penalized quasi-likelihood, Laplace, Gauss-Hermite quadrature, ... (Biswas, 2015); best methods needed for large variance, small clusters
- flexibility and speed vs. accuracy

...

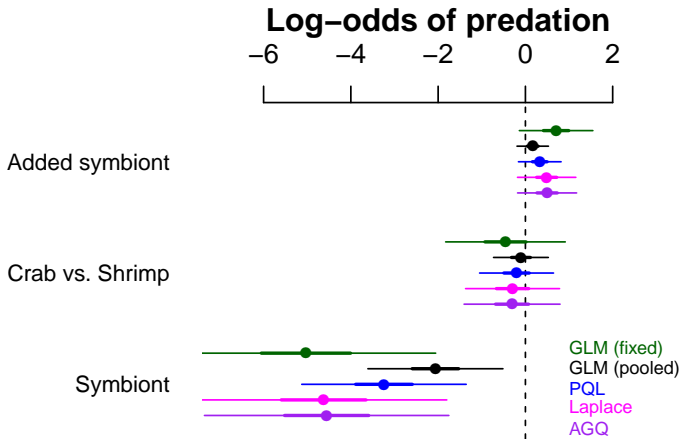
stochastic (Monte Carlo): frequentist and Bayesian (Booth and Hobert, 1999; Ponciano et al., 2009; Sung and Geyer, 2007)

- usually slower but flexible and accurate

Laplace-approximation diagnostics



Estimation: *Culcita* (McKeon et al., 2012)

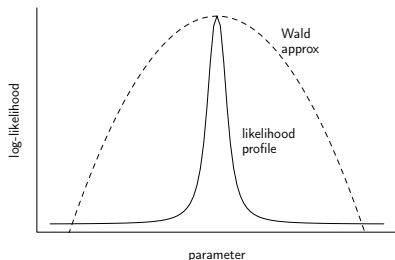


Outline

- 1 Examples and definitions
- 2 Estimation
 - Overview
 - Methods
- 3 Inference
- 4 Challenges & open questions

Wald tests

- typical results of summary
- exact for ANOVA, regression: approximation for GLM(M)s
- fast
- approximation is sometimes awful (Hauck-Donner effect)



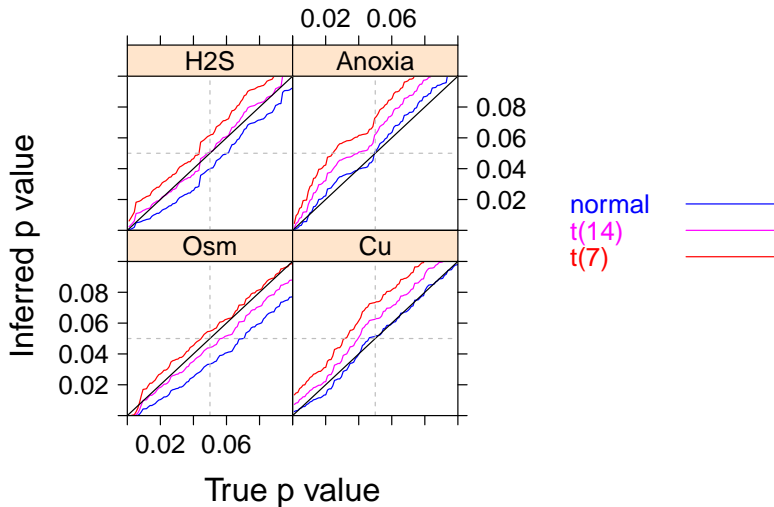
Likelihood ratio tests

- better than Wald, but still have to two problems:
 - “denominator degrees of freedom” (when estimating scale)
 - for GLMMs, distributions are approximate anyway (Bartlett corrections)
 - Kenward-Roger correction? (Stroup, 2014)
- Profile confidence intervals: expensive/fragile

Parametric bootstrapping

- fit null model to data
- simulate “data” from null model
- fit null and working model, compute likelihood difference
- repeat to estimate null distribution
- should be OK but ??? not well tested
(assumes estimated parameters are “sufficiently” good)

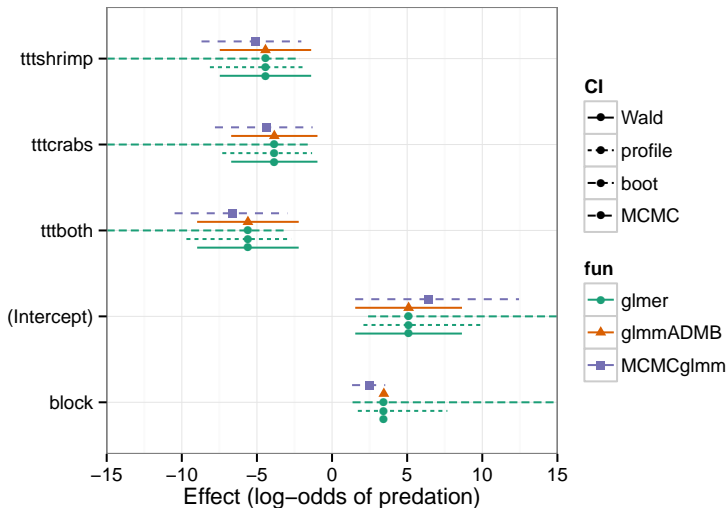
Parametric bootstrap results (*Glycera*)



Bayesian inference

- If we have a good sample from the posterior distribution (Markov chains have converged etc. etc.) we get most of the inferences we want for free by summarizing the marginal posteriors
- *post hoc* Bayesian methods: use deterministic/frequentist methods to find the maximum, then sample around it

Culcita confidence intervals



formula formats

- `fixed`: fixed-effect formula
- `random`: random-effect formula (in `lme4`, combined with `fixed`)
 - generally `x|g` (term|grouping variable)
 - simplest: `1|g`, single intercept term
 - nested: `1|g1/g2`
 - random-slopes: `r|g`
 - independent terms: `(1|g)+(x+0|g)` or `(x||g)`
- `lme`: weights, correlation for heteroscedasticity and residual correlation
- `MCMCglmm`: options for variance structure

Outline

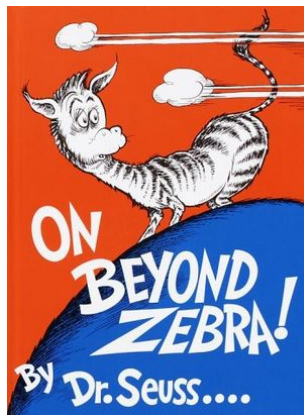
- ① Examples and definitions
- ② Estimation
 - Overview
 - Methods
- ③ Inference
- ④ Challenges & open questions

On beyond lme4

- glmmADMB, glmmTMB: zero-inflated and other distributions
- brms, rstanarm: interfaces to Stan
- INLA: spatial and temporal correlations

On beyond R

- Julia: MixedModels package
- SAS: PROC MIXED, NLMIXED
- AS-REML
- Stata (GLLAMM, xtmelogit)
- AD Model Builder; Template Model Builder
- HLM, MLWiN
- JAGS, Stan, [rethinking package](#)



Challenges

- Small clusters: need AGQ/MCMC
- Small numbers of clusters: need finite-size corrections (KR/PB/MCMC)
- Small data sets: issues with **singular** fits
Barr et al. (2013) vs. Bates et al. (2015)
- Big data: speed!
- Model diagnosis
- Confidence intervals accounting for uncertainty in variances

See also: https://rawgit.com/bbolker/mixedmodels-misc/master/ecostats_chap.html <https://groups.nceas.ucsb.edu/non-linear-modeling/projects>

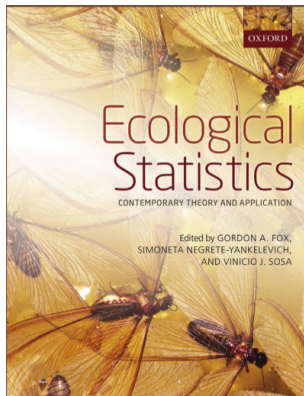
Spatial and temporal correlations

- Sometimes blocking takes care of non-independence ...
- but sometimes there is temporal or spatial correlation **within** blocks
- ... also phylogenetic ... (Ives and Zhu, 2006)
- “G-side” vs. “R-side” effects
- tricky to implement for GLMMs,
but new possibilities on the horizon (Rousset and Ferdy, 2014;
Rue et al., 2009);
<https://github.com/stevencarlislewalker/lme4ord>

Next steps

- Complex random effects:
regularization, model selection, penalized methods
(lasso/fence)
- Flexible correlation and variance structures
- Flexible/nonparametric random effects distributions
- hybrid & improved MCMC methods
- **Reliable** assessment of out-of-sample performance

- <http://ms.mcmaster.ca/~bolker/misc/private/14-Fox-Chap13.pdf>
- https://rawgit.com/bbolker/mixedmodels-misc/master/ecostats_chap.html
- Bolker (2015)



(code ASPROMP8)

References

- Banta, J.A., Stevens, M.H.H., and Pigliucci, M., 2010. *Oikos*, 119(2):359–369. ISSN 1600-0706. doi:10.1111/j.1600-0706.2009.17726.x.
- Barr, D.J., Levy, R., et al., 2013. *Journal of Memory and Language*, 68(3):255–278. ISSN 0749-596X. doi:10.1016/j.jml.2012.11.001.
- Bates, D., Kliegl, R., et al., 2015. *arXiv:1506.04967 [stat]*. ArXiv: 1506.04967.
- Biswas, K., 2015. *Performances of different estimation methods for generalized linear mixed models*. Master's thesis, McMaster University.
- Bolker, B.M., 2015. In G.A. Fox, S. Negrete-Yankelevich, and V.J. Sosa, editors, *Ecological Statistics: Contemporary theory and application*, chapter 13. Oxford University Press. ISBN 978-0-19-967255-4.
- Booth, J.G. and Hobert, J.P., 1999. *Journal of the Royal Statistical Society. Series B*, 61(1):265–285. doi:10.1111/1467-9868.00176.
- Breslow, N.E., 2004. In D.Y. Lin and P.J. Heagerty, editors, *Proceedings of the second Seattle symposium in biostatistics: Analysis of correlated data*, pages 1–22. Springer. ISBN 0387208623.
- Ives, A.R. and Zhu, J., 2006. *Ecological Applications*, 16(1):20–32.
- McKeon, C.S., Stier, A., et al., 2012. *Oecologia*, 169(4):1095–1103. ISSN 0029-8549. doi:10.1007/s00442-012-2275-2.
- Ponciano, J.M., Taper, M.L., et al., 2009. *Ecology*, 90(2):356–362. ISSN 0012-9658.
- Rousset, F. and Ferdy, J.B., 2014. *Ecography*, page no–no. ISSN 1600-0587. doi:10.1111/ecog.00566.
- Rue, H., Martino, S., and Chopin, N., 2009. *Journal of the Royal Statistical Society, Series B*, 71(2):319–392.
- Stroup, W.W., 2014. *Agronomy Journal*, 106:1–17. doi:10.2134/agronj2013.0342.
- Sung, Y.J. and Geyer, C.J., 2007. *The Annals of Statistics*, 35(3):990–1011. ISSN 0090-5364. doi:10.1214/009053606000001389.