

mixed models: introduction

Ben Bolker
18 November 2025

Table of contents

- Overview
- Examples
- Definitions and descriptions
- Model setup
- Estimation
- Inference post fitting

Overview

Acknowledgements

- shoulders of giants
- especially
 - Doug Bates, Martin Mächler, Steve Walker (`lme4`)
 - Kasper Kristensen, Mollie Brooks (`glmmTMB`)
- the R community, `r-sig-mixed-models` denizens, bug reporters ...
- McMaster, NSERC, Univ Florida, US NSF, BIRS

Mixed models

- *tabular* data (row = observation, column = variable)
- *regression* models
 - continuous and categorical *predictor* variables
 - single (univariate) numeric *response* variable
- with one or more *grouping variables* ('blocks', 'clusters', ...)
- for prediction or inference
- typically for 'medium-sized' data (e.g. 10^6 rows, 10^3 clusters, 10-20 predictors)

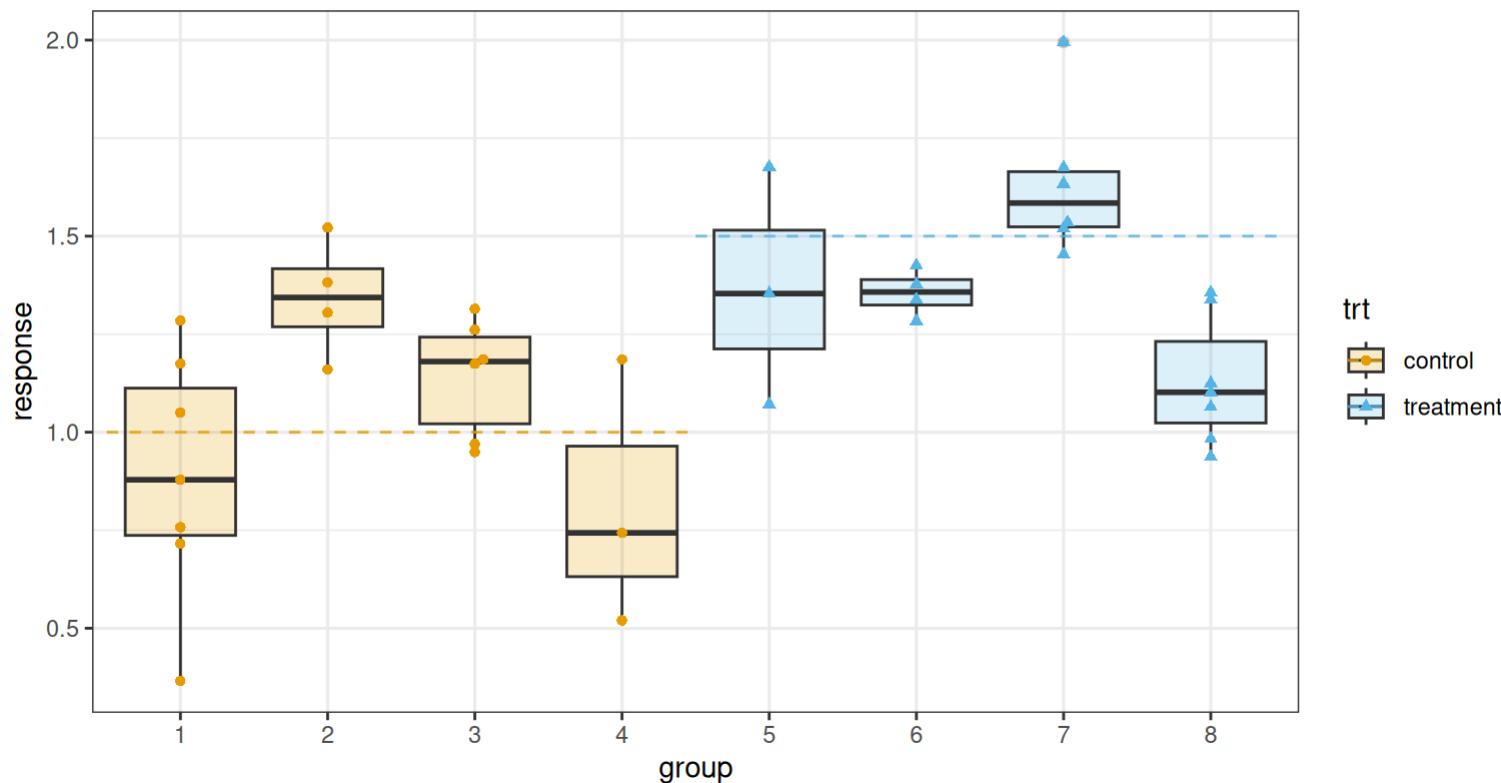
Mixed models (easy version)

- take a standard linear-type model (e.g. $y_i = a + bx_i + \epsilon_i$)
- ... but observations fall in *groups*
- breaks assumption of independence (*pseudoreplication*)
- add *between-group* (random) variation:

$$y_{ij} = a + bx_{ij} + \alpha_i + \epsilon_{ij}$$

single group, categorical predictor (= one-way ANOVA)

$$\beta = (1, 0.5), \sigma_\alpha = 0.4, \sigma_r = 0.2$$



The mixed model universe

- more complex *fixed effects* components (multivariate regression, multiway ANOVA, ANCOVA, ...)
- multiple grouping variables (random effects *terms*); multiple effects per group ("*random slopes*")
- interesting *conditional distributions*
 - exponential dispersion family (à la GLMs): binomial/Poisson/etc.
 - weirder: Tweedie, *t* distribution, Beta, ...
- nonlinearity: *link functions* (log, logit, etc.)
- structured covariances: autoregressive, compound symmetry, regression splines ...

linear regression

ANOVA

analysis of covariance

multiple linear regression

} **general
linear models**

correlation

repeated-measures
models; time series (ARIMA)

(non-normal errors)
(nonlinearity)

nonlinearity

random
effects

**nonlinear
least-squares**

nonlinearity

mixed models

correlation

(nonlinearity)
(non-normal
errors)

nonlinear
time series
models

logistic regression
binomial regression
log-linear models

} **generalized
linear models**

smooth
nonlinearity

scaled
variance

GLMM

generalized
additive
models

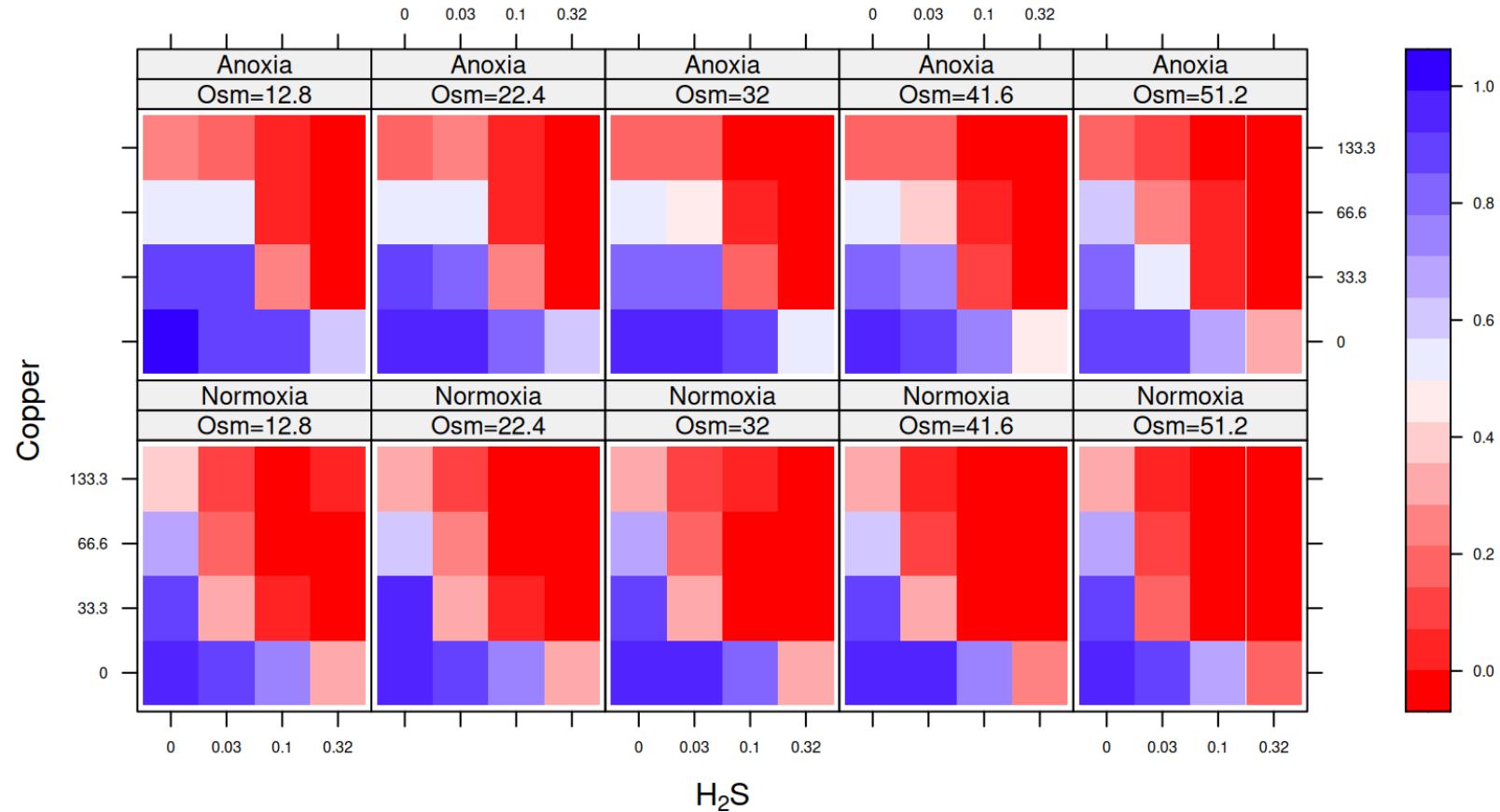
quasilielihood

**negative
binomial models**

thresholds;
mixtures;
compound distributions;
etc. etc. etc.

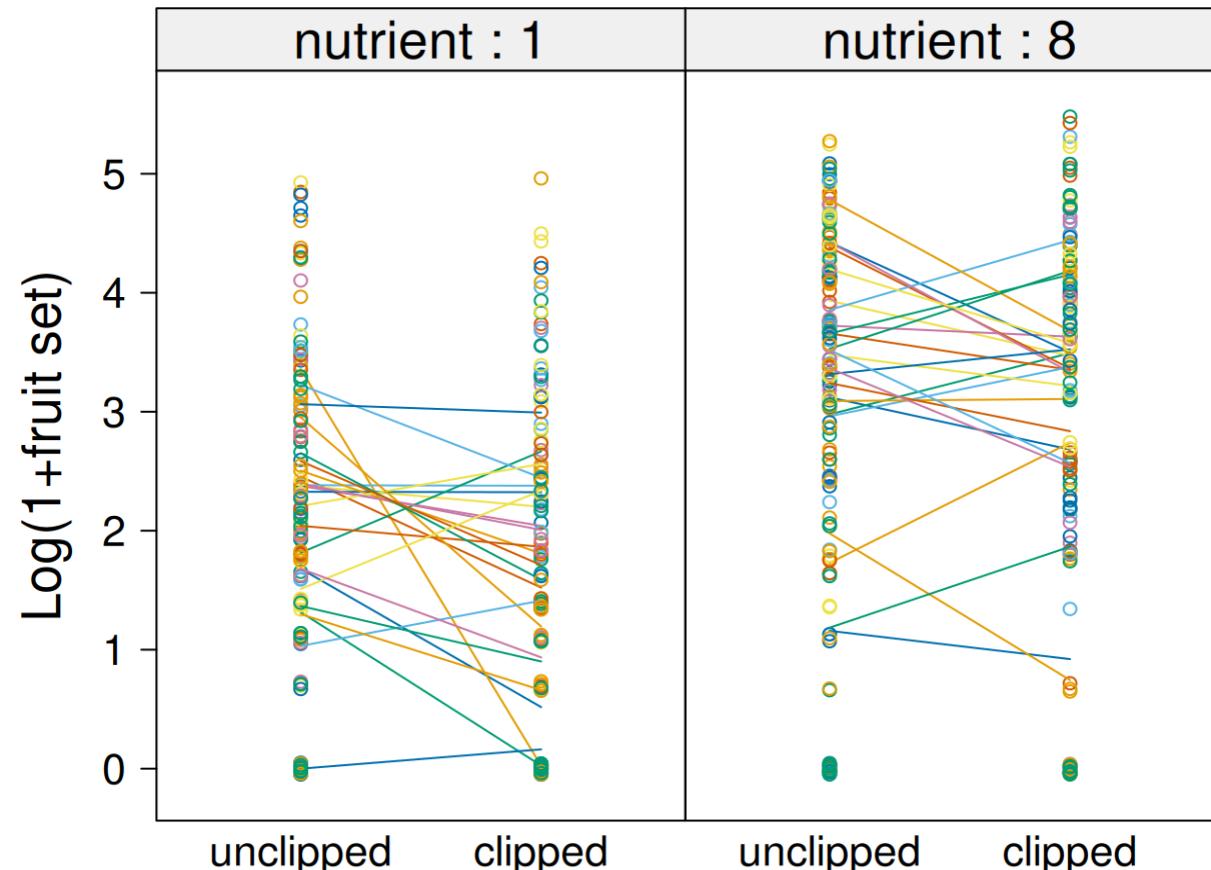
Examples

Experimental: bloodworm cell survival



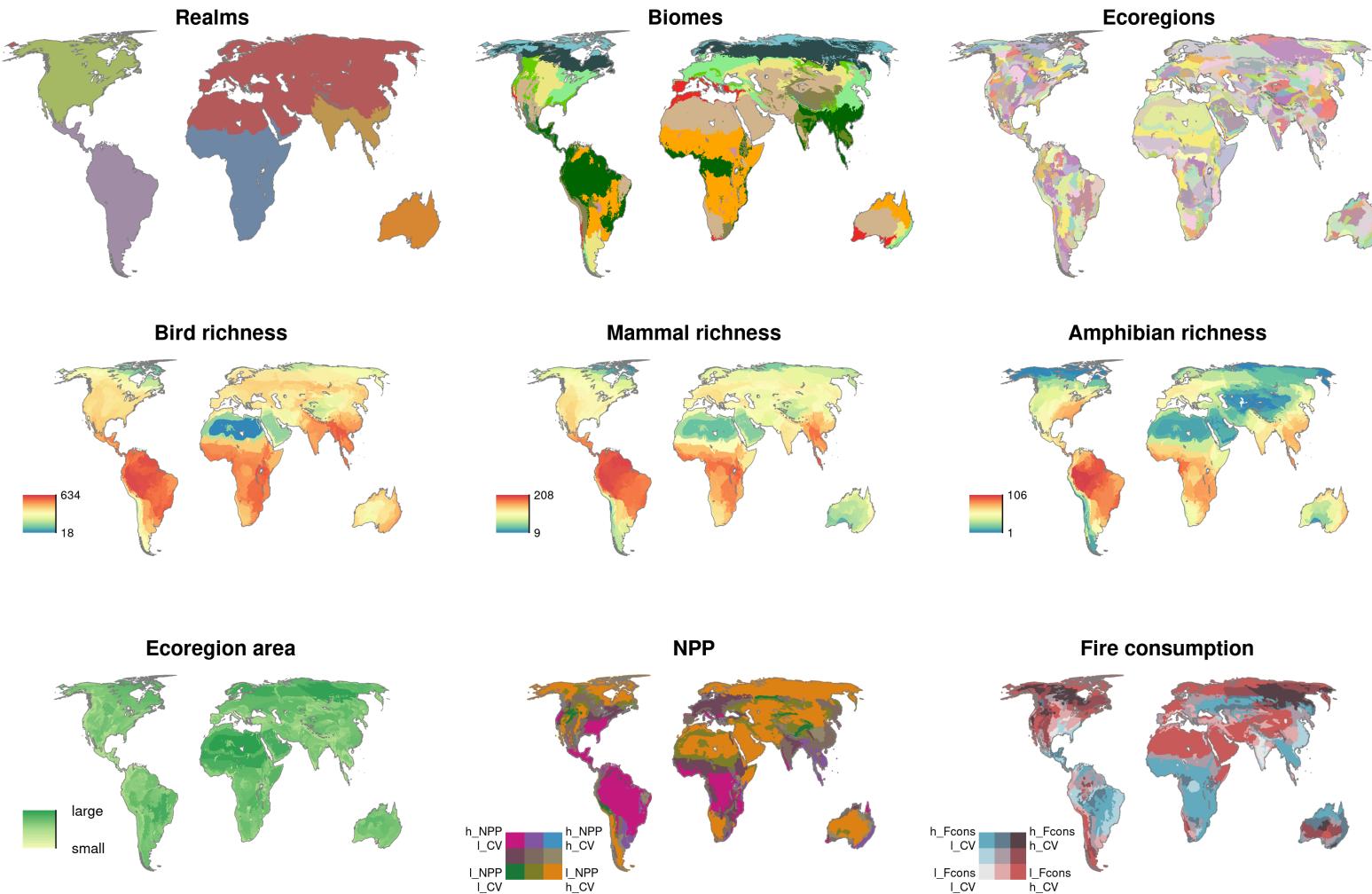
D. Julian, unpublished data

Arabidopsis: fertilization & herbivory



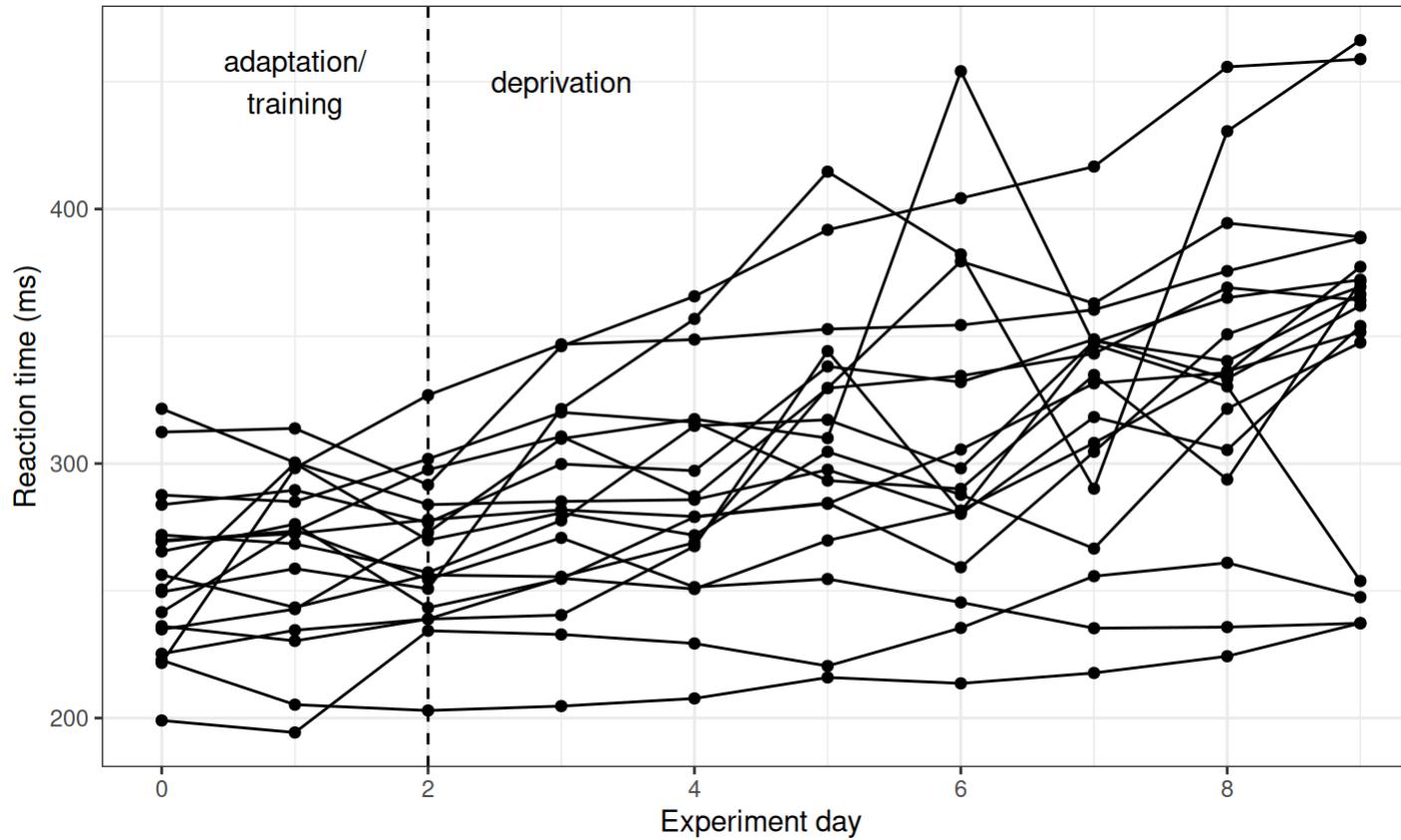
Banta et al. 2010

Wildfire/diversity (Moritz et al. 2023)



Sleep deprivation study

Belenky et al. (2003)



Definitions and descriptions

Simple mixed models (scalar, intercept-only)

$$\begin{aligned}y_{ij} &= \beta_0 + \beta_1 x_{ij} + \epsilon_{0,ij} + \epsilon_{1,j} \\&= (\beta_0 + \epsilon_{1,j}) + \beta_1 x_{ij} + \epsilon_{1,j} \\\epsilon_{0,ij} &\sim \text{Normal}(0, \sigma_0^2) \\\epsilon_{1,j} &\sim \text{Normal}(0, \sigma_1^2)\end{aligned}$$

- Could have multiple, nested levels of random effects (genotype within population within region ...), or *crossed* REs

Random-slopes model

$$\begin{aligned}y_{ij} &= \beta_0 + \beta_1 x_{ij} + \epsilon_{0,ij} + \epsilon_{1,j} + \epsilon_{2,j} x_{ij} \\&= (\beta_0 + \epsilon_{1,j}) + (\beta_1 + \epsilon_{2,j}) x_{ij} \\ \epsilon_{0,ij} &\sim \text{Normal}(0, \sigma_0^2)\end{aligned}$$

$$\{\epsilon_{1,j}, \epsilon_{2,j}\} \sim \text{MVN}(0, \Sigma)$$

- variation in the *effect* of a treatment or covariate across groups
- correlated variation in intercepts/slopes

Terminology

- **grouping variables** define sets of clusters (e.g. school, year, country ...)
- **varying effects**: predictors whose effects vary across clusters (e.g. intercept, herbivory, time)
- **random effects term**: the combination of a clustering variable and a set of varying effects
- **conditional mode** ("BLUP"): the predicted value of the *difference* of a cluster effect from the population mean
- **conditional distribution**: distribution of the data (Gaussian, Poisson, binomial, Beta, Tweedie, ...)

General definition

$$\begin{aligned} Y_i &\sim \text{Distr} \left(g^{-1}(\eta_i), \phi \right) \\ \eta &= X\beta + Zb \\ b &\sim \text{MVN}(\mathbf{0}, \Sigma(\theta)) \end{aligned}$$

conditional distribution
inverse link function
scale parameter

response
linear predictor
conditional modes

fixed effects
random effects

variance-covariance matrix

What are random effects?

A method for ...

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

Random-effect myths

- clusters 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 clusters you have

Reasons for random effects (inferential/philosophical)

- don't want to
 - test hypotheses about differences between responses at particular levels of the grouping variable;
- do want to
 - quantify variation among groups
 - make predictions about unobserved groups
- have levels that are randomly sampled from/representative of a larger population (*exchangeable*)

Reasons for random effects (practical)

- want to combine information across groups
- have variation in information per level (number of samples or noise level)
- have a categorical predictor that is a nuisance variable (i.e., not of direct interest, but needs to be controlled for)
- have more than 5-10 levels¹ (Clarke 2008; Austin et al. 2018)

See also Crawley (2002); Gelman (2005)

¹. Economists say 30-50

Avoiding mixed models

- average: for *nested* (balanced) LMMs (Murtaugh 2007)
- use fixed effects (or *two-stage models*) instead of random effects when there are
 - many observations per cluster (shrinkage unimportant)
 - few clusters (little advantage in # parameters; bad variance estimates)

Model setup

Formulas

- random effects specified with | ("pipe"/"bar")
as $(1+a|g1) + (1+b|g2) + \dots$
- right-hand side ($g1, g2$) is the *grouping variable* (always treated as categorical, usually a factor)
- left-hand side describes the *varying terms* (1 [intercept], a, b)
- $+$ separates independent terms
- formulas: empower but conceal details (McElreath 2015)

Formula definitions

equation

$$\beta_0 + \beta_1 X_i + e_{si}$$

$$(\beta_0 + b_{S,0s}) + \beta_1 X_i + e_{si}$$

$$(\beta_0 + b_{S,0s}) + (\beta_1 + b_{S,1s})X_i + e_{si}$$

$$(\beta_0 + b_{S,0s} + b_{I,0i}) + (\beta_1 + b_{S,1s})X_i + e_{si}$$

As above, but S_{0s}, S_{1s} independent

$$(\beta_0 + b_{S,0s} + b_{I,0i}) + \beta_1 X_i + e_{si}$$

$$(\beta_0 + b_{I,0i}) + (\beta_1 + b_{S,1s})X_i + e_{si}$$

formula

n/a (Not a mixed-effects model)

$\sim X + (1|Subject)$

$\sim X + (1 + X|Subject)$

$\sim X + (1 + X|Subject) + (1|Item)$

$\sim X + (1|Subject) + (1|Item)$

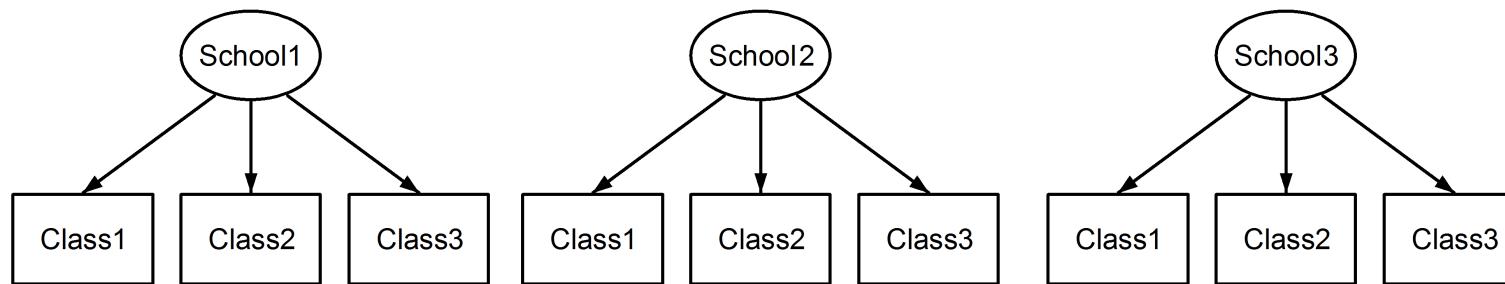
$\sim X + (1|Subject) + (1|Item)$

$\sim X + (0 + X|Subject) + (1|Item)$

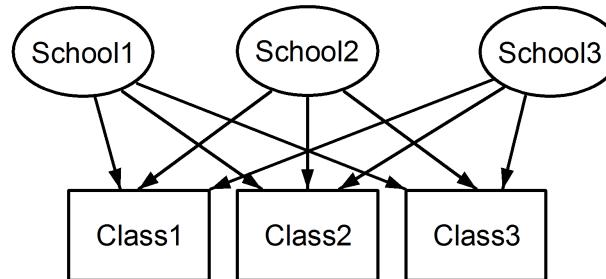
Modified from: <http://stats.stackexchange.com/questions/13166/rs-lmer-cheat-sheet?lq=1> (Livius). Subscripts $\{S, I\}$ refer to Subject vs Item effects. Lower-case $\{s, i\}$ indicate particular subjects/items. $\{0, 1\}$ refer to intercept vs slope effects.

Nesting vs crossing

Nested: sub-unit IDs only measured within a single larger unit. e.g.: Class1 in School1 independent of Class1 in School2



Crossed: sub-unit IDs can be measured in multiple larger units. e.g.



Unique coding: removes ambiguity



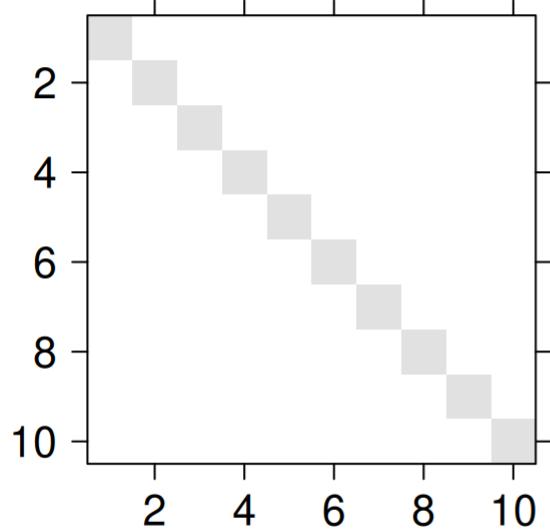
Formulas, interactions, nesting etc.

- Nested: $(1|f/g) \equiv (1|f) + (1|f:g)$. Subplots vary within plots, no commonality across plots
- Crossed: $(1|f) + (1|g)$. Years vary; plots vary independently
- Crossed+: $(1|f*g) \equiv (1|f) + (1|g) + (1|f:g)$. Years vary, plots vary independently, plots vary within years (need >1 observation per plot/year combination).

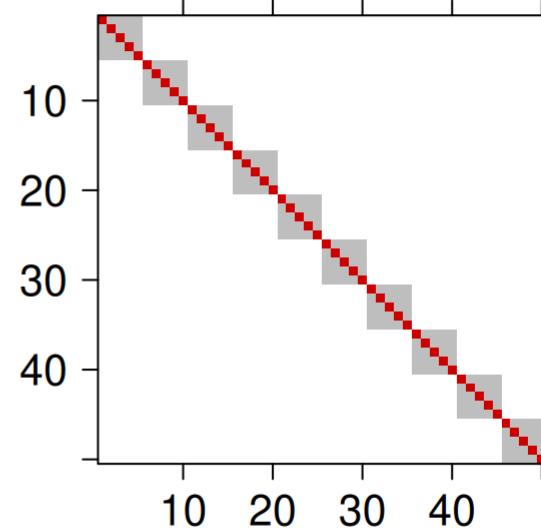
Don't need explicit nesting if sub-groups are uniquely labeled (i.e. A1, A2, ..., B1, B2, ...)

example covariance matrices

- 10 groups, 5 obs/group, intercept RE



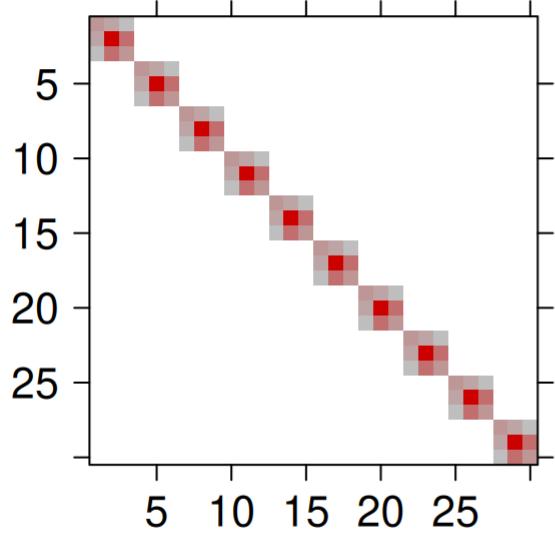
random effects Σ



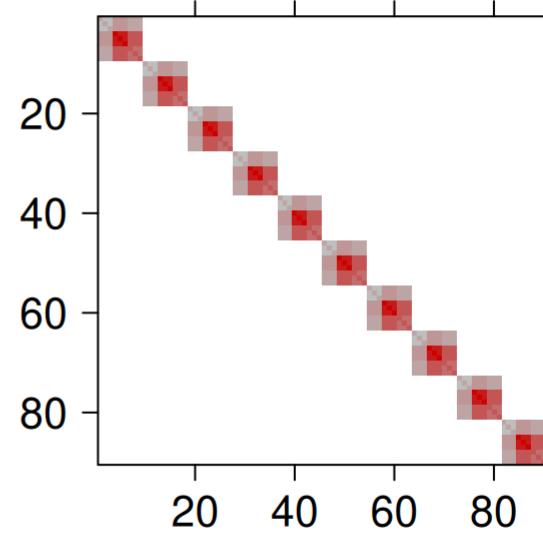
observation Σ

example covariance matrices

- 10 groups, 5 obs/group, 3 effects/group



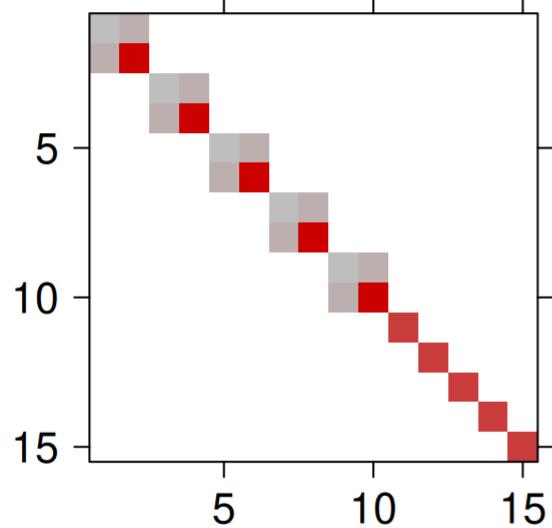
random effects Σ



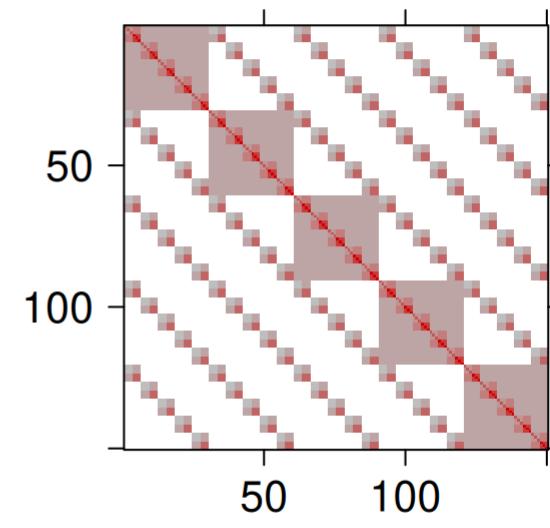
observation Σ

example covariance matrices

- 5 groups, 2 effects (v_1) \times 5 groups, 1 effect (v_1)



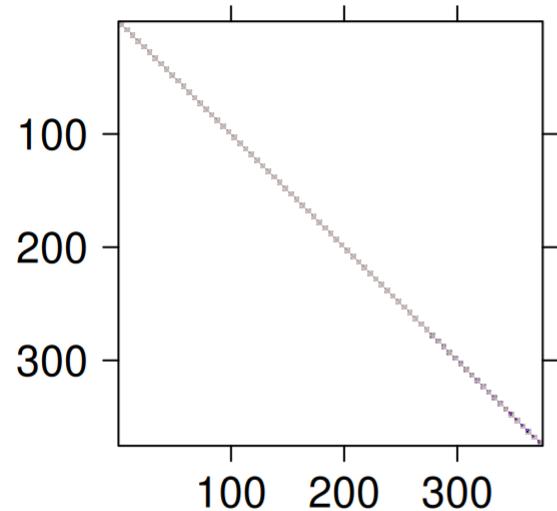
random effects Σ



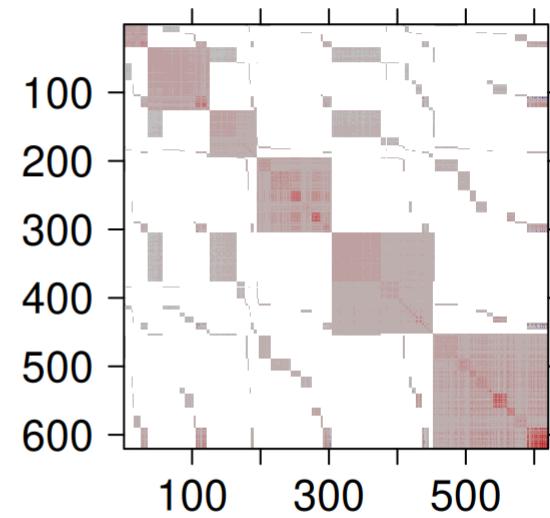
observation Σ

example covariance matrices

Moritz et al. (2023) (fire/diversity; blocks are geographic regions)



random effects Σ



observation Σ

Factors varying across groups

Can generate large variance-covariance matrices. For a four-level factor, we have β_0 (intercept), β_1 (level 2 - level 1), β_2 (level 3 - level 1), β_3 (level 4 - level 1).

$$\Sigma = \begin{bmatrix} \sigma_{\{b|1\}}^2 & \cdot & \cdot & \cdot \\ \sigma_{\{b|1\},\{b|a_{21}\}} & \sigma_{\{b|a_{21}\}}^2 & \cdot & \cdot \\ \sigma_{\{b|1\},\{b|a_{31}\}} & \sigma_{\{b|a_{21}\},\{b|a_{31}\}} & \sigma_{\{b|a_{31}\}}^2 & \cdot \\ \sigma_{\{b|1\},\{b|a_{41}\}} & \sigma_{\{b|a_{21}\},\{b|a_{41}\}} & \sigma_{\{b|a_{31}\},\{b|a_{41}\}} & \sigma_{\{b|a_{41}\}}^2 \end{bmatrix}$$

What is the maximal model?

- Which effects vary *within* which groups?
- If effects don't vary within groups, then we *can't* estimate among-group variation in the effect
 - e.g. in typical clinical trials each patient gets only one treatment (placebo or drug)
 - convenient
 - maybe less powerful (among-group variation is lumped into residual variation)
 - can't evaluate variation in effects

Maximal model is often impractical

- for a RE where n effects vary across clusters we need to estimate $p = n(n + 1)/2$ covariance parameters
- e.g. bloodworm example: 5×4×4×2 factorial design (160 combinations), all measured in each of 10 individuals.
 - Treating covariates as numeric, allow all interactions:
16 effects → $p = 136$
 - Categorical predictors, no interactions:
12 effects → $p = 78$
 - Numeric predictors, no interactions: 4 effects → $p = 10$
- BUT ignoring within-cluster effects is also dangerous (Schielzeth et al. 2009; Heisig et al. 2019; Oberauer 2022)

Estimation

Maximum likelihood estimation

- Best fit is a compromise between
 - consistency of data with fixed effects + conditional modes
 - consistency of random effect with RE distribution
- Goodness-of-fit *integrates* over the RE distribution

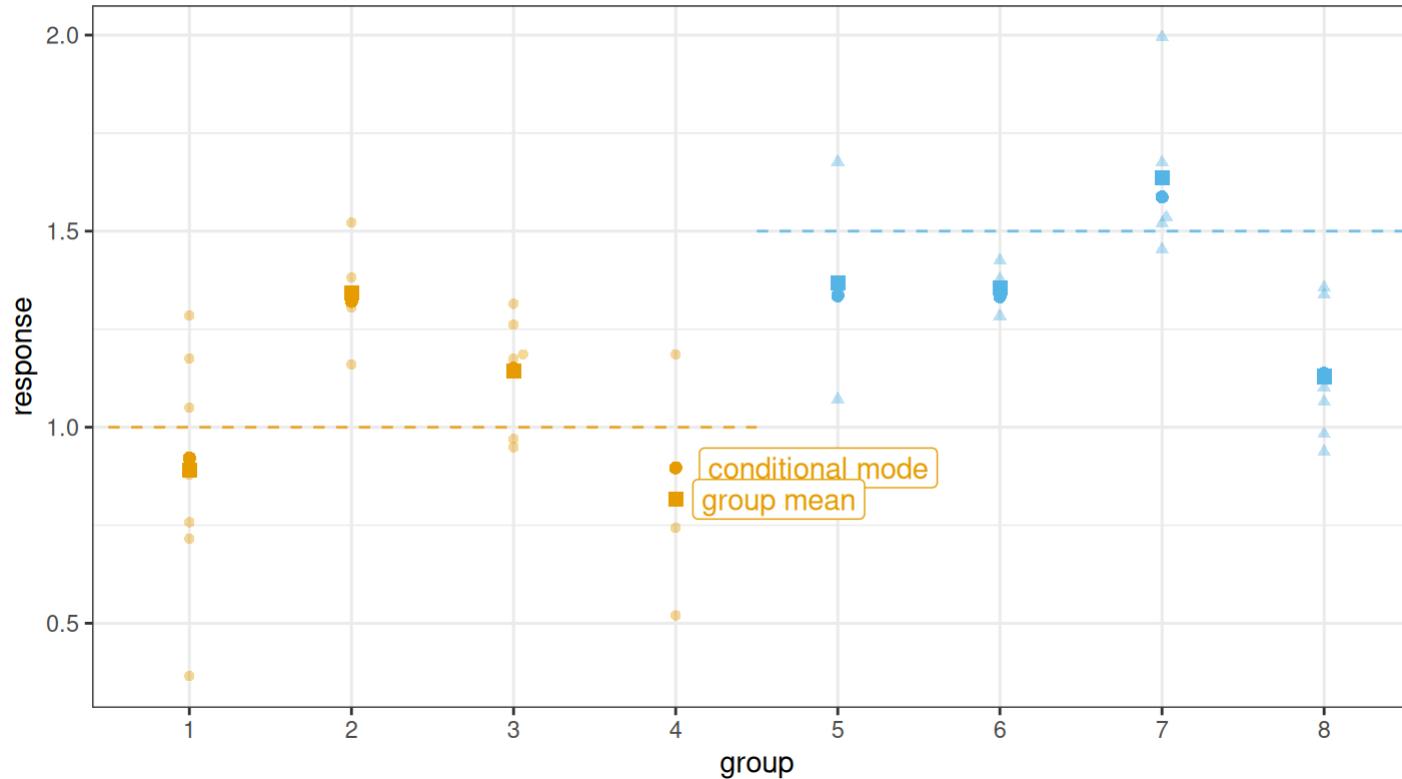
$$\mathcal{L}(x|\beta, \theta) = \iint \mathcal{L}(x|\beta, \mathbf{b}) \cdot \mathcal{L}(\mathbf{b}|\Sigma(\theta)) d\mathbf{b}$$

Restricted maximum likelihood (REML)

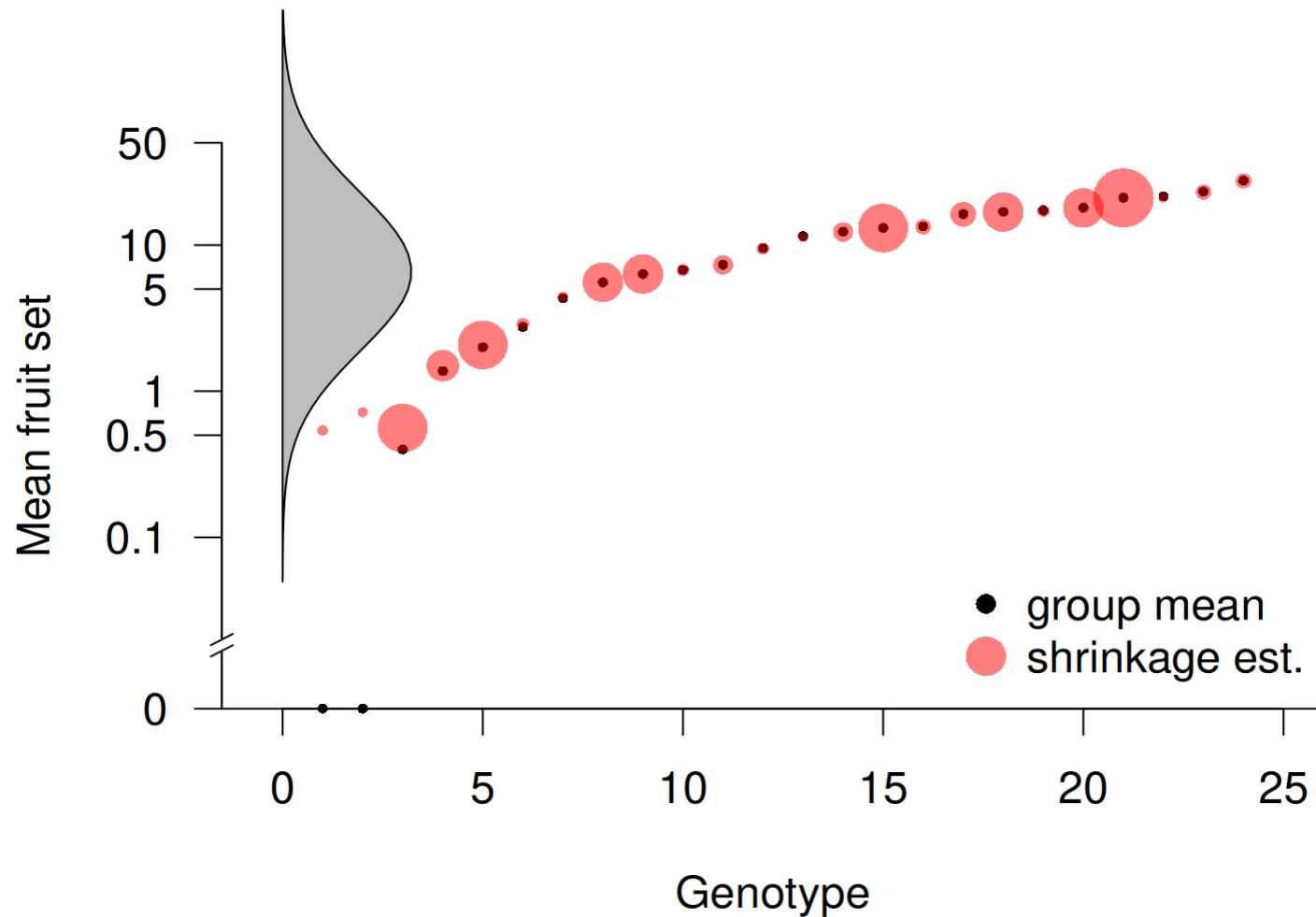
- factor out fixed effects when estimating variances
- analogous to estimating $s^2 = \frac{\text{SSQ}}{n-1}$ rather than $\frac{\text{SSQ}}{n}$
- typically better variance estimates (unbiased in simple cases)
- defaults differ across packages
- never compare models with differing fixed effects fitted by REML 

Shrinkage

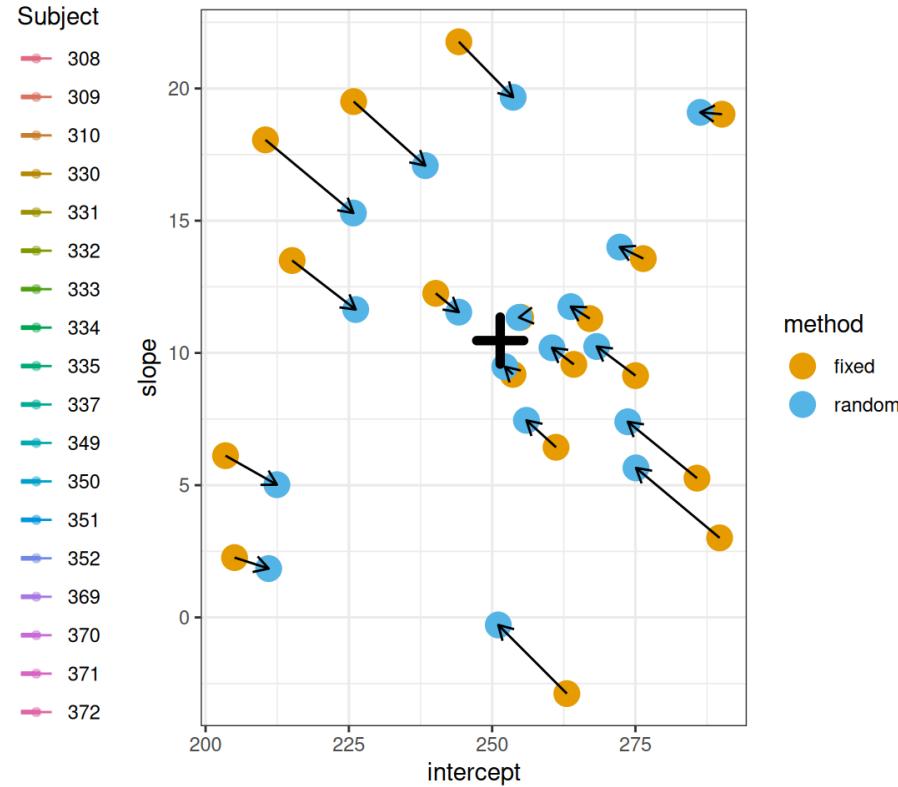
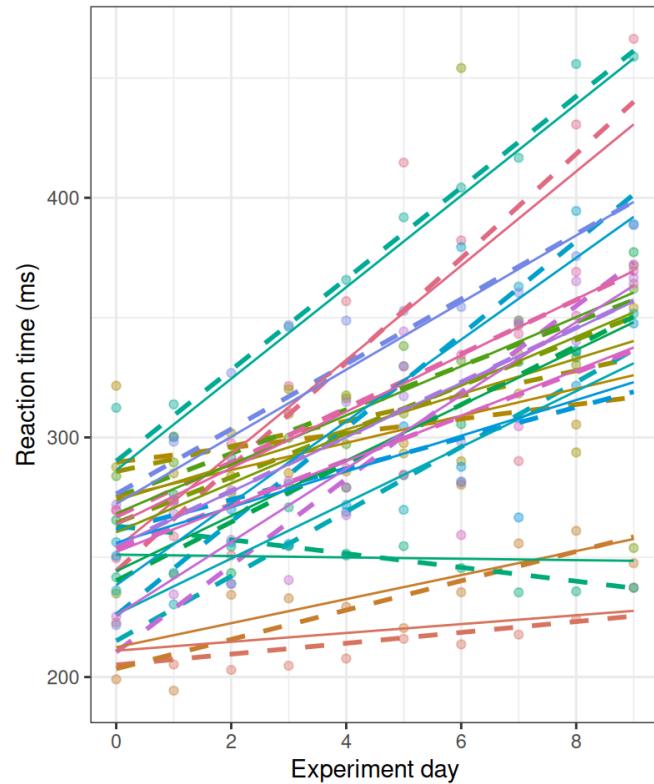
$$\beta = (1, 0.5), \sigma_\alpha = 0.4, \sigma_r = 0.2$$



Arabidopsis shrinkage



Shrinkage in the sleep-study model



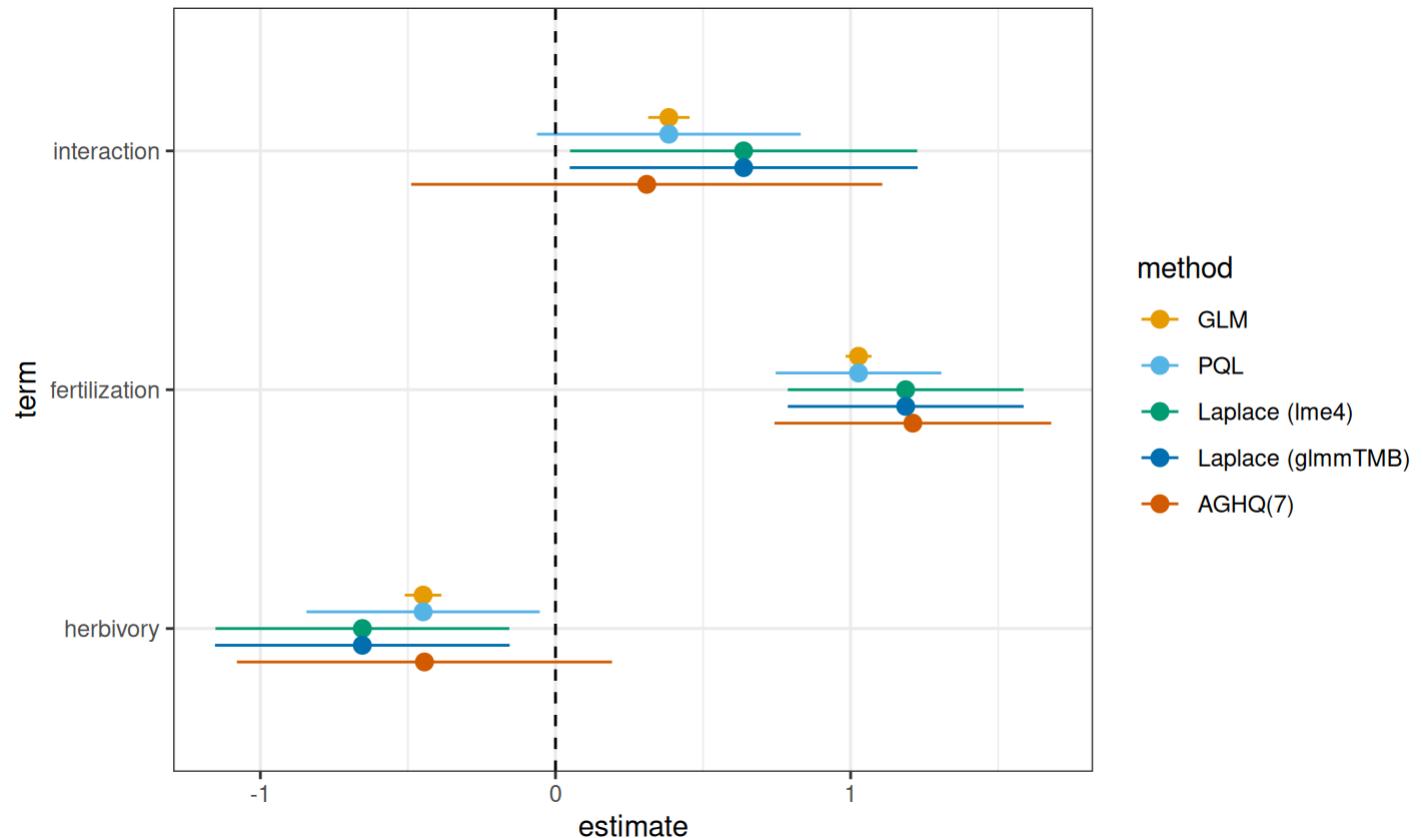
Methods

Estimation methods

Integrating the likelihood is hard (for GLMMs)

- **deterministic**
 - penalized quasi-likelihood (Breslow 2004), Laplace, Gauss-Hermite quadrature, ... (Biswas 2015; Bilodeau et al. 2024)
 - need best methods for small clusters (binary, small # of observations)
 - available methods depend on problem structure
 - accuracy (GHQ) vs. flexibility and speed (PQL, Laplace)
- **stochastic** (Monte Carlo): mostly Bayesian, some frequentist
 - Booth et al. (1999); Gelman et al. (2006); Sung et al. (2007); Ponciano et al. (2009); McElreath (2015)
 - much slower, but flexible and accurate

Estimation: *Arabidopsis*



Inference/post-fitting

Diagnostics

- make sure model is working OK (**singular fits** etc.)
- check model assumptions
 - generalizations of linear/GLM diagnostics: residuals vs fitted, scale-location, etc.
 - conditional modes should be Normally distributed
 - simulation-based residuals (**DHARMa**)

Parameter/model interpretation

- fixed effects: same as linear models (population level trends/effects)
- conditional modes/group-level effects
 - deviations from population-level effects
 - can get *conditional standard deviations*, but can't test significance!
- interpreting random effects
 - standard deviations (on same scale as corresponding fixed effects)
 - variances (for population geneticists)

references

- Austin, PC et al. 2018. *Journal of Statistical Computation and Simulation* 88 (16) (November): 3151–3163. doi:[10.1080/00949655.2018.1504945](https://doi.org/10.1080/00949655.2018.1504945).
- Banta, JA et al. 2010. *Oikos* 119 (2) (February): 359–369. doi:[10.1111/j.1600-0706.2009.17726.x](https://doi.org/10.1111/j.1600-0706.2009.17726.x).
- Belenky, G et al. 2003. *Journal of Sleep Research* 12 (1): 1–12. doi:[10.1046/j.1365-2869.2003.00337.x](https://doi.org/10.1046/j.1365-2869.2003.00337.x).
- Bilodeau, B et al. 2024. *Journal of the American Statistical Association* 119 (545) (January): 690–700. doi:[10.1080/01621459.2022.2141635](https://doi.org/10.1080/01621459.2022.2141635).
- Biswas, K. 2015. Master's thesis, McMaster University.
https://macsphere.mcmaster.ca/bitstream/11375/17272/2/M.Sc_Thesis_final_Keya_Biswas.pdf.
- Booth, JG et al. 1999. *Journal of the Royal Statistical Society. Series B* 61 (1): 265–285. doi:[10.1111/1467-9868.00176](https://doi.org/10.1111/1467-9868.00176).
- Breslow, NE. 2004. In *Proceedings of the second Seattle symposium in biostatistics: Analysis of correlated data*, ed by. Danyu Y. Lin et al., 1–22. Springer.
- Clarke, P. 2008. *Journal of Epidemiology & Community Health* 62 (8) (August): 752–758. doi:[10.1136/jech.2007.060798](https://doi.org/10.1136/jech.2007.060798).
- Crawley, MJ. 2002. *Statistical computing: An introduction to data analysis using S-PLUS*. John Wiley & Sons.
- Gelman, A. 2005. *Annals of Statistics* 33 (1): 1–53. doi:[doi:10.1214/009053604000001048](https://doi.org/10.1214/009053604000001048).
- Gelman, A et al. 2006. *Data analysis using regression and Multilevel/Hierarchical models*. Cambridge, England: Cambridge University Press. <http://www.stat.columbia.edu/~gelman/arm/>.
- Heisig, JP et al. 2019. *European Sociological Review* 35 (2) (April): 258–279. doi:[10.1093/esr/jcy053](https://doi.org/10.1093/esr/jcy053).
- McElreath, R. 2015. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton: Chapman; Hall/CRC.
- Moritz, MA et al. 2023. *Ecology Letters* 26 (4): 563–574. doi:[10.1111/ele.14177](https://doi.org/10.1111/ele.14177).
- Murtaugh, PA. 2007. *Ecology* 88 (1): 56–62. doi:[10.1890/0012-9658\(2007\)88\[56:SACIED\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2007)88[56:SACIED]2.0.CO;2).