Reformulating phylogenetic mixed models to improve flexibility and speed

Michael Li[1*] and Benjamin M. Bolker[1,2,3]

* Corresponding author: Michael Li; lim88@mcmaster.ca. Current address: Public Health Risk Science Division, Public Health Agency of Canada

1 Department of Biology, McMaster University, Hamilton, Ontario, Canada
2 Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada
3 Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada

Abstract

1. Phylogenetic regression is a powerful technique for exploring relationships among characteristics of related species. However, existing procedures may be either insufficiently flexible or too computationally demanding when analyzing large volumes of data.

2. We propose an alternative formulation of phylogenetic generalized linear mixed models that is mathematically equivalent to previous approaches, but is more flexible in practice. We have implemented this formulation in two R statistical packages (`lme4` and `glmmTMB`).

3. Our reformulation of phylogenetic generalized linear mixed models is computationally efficient, operating orders of magnitude faster than existing comparably flexible methods.

4. Our approach can be implemented in any platform for generalized mixed models. Our implementation in `lme4` and `glmmTMB` allows users to fit phylogenetic mixed models to a broad range of previously difficult cases (e.g., large data, unbalanced observational designs, complex random effects).

Keywords: phylogenetic comparative methods, phylogenetic correlation, phyloglmm, species–branch matrix

Introduction

Phylogenetic regressions (PR), a subset of phylogenetic comparative methods, account for evolutionary relatedness when analyzing relationships among morphological, physiological, or ecological characteristics of species. Given a known phylogeny, PRs model the relationships among species traits while incorporating relatedness; they can be used to control statistically for phylogeny, to quantify phylogenetic signal in traits, or both.

In contrast to standard statistical models, where all observations are assumed to be independent, PRs incorporate phylogenetic relationships to account for correlations between observations driven by the unobserved process of trait evolution (Butler and King, 2004; Felsenstein, 1985; Hansen and Bartoszek, 2012). While a wide range of tools is available for PR, most existing procedures are either inflexible or too computationally demanding to analyze large data sets (e.g. random-slopes phylogenetic regressions with hundreds of species). When faced with these constraints, researchers often simplify their analyses, at the risk of neglecting important processes.

Challenges in modeling phylogenetic processes

In classic PRs, phylogenetic correlation in the residuals from a regression between two species-level traits arises because the residual variation in the response trait evolves along the branches of the phylogeny according to a Brownian-motion (BM) evolutionary model (Felsenstein, 1985). If the residuals are normally distributed and observed without additional error or within-species variation, Felsenstein's method of phylogenetically independent contrasts (PICS: Felsenstein, 1985) is sufficient to account for the phylogenetic correlation. More recent approaches — including phylogenetic generalized linear mixed models (PGLMM: Housworth et al., 2004; Ives and Helmus, 2011), Pagel's λ (Pagel, 1999), and Blomberg's K (Blomberg et al., 2003) — extend PICs

by considering different (non-Gaussian) response distributions and by accounting for evolutionary models other than BM. These methods partition residual variation into two components: (1) independent residual variation (tip variation, which may be confounded with observation error) and (2) phylogenetic signal (evolutionary process error: Hansen and Bartoszek, 2012; Housworth et al., 2004). If each species' traits are measured multiple times, we can distinguish a third level of variation; in this case, phylogenetic variation and tip variation both contribute to the evolutionary variation while the observation error can be independently identified from within-species variation (de Villemereuil et al., 2012; Kostikova et al., 2016).

Classic PRs allow the residuals to evolve along the phylogeny, but the effects of the predictor variables may evolve along the phylogeny as well; extending PRs to allow for multiple sources of variation leads to the *phylogenetic mixed model* (Housworth et al., 2004). For example, suppose we wish to fit a PR to predict species' brain size from their body size Felsenstein (1985) using a mixed-effect model. The standard PR allows for phylogenetic correlations in the residuals of the relationship between body and brain size; the phylogenetic mixed model can allow to incorporate both variation in brain size among taxa with similar body sizes (random intercepts) and variation in the *relationship* between predictors and responses among taxa (random slopes).

Several recent studies have developed new tools to fit phylogenetic mixed models in community ecology applications (Li et al., 2017; Nowakowski et al., 2018). However, the phylogenetic mixed modeling tools that allow extensions such as random slopes and separation of tip and observation error in a frequentist framework may not allow other extensions like non-normal response distributions or additional random effects; thus, biologists needing to fit more complex models typically turn to more flexible Bayesian approaches, despite their additional computational burden (Bürkner, 2018; Hadfield, 2010; Kostikova et al., 2016) (Table **??**).

We propose an alternative, more flexible formulation of the phylogenetic mixed

Model	Method	Data	Platform
Generalized Linear Model (GLM)	Correlated residual	Single observation per species	`nlme`:gls, `ape`:pic
	Residual + phylogenetic intercept	Single observation per species	Pagel's λ Blomberg's K via `nlme`:gls `phylolm`
Generalized Linear Mixed Model (GLMM)	Random effect	Single observation per species, Balanced design	`pez`
		Unrestricted	`phyloglmm/lme4`, `phyloglmm/glmmTMB`, `phyr`
Bayesian GLMM	Random effect	Balanced design	`MCMCglmm`
		Unrestricted	`brms`

model that is mathematically equivalent to previous approaches. In particular, it allows for complex phylogenetic effects (random intercepts, slopes, and interactions), without the need to implement special correlation structures, by incorporating phylogenetic structures as part of the mean model (Hefley et al., 2017). We compare our technique (built on the R packages `lme4` and `glmmTMB`) with existing R packages, fitting models to simulated data that incorporates random slopes, random intercepts (tip variation), and residual variation.

Materials and Methods

Phylogenetic regression

Suppose a species trait $\mathbf{y}$ is a linear function of some predictors encoded in a model matrix $\mathbf{X}$, where each species is measured exactly once. The standard phylogenetic regression can be expressed as

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim \text{MVN}(0, \sigma^2\mathbf{C}), \end{aligned} \tag{1}$$

where $\mathbf{y}$ is an length-n response vector; $\mathbf{X}$ is an $n \times m$ matrix, describing n observations of m predictor variables; $\boldsymbol{\beta}$ is an m-vector of coefficients; $\boldsymbol{\epsilon}$ is a multivariate normally distributed n-vector with mean 0 and covariance matrix $\sigma^2\mathbf{C}$ where $\mathbf{C}$ is a $n \times n$ phylogenetic correlation (PC) matrix that quantifies the proportion of shared evolution between any pair of taxa in the phylogeny (Garamszegi, 2014).

Phylogenetic generalized linear mixed model

The phylogenetic generalized linear mixed model (PGLMM) framework defines a wider range of models that includes the standard phylogenetic regression as a special case (Lynch, 1991). The PGLMM allows for non-Gaussian responses and incorporates multiple components of variability. The PGLMM has the form:

$$\begin{aligned} \mathbf{y} &\sim \mathcal{D}(\boldsymbol{\mu}, \phi) \\ \boldsymbol{\mu} &= g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) \\ \mathbf{b} &\sim \mathrm{MVN}(0, \boldsymbol{\Sigma}(\theta)) \end{aligned} \tag{2}$$

where in addition to the terms from (1) $\mathbf{Z}$ is an $n \times q$ model matrix for the q-dimensional vector-valued random effects; $\mathbf{b}$ represents the conditional mean (or mode) of the random effect, which is multivariate normally distributed with covariance matrix $\boldsymbol{\Sigma}(\theta)$; and ϕ is a scale parameter for the conditional distribution $\mathcal{D}$. The PGLMM reduces to the simple PR model (1) when $\mathcal{D}$ is Gaussian, g is the identity function, $\mathbf{Z}$ is the identity matrix, $\boldsymbol{\Sigma}(\theta) = \sigma^2\mathbf{C}$, and $\mathcal{D}$ is Gaussian. In addition, we need $\phi = \sigma_r^2 = 0$ so that the residual variance disappears and the only variance comes from the phylogenetic covariance matrix; otherwise, the additional residual variance term corresponds to one implementation of Pagel's λ.

Reformulating the phylogenetic covariance matrix

In general, correlations within statistical models can be integrated either in the covariance matrix $\mathbf{\Sigma}$ or in the structure of the model matrix $\mathbf{Z}$ (Hefley et al., 2017); thus, we can use $\mathbf{Z}$ to incorporate phylogenetic correlations. Suppose evolution follows a BM process, i.e., continuous traits evolve independently at a constant rate, following an unbiased random walk along each branch of the phylogeny. Then the phylogenetic variability of a particular species can be written as the sum of the variances of evolutionary changes that occurred on all of the branches in its history. Thus, modeling the evolutionary history of each species with a sequence of independent errors with species–branch matrix $\mathbf{S}$ is equivalent to imposing a correlation $\mathbf{C}$. For example, for the phylogeny in figure 1, the corresponding $\mathbf{S}$ takes the form:

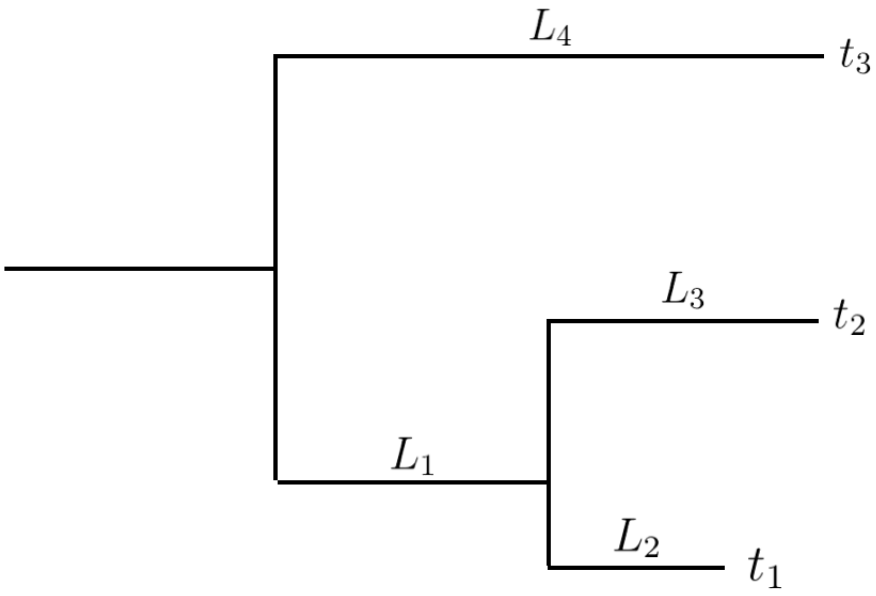

Fig. 1: Three-species phylogenetic tree.

$$
\begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \end{array}
\begin{array}{c}
\begin{array}{cccc} L_1 & L_2 & L_3 & L_4 \end{array} \\
\begin{pmatrix} \ell_1 & \ell_2 & 0 & 0 \\ \ell_1 & 0 & \ell_3 & 0 \\ 0 & 0 & 0 & \ell_4 \end{pmatrix}
\end{array}
\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{pmatrix}
$$

For example, the phylogenetic effect for species 1 is $\ell_1\epsilon_1+\ell_2\epsilon_2$, where $\ell_i = \sqrt{L_i}$, the square root of the branch length L_i in Figure 1, and the ϵ_i are independent Normal deviates with zero mean and variance σ^2 (i.e. the phylogenetic variance for species 1 is $\mathrm{E}[(\ell_1\epsilon_1 + \ell_2\epsilon_2)^2] = (L_1 + L_2)\sigma^2$).

Constructing the species–branch random effects model matrix

The $\mathbf{S}$ matrix is the product of an $m \times b$ indicator matrix $\mathbf{S}_{ind}$ of branch indices and a vector $\boldsymbol{\ell}$ of square roots of branch lengths:

$$
\mathbf{S}_{ind} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \boldsymbol{\ell} = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \end{bmatrix}.
$$

$\mathbf{S}_{ind}$ is a binary matrix that describes whether a particular branch occurs in the history of a focal species. $\mathbf{S}\mathbf{S}^T$ gives the covariance matrix of the phylogeny.

In general, the random-effect model matrix $\mathbf{Z}$ for a mixed model can be decomposed into term-wise model matrices $\mathbf{Z}_i$ as described in Bates et al. (2015). Analogous to this procedure, the phylogenetic random-effect matrix $\mathbf{Z}_i^C$ is

$$
\mathbf{Z}_i^C = (\mathbf{S}^\top \mathbf{J}_i^\top * \mathbf{X}_i^\top)^\top, \tag{3}
$$

where $\mathbf{S}$ is the $m \times b$ species–branch matrix; $\mathbf{J}_i$ is the $n_i \times m$ indicator matrix of grouping factors; $\mathbf{X}_i$ is the $n \times p_i$ raw random-effects model matrix; and $*$ is the Khatri-Rao product (Khatri and Rao, 1968) partitioned at the observation level (n).

For example, using the phylogeny above (figure 1), if we begin with a model matrix corresponding to intercept and slope terms,

$$\mathbf{X} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{bmatrix}$$

then the term-wise phylogenetic random effects model matrix is,

$$\begin{aligned} \mathbf{Z}_i^C = (\mathbf{S}^\top \mathbf{J}_i^\top * \mathbf{X}_i^\top)^\top &= \left[\left(\begin{bmatrix} \ell_1 & \ell_1 & 0 \\ \ell_2 & 0 & 0 \\ 0 & \ell_3 & 0 \\ 0 & 0 & \ell_4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) * \begin{bmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_3 \end{bmatrix} \right]^\top \\ &= \begin{bmatrix} \ell_1 & \ell_1 t_1 & \ell_2 & \ell_2 t_1 & 0 & 0 & 0 & 0 \\ \ell_1 & \ell_1 t_2 & 0 & 0 & \ell_3 & \ell_3 t_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ell_4 & \ell_4 t_3 \end{bmatrix}. \end{aligned} \tag{4}$$

Simulation

Single group model

Using the formulation described in (2–4), we generated test data with a single response variable $\mathbf{y}$ and a single normally distributed predictor variable $\boldsymbol{t}$ for $n = 25$, 50, and 100 species. The response variable $\mathbf{y}$ is conditionally normally distributed (i.e., $\mathcal{D}$ is a Gaussian distribution, and g is the identity link function), corresponding to a linear mixed effect model. For the first set of simulations, we simulate one observation per

species. Thus, the full simulation model is as follows:

$$
\begin{aligned}
\mathbf{y} &= \mathcal{D}(g^{-1}(\boldsymbol{\mu}), \phi) \\
\boldsymbol{\mu} &= (\beta_0 + \mathbf{b}_{\text{phy}_{\text{int}}}) + (\beta_1 + \mathbf{b}_{\text{phy}_{\text{slope}}})\boldsymbol{t} + \boldsymbol{\epsilon} \\
(\mathbf{b}_{\text{phy}_{\text{int}}}, \mathbf{b}_{\text{phy}_{\text{slope}}}) &\sim \text{MVN}\left(0, \begin{bmatrix} \sigma^2_{\text{phy}_{\text{int}}} & \sigma_{\text{phy}_{\text{int-slope}}} \\ \sigma_{\text{phy}_{\text{int-slope}}} & \sigma^2_{\text{phy}_{\text{slope}}} \end{bmatrix}\right) \\
\boldsymbol{\epsilon} &\sim \text{N}(0, \sigma^2_{\epsilon}) \quad .
\end{aligned}
\tag{5}
$$

The model has two fixed effect parameters (β_0 and β_1), three random effect parameters (phylogenetic random intercept variance ($\sigma^2_{\text{phy}_{\text{int}}}$), phylogenetic random slope variance ($\sigma^2_{\text{phy}_{\text{slope}}}$) and covariance between phylogenetic random intercept and slope ($\sigma_{\text{phy}_{\text{int-slope}}}$) and residual variance (σ^2_{ϵ}). The covariance between phylogenetic random intercept and slope measures the correlation of phylogenetic effects on the slope ($b_{\text{phy}_{\text{slope}}}$) and intercept ($b_{\text{phy}_{\text{int}}}$) for each branch of the phylogeny; i.e. a positive correlation indicates that species with similar intercepts also have similar slopes. Predictor-level and intercept-level random effects of species are not applicable in this simulation setting because there is only a single observation per species, so within-species variation cannot be separated from tip variation.

Multi-group model

We extend the simulation model by adding multiple groups where each group has one observation per species. The multi-group model is a generalization of multiple-site models used in community ecology to model phylogenetic attraction and repulsion (Helmus et al., 2007). The full multi-group model is as follows:

$$
\begin{aligned}
\mathbf{y} &= \mathcal{D}(g^{-1}(\boldsymbol{\mu}), \phi) \\
\boldsymbol{\mu} &= (\beta_0 + \mathbf{b}_{\mathrm{phy}_{\mathrm{int}}} + \mathbf{b}_{\mathrm{sp}_{\mathrm{int}}} + \mathbf{b}_{\mathrm{group}}) + (\beta_1 + \mathbf{b}_{\mathrm{phy}_{\mathrm{slope}}} + \mathbf{b}_{\mathrm{sp}_{\mathrm{slope}}})\boldsymbol{t} + \mathbf{b}_{\mathrm{sp:group}} + \boldsymbol{\epsilon} \\
(\mathbf{b}_{\mathrm{phy}_{\mathrm{int}}}, \mathbf{b}_{\mathrm{phy}_{\mathrm{slope}}}) &\sim \mathrm{MVN}\left(0, \begin{bmatrix} \sigma^2_{\mathrm{phy}_{\mathrm{int}}} & \sigma_{\mathrm{phy}_{\mathrm{int-slope}}} \\ \sigma_{\mathrm{phy}_{\mathrm{int-slope}}} & \sigma^2_{\mathrm{phy}_{\mathrm{slope}}} \end{bmatrix}\right) \\
(\mathbf{b}_{\mathrm{sp}_{\mathrm{int}}}, \mathbf{b}_{\mathrm{sp}_{\mathrm{slope}}}) &\sim \mathrm{MVN}\left(0, \begin{bmatrix} \sigma^2_{\mathrm{sp}_{\mathrm{int}}} & \sigma_{\mathrm{sp}_{\mathrm{int-slope}}} \\ \sigma_{\mathrm{sp}_{\mathrm{int-slope}}} & \sigma^2_{\mathrm{sp}_{\mathrm{slope}}} \end{bmatrix}\right) \\
\mathbf{b}_{\mathrm{group}} &\sim \mathrm{MVN}(0, \sigma^2_{\mathrm{group}}) \\
\mathbf{b}_{\mathrm{sp:group}} &\sim \mathrm{MVN}(0, \mathrm{I}_{\mathrm{group}} \otimes \sigma^2_{\mathrm{phy}}) \\
\boldsymbol{\epsilon} &\sim \mathrm{N}(0, \sigma^2_{\epsilon}),
\end{aligned}
\tag{6}
$$

where $\mathrm{I}_{\mathrm{group}}$ is a indicator matrix for groups and $\otimes$ is the Kronecker product.

Compared to the single–group model, the multi-group simulation model has five additional random effect parameters: predictor-level ($\sigma^2_{\mathrm{sp}_{\mathrm{slope}}}$) and intercept-level ($\sigma^2_{\mathrm{sp}_{\mathrm{int}}}$) among-species variances; their covariance ($\sigma_{\mathrm{sp}_{\mathrm{int-slope}}}$); among-group variation in the intercept ($\sigma^2_{\mathrm{group}}$); and variation among species-group combinations in the intercept (σ^2_{phy}). Because each species has multiple observations we can distinguish variation among species from residual variation, and thus we can include predictor-level and intercept-level random effects of species. Variance in the intercept of species–group interactions (σ^2_{phy}) describes whether the species within a group have more similar responses on average than expected by chance, equivalent to phylogenetic attraction (Helmus et al., 2007).

Platforms

We compare our approach with five other R packages that can fit PRs: `nlme` (Pinheiro et al., 2019), `phylolm` (Ho and Ané, 2014), `pez` (Pearse et al., 2015), `phyr`

(Ives et al., 2019) and `brms` (Bürkner, 2018). Phylogenetic generalized least squares (PGLS) (`gls` in `nlme`) is one of the most widely used PR models; it fits a linear model with a covariance structure that assumes an evolutionary process on the tree (typically BM, but other processes can be used) instead of treating the residual error for each species as independent. Phylogenetic generalized linear models (PGLM) (`phyloglm` in the `phylolm` package) extend PGLS by allowing for both phylogenetic and residual variation, as well as non-Gaussian response variables. Both `gls` and `phylolm` can model non-Brownian evolutionary processes and different correlation structures (e.g., Pagel's λ or Blomberg's K), but we restrict our PGLS fits to the simple BM correlation. Neither PGLS nor PGLM can handle random slopes or multiple observations within a species. Among the few packages that currently fit phylogenetic slopes to predictor level variation are `pez` and `phyr`, which can handle random-slope models ($\sigma^2_{\text{phy}_{\text{slope}}}$) and random intercepts of species-group interactions ($\mathbf{b}_{\text{sp:group}}$) but do not incorporate covariation between phylogenetic random slopes and intercepts ($\sigma_{\text{phy}_{\text{int}-\text{slope}}}$). Lastly, Bayesian PGLMMs using Markov chain Monte Carlo (MCMC) can handle all of the cases described above. However, MCMC is usually much more computationally expensive for GLMMs than platforms using deterministic optimization. `MCMCglmm` (Hadfield and Nakagawa, 2010) is the most widely used Bayesian phylogenetic GLMM; the more recent`brms` package is extremely flexible and uses Hamilton Monte Carlo, which is often more computationally efficient (although below we find that `MCMCglmm` is faster in this application).

Simulation and evaluations

Using the R package `ape` (Paradis and Schliep, 2018), we simulated 100 random phylogenetic trees for each sample size (n = 25, 50, 100 and an additional n = 500 for the multi-group model) and then simulated the responses for each tree (5, 6). Each realization was fitted using all model variants. All simulation parameters are shown

	nlme	phylolm	phyloglmm (this paper)	pez	phyr	brms	MCMCglmm
Single Group	✓	✓	✓			✓	✓
Phylo Intercept	✓	✓	✓			✓	✓
Phylo Slope			✓			✓	✓
Phylo Slope-Intercept correlation			✓			✓	✓
Residual		✓	✓			✓	✓
Multi-group			✓	✓	✓	✓	✓
Phylo Intercept			✓	✓	✓	✓	✓
Phylo Slope			✓	✓	✓	✓	✓
Phylo Slope-intercept correlation			✓			✓	✓
Phylo Species-group interaction			✓	✓	✓	✓	
Species intercept			✓	✓	✓	✓	✓
Species Slope			✓	✓	✓	✓	✓
Species Slope-intercept correlation			✓			✓	✓
Residual			✓	✓	✓	✓	✓

Table 1: List of estimable models for each R package.

in Figure 2 and Figure 5. Table 1 shows the parameters that are estimable for each platform. We only evaluated the goodness of fit for model fits that passed the convergence tests implemented by the package. For Bayesian fits, we evaluate realizations with Gelman-Rubin statistic < 1.1. Based on recent concerns about Gelman-Rubin thresholds (Vats and Knudson, 2021), we additionally restricted results to fits with effective sample size > 1000 for the fixed effect parameters (β_0 and β_1: Vehtari et al., 2021). For each replicate, we sample two chains starting with 10000 iterations. We first evaluate our estimates by looking at the distribution of the estimated values (maximum likelihood estimates for non-Bayesian platforms and posterior medians for Bayesian platforms) to quantify bias and variance (i.e., quality of the point estimate). We computed 95% Wald confidence intervals for frequentist methods and quantile-based intervals for Bayesian methods, then computed coverage — the proportion of simulations in which the computed confidence intervals include the true values of parameters — to assess the quality of the confidence intervals. We also compare computational speed between different platforms.

Results

In supplementary materials, we reproduce the examples in chapter 11 of Garamszegi (2014) using phylogenetic GLMMs based on `lme4` and `glmmTMB`.

Single Group model simulations

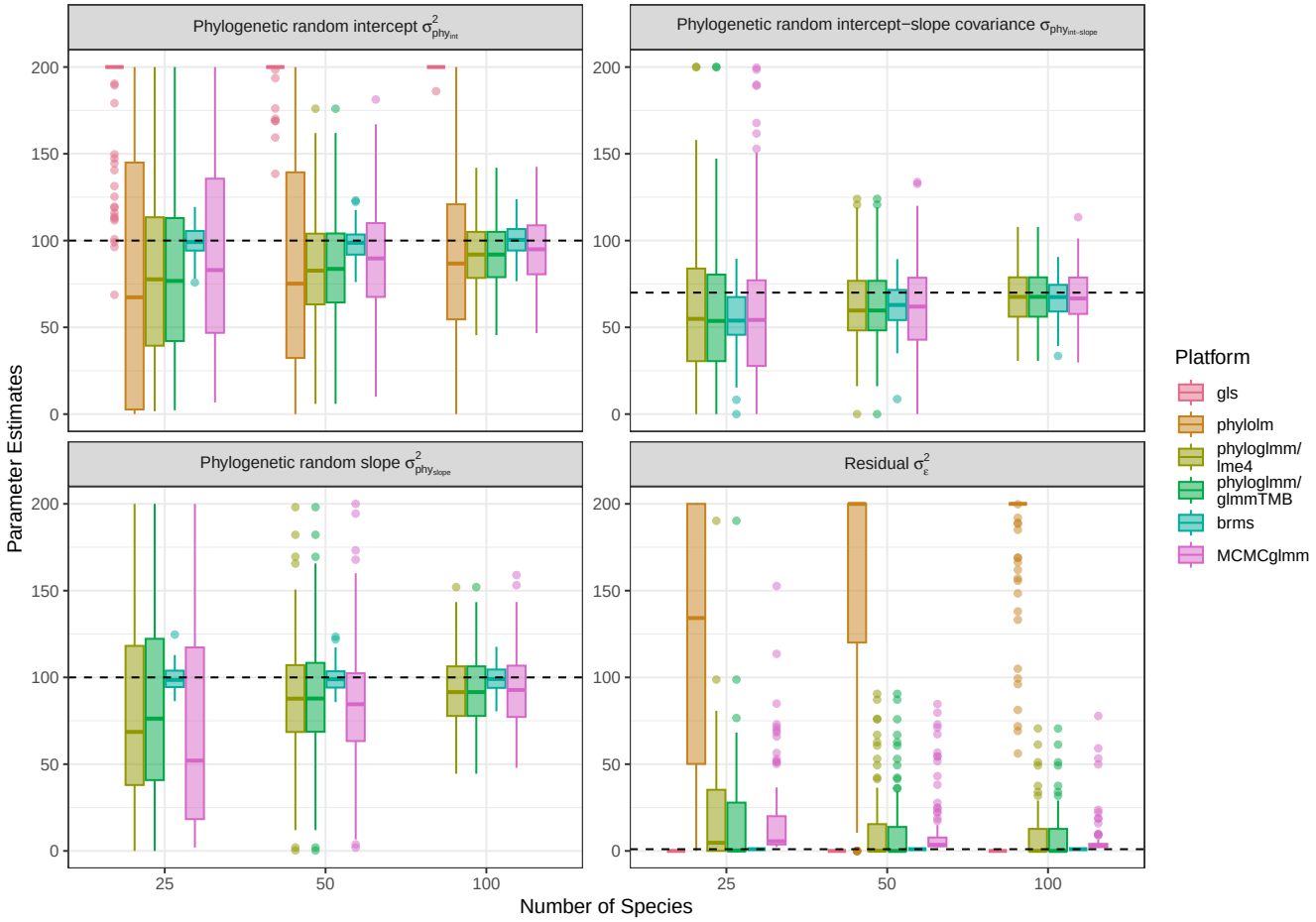

Fig. 2: Comparison of single group model parameter estimates across different R packages in Table 1. Total simulations $n = 100$ for each category. The horizontal line shows the true value of the parameters in the simulation model. Models capable of fitting all parameters (`phyloglmm/lme4`, `phyloglmm/glmmTMB`, and `brms`) fit well for all parameters. Values above 200 (very common for `gls` and `phylolm`, less than 1% of results otherwise) are censored to 200.

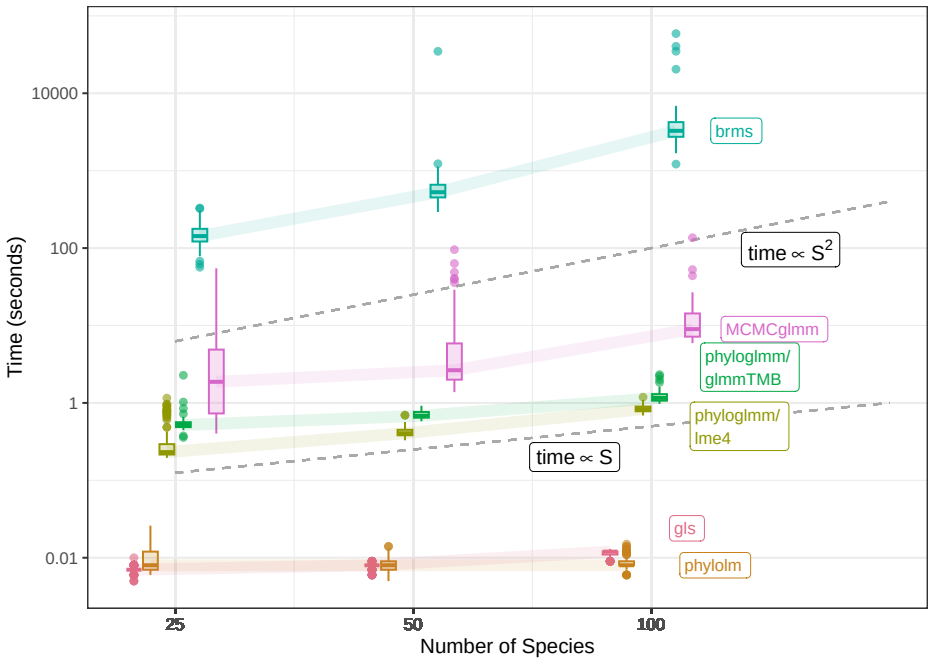

Fig. 3: Single-group model computational speed. Dashed lines indicate scaling relationships: lines parallel to the lower line would indicate linear scaling of computational time with number of species, while those parallel to the upper line would indicate quadratic scaling.

214

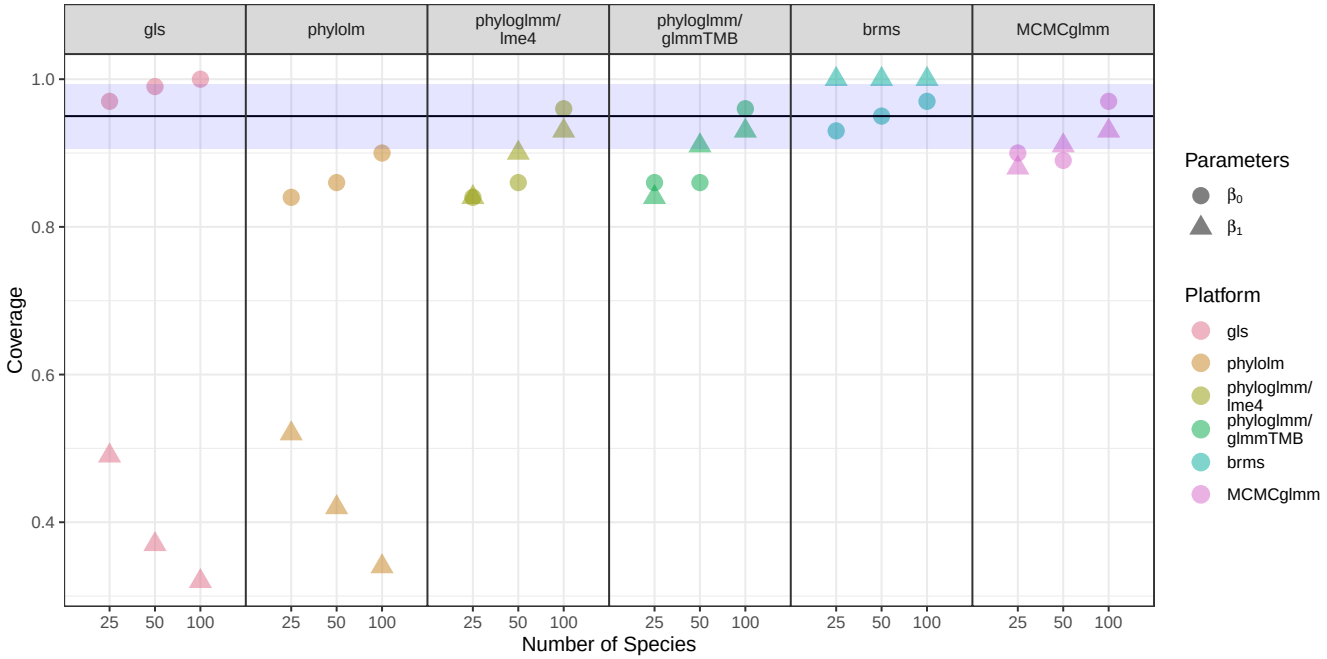

Fig. 4: Coverage probability for fixed effect parameters. Models matching the simulation model (`phyloglmm/lme4`, `phyloglmm/glmmTMB` and `brms`) have coverage near the nominal value of 0.95. The black line shows the nominal coverage; the blue ribbon shows the 95% binomial confidence interval around the nominal coverage based on fits to 100 simulated data sets.

The full fitted model (which matches the simulation model that incorporates phylogenetic intercept, slope, and correlation) provides estimates with low bias (average difference between the estimated parameters and the true simulation parameters) for all parameters. Estimates for fixed effect parameters (β_0 and β_1) approach nominal coverage as the number of species increases for `lme4` and `glmmTMB` but not for other packages. `brms` has higher than nominal coverage for the slope parameter β_1 (i.e., its confidence intervals are too conservative) because the prior distributions for the simulation parameters are centered at the true values (Li et al., 2018).

In general, models that are insufficiently flexible to match the data (PGLM and PGLS) will lead to bias in some parameters. PGLM (which lacks the phylogenetic slope parameter) provides reasonably good estimates for the phylogenetic intercept

standard deviation parameter ($\sigma_{\text{phy}_{\text{int}}}$) but overestimates the residual standard deviation; the estimates for the intercept (β_0) are slightly overconfident (coverage $\approx 90\%$ with 100 species) and the fixed slope parameter (β_1) has extremely poor coverage ($< 60\%$). PGLS, which uses only one parameter, combines all variation (phylogenetic intercept, slope and residual variation) into the phylogenetic intercept parameter; as a result, it overestimates the phylogenetic intercept variation, over-covers for β_0, and under-covers for β_1.

Multi-group model simulations

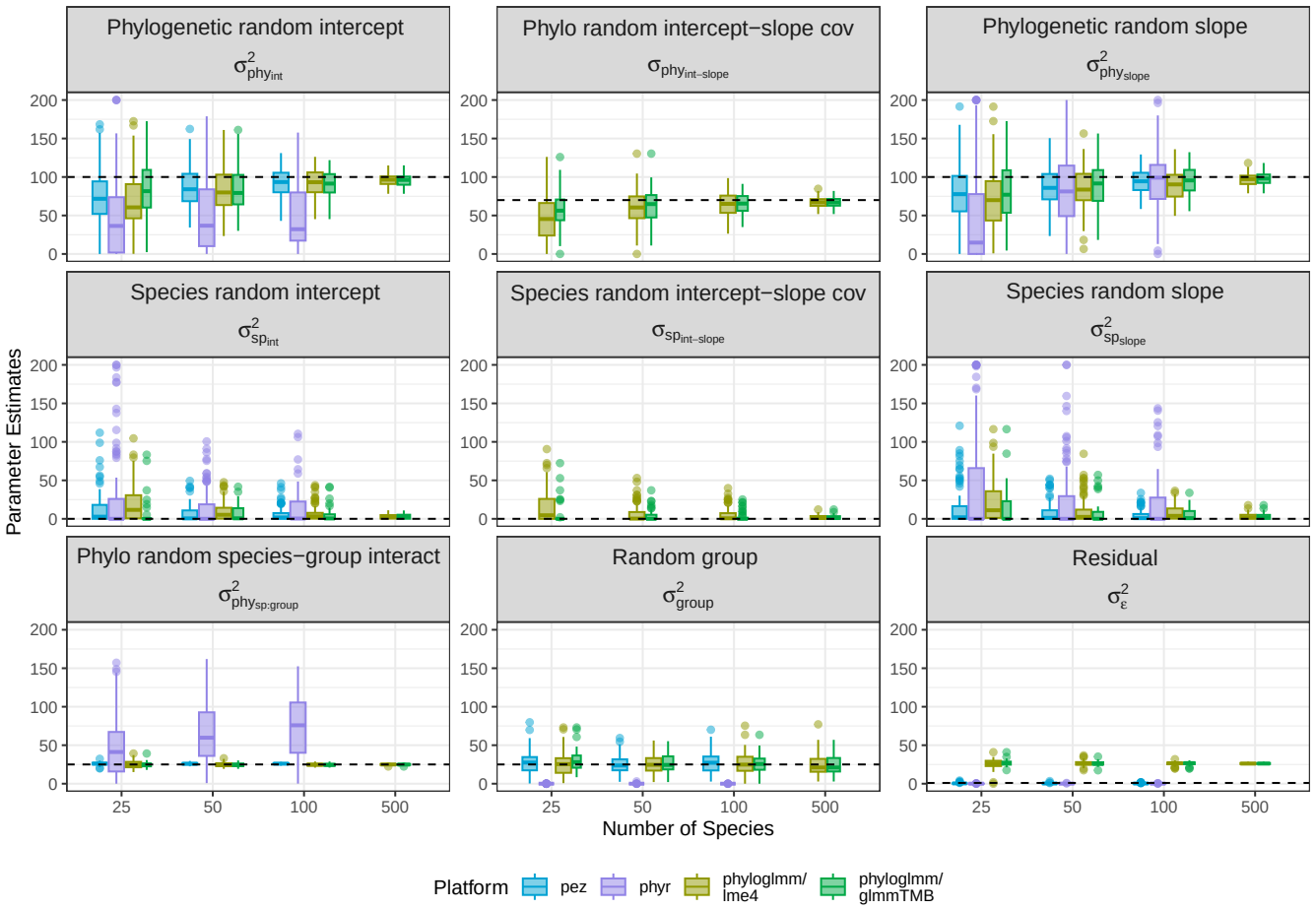

Fig. 5: Multi-group model parameter estimates. The horizontal line shows the true value of the parameters in the simulation model. Models capable of fitting all parameters (`phyloglmm/lme4` and `phyloglmm/glmmTMB`) fit well for all parameters. `pez` and `phyr` estimates for $n = 500$ are missing because the models did not converge within 30 minutes.

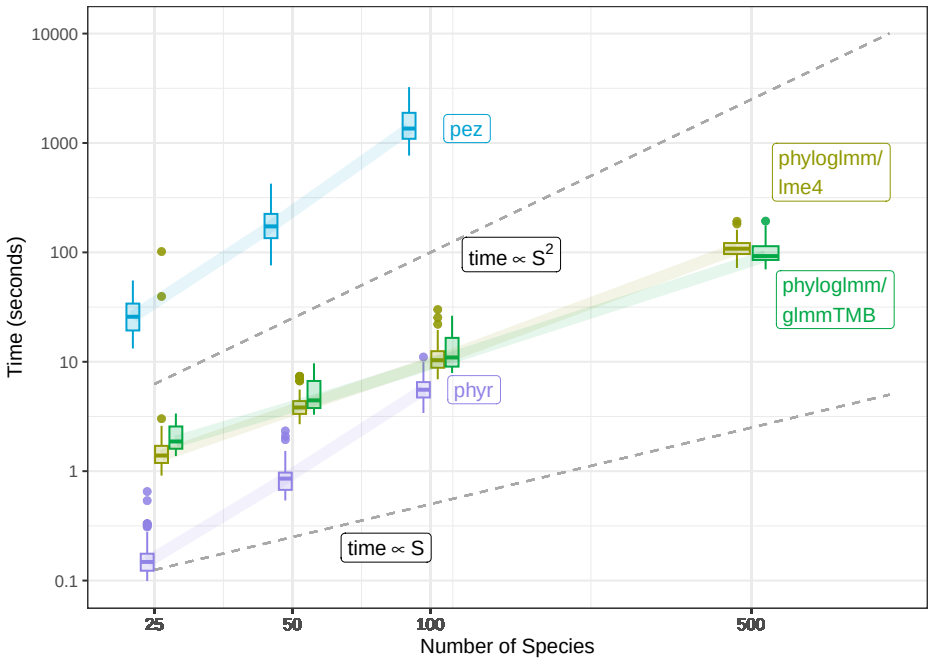

Fig. 6: Multi-group model computational speed. Linear and quadratic scaling lines as in Figure 3.

236

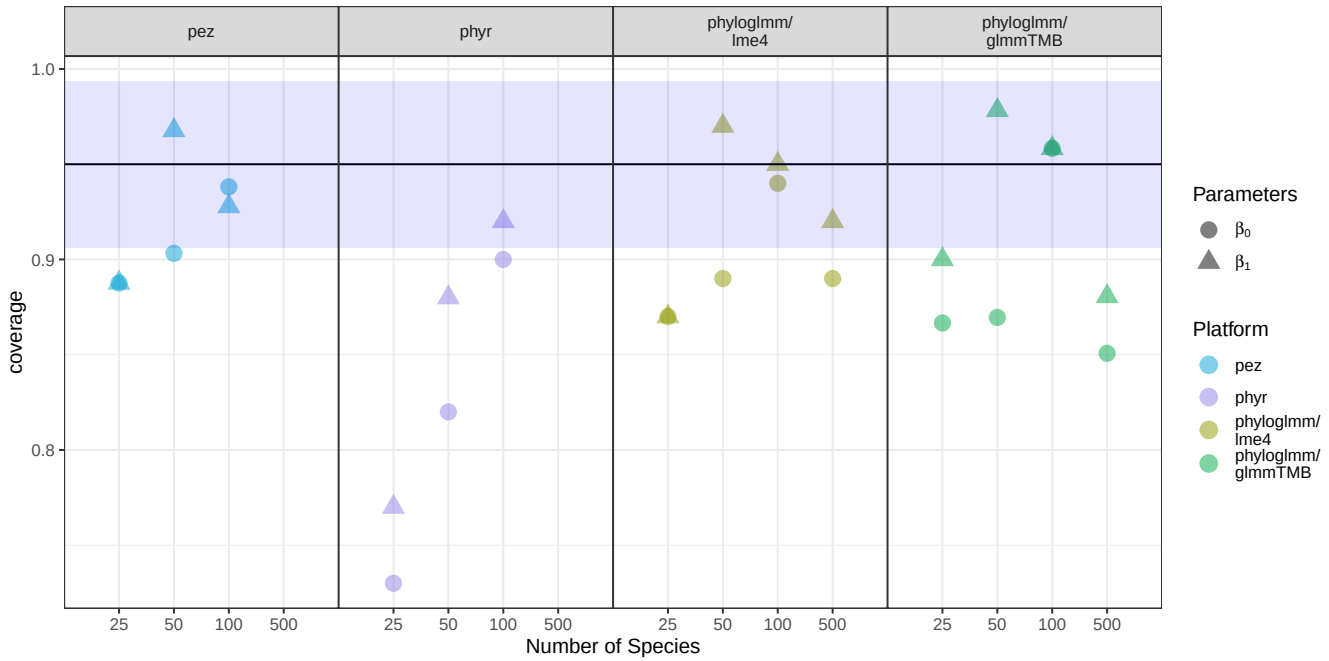

Fig. 7: Comparison of multi-group model coverage. The black line shows the nominal coverage, and the blue ribbon the 95% binomial confidence interval based on 100 simulated fits.

The multi-group model estimates are much more similar across platforms (only a subset of platforms can fit these models at all, and the fitting models are closer to the true model). As with the single-group fits, `lme4` and `glmmTMB` match the simulation model well and provide good estimates for all parameters, except the correlation ($\sigma_{\text{sp}_{\text{int}-\text{slope}}}$) for small numbers of species (i.e. n = 25 and 50). The absence of correlations in `pez` and `phyr`'s statistical models has little effect on the other parameter estimates but leads to underestimates of the residual standard deviation (Figure 5).

Although the parameter estimates are similar across platforms, computational efficiency varies enormously across platforms and sample size. Comparing our two implementations, `glmmTMB` is almost an order of magnitude faster than `lme4`; `glmmTMB` is expected to out perform `lme4` as it is designed to maximize flexibility and speed

(Brooks et al., 2017). `glmmTMB` is also faster than `pez` and `phyr`: the median time for `glmmTMB` to fit a 50-species model is ≈ 9, versus ≈ 200 seconds for `pez` and `phyr`. `glmmTMB` takes ≈ 125 seconds to fit a 500-species model; it was impractical to fit 500-species models with `pez` and `phyr`.

Discussion

We have simulated complex models containing phylogenetic variation in species intercepts and slopes, as well as within-species variation. Comparing our fits with simple platforms for phylogenetic regression that cannot handle these complexities may seem unfair; nevertheless, our models are certainly less complex than evolutionary processes occurring in nature. Models that cannot match the full simulations perform poorly even for the parameters they do estimate.

Even the relatively simple multi-group model described in (6) can incorporate many layers of complexity. In theory, as long as we have enough data and enough computational power, models that can incorporate more of the complexity will always describe a biological system better. However, real applications are always data-constrained. Establishing the practical level of model complexity for a given problem and data set is an open and difficult general problem throughout statistical modeling, not just in phylogenetic studies. While information-theoretic or stepwise selection methods are often used to decide the complexity of models in ecology and evolution (Darriba et al., 2020; Matuschek et al., 2017), sufficiently complex models may be impossible to fit without some form of regularization (Uriarte and Yackulic, 2009), and data-driven model selection may affect inference in unexpected and unwanted ways (Hurvich and Tsai, 1990; Morin et al., 2020).

Incorporating multiple levels of variation

Random-slopes models require appropriate observational or experimental designs (i.e., multiple measurements of traits and responses within each group) and generally require more data overall, but they are relevant over a wide range of scenarios (Cleasby et al., 2015; Ord et al., 2010; Schielzeth and Forstmeier, 2008). Neglecting random slopes can lead to biased fixed-effect estimates with inadequate coverage and inflated type I errors (Schielzeth and Forstmeier, 2008).

Nevertheless, it is impossible to account for all possible complexities. The best model — whether using phylogenetic random effects, simple grouping, or both — depends on the experimental design and whether there is enough data to separate different levels of variation, which can be strongly confounded. If multiple observations are available per species, then simple methods like Pagel's λ will confound tip variation with residual variation (Boettiger, 2013). Multiple observations can be collapsed to a single value (such as the mean) per species, with analyses weighting each species by its number of observations (Murtaugh, 2007). Alternatively, if the within-species variance is of interest, a phylogenetic mixed model can separate tip variation from within-species variation and measurement error (Kostikova et al., 2016). More generally, phylogenetic mixed models can be simplified to ordinary mixed models, at the cost of taxonomic detail, by simplifying the phylogenetic tree to a strictly hierarchical set of nested higher-level taxa (Bunnefeld and Phillimore, 2012)). Users should be aware of two essential questions when fitting random-slope models: do they have enough information to reliably estimate the random slopes, and what are the potential costs of ignoring them (Schielzeth and Forstmeier, 2008)?

Extensions and alternatives

The range of PRs presented here can incorporate many levels and types of phylogenetic variation. Of course, we have neglected further biological complexities, such as

multivariate responses; non-Brownian evolutionary processes such as the Ornstein-Uhlenbeck model (Butler and King, 2004)); and variable-rate models, which allow evolutionary rates to vary across the phylogeny. While it cannot easily incorporate these complexities, the approach here does offer an efficient way to analyze a wide range of evolutionary scenarios. The general principle of encoding phylogenetic structure in the random effects model matrix can be implemented with any platform that supports continuous latent variables. Our implementation allows users to explore new ideas by fitting phylogenetic mixed models with complex random effects to large data sets.

Achnowledgements

We would like to thank Jonathan Dushoff for thoughtful comments. This study was funded by an NSERC Discovery Grant.

Authors' contributions

ML and BMB conceived the ideas and designed methodology; ML and BMB implemented the code in `lme4` and `glmmTMB`; ML ran all simulations; ML and BMB analyzed the results; ML wrote the first draft. Both authors contributed critically to the drafts and gave final approval for publication.

Data Availability

All codes are available at DOI:10.5281/zenodo.2639887.

References

Bates, D., M. Mächler, B. Bolker, S. Walker, et al. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(i01).

Blomberg, S. P., T. Garland Jr, and A. R. Ives (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution 57*(4), 717–745.

Boettiger, C. (2013, October). Is it time to retire Pagel's lambda?

Brooks, M. E., K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Maechler, and B. M. Bolker (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal 9*(2), 378–400.

Bunnefeld, N. and A. B. Phillimore (2012). Island, archipelago and taxon effects: mixed models as a means of dealing with the imperfect design of nature's experiments. *Ecography 35*(1), 15–22.

Butler, M. A. and A. A. King (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist 164*(6), 683–695.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal 10*(1), 395–411.

Cleasby, I. R., S. Nakagawa, and H. Schielzeth (2015). Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution 6*(1), 27–37.

Darriba, D., D. Posada, A. M. Kozlov, A. Stamatakis, B. Morel, and T. Flouri (2020, January). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution 37*(1), 291–294.

de Villemereuil, P., J. A. Wells, R. D. Edwards, and S. P. Blomberg (2012). Bayesian models for comparative analysis integrating phylogenetic uncertainty. *BMC Evolutionary Biology 12*(1), 102.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist 125*(1), 1–15.

Garamszegi, L. Z. (2014). *Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice*. Springer.

Hadfield, J. and S. Nakagawa (2010). General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of evolutionary biology 23*(3), 494–508.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software 33*(2), 1–22.

Hansen, T. F. and K. Bartoszek (2012). Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Systematic Biology 61*(3), 413–425.

Hefley, T. J., K. M. Broms, B. M. Brost, F. E. Buderman, S. L. Kay, H. R. Scharf, J. R. Tipton, P. J. Williams, and M. B. Hooten (2017). The basis function approach for modeling autocorrelation in ecological data. *Ecology 98*(3), 632–646.

Helmus, M. R., K. Savage, M. W. Diebel, J. T. Maxted, and A. R. Ives (2007). Separating the determinants of phylogenetic community structure. *Ecology letters 10*(10), 917–925.

Ho, L. S. T. and C. Ané (2014). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology 63*, 397–408.

Housworth, E. A., E. P. Martins, and M. Lynch (2004). The phylogenetic mixed model. *The American Naturalist 163*(1), 84–96.

Hurvich, C. M. and C. Tsai (1990, August). The Impact of Model Selection on Inference in Linear Regression. *The American Statistician 44*(3), 214–217.

Ives, A., R. Dinnage, L. A. Nell, M. Helmus, and D. Li (2019). *phyr: Model Based Phylogenetic Analysis*. R package version 1.0.2.

Ives, A. R. and M. R. Helmus (2011). Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs 81*(3), 511–525.

Khatri, C. and C. R. Rao (1968). Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhyā: The Indian Journal of Statistics, Series A*, 167–180.

Kostikova, A., D. Silvestro, P. B. Pearman, and N. Salamin (2016). Bridging inter- and intraspecific trait evolution with a hierarchical Bayesian approach. *Systematic biology 65*(3), 417–431.

Li, D., A. R. Ives, and D. M. Waller (2017). Can functional traits account for phylogenetic signal in community composition? *New Phytologist 214*(2), 607–618.

Li, M., J. Dushoff, and B. M. Bolker (2018). Fitting mechanistic epidemic models to data: a comparison of simple Markov chain Monte Carlo approaches. *Statistical methods in medical research 27*(7), 1956–1967.

Lynch, M. (1991). Methods for the analysis of comparative data in evolutionary biology. *Evolution 45*(5), 1065–1080.

Matuschek, H., R. Kliegl, S. Vasishth, H. Baayen, and D. Bates (2017, June). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language 94*, 305–315.

Morin, D. J., C. B. Yackulic, J. E. Diffendorfer, D. B. Lesmeister, C. K. Nielsen, J. Reid, and E. M. Schauber (2020). Is your ad hoc model selection strategy affecting your multimodel inference? *Ecosphere 11*(1), e02997.

Murtaugh, P. A. (2007). Simplicity and Complexity in Ecological Data Analysis. *Ecology 88*(1), 56–62.

Nowakowski, A. J., L. O. Frishkoff, M. E. Thompson, T. M. Smith, and B. D. Todd (2018). Phylogenetic homogenization of amphibian assemblages in human-altered habitats across the globe. *Proceedings of the National Academy of Sciences*, 201714891.

Ord, T. J., J. A. Stamps, and J. B. Losos (2010). Adaptation and plasticity of animal communication in fluctuating environments. *Evolution 64*(11), 3134–3148.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature 401*(6756), 877.

Paradis, E. and K. Schliep (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics 35*, 526–528.

Pearse, W. D., M. W. Cadotte, J. Cavender-Bares, A. R. Ives, C. M. Tucker, S. C. Walker, and M. R. Helmus (2015). Pez: Phylogenetics for the environmental sciences. *Bioinformatics 31*(17), 2888–2890.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team (2019). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-141.

Schielzeth, H. and W. Forstmeier (2008). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology 20*(2), 416–420.

Uriarte, M. and C. B. Yackulic (2009). Preaching to the Unconverted. *Ecological Applications 19*(3), 592–596. Publisher: [Wiley, Ecological Society of America].

Vats, D. and C. Knudson (2021, November). Revisiting the Gelman–Rubin Diagnostic. *Statistical Science 36*(4), 518–529.

Vehtari, A., A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner (2021, June). Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis 16*(2), 667–718. Publisher: International Society for Bayesian Analysis.