

glmmTMB

Mollie Brooks

Dept. of Evolutionary Biology and Environmental Studies
Univ. of Zurich

20 June 2016

ADMB/TMB Developer's Workshop Seattle

A brief history

- glmmADMB is popular because it's flexible.
 - fits LMs, GLMs, ZI-GLMs, LMMs, GLMMs, ZI-GLMMs
 - zero-inflated Poisson and negative binomial
 - hurdle models
 - 253 citations since 2012
- glmmADMB can't handle predictors on zero-inflation.
- I got grants from the Swiss NSF and the ADMB Foundation to write glmmTMB to handle predictors of zero-inflation.
- We wrote glmmTMB in August 2015.

Contributors (so far)

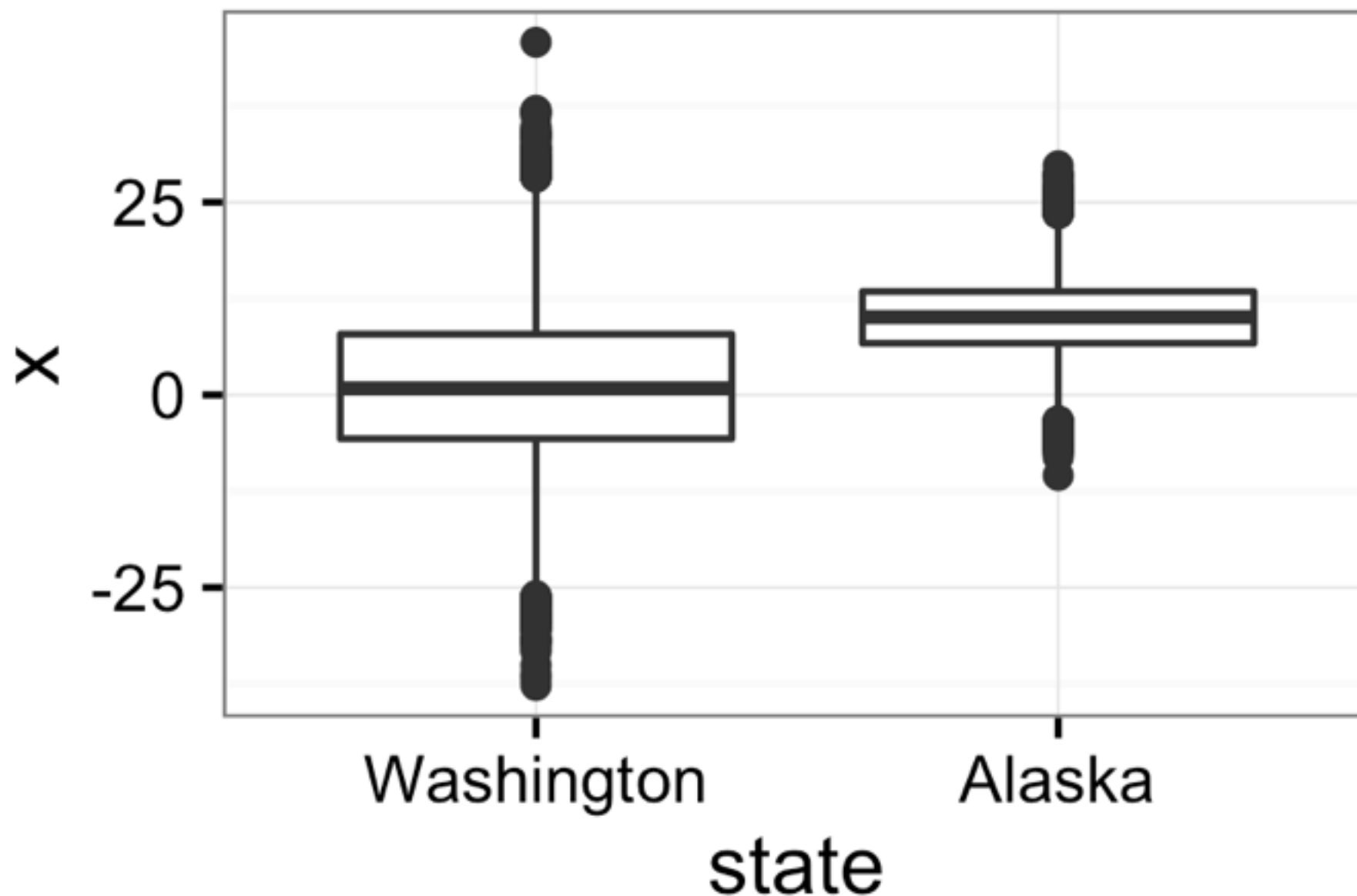
- **Casper Berg**, Technical Univ. of Denmark, Institute of Aquatic Resources
- **Ben Bolker**, McMaster Univ., Depts. of Mathematics & Statistics and of Biology
- **Mollie Brooks**, Univ. of Zurich, Dept. of Evolutionary Biology & Environmental Studies
- **Kasper Kristensen**, Technical Univ. of Denmark, Dept. of Applied Mathematics and Computer Science
- **Martin Mächler**, ETH Zurich, Seminar for Statistics
- **Arni Magnusson**, Marine Research Institute of Iceland
- **Anders Nielsen**, Technical Univ. of Denmark, Institute of Aquatic Resources
- **Hans Skaug**, Univ. of Bergen, Dept. of Mathematics
- **Koen Van Benthem**, Univ. of Zurich, Dept. of Evolutionary Biology & Environmental Studies

Features of glmmTMB

- Flexible zero-inflation formulas
- Flexible dispersion formulas
- Link functions: log, logit, probit, inversre, cloglog
- Weights (need testing)
- Correlation structures: diag, us, cs, AR1 (need testing)
- Truncated count distributions (need testing)
- Convenient functions: summary, predict, residuals, update

Dispersion example

```
m0=glmmTMB(x~state+(1|t)+(1|loc),  
dispformula=~state, dat)
```



Dispersion example (cont.)

```
> m0=glmmTMB(x~state+(1|t)+(1|loc), dispformula=~state, dat)
> summary(m0)
```

```
Family: gaussian ( identity )
Formula: x ~ state + (1 | t) + (1 | loc)
Data: dat
```

AIC	BIC	logLik	deviance	df.resid
269481.3	269532.9	-134734.7	269469.3	39994

Random effects:

Conditional model:

Groups	Name	Variance	Std.Dev.
t	(Intercept)	1.407e-04	0.01186
loc	(Intercept)	1.413e-02	0.11887
Residual		9.858e+01	9.92884

Number of obs: 40000, groups: t, 1000; loc, 40

Conditional model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.04616	0.07508	13.93	<2e-16 ***
stateAlaska	9.01887	0.08706	103.60	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

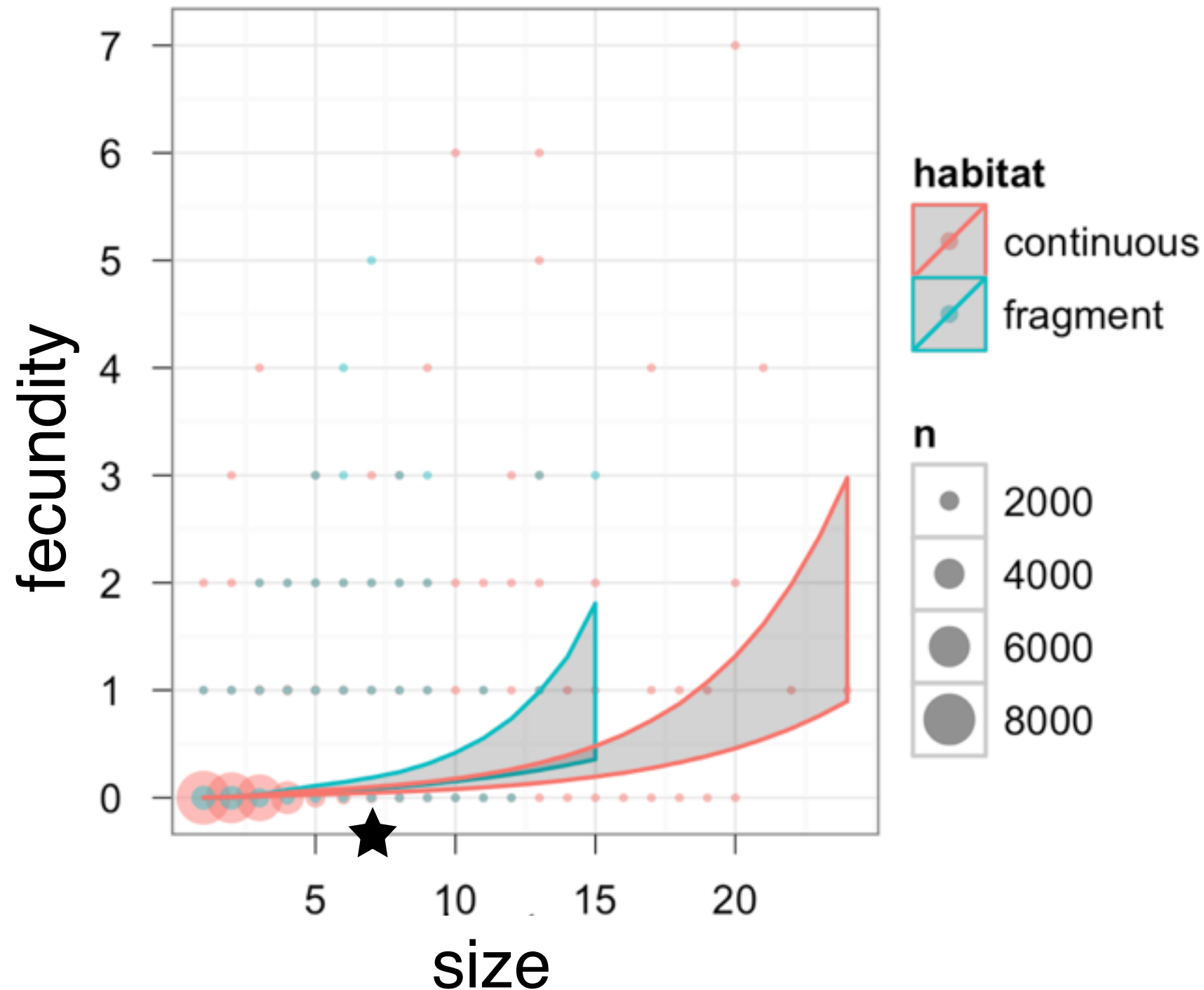
Dispersion model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.59089	0.01000	459.0	<2e-16 ***
stateAlaska	-1.38466	0.01415	-97.9	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Zero-inflated example

```
m2=glmmTMB(fecundity ~size*habitat+(1|ID)+(1|year),  
  zi=~size*habitat, hdat, family='poisson')
```



Zero-inflated example

```
> summary(m2)
```

```
Family: poisson ( log )  
Formula: fecundity ~ size * habitat + (1 | ID) + (1 | year)  
Data: hdat
```

AIC	BIC	logLik	deviance	df.resid
10946.4	11031.0	-5463.2	10926.4	34754

Random effects:

Conditional model:

Groups	Name	Variance	Std.Dev.
ID	(Intercept)	0.3430	0.5857
year	(Intercept)	0.2798	0.5289

Number of obs: 34764, groups: ID, 1742; year, 11

Conditional model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.00261	0.30054	-9.991	< 2e-16 ***
size	0.20170	0.03628	5.559	2.71e-08 ***
habitatCF	-0.34758	0.27169	-1.279	0.201
size:habitatCF	-0.01175	0.03857	-0.305	0.761

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Zero-inflation model:

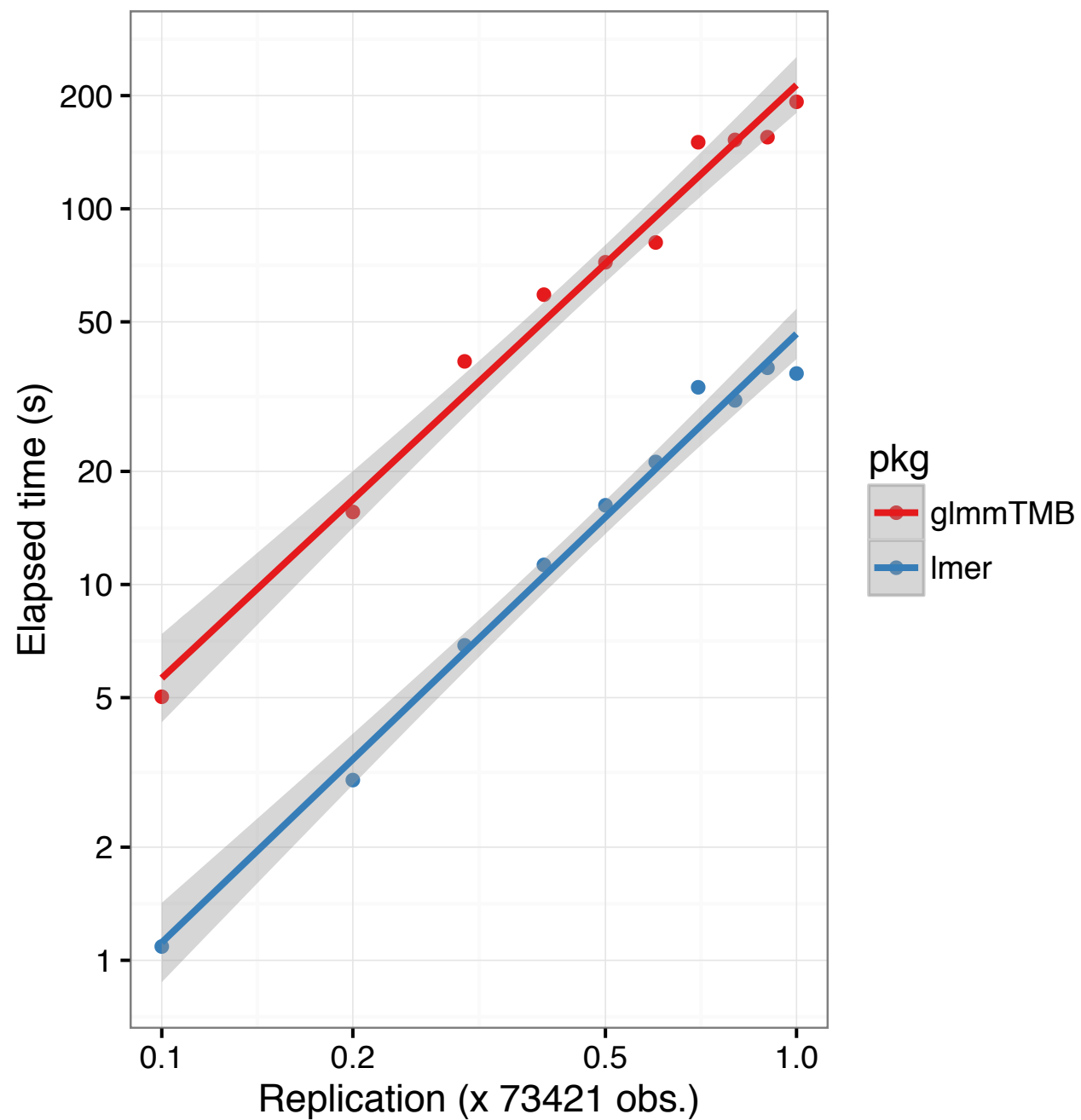
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.84124	0.50727	11.515	<2e-16 ***
size	-1.61232	0.17764	-9.076	<2e-16 ***
habitatCF	-1.09337	0.58727	-1.862	0.0626 .
size:habitatCF	0.05862	0.21620	0.271	0.7863

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Timings

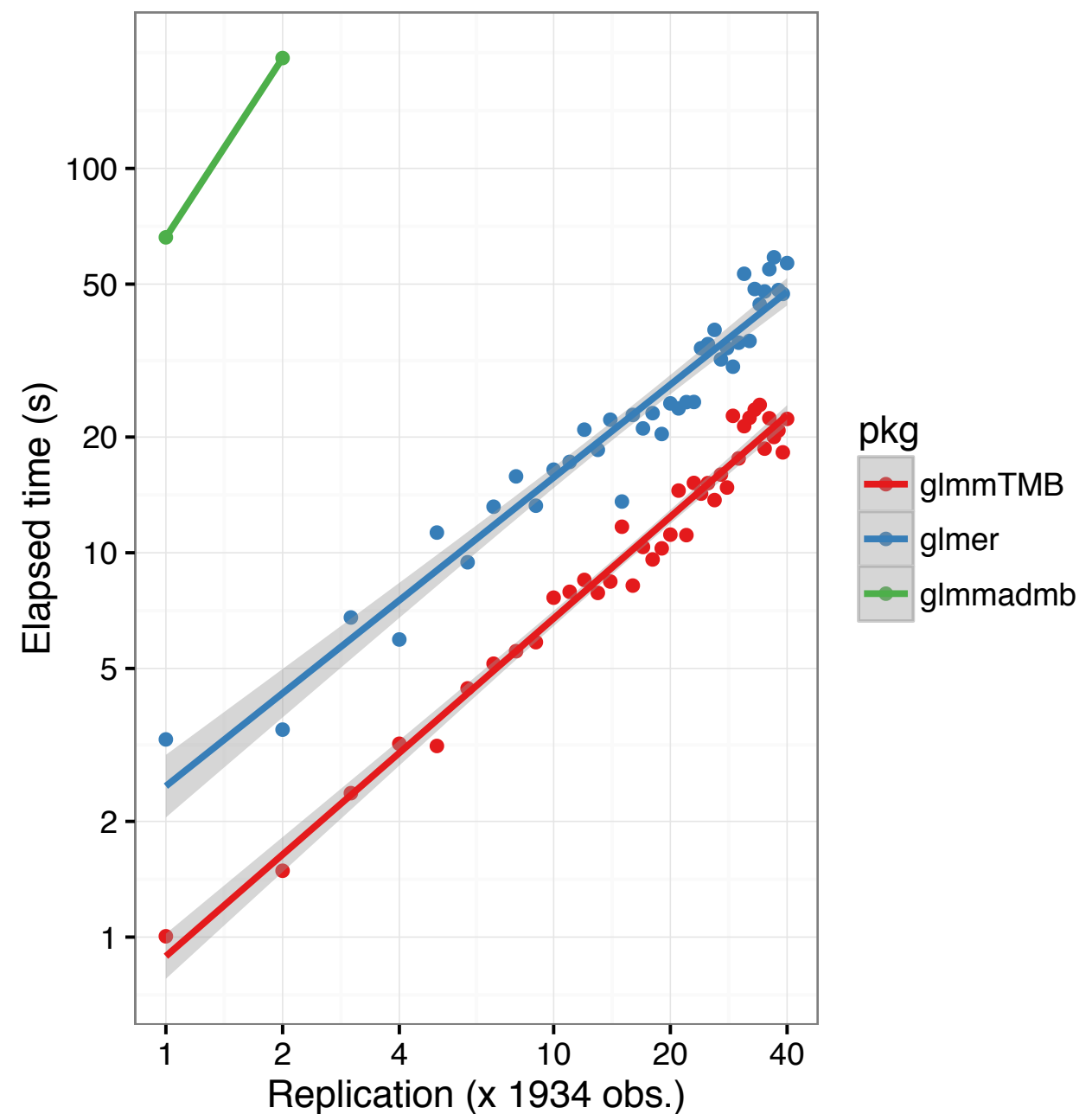
(from glmmTMB vignette written by Bolker)

`lmer` is 5 times faster than `glmmTMB`



InstEval dataset

`glmmTMB` is 2.3 times faster than `glmer`



Contraception dataset

Future additions

- hurdle models
- generalized Poisson distribution
- beta distribution
- t distribution (with option to map shape parameter)
- increase stability?

generalized Poisson

- Better than Poisson and negative binomial for modeling individual annual reproduction (Kendall and Wittmann. 2010. A Stochastic Model for Annual Reproductive Success. The American Naturalist).
- Flexible enough to handle underdispersed and overdispersed count data.
- When underdispersion is not handled correctly, statistical tests are prone to false-negatives (Harris, Yang, and Hardin. 2012. Modeling underdispersed count data with generalized Poisson regression. Stata Journal).
- Currently unavailable in GLMMs in R.

generalized Poisson

Distribution definition

$$(1) \quad f(y_i; \theta_i; \delta) = \frac{\theta_i(\theta_i + \delta y_i)^{y_i-1} e^{-\theta_i - \delta y_i}}{y_i!}, y_i = 0, 1, 2, \dots$$

where $\theta_i > 0$ and $\max(-1, -\theta_i/4) < \delta < 1$. When $\delta > 0$ the distribution is overdispersed and when $\delta < 0$ the distribution is underdispersed. It has mean and variance

$$(2) \quad \mu_i = \frac{\theta_i}{1 - \delta}$$

$$(3) \quad \sigma_i^2 = \frac{\mu_i}{(1 - \delta)^2}.$$

Parameterization for glmmTMB

$$(4) \quad \theta_i = \mu_i \sqrt{\phi}$$

$$(5) \quad \delta = 1 - \frac{1}{\sqrt{\phi}}$$

This should be a numerically stable parameterization since $\phi > 0$ is already bounded by glmmTMB and $\mu_i > 0$ will be bounded by a log link.

Thanks!



FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION



**University of
Zurich^{UZH}**

<https://github.com/glmmTMB/glmmTMB>