lme4: past, present, future

Ben Bolker

McMaster University, Mathematics & Statistics and Biology

12 August 2013

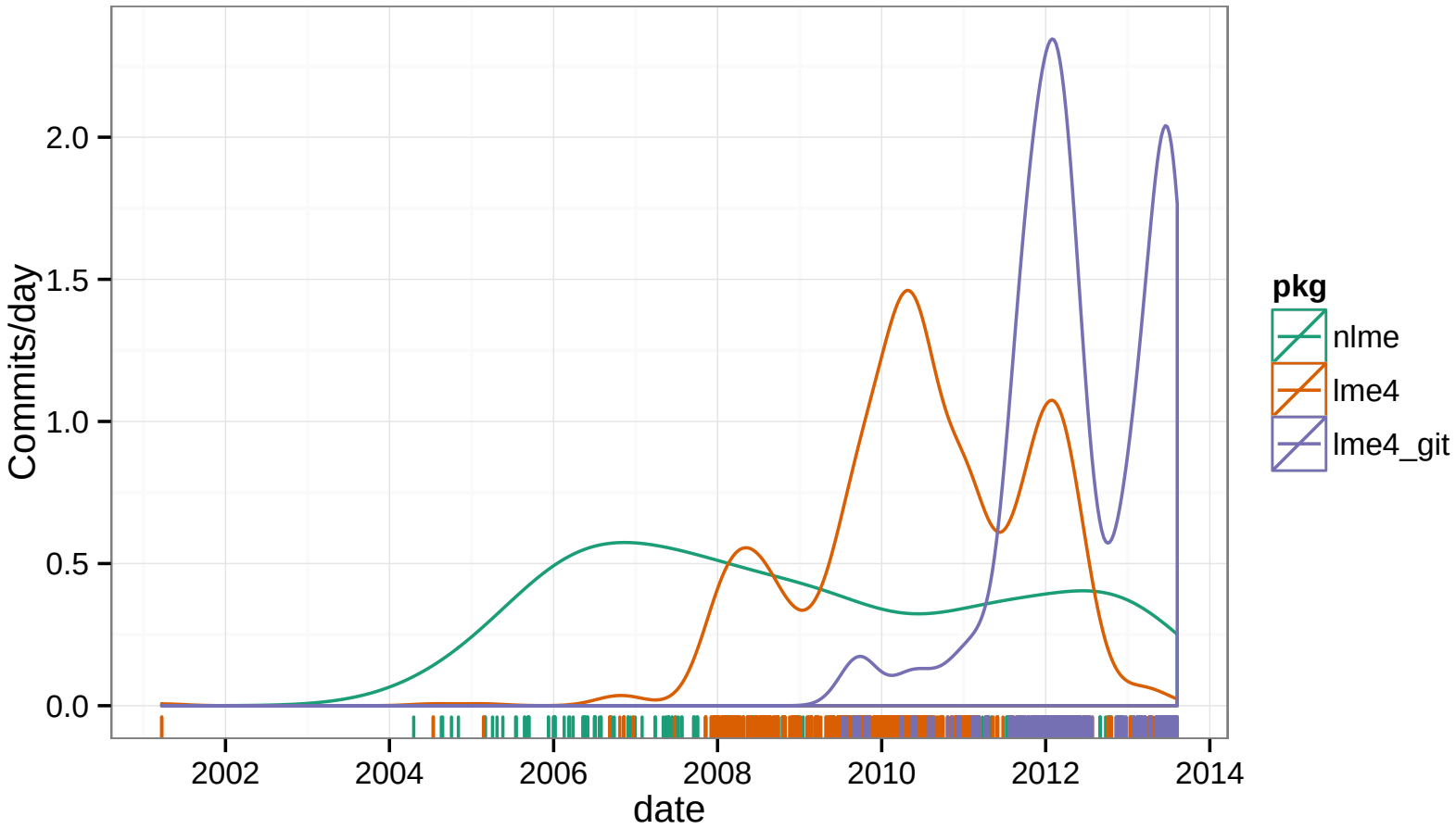
Outline

Outline

History

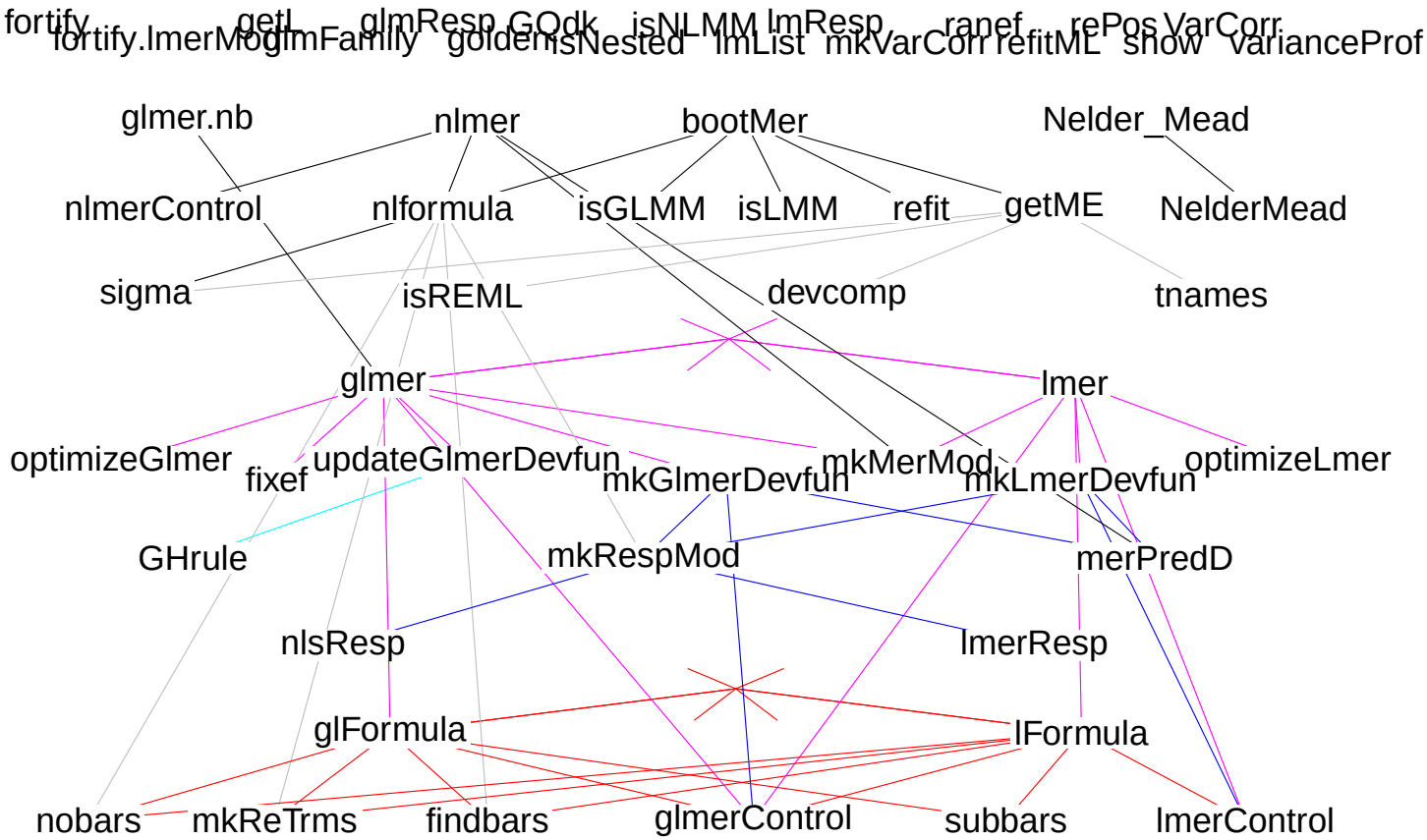
- `nlme`:
  - v. 3.1.1, Nov-1999
  - Pinheiro and Bates (2000)
  - v. 3.1.50: move to R SVN, Apr-2004
  - becomes Recommended 2007?
  - current 3.1.109, Apr-2013
- `lme4`:
  - initial `ChangeLog` entry, Jul-2002
  - v. 0.2.1, Jun-2003
  - move to R-forge SVN, Nov-2007
  - `lme4a` ca. Aug-2009
  - `lme4Eigen` $\rightarrow$ `lme4`: Mar-2012
  - move to github, Nov-2012
  - v. 1.0-0 Aug-2013 ???

Activity


Commits/day
2.0
1.5
1.0
0.5
0.0
2002
2004
2006
2008
2010
2012
2014
date
pkg
nlme
lme4
lme4_git

Outline

Code structure (mvbutils::foodweb)



Tiers

I low-level computational/matrix representation; PWRSS solutions (`Eigen` package; Julia)

II low-level computation: PWRSS iteration/step-halving, Laplace/AGQ

III nonlinear optimization over θ, (β)

IV API; interface, methods, modularization

V downstream packages

Features

I. fill-reducing permutations (speed, for large problems; crossed random effects)

II. GLMMs (step-halving, Laplace, AGQ)

III. fit on constrained scale; default is derivative-free with box constraints (Nelder-Mead, BOBYQA), but flexible

IV. profiling; parametric bootstrap; `predict` and `simulate`; deviance function and modularity

Issues

What have we (`lme4` authors) been doing?

- fixing glitches and achieving consistency . . .
- . . . especially for GLMMs (e.g. updating weights)
- numerical stability of fits (e.g. boundary issues)
- evaluation and environments (e.g. call-modification)

Metrics

- 7600 lines of R code, 3270 of C++ code
- $\approx$ 80 downstream packages (Imports/Suggests/Depends); 433 recursive links (`tools::dependsOnPkgs`)
- for 9 June – 4 August 2013, 4020 downloads from http://probability.ca/cran (rank 44/4866)

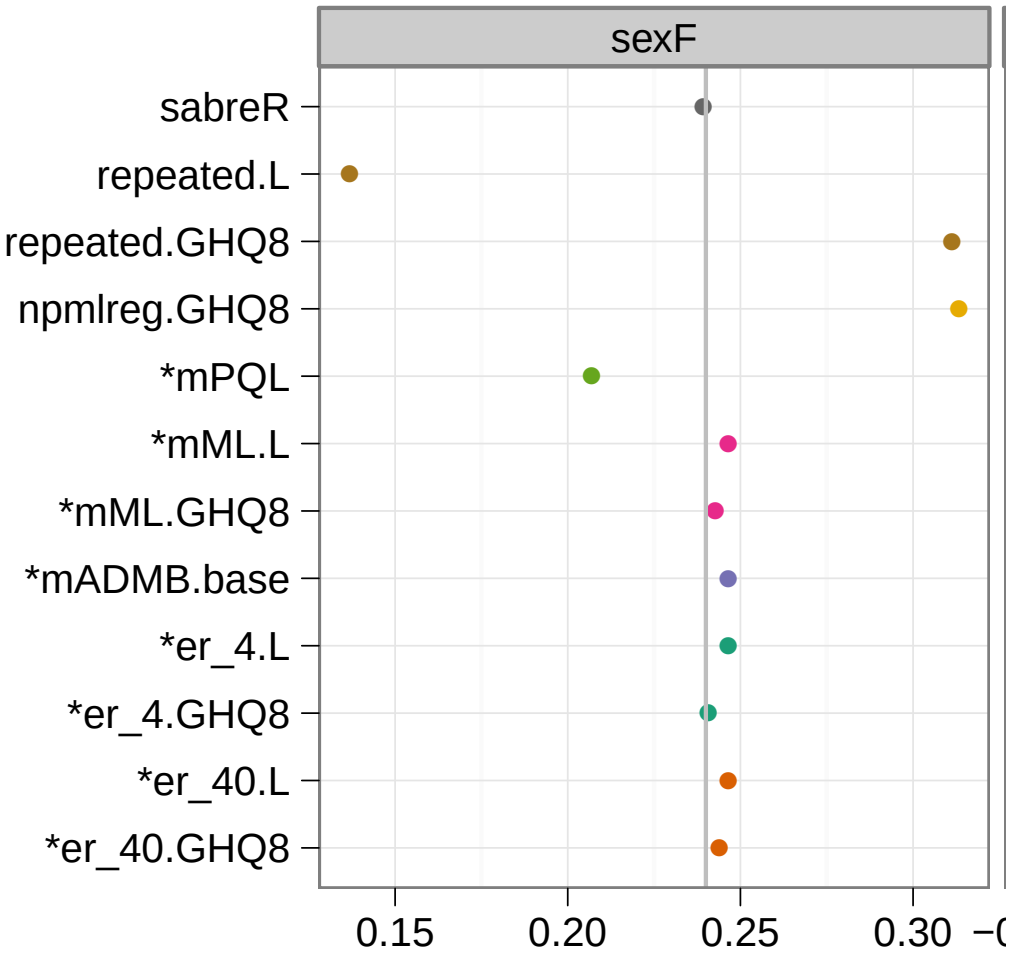
Competitors (deterministic)

package	features
`glmmADMB*`	negative binomial I/II, zero-inflated/hurdle, ...
`glmmML`	non-Normal RE, bootstrap
`npmlreg`	non-parametric RE (by EM)
`hglm`	hierarchical GLMs
`lmm`	(LMMs only)
`repeated*`	
`sabreR`	
`MASS::glmmPQL`	R-side effects (???)

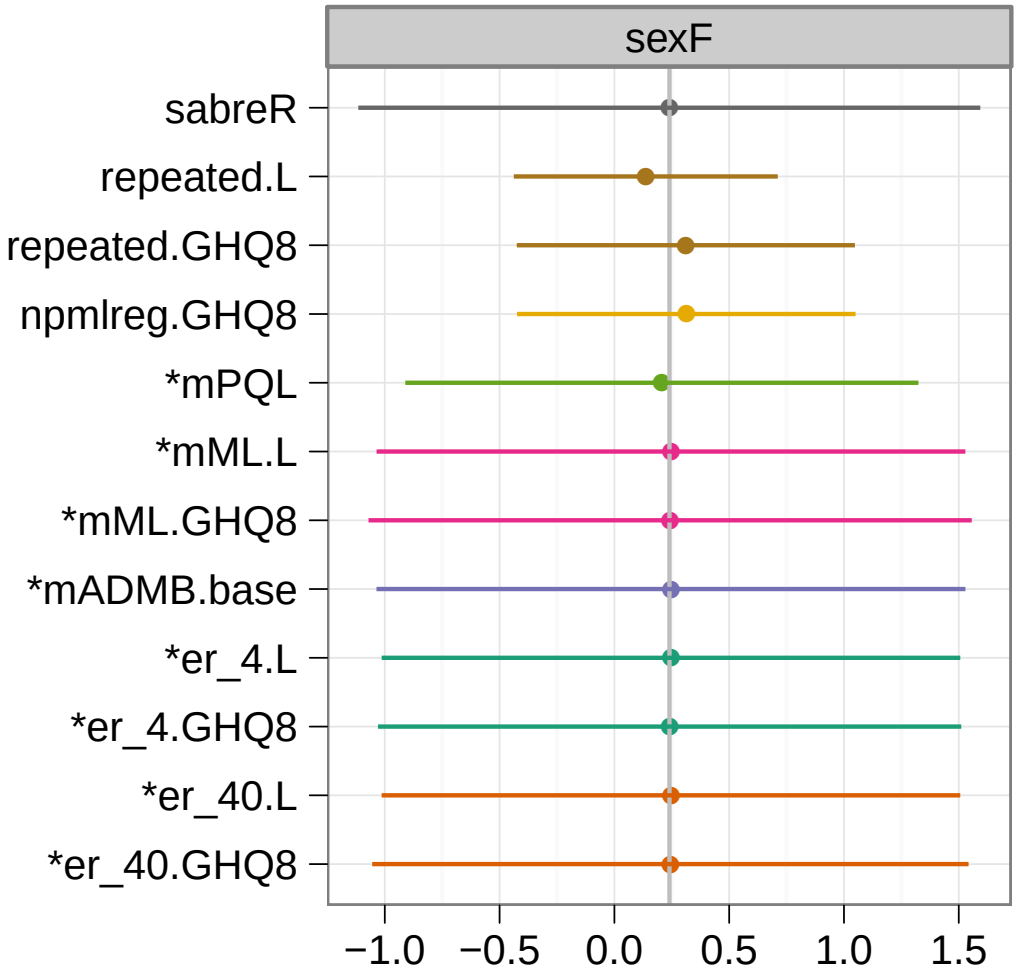
(*=off-CRAN)

Non-R: AS-REML/Genstat (flexible variance), SAS, MLWiN, HLM, GLAAMM/Stata, ...

Comparison with Zhang et al. (2011) (1)



Comparison Zhang et al. (2011) (2)



Competitors (stochastic)

package	features
`bernor*`	Monte Carlo expectation-maximization
`MCMCglmm`	multivariate/multitype, zero-inflated/hurdle, flexible variance, MCMC
`INLA*`	

Non-R: BUGS/JAGS/Stan (and `R2*`)

Environment

- CRAN administrators
  - C++/Rcpp/RcppEigen issues
- Infrastructure maintainers (reference classes, Rcpp)
- downstream package maintainers
  - variation in coding/packaging style, statistical philosophy
  - what should core `lme4` provide?
- users (r-sig-mixed-models, Stack Overflow/Cross Validated)
  - variation in sophistication/needs
  - latent demand?

Outline

Near future

- cleanup and tweaks, e.g.
  - finish neg binom models (IV)
  - better input/output tests (III, IV)
  - low-hanging parallelization
- fix GLMM edge cases (PWRSS failures) (II)
- evaluate/improve performance, esp. on large data (I)
- write JSS paper(s)

Wishlist

- R-side effects (IV)
- flexible RE variance (e.g. spatial/phylogenetic GLMMs) (IV)
- speed optimization (I)
- working MCMC sampler (IV)
- AGQ, postVar for complex models (II, IV)
- improve bootMer (`merBoot` features) (IV)
- improve GLMMs that use estimated scale parameters (III, IV)
- lightweight stored models (IV)
- `simulate`/`predict` with `newdata`, `newpar(a)ms`
- systematic (G)LMM comparison across packages/settings

Pinheiro, J.C. and Bates, D.M., 2000. *Mixed-effects models in S and S-PLUS*. Springer, New York. ISBN 0-387-98957-9.

Zhang, H., Lu, N., et al., 2011. *Statistics in Medicine*. ISSN 1097-0258. doi:10.1002/sim.4265.