Phylogenetic comparative methods for discrete responses in evolutionary biology

---

Master Thesis in Biostatistics (STA495)

by

Sereina Graber
s06914139

supervised by

Prof. Dr. Reinhard Furrer
Dr. Karin Isler Anthroplogical Institute and Museum

Zurich, November 2013

Contents

Abstract

This study investigated and compared four commonly used phylogenetic comparative methods (phylogenetic logistic regression, phylogenetic generalised estimating equations, phylogenetic logistic regression and phylogenetic mixed model) modeling the correlated evolution of simulated data. The phylogenetic mixed model showed the best performance in case of a continuous response variable, the phylogenetic generalised least-squares regression in case of an ordinal scaled variable treated as pseudo-continuous and in case of binary data, the logistic regression by Ives and Garland (2010) and Ho and Ané (in review) showed good statistical abilities. Ignoring the phylogenetic dependencies between species by using a simple non-phylogenetic generalised linear model approach, always results in highly elevated type I error rates and bad estimation abilities.

Furthermore, the methods were tested with respect to different evolutionary and empirical parameters such as phylogenetic tree structure, sample size and strength of correlation. The study shows that the different phylogenetic approaches vary in their statistical abilities depending on the underlying simulated conditions. Therefore, the decision of what method to use does not only depend on the overall statistical performance of a method, but also on the given data and the underlying evolutionary question.

General notes

There are some general notes which need to be made prior to reading this thesis. First, about the mathematical notation: matrices and vectors are always denoted in bold symbols, whereas scalars are shown in normal font. For example: $\mathbf{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. $\boldsymbol{Y}$, $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ are vectors printed in bold font, as well as $\boldsymbol{X}$ representing a matrix. As there is no specific symbolic differentiation between vectors and matrices, in most cases the variables used in mathematical formulas are additionally described in written form and dimension specifications are given. Moreover, MLE stands for the abbreviation of the maximum likelihood estimate. Second, all the methods described are implemented in R (R Development Core Team 2011) and mostly, assumptions and model descriptions are based on their implementations in R. Moreover, functions and packages in R are denoted in the typical `R font`.

Commonly used variables and indices:

- N = number of species $i = \{1, 2, \ldots, N\}$ (indices i and j represent two different species)
- p = number of covariates
- S = number of clusters/units $s = \{1, 2, \ldots, S\}$
- k = rank of a matrix

1 Introduction

1.1 Comparative methods - a key to understand adaptation and correlated evolution

Comparative methods in evolutionary biology are used for testing hypotheses of adaptation which cannot be tested experimentally. They are the key to understand evolution in a very broad sense, to understand evolutionary processes across many different species in contrast to within a single species. Phylogenetic comparative methods, with phylogenetic referring to the evolutionary relationships between species, allow to approach many different kinds of evolutionary questions: the ancestral states of certain biological traits (i.e. specific phenotypic features of a species including morphological, behavioural and life history traits e.g. body size or the age of first reproduction), about how fast certain traits change over evolutionary time or whether two traits show correlated evolution.
To explain the aim and application of a comparative study, the best thing to do is to give an example based on a real evolutionary question, e.g. concerning the evolution of life history across species (i.e. the length of the different episodes of the biological life cycle such as the age at which a species first reproduces). For example, the brain malnutrition risk hypothesis, states that larger brains need longer developmental periods in order to avoid the brain to grow too fast which would result in harmful energetic deficits (Janson and van Schaik 1993). In other words, this hypothesis tries to explain the variation in the length of development (e.g. represented by the age of first reproduction) with the variation in brain size across species predicting a positive correlation between the two traits. For illustration purposes, this example of correlated evolution will be used throughout the whole thesis.
When testing such a correlation, one needs to be aware of the fact that different species share a common evolutionary history leading to the problem of non-independence concerning statistical methods (i.e. for many statistical methods residuals need to be independent and identically distributed). Being precise, species show high similarities in certain traits due to their close phylogenetic relationship (if the great similarity is mainly based on the close phylogenetic relationship, a trait shows a high phylogenetic signal). Many studies have shown that ignoring phylogenetic dependencies between species dramatically increases type I error rates resulting from incorrect variance estimations (e.g. Rohlf 2006, Martins and Garland 1991). Therefore, as spatial clustered data is statistically considered in e.g. mixed models, many different phylogenetic comparative methods have been suggested accounting for the phylogenetic dependency between species (Felsenstein 1985, Cheverud et al. 1985, Gittleman and Kot 1999, Lynch 1991, Hadfield and Nakagawa 2009, Ives and Garland 2010). Because there are several different comparative methods, which claim to do all the same thing (i.e. correcting for phylogenetic dependencies), biologists are often overstrained with the decision of which method to use for what kind of data. In fact, different methods might

respond differently to varying sample sizes or phylogenetic tree structures. Moreover, most of these methods are designed for continuous dependent variables and only few are specifically developed for discrete responses. However, especially discrete data is rather common in evolutionary biology. To exemplify only a few, measures on absence or presence of certain morphological or behavioural traits (binary traits) are typical measures in evolutionary biology.
A lack of knowledge and the discordant and unclear literature probably often result in studies applying wrong methods, violating important assumptions and conditions. Therefore, the aim of this study is to compare four commonly used phylogenetic methods (phylogenetic generalised least-squares regression, phylogenetic generalised estimating equation, phylogenetic logistic regression and phylogenetic generalised mixed model) modeling evolutionary correlations under different scenarios. Based on simulations of correlated evolution between traits with varying number of species, strength of correlations and phylogenetic tree structures, the statistical performances of the four methods are tested and compared. Additionally, to verify and compare the practical applicabilities of these methods on a real evolutionary question, the brain malnutrition risk hypothesis explained above is tested amongst other hypothesis on a data set including 78 primates species.

In a first part, the introduction covers the explanations of specific phylogenetic terms, and second, the specific comparative methods and their mathematical backgrounds are described.

1.2 Phylogenetic terms

Before going into more mathematical details of the specific phylogenetic methods, it is crucial to discuss the underlying phylogenetic terms. Although, the focus in this section is mainly set on the biological perspective, certain links to the mathematical translation are important in order to understand the idea behind comparative methods. In the first section, the phylogenetic tree and its numerical representation in a variance-covariance matrix is explained. The correlations of certain traits between species might be based on their close phylogenetic relationships leading to non-independence of the data. Second, each phylogenetic method makes assumptions about an underlying evolutionary model. Such a model describes the mode and timing of the evolution of a trait. In the last section, the parameters which describe an evolutionary model are discussed.

1.2.1 Phylogenetic tree

A phylogenetic tree represents the evolutionary relationships between species. It represents the phylogenetic dependencies between species which are taken into account in comparative methods. A phylogenetic tree consists of branches, nodes and tips: *branches* link the species in a hierarchical pattern and their lengths represent the evolutionary divergences between them proportional to the expected variances. The divergences, branch lengths, respectively, are either based on absolute evolutionary times since last common ancestors or variable distances of morphological, genetic and

behavioural data (Nunn 2011). The longer the branch length, the higher the evolutionary divergence. *Nodes* represent speciation events, where lineages (species) evolve out of others. *Tips* symbolize extinct or extant species. A graphical illustration is found in Figure 1.1. If more than two branches split from a node is called a *polytomy*. Mostly, such politomies reflect uncertainties about the phylogenetic relationships (Nunn 2011). However, usually, the application of phylogenetic methods requires a complete phylogenetic tree with the underlying structure assumed to be true.

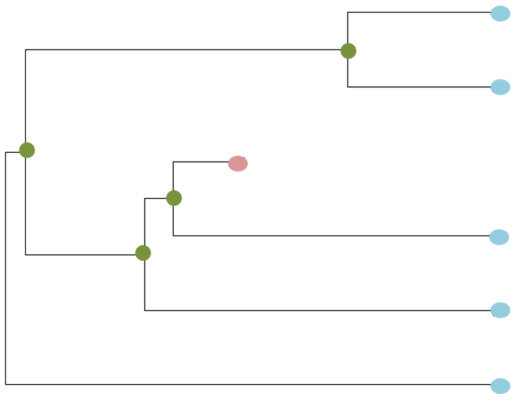
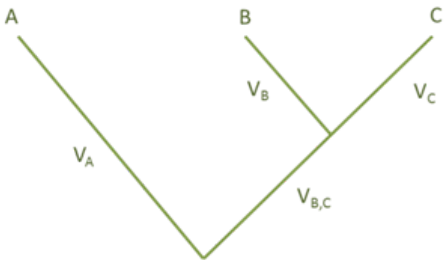


Figure 1.1: **Phylogenetic tree.** The green dots represent nodes, the tips are symbolized by red and blue dots. The red ones stand for extinct species and the blue ones for extant species. The nodes and tips are hierarchically linked by branches.

After having explained the phylogenetic tree mainly from a biological perspective, the following paragraph rather explains the mathematical implementation of a phylogenetic tree. In order to take into account the phylogenetic dependencies between species in a comparative analysis, a phylogenetic tree, including its structure and branch lengths, can be expressed in form of a variance-covariance matrix. The diagonal elements of this phylogenetic variance-covariance matrix, from now on denoted with $\mathbf{V}$, are given by the variance of a trait within a single species, representing the sum of branch length from the root to the tip of the extant species. In mathematical terms these values can be described as $t\sigma^2$ with t representing the total path length from the root to the tip of the corresponding species (Pagel 1999). The off-diagonal elements of $\mathbf{V}$ are the covariances between pairs of species, represented by their shared evolutionary history, the sum of the branch lengths from the root to the last common ancestor of the pair of species. In other words, the longer the (shared) evolutionary time, the higher the variance (covariance) within and between species, respectively. For graphical illustration see Figure 1.2.
Such a variance-covariance matrix builds the basis for all the phylogenetic correlation structures of the described comparative methods in Section 1.3.



	A	B	C
A	V_A	0	0
B	0	$V_{B,C}+V_B$	$V_{B,C}$
C	0	$V_{B,C}$	$V_{B,C}+V_C$

Figure 1.2: **Phylogeny and corresponding vcv matrix.** The phylogenetic tree transformed into a vcv matrix. Diagonal elements represent the variances within species (branch length from last common ancestor to a tip of the tree). Off-diagonal elements represent the covariances between species due to their shared evolutionary history (sum of branch lengths from root to the last common ancestor). (Corresponding to AnthroTree Workshop, Isabella Capellini)

1.2.2 Models of evolution

Different phylogenetic comparative approaches make different assumptions about the underlying evolutionary model. An evolutionary model describes the mode and the timing of the evolution of a trait. Nunn (2011, p. 100) defines it as an "explicit framework for considering how traits change over time". Two famous models of evolution for continuous traits are the *Brownian motion* (Edwards and Cavalli-Sforza 1964) and the *Ornstein-Uhlenbeck* (also known as the *stabilizing selection* model) (originally developed by Uhlenbeck and Ornstein 1930).
The value of a trait modelled according to Brownian motion can randomly increase or decrease showing no directional trend. The model assumes that traits in lineages evolve randomly (thus, also known as a random-walk model) with a constant rate of change and that changes in one lineage are completely independent from changes in another lineage (e.g. Pagel 1997). Additionally, the degree of change is proportional to the length of a branch. Specifically, this means traits modeled by Brownian motion can achieve random unlimited variance. The fact that Brownian motion (e.g Martins and Hansen 1996, Nunn 2011) is very generally applicable and that several comparative methods (e.g. independent contrasts by Felsenstein 1985) assume it to be the underlying evolutionary model, this model has become probably the most famous model of evolutionary change.
Another well known model is the so called Ornstein-Uhlenbeck model which describes stabilizing selection (e.g. Felsenstein 1988, Lavin et al. 2008, Nunn 2011). This model is basically the same as a Brownian motion model, however, an additional incorporated force ("restraining force" α) makes sure the trait value varies around a certain optimum. Thus, a Ornstein-Uhlenbeck model with a restraining force of $\alpha = 0$ is the same as a Brownian motion model. In evolutionary terms, this means that specific traits cannot cross certain threshold values, such as body mass in birds constrained by a physical boundary (Nunn 2011). Simulations of the two evolutionary models are shown in Figure 1.3.

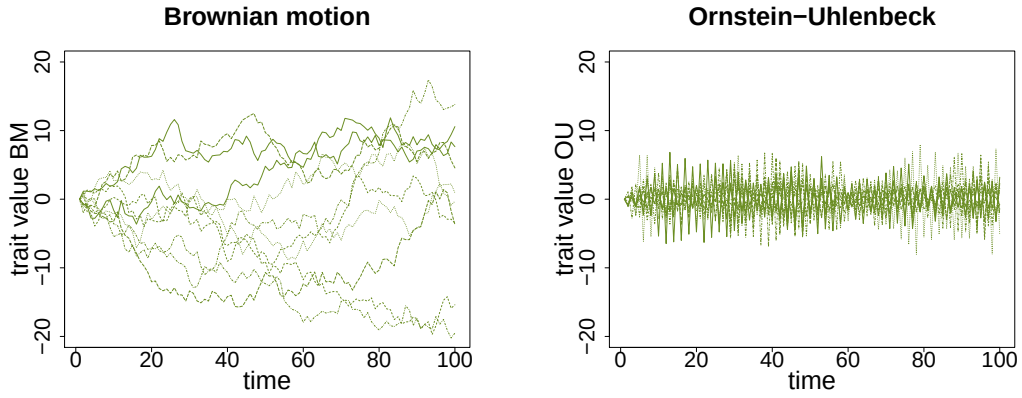


Figure 1.3: **Trait evolution under Brownian motion and Ornstein-Uhlenbeck.** The left figure shows a simulation of 10 evolutionary histories (each line representing a different simulation) of a trait over time according to a Brownian motion model. The right figure shows the same for a Ornstein-Uhlenbeck model ($t_{u+t} = t_u + r_n - t_u \times \alpha$; t_u = trait value at time u, r_n = random normal variable $\overset{iid}{\sim} N(0,1)$, α = restraining force) with a restraining force $\alpha = 0.95$. The trait under Brownian motion shows a much higher variation (variance = 124.1) than under Ornstein-Uhlenbeck (variance = 16.8). Simulations were done in `R` using the package `ape` according to Nunn 2011 and the AnthroTree website http://nunn.rc.fas.harvard.edu/groups/pica/ chapter 5.1 by Charly Nunn.

The two models presented above are designed to describe the evolutionary process of continuous traits. However, in this study, the focus is laid on discrete traits. As an analogue to the Brownian motion model, the Markov-transition process is used to describe the evolution of discrete traits (e.g. Pagel 1994, 1999, Schluter et al. 1997). This model makes the same assumptions as a Brownian motion model, namely that evolutionary changes of one lineage is independent of the changes of other lineages and the rate of evolutionary change is constant over time and along all branches. In that sense it is also a random-walk model. In the case of Markov-transition process, the probabilities that a trait changes from one state to another (transition rates) is represented in a transition matrix (example in Figure 1.4). The transition matrix can be designed according to certain evolutionary assumptions, such as that the changes between states is unordered or ordered, changes might be irreversible etc. (Maddison and Maddison 2000, p. 69-72). These transition rates are estimated by maximum likelihood given the data of extant species and a phylogenetic tree with branch lengths (Pagel 1994, Schluter et al. 1997).

	A	B	C
A	0.7	0.2	0.1
B	0.3	0.5	0.2
C	0.5	0.4	0.1

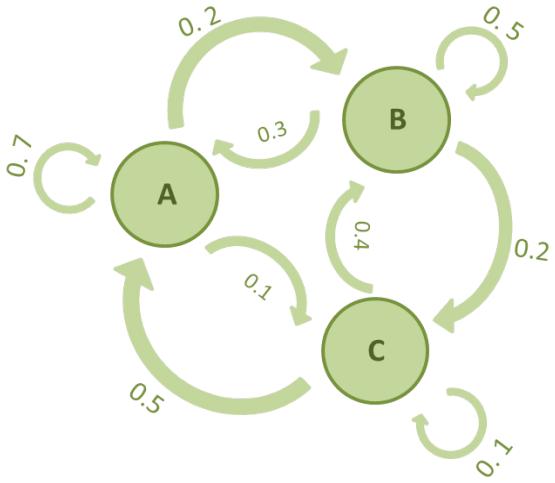


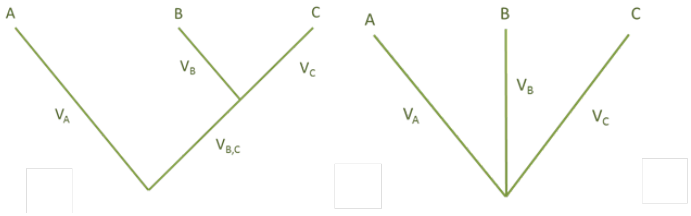
Figure 1.4: **Markov process with corresponding transition matrix.** A Markov process can be used to describe the evolutionary model of discrete data. In this specific case, the discrete trait consists of the three levels A, B and C. The transition matrix contains the probabilities that a trait changes from one state to another (transition rates).

1.2.3 Phylogenetic signal and branch length transformation factors

In order to balance how much of the variation of a trait can be explained by the correlated evolution with an other trait and how much by its phylogenetic predisposition, one needs to measure both, the correlation as well as the phylogenetic predisposition. The latter is measured by a so-called phylogenetic signal. This is a measure for how much of the similarity of a trait between two species can be explained by their phylogenetic relatedness (i.e. common evolutionary history). In particular, going back to the example of the evolution of life history in relation to brain size variation, a high phylogenetic signal means that a large proportion of the similarity in brain size between two closely related species, such as the gorilla and the chimpanzee, can be explained by their close phylogenetic relatedness.

Probably the most widely known measurement of the phylogenetic signal for continuous traits is lambda λ by Pagel 1999 and Freckleton et al. 2002. Switching from the biological to the mathematical perspective, λ can be used to scale the off-diagonal elements (internal branches of a phylogenetic tree) of the phylogenetic variance-covariance matrix ($\boldsymbol{V}$, see former section about phylogenetic trees). λ varies between 0 and 1, 0 meaning that there is no phylogenetic signal at all ending up in a star phylogeny (Figure 1.5) and λ equal to 1 stands for an evolutionary model of Brownian motion (Figure 1.5). Values in between 0 and 1 indicate a smaller phylogenetic signal under the assumption of Brownian motion (e.g. Nunn 2011). The stronger the phylogenetic signal, λ, the larger the off-diagonal elements of the phylogenetic variance-covariance matrix $\mathbf{V}$, the higher the covariances between species due to their shared evolutionary history.
Using a maximum likelihood procedure, an estimate of λ is found by an optimization

algorithm given a model of evolution (Brownian motion) and a phylogeny (further details are found in the subsection about the phylogenetic least-squares regression).



	A	B	C
A	V_A	λ 0	λ 0
B	λ 0	$V_{B,C}+V_B$	λ $V_{B,C}$
C	λ 0	λ $V_{B,C}$	$V_{B,C}+V_C$

Figure 1.5: **Phylogeny and corresponding vcv matrix scaled by λ.** A phylogenetic tree and the corresponding vcv matrix where the off-diagonal elements are scaled by λ. λ equals to 1 is represented by the left phylogenetic tree and λ equal to 0 is represented by the star phylogeny on the left side. (Corresponding to AnthroTree Workshop, Isabella Capellini)

There are other so-called branch length transformation parameters for continuous traits such as delta (δ) and kappa (κ). Delta (ranges between 0 and 3) represents a transformation factor increasing the diagonal (total path length from the root to a tip) and off-diagonal values (shared path length of two species) of $\mathbf{V}$ to the power of the value of delta. δ equal to one corresponds to a Brownian motion model of evolution (Nunn 2011), whereas $\delta > 1$ means that the rate of change increases over time, longer paths contributing more to trait evolution. $\delta < 1$ represents the opposite pattern, shorter paths contributing more to the evolution of a trait (lecture notes by Isabella Capellini, AnthroTree Workshop 2011). In evolutionary biology terms, δ is a measurement for the evolutionary rate (i.e. tempo of evolution), e.g. whether the rate of evolution accelerates or slows down over time (lecture notes by Isabella Capellini, AnthroTree Workshop 2011). In case of kappa (ranges between 0 and 3), the branch lengths are transformed to the power of the kappa value, if $\kappa = 0$, trait evolution is independent of branch lengths, whereas for values larger than one raises long branches more than short branches (Nunn 2011, lecture notes Isabella Capellini, AnthroTree Workshop 2011). Important to note is, that κ acts only on individual branches, but does not scale the phylogenetic vcv matrix. In evolutionary terms κ reveals whether the evolution of a certain trait is rather gradual or punctuational (lecture notes by Isabella Capellini, AnthroTree Workshop 2011).
These branch length transformation factors (λ, κ and δ) can be changed in order to modify and adjust the evolutionary model of certain traits. Also, the different types of trees used for the simulations in this study are constructed using the three described transformation parameters (λ, κ and δ) (see Material & Methods 2.1.1).
Although, λ can only be used to measure the phylogenetic signal of continuous traits assuming a Brownian motion model of evolution, it is often mistakenly applied on discrete data.
For discrete data, especially binary data, also several measurements for the estimation of a phylogenetic signal have been proposed (Abouheif 1999, Fritz and Purvis 2010, Ives and Garland 2010). Fritz and Purvis (2010) for instance estimate ancestral states and then use the scaled sum of the differences between each pair of sister clades as a

measurement for the phylogenetic signal in binary traits. The smaller this scaled sum, the more phylogenetically clumped the two states of a binary trait. Another measurement for a binary phylogenetic signal has been proposed by Ives and Garland (2010), which is explained in detail in the section of phylogenetic logistic regression (Section 1.3.3). For categorical data with more than two states, no elaborated method has been established yet.

1.3 Comparative methods

As in the former sections, comparative analyses were rather explained from an evolutionary biology perspective, this part now covers the mathematical description of the specific methods. However, before describing or using any phylogenetic methods, it is important to state some general assumptions about comparative methods. First, the phylogenetic relationships in a given tree are assumed to be correct, second, more closely related species show larger similarities in the analysed traits due to their shared evolutionary history (phylogenetic signal) and third, the assumption about the underlying evolutionary model must be correct (Hernández et al. 2013).

Four commonly used phylogenetic methods (phylogenetic generalised least-squares regression by Grafen (1989) and Martins and Hansen (1997), phylogenetic generalised estimating equations by Paradis and Claude (2002), phylogenetic logistic regression by Ives and Garland (2010), phylogenetic generalised linear mixed model by Hadfield and Nakagawa (2010)) designed for continuous as well as discrete data are discussed below. Each subsection covers one method including the model structure, parameter estimation, the underlying assumptions and some general notes.

1.3.1 Phylogenetic least-squares regression

Phylogenetic generalised least-squares regression (PGLS) (Grafen 1989, Martin and Hansen 1997, Pagel 1997; Garland and Ives 2000, Rohlf 2001) is probably the most famous and widely used approach in current comparative studies. The reason why it became so famous is its huge flexibility compared to older methods (e.g. independent contrasts, Felsenstein 1985). The PGLS was first proposed by Grafen (1989) as an extension of the independent contrast method (Felsenstein 1985). Grafen (1989) developed the idea of translating the phylogeny into a variance-covariance matrix and to account this way for the non-independence between the species. The PGLS method is implemented by the function `pgls()` in the package `caper` (Orme et al. 2012).

Model structure

As stated by Martins and Hansen 1997, an evolutionary hypothesis needs to be translated into a statistical model. In a PGLS model, the simplest linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ can be used to model different evolutionary processes such as the correlated evolution of two or more traits, rates of evolution, phylogenetic effects and reconstruction of ancestral states (Martins and Hansen 1997). Such a simple linear regression model consists basically of two parts: $\mathbf{X}\boldsymbol{\beta}$ is denoted as the mean structure and $\boldsymbol{\epsilon}$ as the error structure. The mean structure ($\mathbf{X}\boldsymbol{\beta}$) is used to determine what is modeled, either evolutionary rates, phylogenetic effects, the correlated evolution of two traits or ancestral states. For the purpose of this study, the focus is laid on the mean structure describing the correlated evolution of two or more traits (e.g. referring to the formerly used example: the correlated evolution between the length of development and brain size). The coefficients ($\boldsymbol{\beta}$: $(p+1) \times 1$ vector) of the linear regression model

describe the relationship between the response ($\mathbf{Y}$: $N \times 1$ vector with N observations) and the p explanatory variables ($\mathbf{X}$ = design matrix with the dimensions $N \times (p+1)$) and $\boldsymbol{\epsilon}$ represents the $N \times 1$ error structure:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{1.1}$$

The error structure ($\boldsymbol{\epsilon}$), is where the dependence of the data (species sharing evolutionary history) in a phylogenetic study is taken into account. The phylogeny is transformed into a variance-covariance matrix $\mathbf{V}$, as explained earlier, and is expressed in a complex error term.

Martins and Hansen (1997) divide the error structure into three sources of variation: 1) within-species variation and measurement error ($\boldsymbol{\epsilon}_m$); 2) error due to random evolutionary change of species along a phylogeny ($\boldsymbol{\epsilon}_s$); 3) error due to unknown or incomplete phylogenetic relationships ($\boldsymbol{\epsilon}_p$). Although it is possible to combine and include the three sources of errors into one error structure (see Martins and Hansen 1997), often only the data for the phylogenetic dependence between species is taken into account ($\boldsymbol{\epsilon}_s$). In fact, normally the variance of a trait within a species, which would allow to account for the measurement error ($\boldsymbol{\epsilon}_m$), is not given in the literature (if the variation or standard errors are given: package `phytools` in R allows to do a phylogenetic regression with intraspecific sampling error according to Ives et al. 2007).

The error structure based on phylogenetic relationships ($\boldsymbol{\epsilon}_s$) follows a multivariate normal distribution with an expectation of $\mathbf{0}$ and the phylogenetic variance-covariance matrix ($\mathbf{V}$) with an overall phylogenetic variance of σ^2 ($\boldsymbol{\epsilon}_s \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{V})$). The key issue of the PGLS approach is the phylogenetic variance-covariance matrix $\mathbf{V}$ allowing for huge flexibility concerning the model of evolution. This means, the matrix (i.e. the phylogenetic tree) can be adjusted based on a certain model of evolution by correspondingly adjusting branch lengths of the phylogenetic tree by one of the described branch length transformation parameters (λ, κ, δ) (e.g. Pagel 1999, Freckleton et al. 2002). The adjustments with these parameters enable to avoid over - or undercorrection of phylogenetic dependency between species (Lavin et al. 2008). For example, λ is used to scale the off-diagonal elements of the variance-covariance matrix of the phylogenetic tree as explained in the former Section 1.2.3.

Parameter Estimation

The estimation of the parameter and its variation is very simple and comes from the usual formulas of a least-squares regression (please note that the maximum likelihood estimation procedure results in the same parameter estimators (e.g. Pagel 1997)). The advantage of such estimators is that they are the best linear unbiased estimators having the smallest variance among all unbiased estimators (Gauss-Markov estimator) (e.g. Freckleton et al. 2011). The estimation of the GLS approach compared to the OLS estimation is down-weighted by the variance-covariance matrix, in order to correct for the dependence between the observations. $\hat{\boldsymbol{\beta}}$ is a vector ($(p+1) \times 1$) containing

the best unbiased linear estimators for the intercept und the coefficients explaining the relationship between the response variable and the p explanatory variables. $\mathbf{Y}$ is the vector $(N \times 1)$ of the response values for each species (number of species = N; $i = 1, 2, \ldots, N$) and $\mathbf{X}$ is the design matrix $(N \times (p+1))$ of the explanatory variables and the its rank (k). $\mathbf{V}$ $(N \times N)$ is the variance covariance matrix describing the phylogenetic tree. σ^2 and its unbiased estimator s^2 stand for the scaling constant (residual variance) (Rao and Toutenberg 1995, Blomberg et al. 2011), in comparative terms also known as the overall rate of evolution (Ives et al. 2007, Garland and Ives 2000). $\mathbf{W}$ is the variance-covariance matrix of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}. \tag{1.2}$$

$$s^2 = \frac{1}{N-k}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{1.3}$$

$$\hat{\mathbf{W}}(\hat{\boldsymbol{\beta}}) = s^2(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{1.4}$$

It is also possible to model non-Gaussian responses by extending the method to a GLM using some kind of link function which is not the identity function. The following iterative estimation procedure, also known as Fisher scoring, is used with the corresponding approximate estimate of the variation:

$$\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{\beta}}_0 + \left\{ \left(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}}\right)^T \mathbf{V}^{-1} \left(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}}\right) \right\}^{-1} \left\{ \left(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}}\right)^T \mathbf{V}^{-1}(\mathbf{Y} - \boldsymbol{\mu}_0) \right\}. \tag{1.5}$$

$$\hat{\mathbf{W}}(\hat{\boldsymbol{\beta}}_1) \approx \left(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}}\right)^T \mathbf{V}^{-1} \left(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}}\right)^{-1}. \tag{1.6}$$

The $\hat{\boldsymbol{\beta}}_0$ corresponds to an initial value of the coefficient calculated by a non-phylogenetic least-squares regression. This value is used to calculate $\hat{\boldsymbol{\beta}}_1$ in a second iterative step. $\boldsymbol{\mu}_0$ is an initial estimate of the mean ($\boldsymbol{\mu}_0 = g^{-1}(\mathbf{X}\boldsymbol{\beta}_0)$) and $d\boldsymbol{\mu}/d\boldsymbol{\beta}$ is a matrix containing the derivatives of the inverse link function with respect to $\boldsymbol{\beta}$ and evaluated at $\hat{\boldsymbol{\beta}}_0$ (Martins and Hansen 1997). The iterative process is continued until convergence of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$.

Further, to scale the variance-covariance matrix of the phylogenetic tree, Freckleton et al. (2002) presented a maximum likelihood estimation of λ given the data and the phylogenetic tree.
The model for the evolution of the response variable (Y_i) under Brownian motion is

given by the following formula with α standing for the state of the trait at time 0 (trait in the ancestor 0) and $\boldsymbol{\epsilon}$ representing a normally distributed random effect with a mean of 0 and variance of σ^2. t represents the time since the origin 0 and $\sum_{l=1}^{T_i}$ summates across the T branches with the lengths j from the root to the tip of species i:

$$Y_i = \alpha + \sum_{l=1}^{T_i} \epsilon_i \mathrm{t}_{ij}. \tag{1.7}$$

The multivariate normal probability density of the response variable $\mathbf{Y}$ incorporating the phylogenetic variance covariance matrix $\sigma^2\mathbf{V}$:

$$p(\mathbf{Y}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} |\mathbf{V}|^{\frac{-1}{2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\alpha)^T \boldsymbol{V}^{-1}(\mathbf{Y} - \mathbf{X}\alpha)\right]. \tag{1.8}$$

An estimation of α with the phylogenetic variance-covariance matrix scaled by λ is given by (analogue to the estimation of $\boldsymbol{\beta}$ in the PGLS estimation):

$$\hat{\alpha} = (\mathbf{X}^T\mathbf{V}(\lambda)^{-1}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{V}(\lambda)^{-1}\mathbf{Y}). \tag{1.9}$$

and the unbiased estimate of σ^2 is:

$$\hat{\sigma^2} = \frac{1}{N-1}(\mathbf{Y} - \hat{\alpha}\mathbf{X})^T\mathbf{V}(\lambda)^{-1}(\mathbf{Y} - \hat{\alpha}\mathbf{X}). \tag{1.10}$$

In the end, an estimate of λ is given by the maximum of the log-likelihood of the probability density (Equation 1.8) evaluated for a wide range of values for λ, scaling the variance-covariance matrix $\mathbf{V}$.
To test whether the estimate of λ is significantly different from 0 or 1, a log-likelihood ratio test is used with $L(\hat{\lambda})$ representing the log-likelihood at the maximum likelihood estimate of λ and $L(\lambda`)$ representing the log-likelihood at λ equal to 0 or 1. The result is approximately χ^2 distributed with 1 degree of freedom:

$$\chi^2 = -2[L(\hat{\lambda}) - L(\lambda`)]. \tag{1.11}$$

Assumptions

The pure statistical assumptions of a PGLS are the known from a common GLS model: 1) The GLS model assumes that the response variable ($\mathbf{Y}$) can be expressed by a linear combination of explanatory variables ($\mathbf{X}$) resulting from the " linear" term in a linear model. 2) The predictor variables (the columns of $\mathbf{X}$ excluding the first) are assumed to be independent of each other and of the error structure. 3) The errors are assumed to be multivariate normally distributed $\sim \text{MNV}(\mathbf{0}, \sigma^2\mathbf{V})$) with a mean of 0 and and a variance-covariance structure of $\sigma^2\mathbf{V}$ and $\mathbf{V} \neq \mathbf{I}$ ($\mathbf{V}$ is assumed to be known and σ^2 needs to be estimated). Thus, in contrast to an ordinary least-squares model (OLS), the assumption of uncorrelated errors and homoscedasticity is relaxed.

The assumptions from an evolutionary and phylogenetic perspective, some of them which are relaxed in a PGLS compared to IC approach, are the following: 1) Although, one would be able to incorporate the within-species variation or measurement error (Martins and Hanse 1997; Ives et al. 2007) in a PGLS model, standard program implementations do not. Specifically, the `pgls()` function in the package `caper` (Orme 2012) does not take into account a within-species variation. 2) Compared to the Independent Contrast method (Felsenstein 1985), the model of evolution in GLS approach is not strictly restricted to a Brownian motion model of random walk with $\lambda = 1$ (e.g. Martins and Hansen 1996). Using different branch length transformation factors, such as λ, δ or κ which can be estimated or fixed in the `pgls()` function in `caper`, one is able to adjust or rather force the evolutionary model to given data.

General notes

Compared to the old fashioned method of independent contrasts (Felsenstein 1985), it offers a much more flexible way of appliance in terms of nonstandard assumptions (correlation among data points), underlying models of evolution and intercepts which are not forced to be zero (Nunn 2011). Moreover, compared to Felsenstein's independent contrasts method (1985), the PGLS approach by Grafen (1989) can also make use of discrete explanatory variables. In fact, under the assumption of a phylogenetic signal, the PGLS method assumes only the errors to be multivariate normally distributed (and not the explanatory variables as in the IC method according to a Brownian motion model) and although a discrete or dichotomous variable cannot undergo Brownian motion evolution and do not follow a multivariate normal distribution, the error structure does (Martins and Hansen 1996).

Although, the mean structure of this method could theoretically be extended for special link functions (e.g. logit-link) so far no further specifications of an appropriate variance-covariance structure for discrete characters in the function `pgls()` in `caper` has been implemented. The `pgls()` function is only designed for continuous responses. However, many studies apply this function to discrete and ordinal data (Schuppli et al. 2012; Matthews et al. 2010). But, when using the PGLS approach including λ, κ, δ as a branch length transformation parameters, as implemented in `caper` (Freckleton et al. 2002), it is pretty straightforward why it should be applied only on continuous varying

traits. In fact, e.g. using λ assumes a Brownian motion model of evolution which is only applicable for continuous data (a Brownian motion model produces infinite variation which does not match the finite number of states of a discrete trait).

1.3.2 Phylogenetic generalised estimating equations

The generalised estimating equations (GEE) method is an extension of a generalised linear model (GLM) taking correlated data (repeated measures or clustered data) into account. Thus, the GEE approach can be used to model not only Gaussian responses, but also binomial, poisson and other distributions belonging to the exponential family. In contrast to the cluster or subject specific models (e.g. random effects model), the GEE approach estimates population-averaged parameters representing "the averaged effect of a unit change in the predictors for the whole population" (Ghisletta and Spini 2004, p. 423). On the contrary, cluster or subject specific approaches focus on the individual coefficients. In other words, the coefficient is represented by a distribution based on cluster or subject specific regression coefficients (Ghisletta and Spini 2004). However, for a comparative method taking phylogenetic relationships into account, the GEE approach seems to be perfectly appropriate as the main interest lies not on a species specific, but rather on an averaged effect within a certain taxa group. The phylogenetic generalised estimating equations approach is implemented by the function `compar.gee()` in `ape`.

Model structure in general (non-phylogenetic)

The GEE (Liang and Zeger 1986; Zeger and Liang 1986; Zorn 2001) approach is an extension of GLM and is based on the quasi-likelihood methods (Wedderburn 1974; Nelder and Wedderburn 1972; McCullagh and Nelder 1983, 1989). A quasi-likelihood approach is used in case one has an idea on the mean and variance structure of the data, but the distributional assumption of the response is unclear (lecture notes generalised linear regression models STA 406). In the quasi-likelihood method only the relationship between the expectation of the response variable and the predicting variables (i.e. the inverse of the link function, Equation 1.12) and between the mean and the variance (Equation 1.13) are required, in contrast to the full form of the distribution of the response variable (e.g. Zeger and Liang 1986). Specifically, in a quasi-likelihood approach the variance is expressed as a function of the expectation.
The notation for the following formulas is given by S as the number of units/subjects/clusters $s = \{1, 2, \ldots, S\}$, N as the number of time points $i = \{1, 2, \}$ representing the dependent measurements. Here, N is used for the number of time points because in the phlyogenetic case, this corresponds to the number of species. $\mathbf{Y}_s$ represents a column vector of the response variable, $\mathbf{Y}_s = [Y_{s1}, Y_{s2}, \ldots, Y_{sN}]$. $\mathbf{X}_s$ is the design matrix $(N \times (p+1))$ of p covariates for cluster s. $\boldsymbol{\beta}$ is the column vector $((p+1) \times 1)$ containing the regression coefficients for the covariates. ϕ represents the overdispersion parameter. $\mathbf{V}_s$ represents the variance-covariance matrix for cluster s and it is important to note, that this matrix is not equal to the phylogenetic variance-covariance matrix ($\mathbf{V}$) defined for the PGLS approach.

The simple quasi-likelihood approach assuming no correlated data in the observed response variable is specified by the following formulas.
The inverse of the link function g^{-1} describes the relationship between the expectation μ_s and the linear predictor $\mathbf{X}_s\boldsymbol{\beta}$:

$$\boldsymbol{\mu}_s = g^{-1}(\mathbf{X}_s\boldsymbol{\beta}). \tag{1.12}$$

Under the assumption of uncorrelated data (i.e. $N = 1$: independence between observations s), the variance V_s of $\mathbf{Y}_s$ is given by a function f of the expectation:

$$V_s = \frac{f(\mu_s)}{\phi}. \tag{1.13}$$

Specifically, in case of a normal distribution, $f(\mu_s) = 1$, in case of a binary response the variance function has the form $f(\mu_s) = \mu_s(1 - \mu_s)$ and for a poisson distribution the variance is equal to the expectation $f(\mu_s) = \mu_s$.

For correlated response values (i.e. $N > 1$), Liang and Zeger (1986) extended the quasi-likelihood approach. They specified the $\mathbf{V}_s$ in a new way, introducing a $N \times N$ correlation matrix $\mathbf{R}_s(\gamma)$:

$$\mathbf{V}_s = \frac{\mathbf{A}_s^{0.5}\mathbf{R}_s(\gamma)\mathbf{A}_s^{0.5}}{\phi}. \tag{1.14}$$

$\mathbf{A}_s$ represents a $N \times N$ diagonal matrix with $f(\boldsymbol{\mu}_s)$ (i.e. the variance of $\mathbf{Y}_s$ without the dispersion parameter ϕ) on the diagonals and $\mathbf{R}(\gamma)$ stands for the correlation matrix of $\mathbf{Y}_s$. This correlation matrix $\mathbf{R}_s(\gamma)$, the parameter γ, respectively, can be specified accordingly to a particular dependency structure (independent, exchangeable, autoregressive or unstructured correlation; for further details see Liang and Zeger 1986 or Zorn 2001) of the response values. This flexibility in the within-cluster correlation structure represents one of the main advantages of the GEE approach (e.g. Ghisletta and Spini 2004).

The quasi-likelihood score statistics for the p covariates represented by the derivatives of the log-likelihoods is given by:

$$U_p(\boldsymbol{\beta}) = \sum_{s=1}^{S}\Big(\frac{d\boldsymbol{\mu}_s}{d\boldsymbol{\beta}_S}\Big)^T\mathbf{V}_s^{-1}(\mathbf{Y}_s - \boldsymbol{\mu}_s) = 0. \tag{1.15}$$

All in all, the generalised estimating equation for $\boldsymbol{\beta}$ is a combination of the Equations 1.15 and 1.14.

Parameter estimation in general (non-phylogenetic)

The regression coefficients ($\hat{\boldsymbol{\beta}}$) and both, the correlation (γ) and dispersion parameter (ϕ) of this estimating equation can be found using an iterative process, such as Fisher scoring:

$$\boldsymbol{\beta}^{(\text{step}+1)} = \boldsymbol{\beta}^{(\text{step})} - \left\{ \Big(\frac{d\boldsymbol{\mu}_s}{d\boldsymbol{\beta}}\Big)^T \mathbf{V}_s^{-1} \Big(\frac{d\boldsymbol{\mu}_s}{d\boldsymbol{\beta}}\Big) \right\}^{-1} \left\{ \Big(\frac{d\boldsymbol{\mu}_s}{d\boldsymbol{\beta}}\Big)^T \mathbf{V}_s^{-1} (\mathbf{Y}_s - \boldsymbol{\mu}_s) \right\}. \tag{1.16}$$

The iterative process can be started with $\mathbf{R}_s(\gamma) = \mathbf{I}$ and $\phi = 1$ leading to a fist estimation of the coefficients $\boldsymbol{\beta}^{(\text{step})}$.The estimated $\boldsymbol{\beta}^{(\text{step})}$ can be used to calculate the fitted values $\boldsymbol{\mu}_s = g^{-1}(\mathbf{X}_s\boldsymbol{\beta})$ and subsequently for calculating the standardized residuals $\hat{r}_{si} = (Y_{si} - \mu_{si})/\sqrt{V_{ii}}$. These residuals are then required for the estimation of $\mathbf{A}_s$, $\mathbf{R}_s(\gamma)$, γ, respectively, and ϕ. The dispersion parameter ϕ can be estimated using the following formula:

$$\hat{\phi}^{-1} = \sum_{s=1}^{S} \sum_{i=1}^{N} \frac{\hat{r}_{si}^2}{(N - (p+1))}. \tag{1.17}$$

γ can either be estimated from the data by the specific estimators which vary depending on your dependence structure, more details are given in Liang and Zeger (1986, where α corresponds to γ) or can be fully specified. Given the estimates of $\mathbf{A}_s$, $\mathbf{R}_s(\gamma)$ and ϕ, $\boldsymbol{\beta}^{(\text{step}+1)}$ can be calculated in a second iterative step. The last two steps are iterated until convergence. The main advantage of the GEE approach is the consistent estimate of the regression coefficients $\boldsymbol{\beta}$ depending only on the mean structure and most importantly, being robust against misspecifications of the correlation matrix $\mathbf{R}_s(\gamma)$ (Liang and Zeger 1986).

For the sake of completeness, also the following formulas on the variance structures of $\boldsymbol{\beta}$ are given. A naive or "model" based estimate of the variance-covariance matrix $\mathbf{W}_{naive}$ of $\boldsymbol{\beta}$ is given by:

$$\hat{\mathbf{W}}_{naive}(\boldsymbol{\beta}) = \left\{ \Big(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}}\Big) \mathbf{V}^{-1} \Big(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}}\Big) \right\}^{-1}. \tag{1.18}$$

A robust estimate of $\mathbf{W}_{robust}$ of $\boldsymbol{\beta}$ is given by the "sandwich estimator":

$$\hat{\mathbf{W}}_{robust}(\boldsymbol{\beta}) = \left\{ \sum_{s=1}^{S} \Big(\frac{d\boldsymbol{\mu}_s}{d\boldsymbol{\beta}}\Big)^T \mathbf{V}_s^{-1} \Big(\frac{d\boldsymbol{\mu}_s}{d\boldsymbol{\beta}}\Big) \right\}^{-1} \left\{ \sum_{s=1}^{S} \Big(\frac{d\boldsymbol{\mu}_s}{d\boldsymbol{\beta}}\Big)^T \mathbf{V}_s^{-1} (\mathbf{Y}_s - \boldsymbol{\mu}_s)(\mathbf{Y}_s - \boldsymbol{\mu}_s)^T \mathbf{V}_s^{-1} \Big(\frac{d\boldsymbol{\mu}_s}{d\boldsymbol{\beta}}\Big) \right\} \left\{ \sum_{s=1}^{S} \Big(\frac{d\boldsymbol{\mu}_s}{d\boldsymbol{\beta}}\Big)^T \mathbf{V}_s^{-1} \Big(\frac{d\boldsymbol{\mu}_s}{d\boldsymbol{\beta}}\Big) \right\}^{-1}. \tag{1.19}$$

$\hat{\mathbf{W}}_{robust}(\boldsymbol{\beta})$ is also robust against misspecifications of $\mathbf{R}_s(\gamma)$. The middle term of this robust variance estimator serves as correction factor in case of a misspecified correlation structure (Norton et al. 1996). The standard error of $\hat{\boldsymbol{\beta}}$ is simply the square root of the diagonal elements of $\hat{\mathbf{W}}(\boldsymbol{\beta})$.

Phylogenetic GEE

In the case of a phylogenetic approach using GEE (PGEE) (Paradis and Claude 2002), only a single cluster is assumed with the phylogenetic tree representing that cluster (i.e. $S = 1$, model indices s drop out) and the number of time points (N) corresponding to the number of species (compare to section "Model structure in general (non-phylogenetic)"). Furthermore, $\mathbf{R}_s(\gamma)$ is the $N \times N$ correlation matrix containing the expected correlations of the response variable $\mathbf{Y}$ between the species based on their phylogenetic relationships. In other words, the correlations are based on the species shared path lengths on their phylogenetic tree.
With these specifications, $\boldsymbol{\beta}$ can be estimated in the way described above using the iterative process in Equation 1.16. However, for the estimation of the variance-covariance matrix of $\boldsymbol{\beta}$, Paradis and Claude (2002) suggested some additional specifications.

For a continuous response variable, Paradis and Claude (2002) use the naive estimator of $\mathbf{W}_{naive}(\boldsymbol{\beta})$ because the robust estimator $\hat{\mathbf{W}}_{robust}(\boldsymbol{\beta})$ is not good in case of a small number of independent clusters (note that using a phylogenetic approach there is only a single cluster) (Horton and Lipsitz 1999).
Although, this naive estimator of the variance-covariance matrix of $\boldsymbol{\beta}$ shows good properties for a continuous response, as shown by the simulation study by Paradis and claude 2002, it is not appropriate for discrete response variables (Mancl and DeRouen 2001). Thus, for binary response variable they chose a quasi-likelihood estimator of $\mathbf{W}_{quasi}(\boldsymbol{\beta})$ with Q representing the quasi-likelihood function:

$$\hat{\boldsymbol{W}}_{quasi}(\boldsymbol{\beta}) = \left\{ -\frac{d^2 \ln Q}{d\boldsymbol{\beta}^2} \right\}^{-1}. \tag{1.20}$$

The quasi-likelihood function (Wedderburn 1974) is given by:

$$\frac{dQ}{d\boldsymbol{\mu}} = \sum_{i=1}^{N} \frac{Y_i - \mu_i}{\phi \mu_i (1 - \mu_i)}. \tag{1.21}$$

Assumptions

The general assumptions for a GEE approach are that (1) the response variable should be able to be expressed by a linear combination of the explanatory variables (a specific link function needs to be determined if the response variable is not normally

distributed), (2) the number of clusters should be relatively large and (3) the observations between clusters should be uncorrelated (Norton et al. 1996). However, in the case of a phylogenetic GEE approach, there is only a single cluster, meaning that the second assumption given above does not hold anymore. To avoid any bad influence on the inference, Paradis and Claude (2002) suggest to use the naive estimator the variance-covariance matrix of $\boldsymbol{\beta}$.

General notes

Compared to the PGLS method, where additionally the phylogenetic signal of a trait is taken into account, this approach does not do that. Furthermore, as indicated by Ives and Garland (2010), the correlation structure $\mathbf{R}(\gamma)$ based on the phylogenetic tree is not appropriate for discrete traits as it assumes a Brownian motion model of evolution. However, the simulation study by Paradis and Claude (2002) showed that for binary traits, the type I error was not significantly larger than 5%, whereas for continuous traits this was the case. This means, that the GEE approach shows better properties for binary characters, however, one needs to keep in mind that the variance-covariance structure is not really appropriate.

1.3.3 Phylogenetic logistic regression

A very recent proposed approach is the phylogenetic logistic regression (PLR) for binary response variables by Ives and Garland (2010). This method is related to the GEE approach by Paradis and Claude (2002), but they adopted the correlation matrix especially for binary traits. Additionally, this approach is able to give an estimation of the phylogenetic signal of a binary trait simultaneously to the estimation of other parameters, giving an advantage over the method of the phylogenetic GEE by Paradis and Claude (2002) (Ives and Garland 2010). This method is implemented in MATLAB but also very recently, it has been implemented in `R` with the function `phyloglm()` in the package `phylolm` by Ho and Ané (in review).

Model structure

The PLR approach by Ives and Garland (2010) is basically splitted into two parts: First, an univariate case which is used to measure the phylogenetic signal of a binary trait (i.e. no explanatory variables), and second, the approach is extended for the case where one or more explanatory variables are modeled to explain the variation in the binary response (multivariate case). In the univariate case the correlation of the binary response between species depends on the transition rates between the two states, whereas in a multivariate case, the expectation ($\boldsymbol{\mu}$) depends on the explanatory variables.

For a binary response variable (taking either state 0 or 1), the correlation structure specified as in the GEE approach ($\mathbf{R}(\gamma)$) by Paradis and Claude (2002) based on the assumption of a Brownian motion model of evolution is not appropriate (Ives and Garland 2010). For the PLR, Ives and Garland (2010) suggest a specific formulation for the correlation structure ($\mathbf{C}(\alpha)$) for a binary response based on a Markov process (see Equations 1.22 and 1.23). For notation, $\mathbf{Y}$ represents the vector of response variable $\in \{0, 1\}$ for N species, $i = \{1, 2, \ldots, N\}$; $\mathbf{V}$ is basically the variance-covariance matrix representing the phylogenetic tree as in the PGLS approach, however, with all diagonal elements set to one (under the assumption of an ultrametric tree, meaning that the tips of the tree are equally distant from the root). Important to note is that this matrix is not the same as the correlation matrix $\mathbf{R}(\gamma)$ used in the phylogenetic GEE approach. $\boldsymbol{J}$ is a $N \times N$ matrix with all elements equal to 1. α corresponds to the sum of the transition rates from state 1 to 0 (α_0) and the transition rate for the opposite direction (0 to 1) α_1 ($\alpha = \alpha_0 + \alpha_1$).

For the *univariate case* (i.e. phylogenetic signal) the following formula, where the exponential refers to an element wise operation, gives the correlation structure of $\mathbf{Y}$ with the matrix $2(\boldsymbol{J} - \mathbf{V})$ representing the pairwise distances of two species:

$$\mathbf{C}(\alpha) = \exp(-2\alpha(\boldsymbol{J} - \mathbf{V})). \tag{1.22}$$

The higher the transitions rates α_0 and α_1, the higher α leading to lower off-diagonal elements in $\mathbf{C}(\alpha)$. From an evolutionary intuitive perspective, this means that higher transition rates result in lower phylogenetic correlations. This is why α can be regarded as a measurement for the phylogenetic signal for a binary trait. For very high values of α_0 and α_1, the probability of state 1 approaches $\mu = \alpha_1/(\alpha_0 + \alpha_1)$.
Generally, the matrix $\mathbf{C}(\alpha)$ represents an analogue to the variance-covariance matrix $\mathbf{V}$ used for a continuous response under the assumption of a Ornstein-Uhlenbeck model of evolution (e.g. PGLS). In the case where α equals to 1, Ives and Garland (2010) compare $\mathbf{C}(\alpha)$ to $\mathbf{V}$ for a continuous response under the assumption of a Brownian motion model, where the degree of change is proportional to the length of a branch. In other words, α equal to 1 for a binary trait is comparable to the continuous case whit a λ of 1. However, $\mathbf{V}$ and $\mathbf{C}(\alpha)$ are of course never identical. Moreover, instead of using α directly as a measurement for the phylogenetic signal, Ives and Garland propose to use a with $a = -\log(\alpha)$. This leads to a more intuitive interpretation of the phylogenetic signal as with increasing a the phylogenetic correlation of the response variable between species increases. They use $a = -4$ as a cut point: everything smaller than that indicates the absence of a phylogenetic signal, as then the $\mathbf{C}(\alpha)$ is basically equivalent to the identity matrix $\mathbf{I}$.

For the *multivariate case*, the correlation structure is the following with $\mathbf{M}$ as a diagonal matrix with the components $m_{ii} = (1 - \bar{\mu})[\mu_i/(1 - \mu_i)]^{0.5}$ for $\mu_i < \bar{\mu}$ and $m_{ii} = \bar{\mu}[(1 - \mu_i)/\mu_i]^{0.5}$ for $\mu_i > \bar{\mu}$ with μ_i representing the expected probability of state 1 for species i and $\bar{\mu}$ representing the average of the expectations of all species (expectation see formula 1.24) (for further details see Ives and Garland 2010):

$$\widetilde{\mathbf{C}}(\alpha) = \mathbf{MC}(\alpha)\mathbf{M} - \mathrm{diag}(\mathbf{MC}(\alpha)\mathbf{M}) + \mathbf{I}. \tag{1.23}$$

In the case of independent data points, the analysis corresponds to a common logistic regression where $\widetilde{\mathbf{C}}(\alpha)$ equals to the identity matrix $\mathbf{I}$.

In the end, there are two parameters which need to be estimated: α as a measurement of the phylogenetic signal using the univariate approach and $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ (see Equations 1.24), respectively, using the multivariate approach.

Parameter estimation

Ives and Garland (2010) propose to use alternately a quasi-likelihood function to estimate $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, respectively, given α, and least-squares estimation for estimating α given $\boldsymbol{\mu}$ until convergence of both parameters. To find the MLE of $\boldsymbol{\mu}$, they do not use the Newton-Raphson algorithm as in the GEE approach, however, they use another optimizing algorithm, the simplex minimization procedure, with higher flexibility and efficiency.
For estimating the phylogenetic signal (i.e. univariate case), $\mathbf{X}$ is a $N \times 1$ vector of

ones, as there are no covariates, and the expectation of state 1 is a scalar μ. In the case of a multivariate analysis, $\mathbf{X}$ represents the $N \times (p+1)$ design matrix (see Equations 1.24) and $\boldsymbol{\mu}$ is a $N \times 1$ column vector.
The formulas for the parameter estimation are given below, however, those are not explained in detail as this is beyond the scope of this study. Further and more detailed information can be found in Ives and Garland (2010).

The following two formulas are the logit function and and its inverse representing the asymptotic expectation of the probability of state 1 of Y_i ($\boldsymbol{\mu}$ in bold font represents the $N \times 1$ vector containing the expectations μ_i for each species i for the multivariate case, and the logarithm as well as the exponential are to be understood as element wise operators):

$$g(\boldsymbol{\mu}) = \log \frac{\boldsymbol{\mu}}{\mathbf{1} - \boldsymbol{\mu}} \qquad \boldsymbol{\mu} = \frac{\exp(\mathbf{X}\boldsymbol{\beta})}{\mathbf{1} + \exp(\mathbf{X}\boldsymbol{\beta})}. \tag{1.24}$$

The variance matrix $\mathbf{V}(\alpha)$ of $\mathbf{Y}$ with $\mathbf{C}(\alpha)$ (univariate case) or $\widetilde{\mathbf{C}}(\alpha)$ (multivariate case) as the correlation matrix and the diagonal matrix $\mathbf{A}$ known from the GEE approach is given by:

$$\mathbf{V}(\alpha) = \mathbf{A}^{0.5}\widetilde{\mathbf{C}}(\alpha)\mathbf{A}^{0.5}. \tag{1.25}$$

The quasi-likelihood score function for estimation of $\boldsymbol{\beta}$ is given by (McCullagh and Nelder 1989, p. 333):

$$\mathrm{U}(\hat{\boldsymbol{\beta}}(\alpha)|\alpha) = \sum_{p+1} \Big\{ (\mathbf{AX})^T \mathbf{V}(\alpha)^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \Big\} = 0. \tag{1.26}$$

As it is well known that the estimates of $\boldsymbol{\beta}$ in logistic regressions are biased, Ives and Garland (2010) use a penalized equation for the estimation of $\boldsymbol{\beta}$ (penalization procedure by Firth 1993; see also Heinze and Schemper 2002), where the information matrix $\boldsymbol{I}(\boldsymbol{\beta})$ is given by $(\mathbf{AX})^T\mathbf{V}(\alpha)^{-1}(\mathbf{AX}) = \mathbf{X}^T\mathbf{A}^{0.5}\mathbf{C}(\alpha)^{-1}\mathbf{A}^{0.5}\mathbf{X}$ and β_r is the r-th regression coefficient of $\boldsymbol{\beta}$:

$$\mathrm{U}^*_r(\hat{\boldsymbol{\beta}}(\alpha)|\alpha) = \mathrm{U}_r(\hat{\boldsymbol{\beta}}(\alpha)|\alpha) + \frac{1}{2}\,\mathrm{tr}\left\{ \boldsymbol{I}(\beta_r)^{-1} \Big[\frac{d\boldsymbol{I}(\beta_r)}{d\beta_r} \Big] \right\} = 0. \tag{1.27}$$

For estimating α the generalised least-squares formula is used:

$$\mathrm{SS}(\hat{\alpha}(\boldsymbol{\mu})|\boldsymbol{\mu}) = -\frac{1}{2}(\log|\mathbf{V}(\alpha)| + (\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{V}(\alpha)^{-1}(\mathbf{Y}-\boldsymbol{\mu})). \tag{1.28}$$

The variance of $\boldsymbol{\beta}$ is estimated by the naive estimator also used in the GEE approach, except that now the variance $\mathbf{W}(\boldsymbol{\beta})$ is based on the fixed MLE of α:

$$\hat{\boldsymbol{W}}_{naive}(\hat{\boldsymbol{\beta}}|\hat{\alpha}) = \Big(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}}\Big)^T\mathbf{V}(\hat{\alpha})^{-1}\Big(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\beta}}\Big). \tag{1.29}$$

Assumptions

Generally, the ordinary logistic regression is part of generalised linear models and thus, it assumes no normality, no homoscedasticity and no linear relationship between the response and the explanatory variables. However, the response variable needs to be binary and the log odds should be linearly related to the explanatory variables.
Specifically for the phylogenetic case, the PLR method by Ives and Garland (2010) assumes the process to be at stationarity, meaning that the probability for state 1 is the same over the whole tree, on the tips as well as on the root of the tree (Ives and Garland 2010). Another method for measuring phylogenetic signal in binary traits has been suggested by Fritz and Purvis (2010). This method is based on the sum of sister-clade (two direct related phylogenetic groups such as the birds and crocodiles) differences and the assumption about stationarity is relaxed.

General notes

The estimate of a as a measurement for the phylogenetic signal is only precise (i.e. showing low variability) and unbiased if the binary data shows a balanced number of zeros and ones. As soon as there are only very few zeros and or very few ones, the estimate of the phylogenetic signal shows rather poor properties (i.e. low precision and downward bias) (see Figures 1 and 2 in Ives and Garland 2010). In real data sets, however, this kind of unbalanced data is rather common and thus, one needs to be especially careful using this kind of phylogenetic signal for binary traits. Moreover, in the multivariate case, the simulations of Ives and Garland (2010) show that the estimates ($\boldsymbol{\beta}$) are upward biased, with the bias being stronger in case of higher phylogenetic signals (a). However, they still perform better than an ordinary logistic regression where dependence between species is not taken into account. Parametric bootstrapping in the MATLAB code allows for detecting possible bias in the estimators.

1.3.4 Phylogenetic generalised linear mixed model

The phylogenetic generalised linear mixed model (PGLMM), which is based on quantitative genetic methods, can be applied for a wide range of phylogenetic questions and is able to analyse non-Gaussian and discrete dependent variables. The phylogenetic covariance matrix, the analogue to the relationship matrix of a pedigree in quantitative genetics, is included as a random effect in the mixed model. This method makes use of Bayesian inference using the Markov Chain Monte Carlo (MCMC) technique (Hadfield and Nakagawa 2010). The phylogenetic mixed model is implemented with the function `MCMCglmm` in the equally named package `MCMCglmm`.

Model structure

The phylogenetic mixed model is based on a quantitative genetic method, the so-called animal model. This model is used to distinguish between genetic and environmental effects on a certain phenotype, including all the relationships within a pedigree. An animal model is nothing else than a mixed model including a fixed effect, representing an overall mean (μ), and two random effects, one for the genetic additive effect (heritability resulting from a pedigree) ($\boldsymbol{a}$) and one for the environmental effect ($\boldsymbol{e}$) (Lynch 1991, Postma and Charmantier 2007). In case of a single character, μ is a scalar and $\mathbf{Y}$, $\boldsymbol{a}$, $\boldsymbol{e}$ and $\mathbf{1}$ are all $N \times 1$ column vectors with N representing the number of individuals in a pedigree:

$$\mathbf{Y} = \mu\mathbf{1} + \boldsymbol{a} + \boldsymbol{e}. \tag{1.30}$$

This model built the basis for the phylogenetic mixed model where a phylogenetic tree with N species is the analogue to a pedigree (Lynch 1991, Housworth et al. 2004, Hadfield and Nakagawa 2010). Lynch (1991) and Housworth et al. (2004) describe the phylogenetic mixed model as a way to model two different kinds of evolutionary changes: gradual, long-lasting evolutionary change as an analogue to the heritability in quantitative genetics ($\boldsymbol{a}$), and fast, reversible evolutionary change ($\boldsymbol{e}$), such as phenotypic plasticity, as an analogue to the environmental effects (Housworth et al. 2004). A very recent publication by Hadfield and Nakagawa (2010) further emphasizes the flexibility of the phylogenetic mixed model by making the link to meta-analysis. They demonstrate that many proposed model designs (the original phylogenetic meta-analysis by Adams (2008), the method allowing for additionally including within-species variation by Ives et al. 2007 and the method proposed by Felsenstein (2008) considering multiple measurements per species) can be applied as a mixed model using ASReml (Gilmour et al. 2009), a program made for analysing animal models.

The main advantage of this model framework is its huge flexibility as it allows to include any number of fixed and random effects and even makes it possible to model multiple responses (Hadfield and Nakagawa 2010). The basic model, as described by Hadfield (2010, 2012) and Hadfield and Nakagawa (2010), consists of $\mathbf{X}$ as the design matrix of the fixed effects and $\mathbf{Z}$ as the design matrix for the random effects. $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are vectors containing the fixed effects and random effects, respectively. $\boldsymbol{\epsilon}$ stands for the residuals. For the non-Gaussian case a latent variable ł is introduced, the response

function (inverse of link function) of which is equal to the canonical parameter of a specific distribution (e.g. Poisson distribution: $Y_i \sim Pois(\lambda = exp(l_i))$).
The basic model for a Gaussian and non-Gaussian response is given by:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{\epsilon} \qquad \mathbf{l} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{\epsilon}. \tag{1.31}$$

The fixed and random effects and the residuals follow a multivariate normal distribution with the mean vectors of $\boldsymbol{\beta}_0$ and $\mathbf{0}$. The overall variance-covariance matrix contains the variance-covariance matrix of the fixed effects $\mathbf{B}$, the variance-covariance matrix of the random effects $\sigma_u^2\mathbf{G}$ and the variance-covariance matrix of the residuals $\sigma_\epsilon^2\mathbf{I}$. In the phylogenetic case, where the phylogenetic relationships are included as a random effect, $\mathbf{G}$ corresponds to $\mathbf{V}$, the phylogenetic $N \times N$ variance-covariance matrix. The variance matrix of the residuals is the identity matrix $\mathbf{I}$, assuming the residuals to be independent and homoscedastic.

$$\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{B} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2\mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\epsilon^2\mathbf{I} \end{bmatrix} \right). \tag{1.32}$$

The latent variables, the fixed and random effects as well as the variances σ_u^2 and σ_ϵ^2 need to be estimated. In the package `MCMCglmm` a multivariate normal prior for the fixed effects is assumed, where you need to specify the mean vector (mu) and a variance-covariance matrix ($\mathbf{V}$, note: has nothing to do with the phylogenetic vcv matrix). For the variance components (σ_u^2 and σ_ϵ^2) an inverse-Whishart prior with the parameter V and the degree of belief parameter nu is assumed. Thus, for the function `MCMCglmm`, one needs to specify the priors for the fixed effects, and the variance components σ_a^2 and σ_ϵ^2. The default prior for the fixed effects is a multivariate normal distribution with a zero mean and large variances (10^8) and covariances equal to zero. For the inverse-Wishart priors for the variance components of the random effects and the residuals, the default is a nu of zero and V of 1. Further information about the prior specifications is found in the section about method specifications.

A special and useful application of the phylogenetic mixed model proposed by Hadfield and Nakagawa (2010) is to model multinomial logit models for nominal responses with more than two levels. To keep it short and simple, the idea behind the multinomial model is a parameter reduction. In other words, if the multinomial variable has J levels, the parameters are reduced to $J-1$ by using one level as a reference category. The parameters are represented by log odds ratios (each of the $J-1$ levels in relation to the reference level) (Equation 1.33). The corresponding matrix, known as a contrast matrix $\boldsymbol{\Delta}$, with rows representing the multinomial levels and the columns the $J-1$ log odds ratios (latent variables l_{ij}) looks as follows (example: multinomial variable with three levels 1, 2 and 3 and α_{ij} as the probability for level j of species i):

$$l_{ij} = \log\left(\frac{\alpha_{ij}}{\alpha_{i1}}\right) \qquad \boldsymbol{\Delta} = \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{1.33}$$

A simple model with fixed effects:

$$\exp((\boldsymbol{\Delta}\boldsymbol{\Delta}^T)^{-1}\boldsymbol{\Delta}\mathbf{X}_i\boldsymbol{\beta}) \propto \mathrm{E}\begin{bmatrix}\alpha_{i1}\\ \alpha_{i2}\\ \alpha_{i3}\end{bmatrix}. \tag{1.34}$$

The estimation of variance covariance matrices of the random effects and the residuals is given by $\mathbf{V} = \boldsymbol{\Delta}^T\tilde{\mathbf{V}}\boldsymbol{\Delta}$ where the tilde marked $\mathbf{V}$ represents the variance covariance matrix of the log probabilities ($\log(\alpha_i)$). As the residuals are not estimated in case of multinomial data, the variance-covariance matrix needs to be fixed, basically arbitrarily.

Parameter estimation

The original phylogenetic mixed model proposed by Lynch (1991) uses the iterative expectation-maximization (EM) algorithm and maximum likelihood for parameter estimation, however, this procedure was weigh too computer intensive and for a small number of species problems with convergence arise (Lynch 1991). Moreover, Housworth et al. (2004) point out that small sample sizes, which is often the case in phylogenetic analyses compared to pedigree analyses, may lead to negative variance estimates of σ_u^2 (" genetic additive effect") and σ_ϵ^2 (" environmental effect") due to no mathematical constraints. Thus, Housworth et al. (2004) reparametrized (taking the two variance components together as a total variance: $\sigma^2 = \sigma_u^2 + \sigma_\epsilon^2$) the model of Lynch (1991) in order to overcome this problem and use maximum likelihood and restricted maximum likelihood (REML) estimation procedures. However, due to complexity of the likelihood, these estimation procedures are basically only reliable for Gaussian response variables (Hadfield and Nakagawa 2010). In form of a generalised linear mixed model, Hadfield and Nakagawa (2010) propose a Bayesian approach, the MCMC technique, to estimate parameters of non-Gaussian response variables. MCMC is a class of algorithms, often used in Bayesian inference, drawing random numbers out of probability density functions. In case of the PGLMM, two such algorithms are used: Metropolis-Hastings-algorithm and the Gibbs sampler (see Appendix A1).

As known from the previous section, the following parameters need to be estimated: the latent variables , the fixed and random effects and the variances σ_u^2 and σ_ϵ^2. The latent variables are sampled in blocks using the Metropolis-Hastings algorithm (see Box 1) because the conditional posterior is unrecognizable. The fixed and random effects, and the variances σ_u^2 and σ_ϵ^2 are sampled using the Gibbs-sampler (see Box2) and follow a multivariate normal and a scaled inverted chi-squared distribution, respectively. Further details are found in the appendix of Hadfield (2010) and in Hadfield (2012).

Assumptions

For continuous data, it assumes a Brownian motion model of evolution (Martins et al. 2002). However, as one can also model discrete data using the MCMC technique, which obviously does not follow Brownian motion, this assumption is relaxed for discrete data.

General notes

Using the Bayesian approach for estimation of a phylogenetic mixed model, one needs to be aware of the fact that the priors have higher influence on the posterior if the sample size is small, i.e. small number of species. Thus, having a small sample size, one really needs to make sure to specify appropriate prior distributions (Hadfield and Nakagawa 2010).

The main interest in this study is whether the mixed model approach by Hadfield and Nakagawa (2010) using Bayesian inference gives similar results to PGLS, PGEE and PLG (in this study ignoring intraspecific variation). Housworth et al. (2007) compared their mixed model (using likelihood inference) to PGLS and noted that whether to use PGLS or the mixed model approach basically depends on your evolutionary assumptions. For example, if a trait is assumed to show high phenotypic plasticity, then a mixed model would be more appropriate, whereas the PGLS approach is probably more useful if the trait values of the species vary around a single optimum.

Furthermore, some considerations about the interpretation of the effects of a phylogenetic mixed models are important as they might considerably differ from the interpretations of an animal model. First, the phylogenetic signal in a mixed model approach is represented by the pyhlogenetic heritability from quantitative genetics ($H^2 = \sigma_u^2/(\sigma_u^2 + \sigma_\epsilon^2)$). This is the analogue measurement to λ and also ranges between zero and one (Hadfield and Nakagawa 2010). However, compared to the PGLS approach, the estimated phylogenetic signal is not used to scale/adjust the phylogenetic vcv matrix. Second, the overall mean (μ) in the phylogenetic mixed model basically represents the ancestral state of a trait of a phylogeny. Third, the additive phylogenetic effect (a) represents that part of a trait which is phylogenetically inherited from the clade, but also includes non-genetic adaptations due to a shared environment. And finally, the non-genetic, environmental effects (e) might be also either of genetic or non-genetic nature, however, considered on a much shorter time scale (Housworth et al. 2007).

1.4 Aims and questions

The aim of this study is to compare four different comparative methods (described above: PGLS (Martins and Hansen 1997); PGEE (Paradis and Claude 2002); PLR (Ives and Garland 2010); PGLMM (Hadfield and Nakagawa 2009)) measuring the correlated evolution between two traits with the focus directed on categorical response variables (i.e. ordinal and binary). Besides categorical response data, the methods are also applied to continuous data. Although, individually, these methods have been compared mostly to the original independent contrast (Felsenstein 1985) method as well as to PGLS, an overall comparison, focused on categorical responses, has not been done yet. Moreover, several studies have compared the statistical abilities of different phylogenetic methods, however, mostly using continuous traits (Martins and Garland 1991; Diaz-Uriarte and Garland 1996; Martins et al. 2002).

With the help of simulations the statistical performances and adequateness of these methods are compared and verified. In that respect, the reliability and bias of the estimates is of special interest. Further, this study aims to test how sensible the performances of the different methods are in respect to different scenarios in terms of sample size, tree structure and the strength of correlation. Moreover, comparing those models, it needs to be tested, whether they end up in the same or at least similar results applying them to a real data set. With a data set of 78 primate species, containing data about niche complexity (ordinal and binary response variables) and life history, these models are applied and compared on a real evolutionary correlation question.

In a broader context, the link between the mathematical estimation procedures, also in terms of how phylogenetic relationships are taken into account, and the evolutionary interpretations of results is of interest.

2 Material & Methods

With simulating phylogenetically dependent data and their correlations based on varying parameters (number of species, phylogenetic tree structure and strength of correlation), the performances of the four different phylogenetic comparative methods, always in comparison to the corresponding non-phylogenetic method, were investigated.
It is important to note that the following simulations and analyses are based on a few phylogenetic assumptions. First, it is always assumed that the phylogenetic relationships between species are known, and thus, possible uncertainties of phylogenetic relationships are not taken into account. Third, the phylogeny is assumed to be based on characters others than those studied for phylogenetic correlations (Martins and Garland 1991).

2.1 Simulation

2.1.1 Simulation setup

The main aim of this study is to test how well different phylogenetic methods perform on different types of data. The uncorrelated/correlated evolution of two or three traits was simulated (response and one or two explanatory variables) along a phylogeny and subsequently the different non-phylogenetic and phylogenetic methods were applied to the simulated data. To test the flexibility, sensitivity and robustness of those methods, the simulations were run with different varying parameters including the number of species, the structure of the phylogenetic tree and the strength of the correlation. All these parameters might have an impact on the statistical abilities of the different phylogenetic methods.
The simulation loops, including the data simulation and subsequent analyses using the four phylogenetic methods, for the continuous, ordinal and binary response implemented in `R` are found in the Appendix B1 to B3.

The *number of species* were simulated to vary between 20, 50 and 100 species, which are very realistic and plausible numbers in comparative studies. Below 20 species the power for applying any phylogenetic analyses is probably too low.
Using these three different numbers of species, four different structures of a *phylogenetic tree* were generated using the functions `sim.bdtree()` and `transform()` in `geiger` (Harmon et al. 2008) (Figure 2.1). Tree 1 represents a random ultrametric tree with ultrametric meaning that all tips of a tree are equally distant from the root (tree 1 to 3 in Figure 2.1 are ultrametric). Tree 1 is also the basis for the transformations of tree 2 to 4. Tree 2 simulates a macroevolutionary pattern where most diversifications occur very early in the evolutionary history (λ chosen to be very small, 0.1, reducing the length of the shared branches), whereas the third type of tree represents the opposite pattern, the largest part of the diversification occurring late in the evolutionary history of a clade (branch length transformation factor δ set to 0.1

which means that shorter paths contribute more to the trait evolution). The last tree was simulated to have all branch-lengths equal to one (branch length transformation factor κ set to zero, meaning that trait evolution is independent of branch lengths) resulting in a non-ultrametric tree. This scenario stands for cases where the actual branch lengths are not known and thus, all set to one. The four types of trees are shown in Figure 2.1.

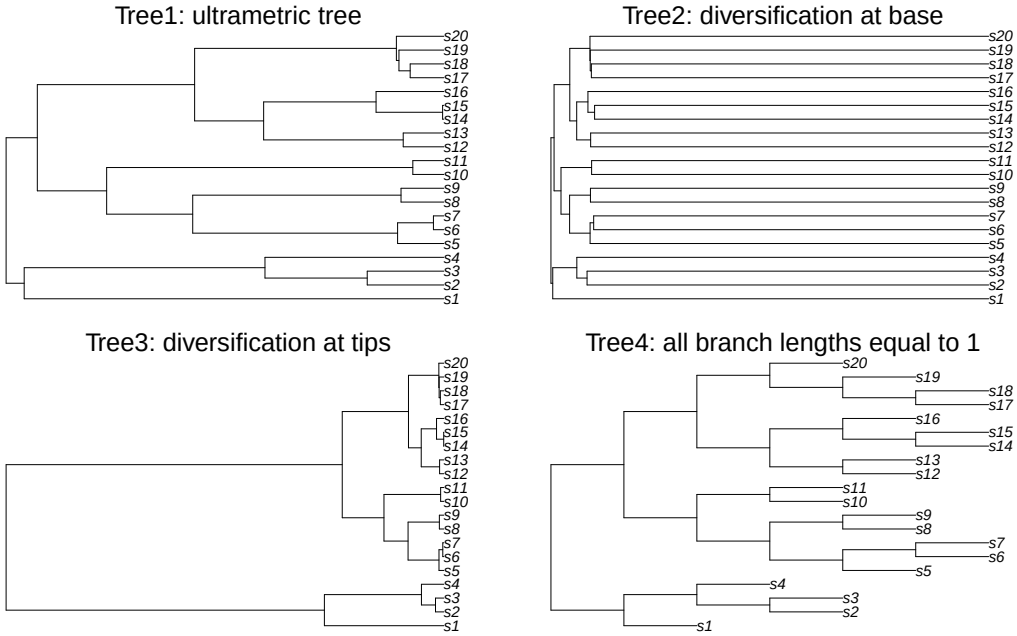


Figure 2.1: **Phylogenetic tree types for simulation setup.** The four different types of phylogenetic trees used in the simulation setup here shown for 20 species: tree 1 is an ultrametric random generated tree; tree 2 represents a transformation of the phylogenetic tree 1 were most of the divergences occur early in the tree (similar to a star phylogeny) with λ=0.1; tree 3 represents the opposite of a star phylogeny, where most of the divergences occur late in the tree (transformation of tree 1 with δ=0.1) and in tree 4 all branch lengths are equal to one (tree 1 transformed with κ=0). For the tree transformations the function `transform()` in `geiger` was used.

Another source of variation is the *strength of correlation*. Each method was applied to two simple models: an univariate model with a single continuous explanatory variable (`Y ~X1`) and a multivariate model with a continuous and a binary predictor variable (`Y ~X1 + X2`). The three traits (one response and two explanatory variables) were simulated to correlate either weakly ($r = 0.01$), moderately ($r = 0.5$) or strongly ($r = 0.9$) with each other using the function `sim.char()` in `geiger`. The correlation coefficients r of 0.01 and 0.9 were chosen in order to calculate the type I error rates and the power of the statistical methods. The correlation coefficient for simulating a strong association was chosen to be 0.9 and not higher in order to avoid a high occurrence of the perfect fit problem. A perfect fit, for instance in case of a binary response, means that all cases ($Y = 1$) of the binary response have higher values in

the explanatory variable than the non-cases ($Y = 0$) or vice versa. This might cause numerical problems especially in case of logistic regression models.
For the `sim.char()` function, if simulating two or more correlated traits, a variance-covariance matrix needs to be specified. The response variable (`Y`) and each of the two explanatory variables (`X1` and `X2`) were simulated to be correlated with the three correlation coefficients mentioned above (variance-covariance matrices shown in Tables 2.1 and 2.2). The two explanatory variables among themselves were simulated not to be correlated (see Table 2.1). However, for a correlation coefficient of $r = 0.9$, the covariances between `X1` and `X2` had to be set to 0.65, the smallest covariance still resulting in a positive definite matrix (see Table 2.2). In fact, `sim.char()` accepts only positive definite matrices.
All in all, this leads to 36 possible parameter combinations (3 numbers of species $\times$ 4 trees $\times$ 3 correlation coefficients).

Table 2.1: **Simulation of correlated traits.** Vcv implemented in the `sim.char()` function for $r = 0.01$ and $r = 0.5$ with `Y` as the response variable and `X1` and `X2` as the two explanatory variables.

	Y	**X1**	**X2**
Y	1	r	r
X1	r	1	0
X2	r	0	1

Table 2.2: **Simulation of correlated traits.** Vcv implemented in the `sim.char()` function for $r = 0.9$ with `Y` as the response variable and `X1` and `X2` as the two explanatory variables.

	Y	**X1**	**X2**
Y	1	r	r
X1	r	1	0.65
X2	r	0.65	1

The *response variables* are either continuous, ordinal (with four levels) or binary. The ordinal and binary response as well as the binary explanatory variable were generated based on the simulated continuous variable. In other words, the simulated continuous response was transformed into an ordinal and binary response variable using quarter quantiles and the mean, respectively, as cut points. More specifically, the ordinal response variable has four different levels, each level corresponding to a quarter cut of the originally continuous variable (`cut()` function). The analogue holds for the binary response and explanatory variable, where the cut point for 0 and 1 is the mean of the simulated continuous variable. A small simulation loop shows that transforming a continuous character into an ordinal or binary character does not affect the input correlation coefficient initially set between two continuous characters. In other words, when looking at the correlation between the continuous and both, the generated ordinal and binary character, the Spearmen correlation coefficients approximate the input correlation coefficients quite well (see Figure 2.2 with an input correlation coefficient of 0.5). This shows that transforming a continuous variable into an ordinal or binary character does not affect the input correlation coefficient between two continuous variables legitimating the way of simulating the correlated evolution between a categorical and continuous character along a phylogeny.

Further, the function `sim.char()` allows to select a specific model of evolution including Brownian motion and speciational models. The simulations of this study follow Brownian motion (i.e. $\lambda = 1$) resulting in multivariate normally distributed response and explanatory variables with the phylogenetic covariance matrix corresponding to the tree structure.

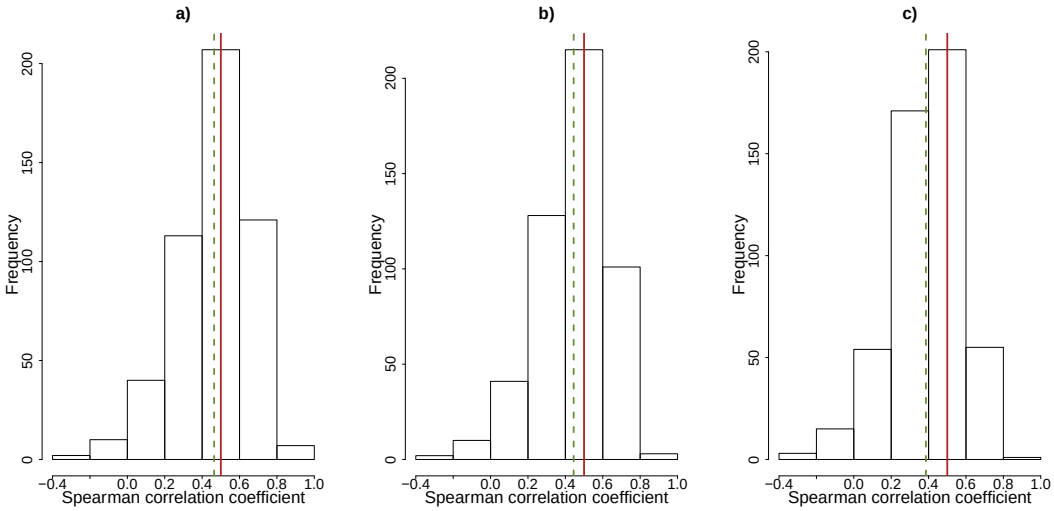


Figure 2.2: **Spearman correlation.** The distributions of 500 Spearman correlation coefficients. Two continuous characters are simulated 500 times with `sim.char()` using an input correlation coefficient of $r = 0.5$. Out of one of these characters an ordinal and a binary variable is generated as described in the text. For the 500 simulations, the Spearman correlation coefficients between the two continuous (a) and the continuous and both, the ordinal (b) and the binary (c) variable are calculated. The dashed green vertical line indicates the empirical mean, and the red solid line the input correlation of $r = 0.5$ of the simulated data.

With the three different types of response variables, continuous, ordinal or binary, the simulation setup ends up in a total of 108 possible combinations. In other words, for a single simulation, there are 108 differently generated data sets (one data set contains five columns: continuous, ordinal and binary response variable and a continuous and binary predictor variable). In order to analyse the statistical properties of the phylogenetic methods, each these 108 possible data sets were simulated a 1000 times, resulting in a total of 180‘000 simulated data sets. The simulation setup is graphically illustrated in Figure 2.3.

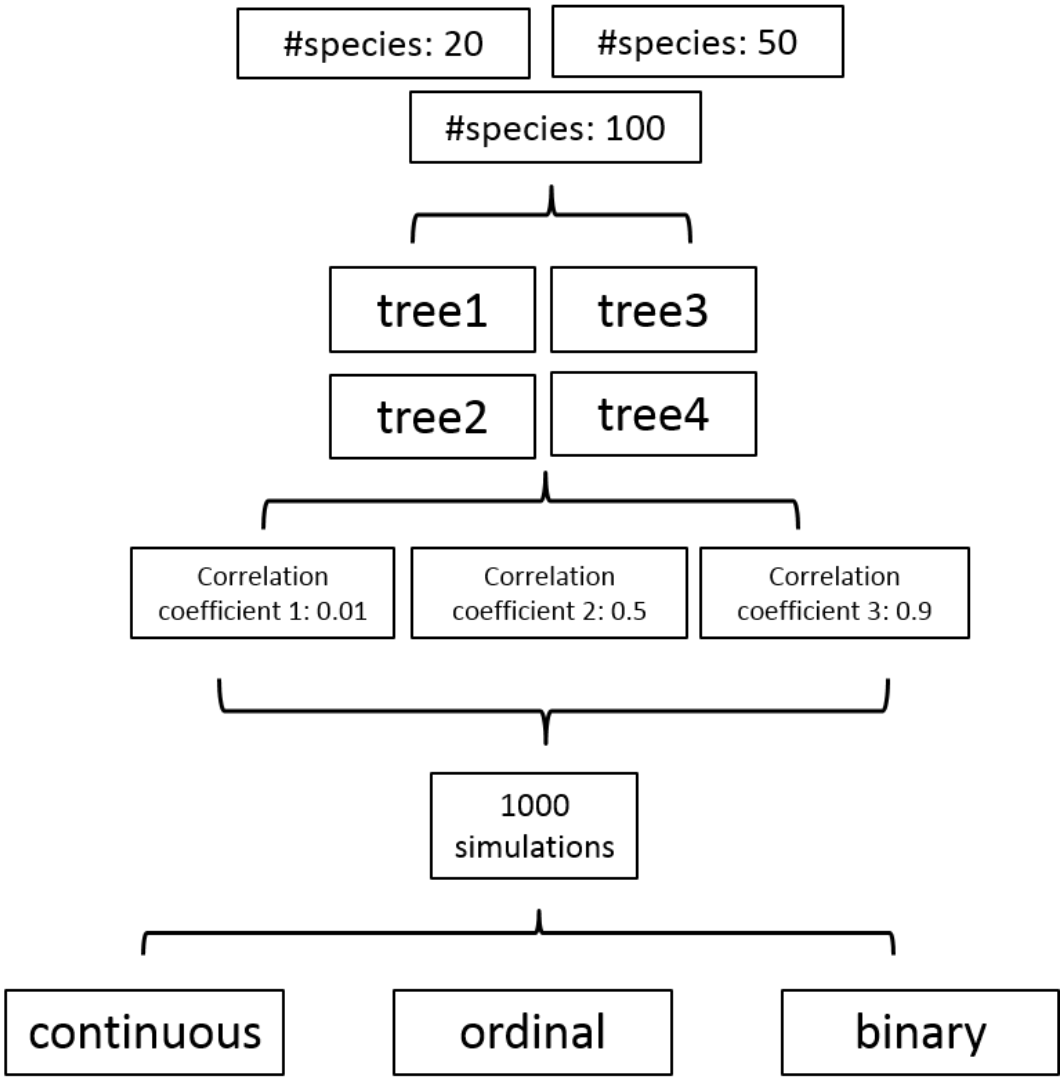


Figure 2.3: **Simulation setup.** Four different types of trees are simulated containing 20, 50 and 100 species. For each of these combinations, three different correlations coefficients are simulated a 1000 times and subsequently the simulated continuous variable is transformed into and ordinal and binary response variable. For each type of response variable (continuous, ordinal and binary) there are 36 possible combinations, in total, for all types of response variables, there are 108 possible combinations. For each parameter condition, 1000 simulations were run. This results in a total of 108,000 simulated data sets.

As a next step in the simulation setup, the simulated data is analysed by the different phylogenetic methods. Specifically, on each data set within that 1000, 180,000, respectively, a non-phylogenetic analysis and the four *phylogenetic methods* are applied (PGLS, PGEE, PLR, and PGLMM). For the non-phylogenetic analysis, Gaussian regression models for continuous and ordinal responses and logistic regressions for binary responses were applied. One needs to keep in mind that the PLR cannot be used for continuous and ordinal data. Therefore, not each method is applied to each kind of response variable. A detailed overview of which method was applied to which type of

Table 2.3: **Methods used for the different types of response variables.** Within the simulation loop, depending on the type of response variable, four of five non-phylogenetic and phylogenetic methods were used. For the continuous response, assumed to follow a Gaussian distribution, the identity link is used. The ordinal response consists of four categories based on an underlying continuous distribution. Here, as commonly done in literature, the ordinal response is assumed to be pseudo-continuous, thus, also the identity link is used. In the case of a binary response variable, which is also based on an underlying continuous distribution, the logit link is used.

	Continuous response	Ordinal response	Binary response
Non-phylogenetic GLM	✓	✓	✓
PGLS	✓	✓	
PGEE	✓	✓	✓
PLR			✓
PGLMM	✓	✓	✓

response is found in Table 2.3.
In the case of the continuous response variable, the Gaussian family and the identity link were used. Assuming the ordinal response to be pseudo-continuous, the Gaussian family with the identity link was also used, which is often done in phylogenetic studies (e.g. Matthews et al. 2010, Schuppli et al. 2012). This way, one can test whether treating ordinal as continuous data leads to elevated type I error rates and whether the new method using generalised linear mixed models gives better results using a multinomial model. In the case of the binary response variable, the binomial family with the logit-link was used.

2.1.2 Method specifications

For each type of response variable, a non-phylogenetic and three phylogenetic methods were applied (see Table 2.3). For the implementation in `R` the functions `glm()`, `pgls()`, `compar.gee()`. `MCMCglmm()` and `phyloglm()` were used from the packages `stats`, `caper`, `ape`, `MCMCglmm` and `phylolm`. For the non-phylogenetic GLM, the PGLS and the PGEE the implementation is straightforward and no special specifications need to be done.

For the PGLMM implemented in `MCMCglmm` some specifications need to be done before a model can be run. As this package uses a Bayesian approach, MCMC, priors for variance-covariance matrix of the fixed effects ($\mathbf{B}$), the variance components of the random effects (σ_u^2) and residuals (σ_ϵ^2) need to be specified (see Section 1.3.5). For the variance-covariance matrix of the fixed effects ($\mathbf{B}$) a multivariate normal prior is used where one has to specify an expected value(s) `mu` and a (co)variance matrix `V` representing the strength of belief. The bigger the variance, the more flat and thus, the weaker the prior. Further, in `MCMCglmm` the priors for the the variance scalars (σ_u^2 and σ_ϵ^2) of the random effects and residuals are assumed to follow an inverse-Wishart distribution with the expected (co)variance `V` and `nu` representing the degree of belief parameter (Hadfield 2012).
In the case of a continuous response the default prior for the fixed effects was used with `mu` $= 0$, large variances `V` $= 10^8$ and covariances set to zero (not listed in prior

example below). For the variance components σ_u^2 (`R`) and σ_ϵ^2 (`G`) a variance of 1 and `nu` set to a small number, 0.002, is a commonly used prior specification (`prior <- list(R=list(V=1,nu=0.002), G=list(G1=list(V=1, nu=0.002)))`). It must be taken care of that the probability distribution integrates to one because the variable must take some value. In case of the inverse-Wishart distribution this is achieved if `V` and `nu` are chosen to be larger than zero, otherwise the prior is improper (Hadfield 2012).
For the multinomial model, things get more complicated. First, because the residual covariance matrix of categorical data cannot be estimated from the data, the choice of its values is arbitrary. Usually a very strong prior is fitted with fixing it at e.g. one (`R = list(V = IJ, fix=1)`). `IJ` represents a covariance matrix of the form $\frac{1}{\mathbf{J}}$ $(\mathbf{I}+\mathbf{J})$ where $\mathbf{I}$ is a identity matrix and $\mathbf{J}$ a unit matrix with the dimensions $J-1$ (J = number of latent variables = number of categories - 1, more details found in Hadfield 2012). This covariance matrix implies that the variances of the probabilities (probabilities of being in category 2, 3 or 4 compared to the baseline category 1) are constant and that these probabilities are independent of each other, conditional on the constraint that they must sum to one (personal communication, Jarrod Hadfield). Second, for the phylogenetic covariance matrix the same prior is chosen, however, this one is not fixed in order to actually estimate it. Third, for the fixed effects $\mathbf{B}$ again the default is used. Moreover, in order to have trait (`trait` indexes the latent variables) specific intercepts and regression coefficients for the covariates the following implementation is used in case of a multinomial model: `ordinal.y~trait-1 + trait:X1`. Additionally, the global intercept is removed (`-1`) in order to have easier interpretable model outputs (personal communication, Jarrod Hadfield).
In case of a binary response, also the residual variance structure (`R`) is fixed because again the residual covariance matrix of categorical data cannot be estimated from the data. Moreover, the default prior for the fixed effects was used and the prior for the random effect (`G`) was specified in the same way as for the continuous response model (`prior <- list(R = list(V = 1, fix=1),G = list(G1 = list(V = 1,nu=0.002)))`).

For the phylogenetic logistic regression implemented in `phylolm`, the function `phyloglm()` needs some specifications concerning the boundary of the searching space for the linear predictor. For specifying this boundary, the argument `btol` is used. In particular, this argument constrains the fitted values, in case of the default value of 10, this means that the probability of the model prediction of "1" lies within the range of $1/(1+\exp(10)) = 0.000045$ and $1/(1+\exp(-10)) = 0.999955$ (Ho and Ané, 2013). On the one hand, a too low value of `btol` causes an error message saying that the current value excludes the estimated coefficients in the absence of phylogenetic signal. In fact, the function `phyloglm()` uses an iterative optimization procedure the results of a non-phylogenetic logistic regression (`glm()`) as a starting point. Thus, the former error message means that the non-phylogenetic model prediction of 1 lies outside of the boundary set by `btol`. On the other hand, it seems that higher `btol` values lead to lower standard errors. However, due to numerical problems, too high values of `btol` may also lead to an error message suggesting to reduce the boundary again. Therefore,

to find an appropriate value somewhere in the middle is rather a trial and error issue. For the simulations of this study, `btol` was set to 30.

2.1.3 Simulation output

For analysing and comparing the statistical performances of the different methods, one needs to save the relevant estimates and values of the model outputs of each single simulation. Generally, for keeping track of error messages, the function `try()` was used in front of each method command. This way, errors were able to be saved in the simulation outputs.

For all methods, except for PGLMM, the estimate of the coefficients, the standard error, t-values and p-values were saved. In case of the phylogenetic mixed model using a Bayesian approach, the posterior mean, the lower and upper 95%-confidence interval of the means, the effective sample size and the MCMC p-value were saved. The p-value saved in that case is not a classical p-value as we know it from frequentist inference. As described by Jarrod Hadfield (R-sig-ME group), this Bayesian p-value in the output of `MCMCglmm()` is given by twice the smaller probability of i) estimate < 0 and ii) estimate > 0. Important to note is that for the analyses of the statistical performances, these Bayesian p-values were treated as the p-values from the frequentist methods. Furthermore, in case of the multinomial model using PGLMM, there are three posterior means representing the effects of the predictor variables on the probability of being in level 2,3,4 compared to level 1. To keep it simple, the estimate of the probability of level 4 compared to level 1 was used in the analyses. In fact, if there is an effect, this should be the most pronounced.
The estimates were used for calculating the mean error and the rooted mean squared error, whereas the p-values were used for evaluating type I error rates and power. More details about the analyses are found in Section 2.3.

Besides these estimates, further output values were relevant for the analyses. Concerning phylogenetic signals, in case of PGLS, λ, and in case of the PLR, α, the phylogenetic signal for binary traits, were saved.
Furthermore, the simulation output contained information about whether the statistical models converged or not in order to exclude those in the analyses which did not converge. This was done by either saving the number of used iterations from the estimation procedure (if they exceed the default of 25, the models was assumed not to have converged) or by explicitly saving an attribute of the model output saying whether the model actually converged or not. In case of the PGLMM using MCMC, the Geweke's convergence diagnostic test was used using the function `geweke.diag()` in the package `coda` (Plummer et al. 2012). Especially for binary response variables using the PGEE, PLR or the PGLMM approaches, converging problems arise rather often. One issue which underlies this problem is the perfect fit between the variables, i.e. all species with the binary response equal to one have higher values of the independent variable than the species with the binary response equal to zero. In case where the simulated correlation coefficient is very high ($r = 0.9$), often the perfect fit problem occurs. If in

a simulated data set a perfect fit occurs, this is saved in the simulation output.

2.2 Real data set

For directly comparing the specific estimates of the four phylogenetic methods and in order to illustrate their application on a real evolutionary question, the four methods were applied to a continuous, ordinal and binary response variable of a real data set.

The data set contains information about the complexity of foraging niches and various morphological and life history traits of 78 primate species. The complexity of the foraging niche is composed of different foraging behaviours, such as extractive foraging, big game hunting and tool use. These elements add up to an overall niche complexity score representing an ordinal variable. The morphological and life history traits include measures on body size, brain size and development approximated by the age at first reproduction (AFR).
The evolutionary questions which are aimed to be tested using that data is about the effects of brain size on the lenght of development (brain malnutrition risk hypothesis by Janson and van Schaik 1993) and on niche complexity based on the studies about how some species manage to evolve into more complex foraging niches (Schuppli et. 2012; Schuppli et al., in preparation; Graber et al., in preparation). Specifically, there are three predictions for the three types of response variables: continuous, ordinal and binary. First, brain size is predicted to be positively correlated with development, as larger brains need more time to develop as stated by the brain malnutrition hypothesis (AFR ~ Brain + Body). The second hypothesis, saying that smarter species with larger brains evolve into more complex niches, predicts a positive association between the niche complexity score and brain size (Niche complexity ~ Brain + Body). This example illustrates a real data application for an ordinal response. Third, also based on the former hypothesis, the occurrence of extractive foraging (0/1 coded) as part of a complex foraging niche is expected to be positively affected by brain size (Extractive foraging ~ Brain + Body). In other words, species with larger brains have a higher probability to show extractive foraging as a part of their behavioural repertoire. All the three models are additionally controlled for body size in order to test the effect of relative rather than absolute brain mass.

2.3 Analyses of simulations

The following measurements representing parts of the statistical performance of a statistical method are all calculated for each combination of the different parameters varied in the simulation (i.e. for each combination of tree, number of species and correlation coefficient).
Before analyses, all the results with errors and the results of non-converged models were excluded.

2.3.1 Type I Error Rates and Power

In connection with hypothesis testing, the different methods were compared by evaluating type I error rates and power calculated from the p-values from the simulations. The type I error represents the probability of rejecting the null hypothesis although it is true and power is the probability or rejecting the null hypothesis under the assumption of a true alternative.
The error rates were simply calculated from the ratio of the number of significant p-values ($p < 0.05$) to the total number of p-values based on the simulations with a correlation coefficient of $r = 0.01$. The analogue was done to calculate the power, however, based on the simulations with a correlation coefficient of $r = 0.9$. To test whether the observed ratios of significant vs. non-significant p-values are significantly different from the expected ratio under the assumption of a type I error rate of 0.05%, Fisher‘s exact tests were used.

2.3.2 Mean Error, Mean Squared Error and Rooted Mean Squared Error

In order to compare the estimation abilities of the different methods, the estimates (x_i) were set in relation to the overall mean of all methods ($\bar{x}$) by calculating the mean error (ME) and the rooted mean squared error (RMSE) ($Nsim$ = number of simulations):

$$\text{ME} = \frac{1}{Nsim} \sum_{i=1}^{Nsim} (x_i - \bar{x}). \tag{2.1}$$

$$\text{MSE} = \frac{1}{Nsim} \sum_{i=1}^{Nsim} (x_i - \bar{x})^2. \tag{2.2}$$

$$\text{RMSE} = \sqrt{MSE} = \sqrt{\frac{1}{Nsim} \sum_{i=1}^{Nsim} (x_i - \bar{x})^2}. \tag{2.3}$$

The ME represents an index for bias, depending on whether it is negative or positive it indicates an over - or underestimation of an estimate. The RMSE is simply the square root of the mean squared error. The mean squared error can be decomposed in the variance and the squared bias of an estimator (e.g. Held and Sabanés Bové 2013), giving information about the accuracy of a parameter estimate weighing larger errors

more. If the bias is equal to zero, then the mean squared error is the same as the variance of an estimate.
For each type of response variable (continuous, ordinal and binary) and each input correlation coefficient ($r = 0.01$, $r = 0.5$ and $r = 0.9$), a separate overall mean was calculated across all methods. In case of the ordinal response, however, two separate overall means were calculated, because the estimates or rather the posterior means of PGLMM have a different meaning compared to the estimates of the other three methods. Thus, an overall mean including the non-phylogenetic GLM, the PGLS and the PGEE was calculated and one for the PGLMM separately.

3 Results

Using simulated data based on different parameter combinations, four phylogenetic comparative methods (PGLS, PGEE, PGLMM and PLR) are compared modeling the correlated evolution between two or three traits a 1000 times. These methods are compared for three types of response variables: continuous, ordinal and binary. For each type of response, the methods are compared and discussed separately in the following three sections based on type I error rates, power, mean errors and rooted mean squared errors. The results in the tables show the means for each combination of the four trees (tree 1 to 4) and three numbers of species ($n = 20$, $n = 50$, $n = 100$) based on the 1000 simulations. Before the analyses, the model runs which resulted in error messages or which did not converge were excluded from the analyses explaining the numbers of simulations smaller than 1000. For illustrating purposes of the statistical performances of the different methods, only the results based on the simulated data of tree structure 1 and 50 species are shown graphically or in case of rooted mean squared error plots, only the ones for the input correlation $r = 0.5$. Unless otherwise stated, the analogues plots for the other parameter combinations of tree types, numbers of species and input correlations look similar, and thus, are not shown.

3.1 Continuous response variable

3.1.1 Non-phylogenetic GLM

Type I Error and Power

For the non-phylogenetic approach using a simple GLM model with the identity link, *type I error rates* generally strongly exceed the commonly assumed error rate of $\alpha = 0.05$. For almost all combinations of the four different trees (tree 1 to 4) and the three numbers of species ($n = 20$, $n = 50$, $n = 100$), the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values is significantly different from the expected ratio (0.05/0.95). In other words, these non-phylogenetic models declare too often statistical significance. However, interestingly, the second tree type, where most of the diversification was simulated to occur early in the phylogenetic history (i.e. short shared evolutionary histories between species), never shows significantly inflated type I error rates and tree 3, with the opposite evolutionary pattern, shows the highest type I error rates. Moreover, the type I error rates tend to increase with increasing sample size.
The *power*, on the other hand, mainly never falls lower than 90% and tends to increase for larger sample sizes. This means that the probability of detecting significance if it really is significant, is very high. The corresponding results are found in Table 3.1 and Figures 3.1, 3.2.

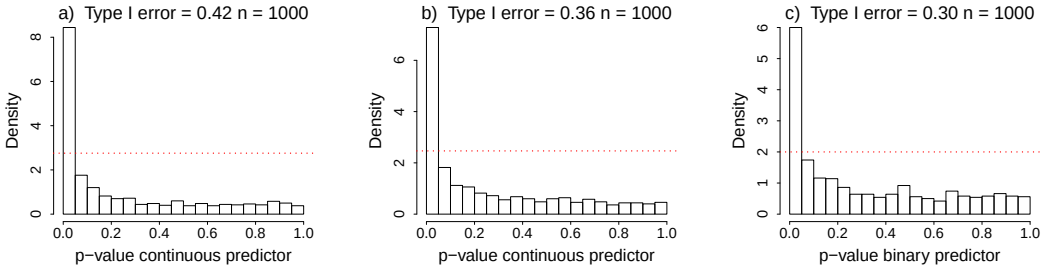


Figure 3.1: **Type I error rates from continuous response models - non-phylogenetic GLM.** Distributions of p-values from data simulated with $r = 0.01$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The horizontal dashed line indicates the density at the critical value of 0.05.

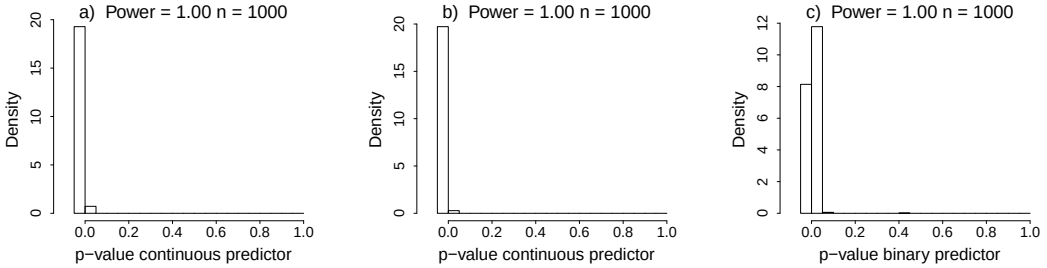


Figure 3.2: **Power from continuous response models - non-phylogenetic GLM.** Distributions of p-values from data simulated with $r = 0.9$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The density bar from -0.1 to 0.0 represents the density of p-values equal to zero.

Table 3.1: **Type I error rates (TIE) and Power from continuous response models - Non-phylogenetic GLM**: Fisher's exact test was used to compare the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values to the expected ratio (0.05/0.95) under the assumption of a type I error of $\alpha = 5\%$. The different significance levels of the Fisher's exact test are indicated by bold font for $p < 0.001$ and in italics for $p < 0.05$. a) univariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.23**	1.00	**0.42**	1.00	**0.39**	1.00
Tree 2	0.06	1.00	0.05	1.00	0.06	1.00
Tree 3	**0.58**	0.94	**0.51**	1.00	**0.75**	0.99
Tree 4	**0.23**	0.99	**0.34**	1.00	**0.47**	1.00

b) multivariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.22**	1.00	**0.36**	1.00	**0.36**	1.00
Tree 2	0.05	1.00	0.04	1.00	0.07	1.00
Tree 3	**0.45**	0.98	**0.51**	0.99	**0.71**	1.00
Tree 4	**0.19**	1.00	**0.31**	1.00	**0.43**	1.00

c) multivariate model - binary predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.20**	0.92	**0.30**	1.00	**0.29**	1.00
Tree 2	0.05	0.97	0.04	1.00	0.05	1.00
Tree 3	**0.41**	0.81	**0.36**	0.94	**0.59**	0.99
Tree 4	**0.14**	0.93	**0.24**	0.99	**0.32**	1.00

Mean estimates (M), Mean Error (ME) and Rooted Mean Squared Error (RMSE)

First of all, the *mean estimates* usually strongly accord to the correlation coefficients used for the simulations of the uncorrelated/correlated data. However, for the multivariate model, the pattern is not that clear compared to the univariate model, which is probably based on the fact that two predictors are included in the model.
The *mean errors* are generally rather moderate especially for the univariate model. The errors increase for the multivariate model, in particular for the binary predictor. This means that bias in the estimates is rather low, however increases in case of multivariate models, especially for binary predictors. Moreover, tree 2 with most diversification happening early in the evolutionary history, shows the lowest estimation bias, whereas the opposite tree structure (i.e. tree 3 with most diversification occurring late in evolution) leads to rather elevated estimation bias.
Considering the *rooted mean squared errors* as an indicator for the accuracy around the overall mean, again especially the binary predictor of the multivariate models shows much higher rooted mean squared errors compared to the continuous predictors of the univariate and multivariate models.
As already shown by the mean errors, tree 2 with most diversifications happening early in the phylogeny, shows generally more accurate estimates compared to the other trees,

and tree 3 generally shows the least accurate parameter estimates. However, for the binary predictor of the multivariate model, the pattern looks slightly different. There, not tree 3 but rather tree 4 shows the least accurate estimates.
Further, with increasing sample size, the rooted mean squared errors decrease implying a higher accuracy of the estimates. The corresponding results and graphical illustrations are found in Tables C.1 to C.3 in the Appendix C.1.1 and Figures 3.3, 3.4.

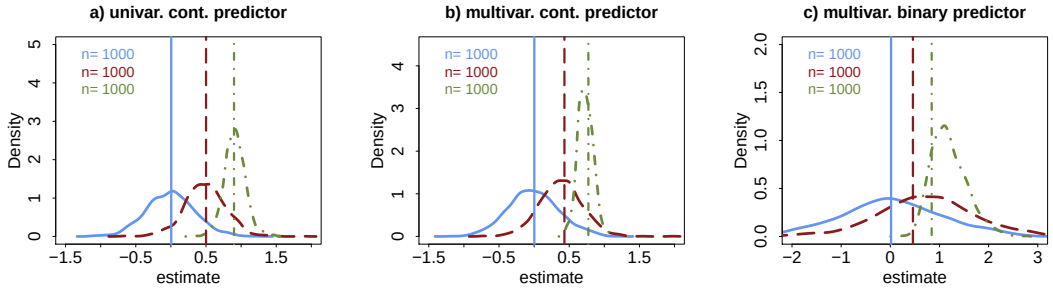


Figure 3.3: **Estimates from continuous response models - Non-phylogenetic GLM.** Distributions of the effect sizes based on the simulated data with an input correlation coefficients of $r = 0.01$ (blue solid line), $r = 0.5$ (red dashed line) and $r = 0.9$ (green dashed-dotted line): a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model binary predictor including the data for tree 1 and the number of species of 50. The vertical lines indicate the means overall methods. The bandwidth of the densities was set to the default (`"nrd0"`).

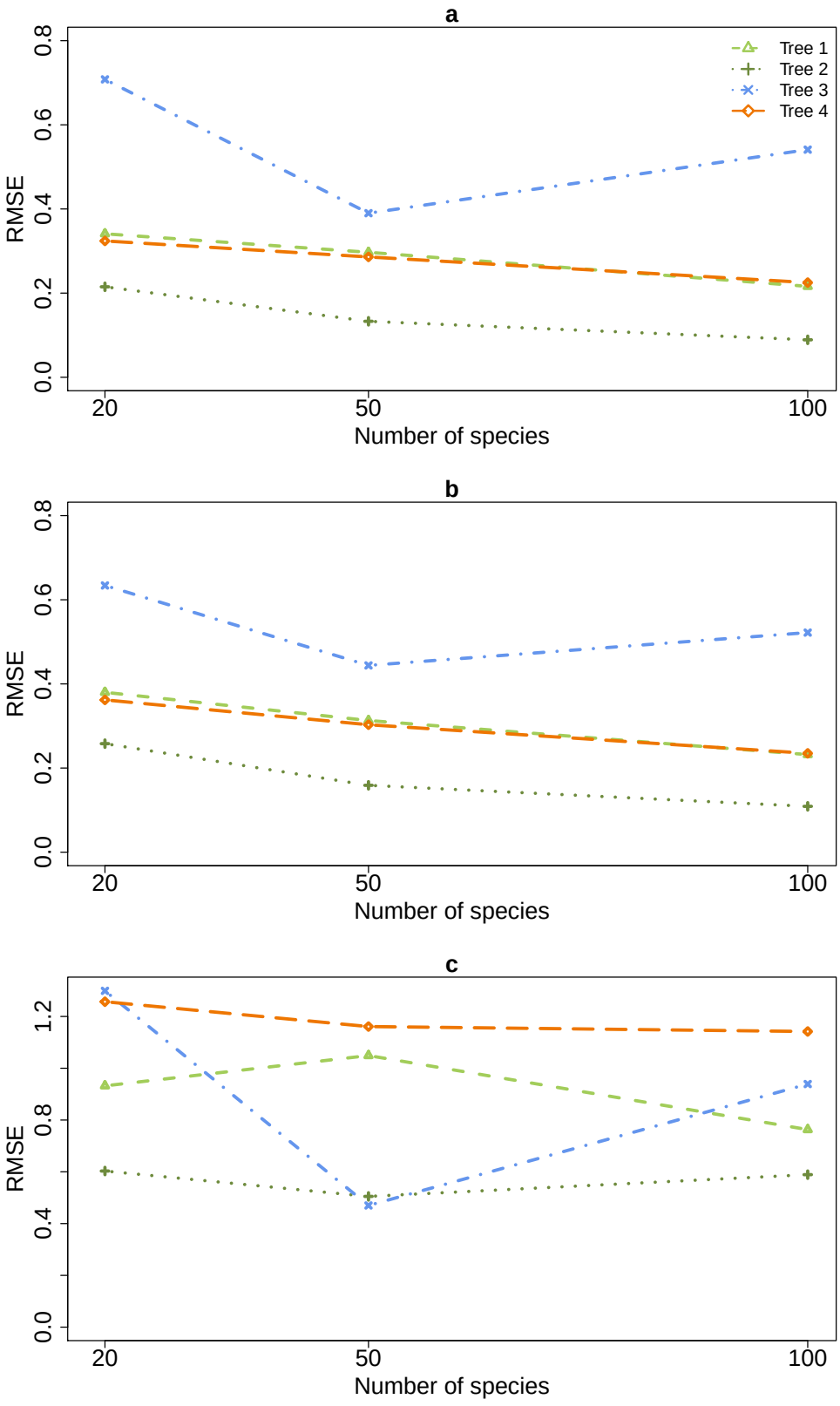


Figure 3.4: **Rooted mean squared errors from continuous response models - Non-phylogenetic GLM.** The rooted mean squared errors (RMSE) are shown for the data simulated with a correlation coefficient of $r = 0.5$. a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

3.1.2 PGLS

Type I Error and Power

Compared to the non-phylogenetic GLM, PGLS shows strongly reduced *type I error rates*. Only the models including smaller samples sizes (i.e. number of species = 20) mainly show elevated type I error rates. And remarkably, there mainly all the tree types except type 2 (diversification early in phylogenetic history) show significantly elevated type I error rates. Moreover, type I error rates show a tendency to decrease with increasing sample size. All in all, except for a few cases, PGLS incorporating pyhlogenetic dependencies performs better with respect to hypothesis testing compared to the non-phylogenetic method.
The *power* again almost never falls below 90%, only for the binary predictor of the multivariate model the values are generally a little lower, but again cross 90% for larger samples sizes. The corresponding results and graphical illustrations are found in Table 3.2 and Figures 3.5 and 3.6.

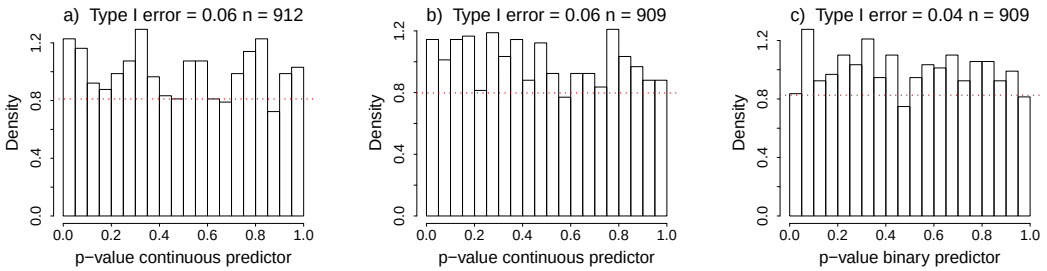


Figure 3.5: **Type I error rates from continuous response models - PGLS.** Distributions of p-values from data simulated with $r = 0.01$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The horizontal dashed line indicates the density at the critical value of 0.05.

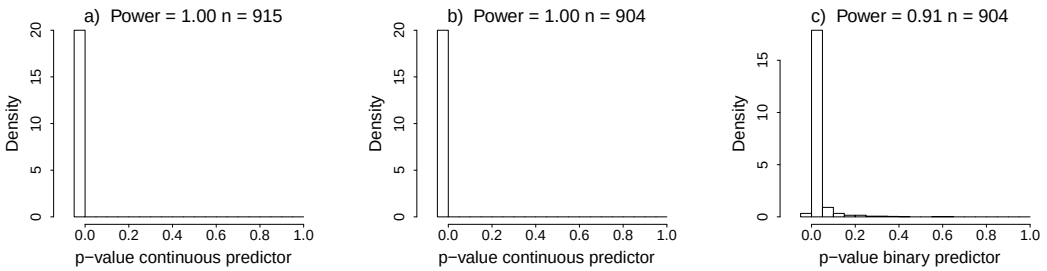


Figure 3.6: **Power from continuous response models - PGLS.** Distributions of p-values from data simulated with $r = 0.9$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The density bar from -0.1 to 0.0 represents the density of p-values equal to zero.

Table 3.2: **Type I error rates (TIE) and Power from continuous response models - PGLS**: Fisher's exact test was used to compare the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values to the expected ratio (0.05/0.95) under the assumption of a type I error of $\alpha = 5\%$. The different significance levels of the Fisher's exact test are indicated by bold font for $p < 0.001$ and in italics for $p < 0.05$. a) univariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	*0.07*	1.00	0.06	1.00	0.05	1.00
Tree 2	0.06	1.00	0.05	1.00	0.06	1.00
Tree 3	**0.12**	1.00	0.05	1.00	0.05	1.00
Tree 4	**0.09**	1.00	0.06	1.00	0.05	1.00

b) multivariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	0.07	1.00	0.06	1.00	0.06	1.00
Tree 2	0.05	1.00	0.05	1.00	0.06	1.00
Tree 3	**0.15**	1.00	0.05	1.00	0.05	1.00
Tree 4	*0.08*	1.00	0.06	1.00	0.05	1.00

c) multivariate model - binary predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	*0.08*	0.74	0.04	0.91	0.06	0.99
Tree 2	0.05	0.97	0.04	1.00	0.05	1.00
Tree 3	**0.14**	0.55	0.05	0.83	0.05	0.86
Tree 4	0.07	0.85	0.05	0.98	0.06	1.00

Mean estimates (M), Mean Error (ME) and Rooted Mean Squared Error (RMSE)

As for the non-phylogenetic models, the mean estimates usually strongly accord to the correlation coefficients used for the simulations of the uncorrelated/correlated data. However, for the multivariate model, the pattern is not that clear compared to the univariate model, which is probably based on the fact that two predictors are included in the model.

The *mean errors* tend to be slightly lower for PGLS compared to the non-phylogenetic GLM. The errors are rather low for the univariate models, however, for the multivariate models, the errors are also elevated especially for the binary predictor. Thus, the bias of the binary predictor is larger compared to the continuous predictors of the univariate and multivariate model.

The *rooted mean squared errors* are also decreased in comparison to the non-phylogenetic approach. They show similar patterns for all the four trees, except in case of the binary predictor in the multivariate model. Strikingly, the models for the data simulated to be highly correlated ($r = 0.9$) tend to have lower mean squared errors compared to the data simulated with $r = 0.01$ and $r = 0.5$. These results are only observable in the tables, as the Figure 3.8, illustrate the RMSEs for $r = 0.5$.

Furthermore, there is a strong decrease of the RMSE with increasing sample size. The

corresponding results and graphical illustrations are found in Tables C.4 to C.6 in Appendix in C.1.2 and Figures 3.7, 3.8.

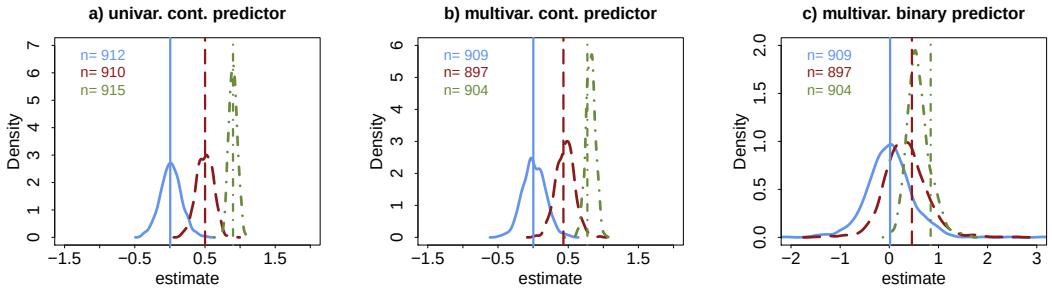


Figure 3.7: **Estimates from continuous response models - PGLS.** Distributions of the effect sizes based on the simulated data with an input correlation coefficients of $r = 0.01$ (blue solid line), $r = 0.5$ (red dashed line) and $r = 0.9$ (green dashed-dotted line): a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model binary predictor including the data of tree 1 and 50 species. The vertical lines indicate the means overall methods. The bandwidth of the densities was set to the default (`"nrd0"`).

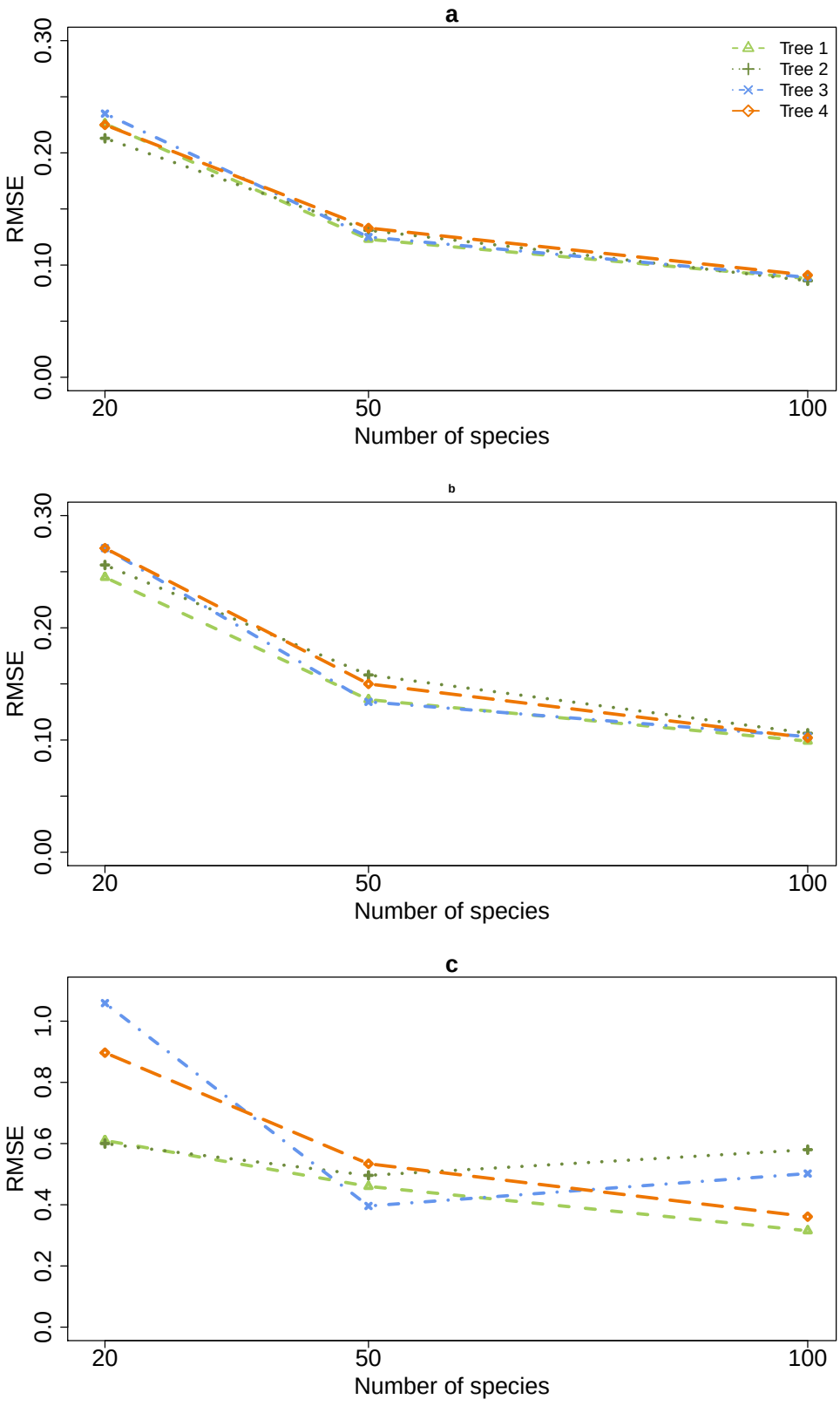


Figure 3.8: **Rooted mean squared errors from continuous response models - PGLS.** The rooted mean squared errors (RMSE) are shown for the data simulated with a correlation coefficient of $r = 0.5$. a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

Phylogenetic signal λ

Additionally to the estimates of the predictor variables, the PGLS method measures the phylogenetic signal λ ($\in [0, 1]$) using a maximum likelihood procedure.
To be remembered, all the data was simulated using the `sim.char()` function under a Brownian motion model of evolution (i.e. the simulated observations can basically reach unlimited variance, see Section 1.2.2) given a certain phylogenetic tree (tree 1 to 4). A phylogenetic signal λ of 1 implies a trait to evolve fully under Brownian motion. Important to note is also that for each model (univariate or multivariate), `PGLS()` estimates only one λ (see estimation procedure in Section 1.3.1).
For the univariate as well as the multivariate models, the measurements of λ are generally very high (i.e. > 0.9), except for tree 2, where most of the diversification occurs early in the phylogenetic history. In other words, the similarity in a certain trait in closely related species can be explained largely by their common evolutionary history. However, for tree 2, showing very short common evolutionary histories between species, only a small part of the similarity between species in a trait can be explained by their phylogenetic relatedness. Please note there also the rather high standard deviations. The corresponding results are found in Table 3.3.

Table 3.3: **Mean phylogenetic signal λ with the standard deviation in brackets from continuous response models - PGLS**:
a) univariate model.

	Species 20	**Species 50**	**Species 100**
tree 1	0.92 (0.19)	1.00 (0.03)	1.00 (0.00)
tree 2	0.31 (0.45)	0.30 (0.43)	0.39 (0.44)
tree 3	0.90 (0.28)	1.00 (0.05)	1.00 (0.00)
tree 4	0.78 (0.33)	0.94 (0.12)	0.98 (0.03)

b) multivariate model.

	Species 20	**Species 50**	**Species 100**
tree 1	0.85 (0.29)	0.98 (0.10)	1.00 (0.02)
tree 2	0.29 (0.45)	0.27 (0.41)	0.35 (0.43)
tree 3	0.78 (0.40)	0.99 (0.10)	1.00 (0.00)
tree 4	0.69 (0.40)	0.90 (0.18)	0.97 (0.05)

3.1.3 PGEE

Type I Error and Power

Compared to the non-phylogenetic GLM, the PGEE method performs better concerning hypothesis testing showing predominantly non-elevated *type I error rates*. Interestingly, for tree 3 and tree 4 in combination with large sample sizes (i.e. number of species = 100), this method shows significantly elevated type I error rates. Tree 3 simulates a phylogenetic history with most of the diversifications occurring late in the evolutionary time and in tree 4 all branch lengths were set to 1. In sum, the error rates are comparable to the PGLS method, however, there they occur rather for small sample size whereas for PGEE they occur counterintuitively rather for large sample sizes.
The *power* based on the simulations with $r = 0.9$ are mainly higher than 90% implying a very good performance, however, for the binary predictor of the multivariate model, the power is again slightly decreased especially for a small sample size. Due to unknown reasons, the analyses of the multivariate model with the simulated data based on tree 3 and 20 species did not produce any p-values, indicated by NA in Table 3.4. The corresponding results and graphical illustrations are shown in Table 3.4 and Figures 3.9, 3.10.

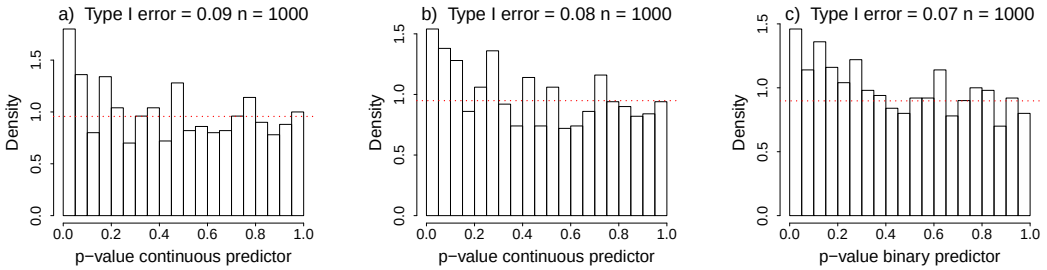


Figure 3.9: **Type I error rates from continuous response models - PGEE.** Distributions of p-values from data simulated with $r = 0.01$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The horizontal dashed line indicates the density at the critical value of 0.05.

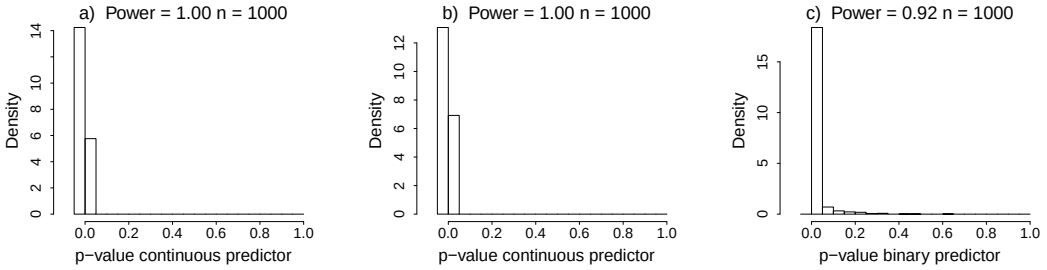


Figure 3.10: **Power from continuous response models - PGEE.** Distributions of p-values from data simulated with $r = 0.9$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The density bar from -0.1 to 0.0 represents the density of p-values equal to zero.

Table 3.4: **Type I error rates (TIE) and Power from continuous response models - PGEE**: Fisher's exact test was used to compare the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values to the expected ratio (0.05/0.95) under the assumption of a type I error of $\alpha = 5\%$. The different significance levels of the Fisher's exact test are indicated by bold font for $p < 0.001$ and in italics for $p < 0.05$. a) univariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	0.06	1.00	**0.09**	1.00	0.07	1.00
Tree 2	0.05	1.00	0.05	1.00	0.06	1.00
Tree 3	0.00	0.76	0.02	1.00	**0.13**	1.00
Tree 4	*0.08*	1.00	*0.08*	1.00	**0.11**	1.00

b) multivariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	0.04	1.00	*0.08*	1.00	0.06	1.00
Tree 2	0.06	1.00	0.05	1.00	0.06	1.00
Tree 3	NA	NA	NA	NA	*0.08*	1.00
Tree 4	0.06	1.00	*0.07*	1.00	**0.12**	1.00

c) multivariate model - binary predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	0.04	0.58	*0.07*	0.92	*0.07*	0.99
Tree 2	0.05	0.97	0.04	1.00	0.05	1.00
Tree 3	NA	NA	NA	NA	*0.08*	0.74
Tree 4	0.04	0.73	*0.07*	0.97	**0.12**	0.99

Mean estimates (M), Mean Error (ME) and Rooted Mean Squared Error (RMSE)

The *mean estimates* by PGEE for the three different correlation coefficients ($r = 0.01$, $r = 0.5$, $r = 0.9$) are basically the same as for the non-phylogenetic GLM and PGLS. Concerning estimation bias, PGEE is comparable to PGLS. Again, the estimation bias seems to be especially pronounced for the binary predictor in case of the multivariate model and tends to decrease for larger sample sizes.
The *rooted mean squared error* as an indicator for the accuracy of the estimates is also similar to PGLS. The four types of trees don‘t show any differences, except for the binary predictor of the multivariate model. Moreover, a very high input correlation coefficient of $r = 0.9$ and increasing sample sizes lead to lower RMSE (i.e. more accurate estimates). Overall, the performance of PGEE is comparable to PGLS, both showing mainly non-elevated type I error rates and similar estimation bias and accuracy. The corresponding results and graphical illustrations are found in Tables C.7 to C.9 in Appendix C.1.3 and Figures 3.11, 3.12.

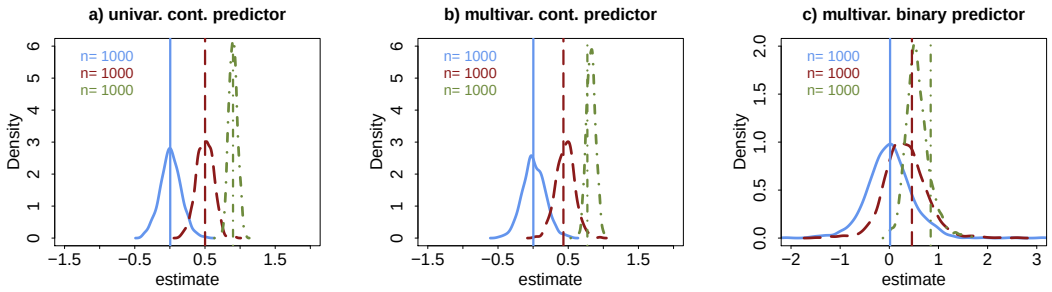


Figure 3.11: **Estimates from continuous response models - PGEE.** Distributions of the effect sizes based on the simulated data with an input correlation coefficients of $r = 0.01$ (blue solid line), $r = 0.5$ (red dashed line) and $r = 0.9$ (green dashed-dotted line): a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model binary predictor including the data of tree 1 and 50 species. The vertical lines indicate the means overall methods. The bandwidth of the densities was set to the default (`"nrd0"`).

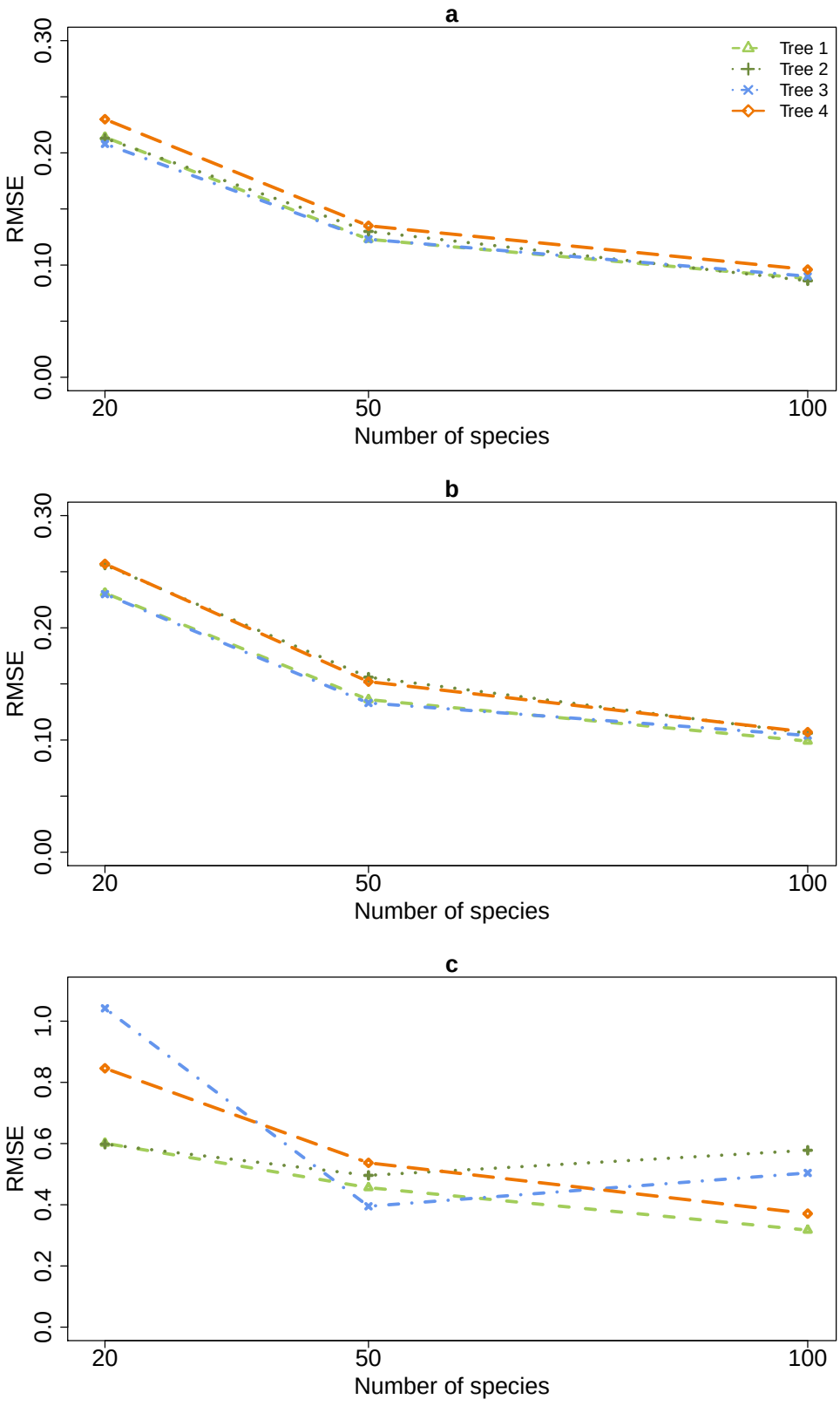


Figure 3.12: **Rooted mean squared errors from continuous response models - PGEE.** The rooted mean squared errors (RMSE) are shown for the data simulated with a correlation coefficient of $r = 0.5$. a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

3.1.4 PGLMM

Type I Error and Power

The PGLMM method shows basically no elevated *type I error rates*, not even for small sample sizes. In other words, the Bayesian approach almost never mistakenly declares statistical significance. Furthermore, detecting significance under the assumption of the alternative, this method shows a good performance with powers of mainly 80% and higher. As for the other phylogenetic methods concerning a continuous response, the power is slightly decreased for the binary predictor of the multivariate model, however, tends to increase for larger sample sizes. The corresponding results and the graphical illustrations are found in Table 3.5 and Figures 3.13 and 3.14.

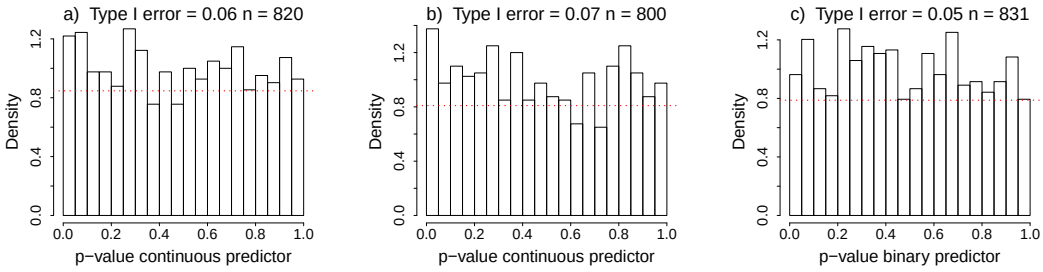


Figure 3.13: **Type I error rates from continuous response models - PGLMM.** Distributions of p-values from data simulated with $r = 0.01$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The horizontal dashed line indicates the density at the critical value of 0.05.

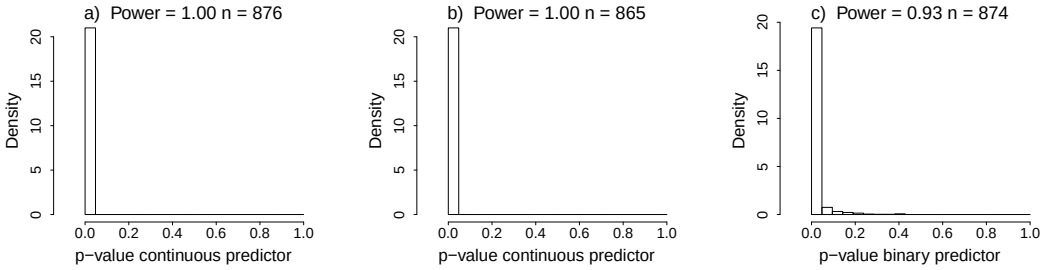


Figure 3.14: **Power from continuous response models - PGLMM.** Distributions of p-values from data simulated with $r = 0.9$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

Table 3.5: **Type I error rates (TIE) and Power from continuous response models - PGLMM**: Fisher's exact test was used to compare the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values to the expected ratio (0.05/0.95) under the assumption of a type I error of $\alpha = 5\%$. The different significance levels of the Fisher's exact test are indicated by bold font for $p < 0.001$ and in italics for $p < 0.05$. a) univariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	0.04	1.00	0.06	1.00	0.06	1.00
Tree 2	0.06	1.00	0.05	1.00	0.05	1.00
Tree 3	0.07	1.00	0.05	1.00	0.07	1.00
Tree 4	0.07	1.00	0.05	1.00	0.05	1.00

b) multivariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	0.04	1.00	0.07	1.00	0.06	1.00
Tree 2	0.05	1.00	0.04	1.00	0.06	1.00
Tree 3	*0.07*	1.00	0.05	1.00	0.08	1.00
Tree 4	0.05	1.00	0.05	1.00	0.06	1.00

c) multivariate model - binary predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	0.06	0.70	0.05	0.93	0.06	1.00
Tree 2	0.05	0.97	0.04	1.00	0.05	1.00
Tree 3	0.06	0.43	0.04	0.85	0.06	0.89
Tree 4	0.04	0.82	0.05	0.98	0.06	1.00

Mean estimates (M), Mean Error (ME) and Rooted Mean Squared Error (RMSE)

In order to prevent to be too much repetitive, this method shows similar results to the PGLS and PGEE in case of a continuous response variable. In other words, the bias, given by the *mean error*, is generally rather low, however, is increased for the binary predictor of the multivariate model. Further, the *rooted mean squared error* indicating the accuracy of the parameter estimates decreases for higher input correlation coefficients (i.e. $r = 0.9$) and increases for larger samples sizes. The corresponding results and graphical illustrations are found in Tables C.10 to C.12 in Appendix C.1.4 and Figures 3.15, 3.16.

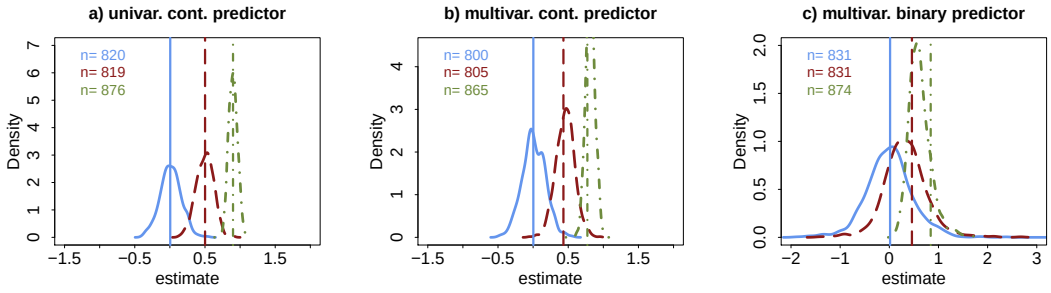


Figure 3.15: **Estimates from continuous response models - PGLMM.** Distributions of the effect sizes based on the simulated data with an input correlation coefficients of $r = 0.01$ (blue solid line), $r = 0.5$ (red dashed line) and $r = 0.9$ (green dashed-dotted line): a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model binary predictor including the data of tree 1 and the number of species 50. The vertical lines indicate the means overall methods. The bandwidth of the densities was set to the default (`"nrd0"`).

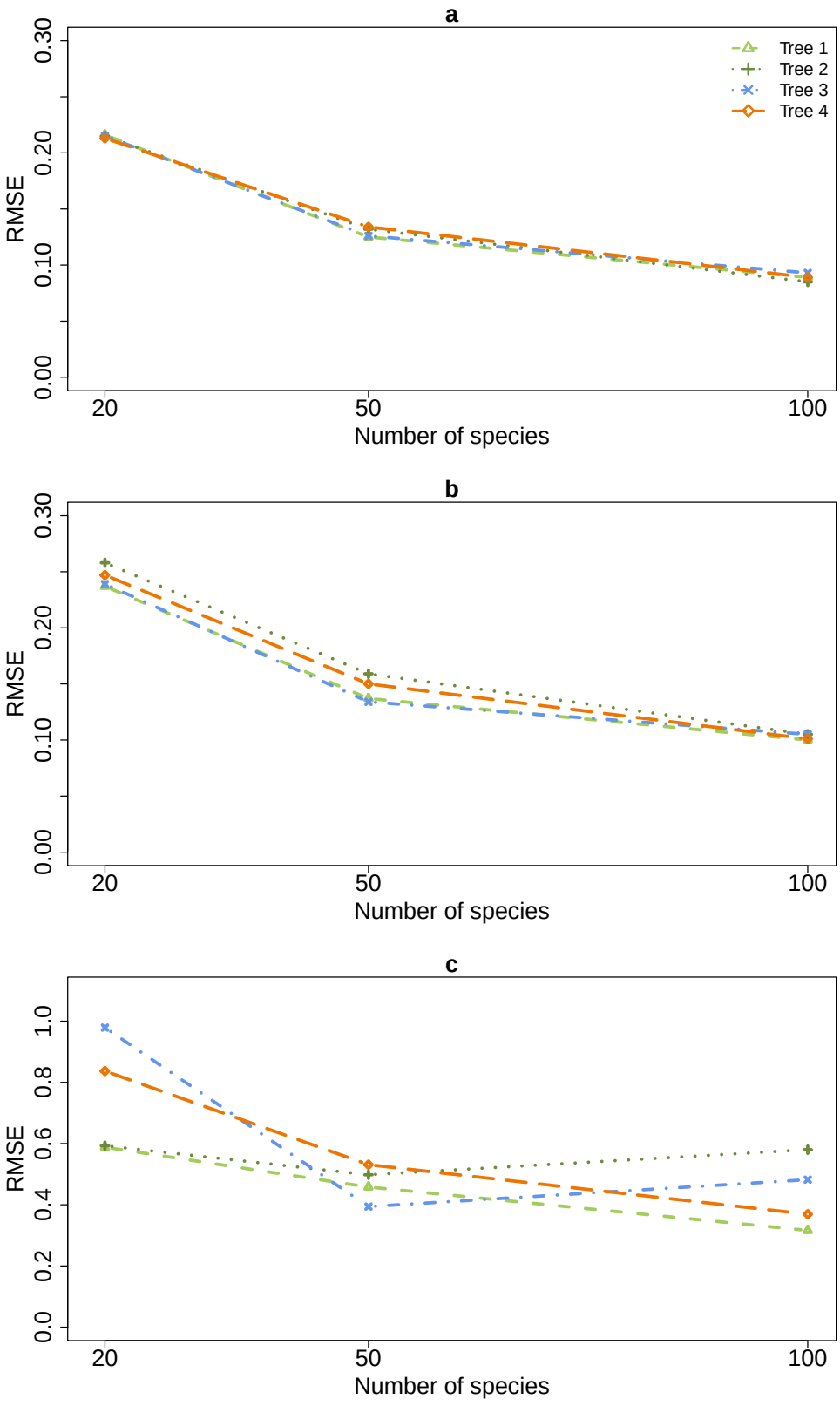


Figure 3.16: **Rooted mean squared errors from continuous response models - PGLMM.** The rooted mean squared errors (RMSE) are shown for the data simulated with a correlation coefficient of $r = 0.5$. a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

3.1.5 The four methods in comparison

Simulated data

All in all, the three phylogenetic comparative methods (PGLS, PGEE and PGLMM) applied on continuous dependent variables, show very similar performances. All show highly reduced type I error rates compared to the non-phylogenetic analysis, whereas the PGLMM method for all parameter conditions of type of tree, sample size and strength of correlation, shows the best hypothesis testing performance. Concerning estimation abilities, all the three methods perform equivalently well.
Figure 3.17 shows the mean estimates with the corresponding 95%-confidence intervals for the data simulated a 1000 times based on tree 1 and 50 species. The analogous plots for the other parameter combinations would show a similar picture, and thus, are not shown. The figure illustrates, that generally the confidence intervals of the estimates for the 1000 simulations are very narrow, meaning the parameter estimates fall all in the same narrow range. Moreover, the non-phylogenetic method as well as the binary predictor of the multivariate model in general shows higher variability in parameter estimates. The data with an input correlation coefficient of $r = 0.5$ and $r = 0.9$ show decreased mean parameter estimates for the multivariate models.
The parameters of the non-phylogenetic and the three phylogenetic methods are plotted against each other pairwise in Figure 3.18 a) to f). This shows that the non-phylogenetic GLM produces a higher variability in parameter estimates compared to the three phylogenetic methods (Figure 3.18 a) to c)). The phylogenetic methods compared to each other show equivalent estimation variabilities (Figure 3.18 d) to f)).

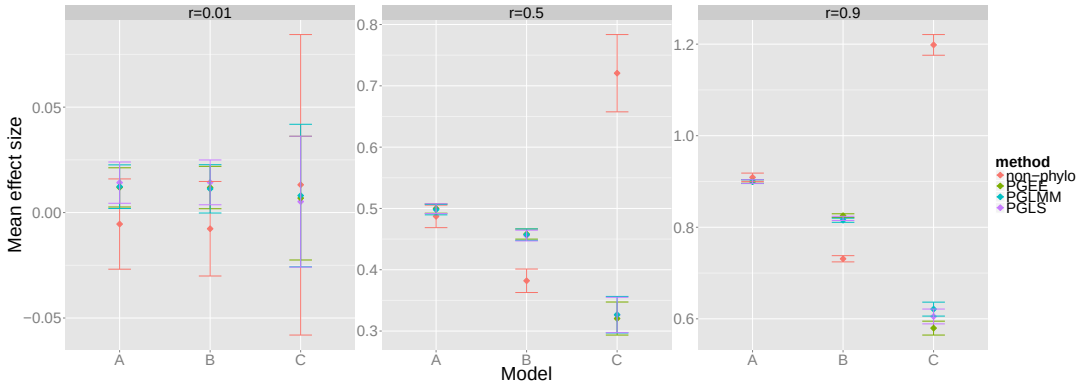


Figure 3.17: **Mean estimates and 95%-confidence intervals from continuous response models.** Comparison of the mean estimates and corresponding 95%-confidence intervals from the four methods for the three input correlation coefficients including the simulated data for tree 1 and 50 species (number of simulations = 1000). A: univariate model - continuous predictor; B: multivariate model - continuous predictor; C: multivariate model - binary predictor.

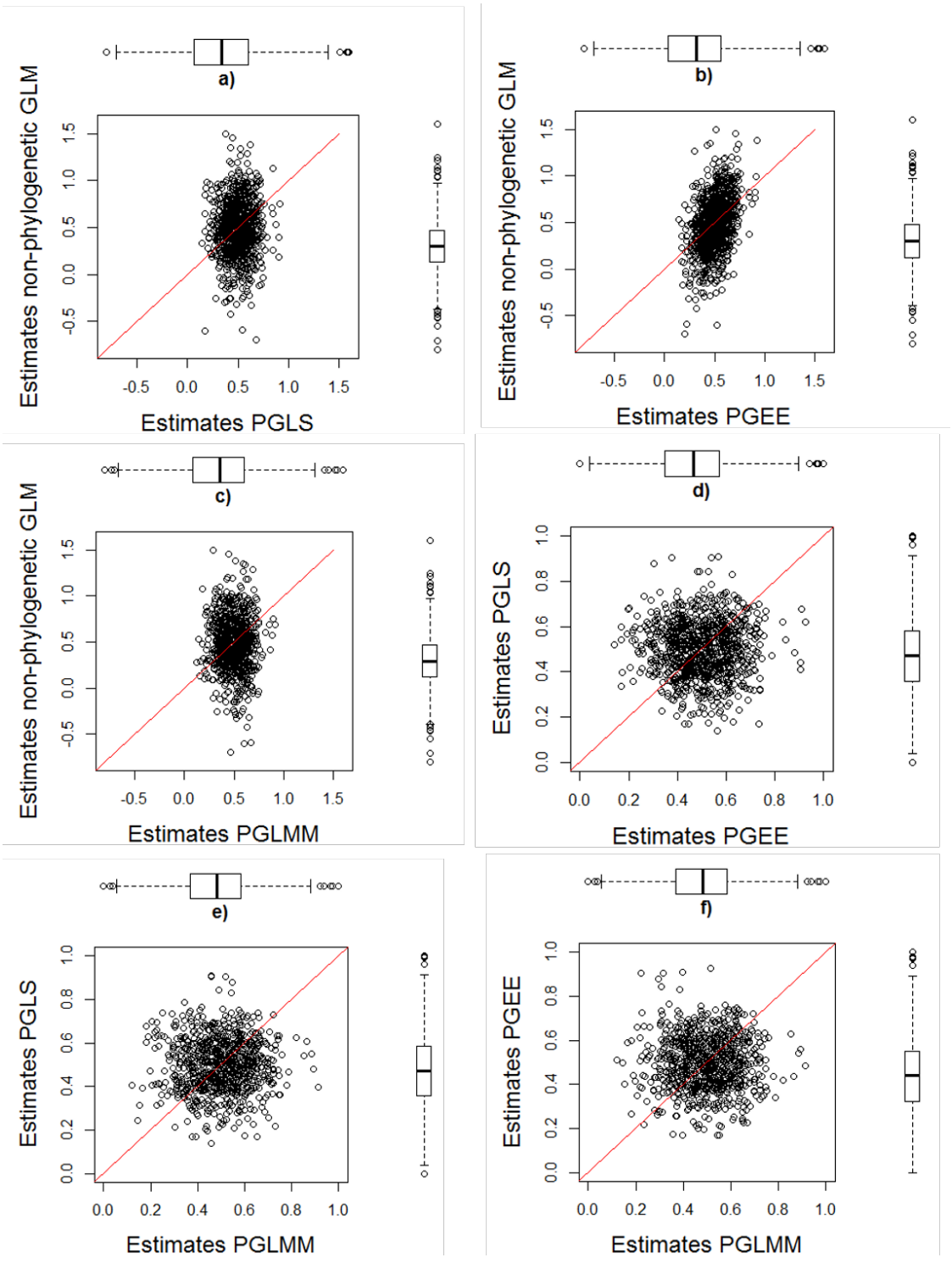


Figure 3.18: **Parameter estimates compared.** The parameter estimates based on the simulated data with $r = 0.5$ and 50 species from the non-phylogenetic and three phylogenetic methods (continuous response models) plotted against each other pairwise with the corresponding boxplots aside: a) Non-phylogenetic GLM vs. PGLS; b) Non-phylogenetic vs. PGEE; c) Non-phylogenetic vs. PGLMM; d) PGLS vs. PGEE; e) PGLS vs. PGLLMM; f) PGEE vs. PGLMM.

Application on a real data set

In order to apply the methods on a real evolutionary question and to directly compare the outcomes of continuous response models, the brain malnutrition hypothesis (Janson and van Schaik 1993) was tested using the example data set including 78 primate species. This hypothesis implies an effect of relative brain size (i.e. brain size corrected for body size) on the length of development (given by age at first reproduction).
Table 3.6 shows the results of the models for the non-phylogenetic GLM and the three phylogenetic methods (PGLS, PGEE and PGLMM). Age at first reproduction represents the continuous response, whereas brain size and body size are the predictor variables. The parameter estimates of brain size from the four methods strongly concur as well as the p-values. For the covariate body size, the estimates also quite accord, however, the p-values are slightly different. All in all, in concordance with the prediction of the brain malnutrition hypothesis, the conclusions under the four different methods are the same. In other words, whatever phylogenetic method is used, brain size shows a positive effect on the length of development, meaning larger brains need more time to develop.
However, simply conducting a statistical analysis is not enough. Further, a model needs to be validated in terms of how good the model fits the data. This is usually done by looking at so-called diagnostic plots: plot of residuals vs. fitted values to check for homoscedasticity of the residuals and the normal quantile-quantile (q-q) plot to test for normality of the residuals. These diagnostic plots are helpful in case of non-phylogenetic GLM, PGLS and PGEE models, where for this specific example, the assumptions of homoscedasticity and normality seem to be fully met (Figures 3.19, 3.20, 3.21). For the Bayesian approach using PGLMM, it needs to be checked whether the MCMC chains converged by plotting the traces of the sampled posteriors along the iterations. For the upper example, the trace plots show no increase or decrease, thus, the estimates seem to have reached convergence (Figure 3.22).
To sum up, in case of a continuous response variable, the four methods result in comparable parameter estimates and p-values and further, the diagnostic plot show that the underlying model assumptions are met. However, for the non-phylogenetic GLM it needs to be kept in mind that it shows elevated type I error rates (Section 3.1.1) and thus, should not be used in case of phylogenetically dependent data.

Table 3.6: **Application of the four methods on a real data set - continuous response model.** Testing the effect of brain size (continuous predictor) on the length of development (continuous response) correcting for body size (continuous predictor). Given are sample size (N), phylogenetic signal lambda (λ) in case of PGLS, and the estimate, standard error and p-value (bold if significant) of the explanatory variables brain size and body size. For PGLMM, the number of iterations is set to 50,000 and the estimates represent the posterior means. According to graphical inspection of the MCMC chains and the Geweke diagnostic test (p>0.05), the MCMC model has converged.

			brain mass			**body mass**		
Method	N	λ	estimate	std. error	p-value	estimate	std. error	p-value
Non-phylogenetic GLM	78		0.543	0.091	**0.000**	-0.060	0.075	0.425
PGLS	78	0.893	0.568	0.130	**0.000**	-0.150	0.096	0.121
PGEE	78		0.631	0.027	**0.000**	-0.161	0.023	**0.000**
PGLMM	78		0.573		**0.000**	-0.172		0.096

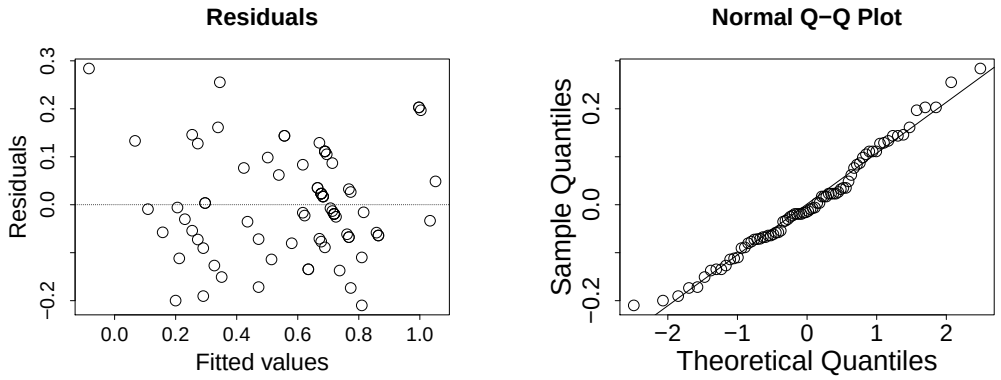


Figure 3.19: **Diagnostic plots from continuous response model - non-phylogenetic GLM.** The diagnostic plots serve to check the assumptions about the normal and homoscedastic distribution of the residuals. The plot on the left shows the residuals as a function of the fitted values. The plot on the right shows a normal quantile plot.

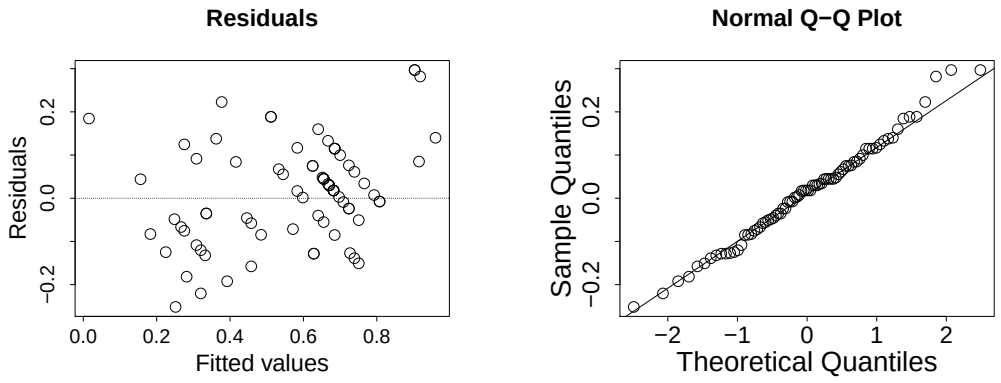


Figure 3.20: **Diagnostic plots from continuous response model - PGLS.** The diagnostic plots serve to check the assumptions about the normal and homoscedastic distribution of the residuals. The plot on the left shows the residuals as a function of the fitted values. The plot on the right shows a normal quantile plot.

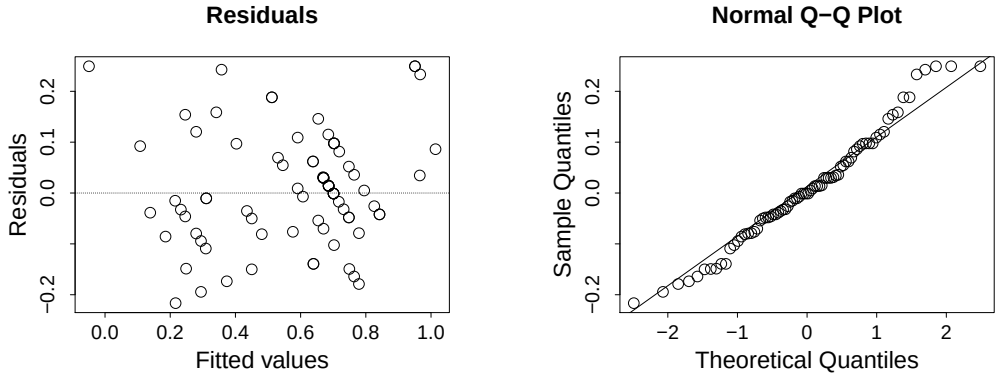


Figure 3.21: **Diagnostic plots from continuous response model - PGEE.** The diagnostic plots serve to check the assumptions about the normal and homoscedastic distribution of the residuals. The plot on the left shows the residuals as a function of the fitted values. The plot on the right shows a normal quantile plot.

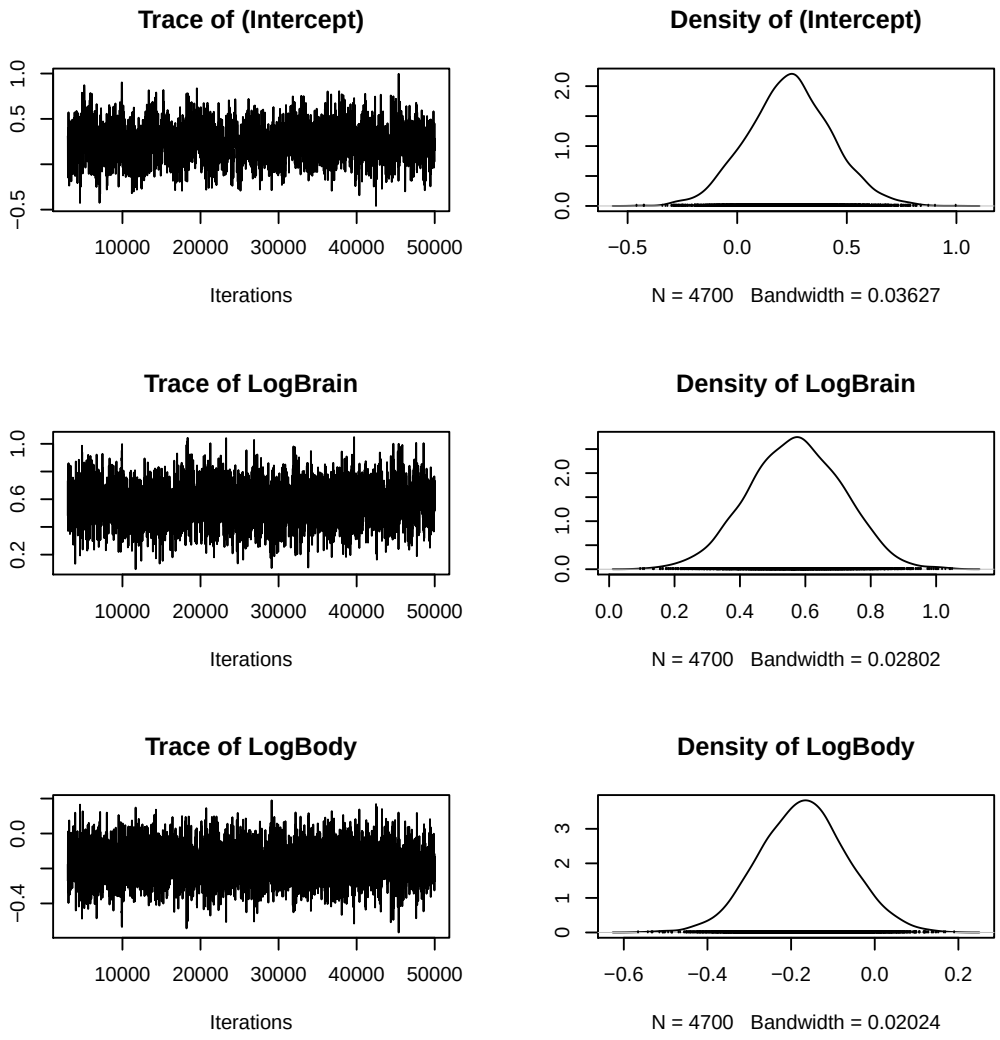


Figure 3.22: **Diagnostic plots from continuous response model - PGLMM.** The diagnostic plots of the Markov chains serve to check for convergence. The plots on the left show the traces of the sampled posteriors along the iterations. The plots on the right show the distributions of the sampled posteriors.

3.2 Ordinal response variable

To test the different phylogenetic methods on a nominal response variable, an ordinal scaled variable with four levels (also called categories) was simulated. Very often, ordinal scaled variables are treated as pseudo-continuous in comparative analyses. On the one hand, this can be explained by a lack of knowledge, and on the other hand, by a lack of available methods. In fact, until the development of the `MCMCglmm` package, no method was available modeling nominal or ordinal scaled variables with more than two levels. Therefore, this study tries to compare the analyses of ordinal variables treated as pseudo-continuous using PGLS and PGEE with the PGLMM, explicitly modeling multinomial logit models.

3.2.1 Non-phylogenetic GLM

The application of the non-phylogenetic GLM on the ordinal response assumes the variable to be pseudo-continuous.

Type I Error and Power

In respect to hypothesis testing, the non-phylogenetic method shows a large number of significantly elevated *type I error rates*, meaning the method declares statistical significance too often assuming the the null hypothesis to be true. As already noted in case of the continuous response variable, the error rates are especially pronounced in case of tree 3, however, the analyses based on tree 2 do not show significantly elevated type I error rates. Further, the type I error rates increase with an increasing sample size (Table 3.7, Figure 3.23).
The analyses of *power* including the simulated data with a input correlation coefficient of 0.9 implies a good performance of the method. The power reaches dominantly more than 90% and even increases for larger sample sizes (Table 3.7, Figure 3.24).

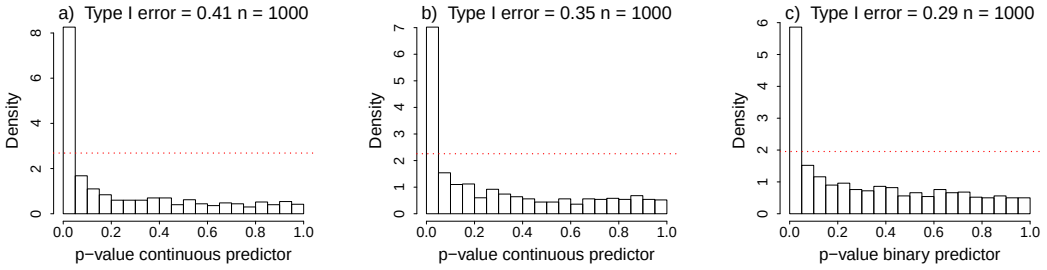


Figure 3.23: **Type I error rates from ordinal response models - non-phylogenetic GLM.** Distributions of p-values from data simulated with $r = 0.01$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The horizontal dashed line indicates the density at the critical value of 0.05.

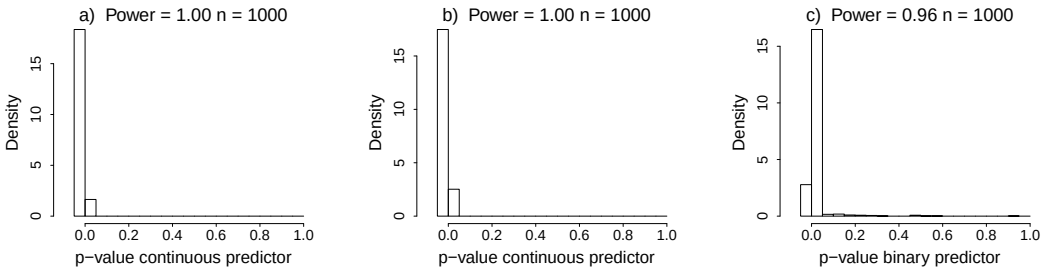


Figure 3.24: **Power from ordinal response models - non-phylogenetic GLM.** Distributions of p-values from data simulated with $r = 0.9$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The density bar from -0.1 to 0.0 represents the density of p-values equal to zero.

Table 3.7: **Type I error rates (TIE) and Power from ordinal response models - Non-phylogenetic GLM**: Fisher's exact test was used to compare the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values to the expected ratio (0.05/0.95) under the assumption of a type I error of $\alpha = 5\%$. The different significance levels of the Fisher's exact test are indicated by by bold font for $p < 0.001$ and in italics for $p < 0.05$. a) univariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.21**	0.99	**0.41**	1.00	**0.36**	1.00
Tree 2	0.05	1.00	0.06	1.00	0.05	1.00
Tree 3	**0.41**	0.97	**0.38**	1.00	**0.79**	0.99
Tree 4	**0.20**	1.00	**0.30**	1.00	**0.44**	1.00

b) multivariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.18**	0.99	**0.35**	1.00	**0.33**	1.00
Tree 2	0.04	1.00	0.06	1.00	0.05	1.00
Tree 3	**0.39**	0.96	**0.35**	1.00	**0.75**	0.98
Tree 4	**0.16**	1.00	**0.27**	1.00	**0.39**	1.00

c) multivariate model - binary predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.14**	0.83	**0.29**	0.96	**0.28**	1.00
Tree 2	0.05	0.87	0.05	0.99	0.05	1.00
Tree 3	**0.22**	0.69	**0.29**	0.90	**0.47**	0.91
Tree 4	**0.13**	0.85	**0.22**	0.98	**0.30**	0.99

Mean estimates (M), Mean Error (ME) and Rooted Mean Squared Error (RMSE)

Under different parameter conditions in terms of tree types and numbers of species, the *mean estimates* show higher variability compared to the case of the continuous response variable, specifically for the multivariate models (i.e. the continuous and binary predictor of the multivariate model).
The *mean errors* indicating the estimation bias, is comparably moderate for all conditions, except for tree type 3, where most species diversification was simulated to occur late in the evolutionary history. A similar pattern is given by the measurement for accuracy of the estimates, the *rooted mean squared error* where also a tree structure with most diversification occurring late in the evolutionary history leads to less accurate estimates. Moreover, tree 2, showing an opposite structure to tree 3, shows the lowest rooted mean squared errors. However, these patterns seen for tree 3 and tree 2 are not that pronounced for the binary predictor of the multivariate model.
Furthermore, as already noted in former analyses, the accuracy of the parameter estimates severely decrease for the binary predictor of the multivariate model and increasing sample sizes generally result in decreased mean and rooted mean squared errors. The corresponding results are found in Tables C.13 to C.15 in Appendix C.2.1 and Figures 3.25, 3.26.
To sum up, concerning hypothesis testing and the estimation abilities, the non-phylogenetic approach assuming pseudo-continuity of the ordinal response variable performs comparably to the continuous response models (Section 3.1.1)

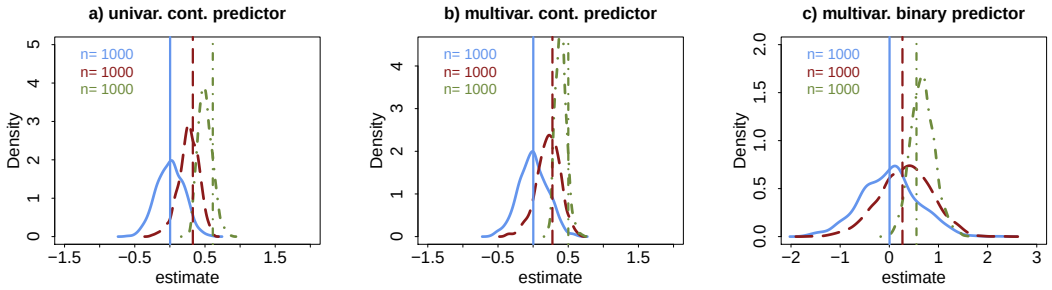


Figure 3.25: **Estimates from ordinal response models - Non-phylogenetic GLM.** Distributions of the effect sizes based on the simulated data with an input correlation coefficients of $r = 0.01$ (blue solid line), $r = 0.5$ (red dashed line) and $r = 0.9$ (green dashed-dotted line): a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model binary predictor including the data of tree 1 and 50 species. The vertical lines indicate the means overall methods. The bandwidth of the densities was set to the default (`"nrd0"`).

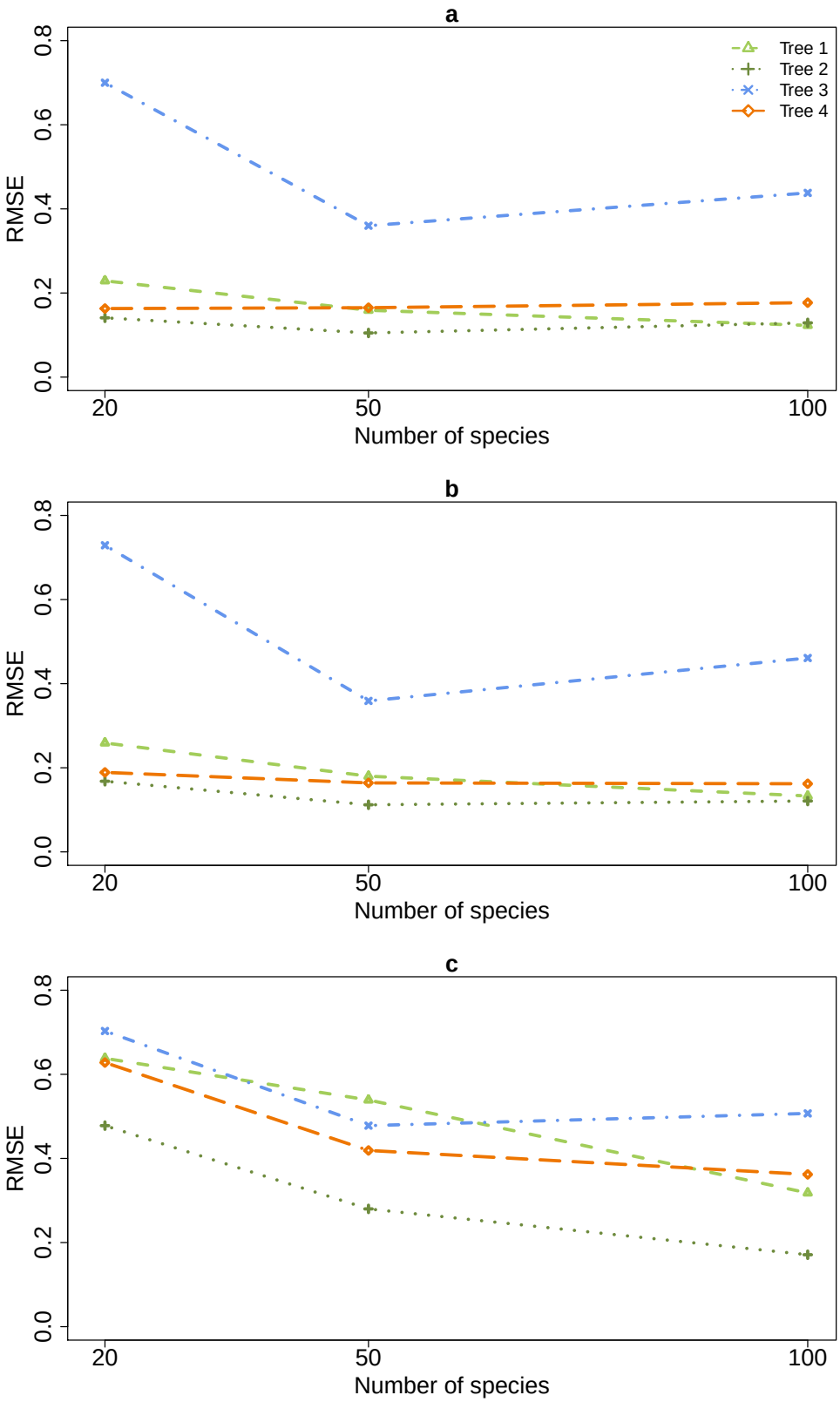


Figure 3.26: **Rooted mean squared errors from ordinal response models - Non-phylogenetic GLM.** The rooted mean squared errors (RMSE) are shown for the data simulated with a correlation coefficient of $r = 0.5$. a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

3.2.2 PGLS

The PGLS regression applied on an ordinal scaled response variable also assumes pseudo-continuity of the dependent variable.

Type I Error and Power

The *type I error rates* are strongly reduced compared to the non-phylogenetic approach, where almost for every parameter condition, α (i.e. type I error) was significantly higher than expected (Table 3.8 and Figure 3.27). However, as for the continuous response models, especially smaller sample sizes of 20 species mostly result in false declared significance. But not so in case of tree structure 2, where the shared evolutionary history between species is rather short.
Power generally reaches 90% and higher implying a good performance concerning hypothesis testing (Table 3.8 and Figure 3.28).

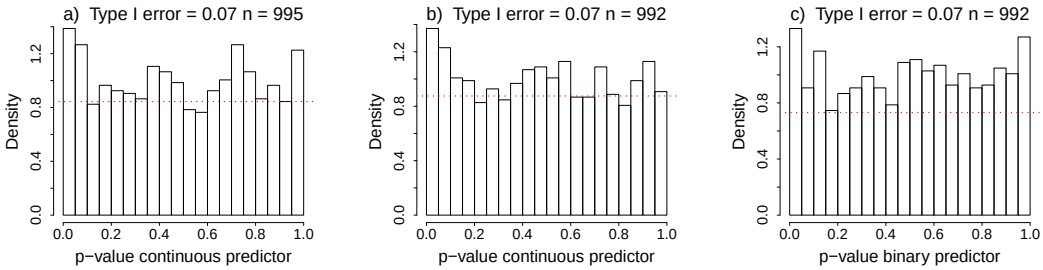


Figure 3.27: **Type I error rates from ordinal response models - PGLS.** Distributions of p-values from data simulated with $r = 0.01$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The horizontal dashed line indicates the density at the critical value of 0.05.

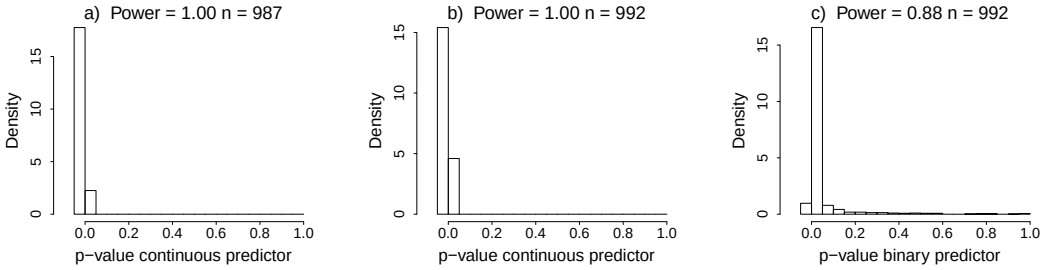


Figure 3.28: **Power from ordinal response models - PGLS.** Distributions of p-values from data simulated with $r = 0.9$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The density bar from -0.1 to 0.0 represents the density of p-values equal to zero.

Table 3.8: **Type I error rates (TIE) and Power from ordinal response models - PGLS**: Fisher's exact test was used to compare the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values to the expected ratio (0.05/0.95) under the assumption of a type I error of $\alpha = 5\%$. The different significance levels of the Fisher's exact test are indicated by bold font for $p < 0.001$ and in italics for $p < 0.05$. a) univariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.10**	1.00	0.07	1.00	0.06	1.00
Tree 2	0.06	1.00	0.06	1.00	0.05	1.00
Tree 3	**0.16**	0.96	0.06	1.00	0.05	0.99
Tree 4	*0.09*	1.00	0.06	1.00	0.05	1.00

b) multivariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.10**	0.99	0.07	1.00	0.06	1.00
Tree 2	0.05	1.00	0.06	1.00	0.06	1.00
Tree 3	**0.17**	0.92	0.07	0.99	0.05	0.97
Tree 4	*0.09*	1.00	0.06	1.00	0.05	1.00

c) multivariate model - binary predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	*0.09*	0.77	0.07	0.88	*0.07*	0.97
Tree 2	0.05	0.87	0.05	0.99	0.05	1.00
Tree 3	**0.12**	0.63	*0.08*	0.71	0.07	0.70
Tree 4	*0.09*	0.76	0.06	0.93	0.05	0.95

Mean estimates (M), Mean Error (ME) and Rooted Mean Squared Error (RMSE)

Concerning the estimation abilities of the PGLS regression for a pseudo-continuous response variable, the pattern looks similar as for the non-phylogenetic approach, however, here phylogenetic dependency is taken into account. Especially conspicuous are the *mean errors* for tree structure 3 with long shared evolutionary paths (i.e. high phylogenetic covariances), however for the binary predictor of the multivariate model this discrepancy disappears. In other words, a high phylogenetic signal leads to a larger estimation bias in continuous predictors. This pattern is also reflected in the *rooted mean squared error* indicating less accurate estimates for tree type 3. Further, also the binary predictor of the multivariate model shows the least accurate parameter estimates and generally, the accuracy increases with increasing sample size. The corresponding results are shown in Tables C.16 to C.18 in Appendix C.2.2 and Figures 3.29, 3.30.

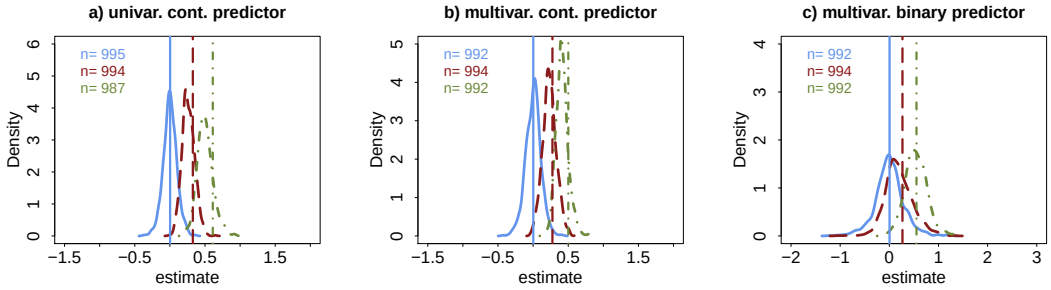


Figure 3.29: **Estimates from ordinal response models - PGLS.** Distributions of the effect sizes based on the simulated data with an input correlation coefficients of $r = 0.01$ (blue solid line), $r = 0.5$ (red dashed line) and $r = 0.9$ (green dashed-dotted line): a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model binary predictor including the data of tree 1 and 50 species. The vertical lines indicate the means overall methods. The bandwidth of the densities was set to the default (`"nrd0"`).

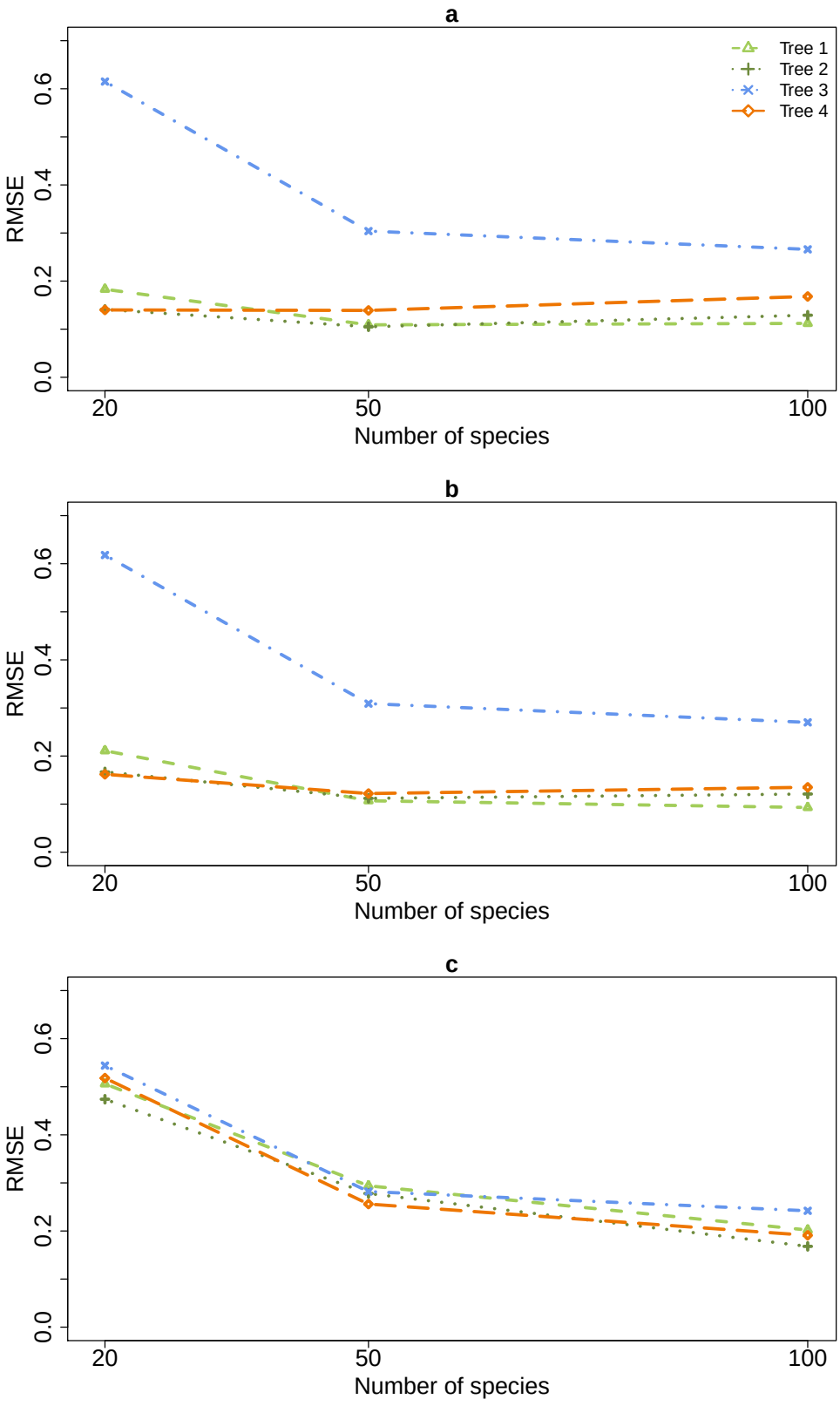


Figure 3.30: **Rooted mean squared errors from ordinal response models - PGLS.** The rooted mean squared errors (RMSE) are shown for the data simulated with a correlation coefficient of $r = 0.5$. a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

Phylogenetic signal λ

The phylogenetic signal λ estimated by the PGLS regression are basically the same as in case of continuous response variable (Section 3.1.2). The phylogenetic models including tree 2, showing short shared evolutionary histories between species, show rather low phylogenetic signals (i.e $\lambda < 0.4$), whereas the opposite tree structure, tree 3, shows the highest phylogenetic signals. In other words, trait variation between species can only be explained by a small extent by common evolutionary history in case of tree 2, whereas the opposite is true for tree 3. The corresponding results are found in Table 3.9.

Table 3.9: **Mean phylogenetic signal λ with the standard deviation in brackets from ordinal response models - PGLS**: a) univariate model.

	Species 20	**Species 50**	**Species 100**
tree 1	0.63 (0.39)	0.84 (0.23)	0.86 (0.17)
tree 2	0.24 (0.42)	0.35 (0.45)	0.37 (0.43)
tree 3	0.71 (0.42)	0.94 (0.17)	0.97 (0.1)
tree 4	0.72 (0.35)	0.77 (0.24)	0.83 (0.17)

b) multivariate model.

	Species 20	**Species 50**	**Species 100**
tree 1	0.53 (0.43)	0.74 (0.34)	0.77 (0.28)
tree 2	0.24 (0.42)	0.32 (0.44)	0.33 (0.42)
tree 3	0.58 (0.46)	0.87 (0.29)	0.94 (0.19)
tree 4	0.62 (0.41)	0.66 (0.34)	0.74 (0.28)

3.2.3 PGEE

Type I Error and Power

PGEE modeling the correlated evolution of a pseudo-continuous response variable shows higher frequencies of significantly elevated *type I error rates* compared to the analogues analysis using PGLS. In particular, these false positive declarations are observed for tree 1 and 4 and sample sizes of 50 and more species (Table 3.10, Figure 3.31).
The *power* analysis shows as always a very good performance (Table 3.10 and Figure 3.32).
Due to unknown reasons, the analyses of the multivariate model with the simulated data based on tree 3 and 20 species did not produce any p-values, indicated by NA in Table 3.10.

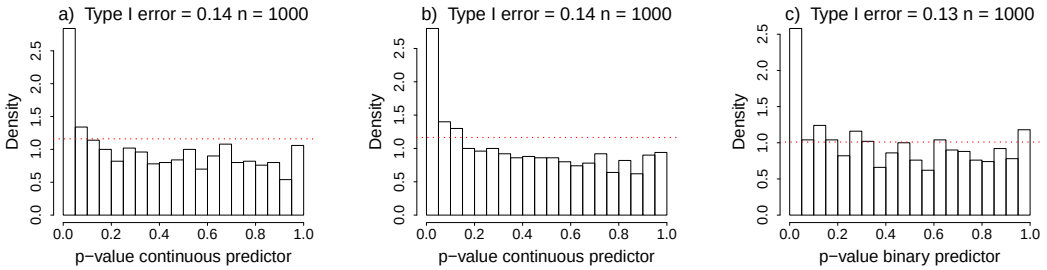


Figure 3.31: **Type I error rates from ordinal response models - PGEE.** Distributions of p-values from data simulated with $r = 0.01$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The horizontal dashed line indicates the density at the critical value of 0.05.

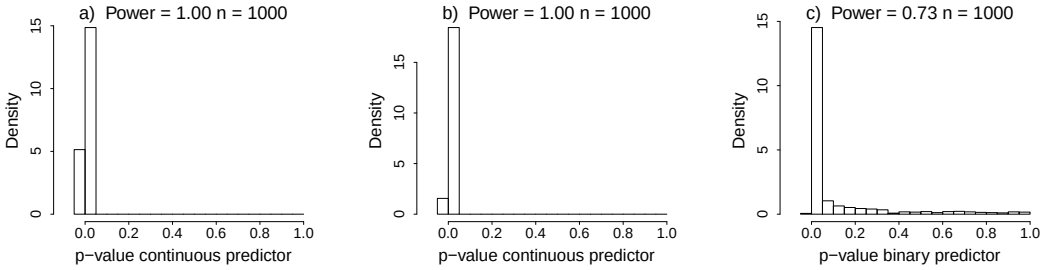


Figure 3.32: **Power from ordinal response models - PGEE.** Distributions of p-values from data simulated with $r = 0.9$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The density bar from -0.1 to 0.0 represents the density of p-values equal to zero.

Table 3.10: **Type I error rates (TIE) and Power from ordinal response models - PGEE**: Fisher's exact test was used to compare the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values to the expected ratio (0.05/0.95) under the assumption of a type I error of $\alpha = 5\%$. The different significance levels of the Fisher's exact test are indicated by bold font for $p < 0.001$ and in italics for $p < 0.05$. a) univariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	*0.06*	1.00	**0.14**	1.00	**0.14**	1.00
Tree 2	0.06	1.00	0.06	1.00	0.06	1.00
Tree 3	0.00	0.05	**0.13**	0.98	**0.17**	0.94
Tree 4	*0.07*	1.00	*0.08*	1.00	**0.14**	1.00

b) multivariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	0.06	1.00	**0.14**	1.00	**0.15**	1.00
Tree 2	0.05	1.00	0.06	1.00	0.06	1.00
Tree 3	NA	NA	0.02	0.87	0.00	0.47
Tree 4	0.07	1.00	*0.08*	1.00	**0.14**	1.00

c) multivariate model - binary predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	*0.07*	0.68	**0.13**	0.73	**0.16**	0.80
Tree 2	0.06	0.87	0.05	0.99	0.05	1.00
Tree 3	NA	NA	0.03	0.25	0.00	0.09
Tree 4	0.06	0.68	**0.09**	0.86	**0.13**	0.91

Mean estimates (M), Mean Error (ME) and Rooted Mean Squared Error (RMSE)

The estimation abilities of PGEE for a pseudo-continuous response is comparable to the PGLS regression. The *mean errors* as well as the *rooted mean squared error* are especially pronounced for tree 3 in case of the continuous predictors of the univariate and multivariate models. Moreover, the binary predictor of the multivariate model shows an increased bias and decreased estimation accuracy especially for small sample sizes. However, generally, increasing sample sizes lead to improved estimation abilities. The corresponding results are shown in Tables C.19 to C.21 in Appendix C.2.3 and Figures 3.33, 3.56.

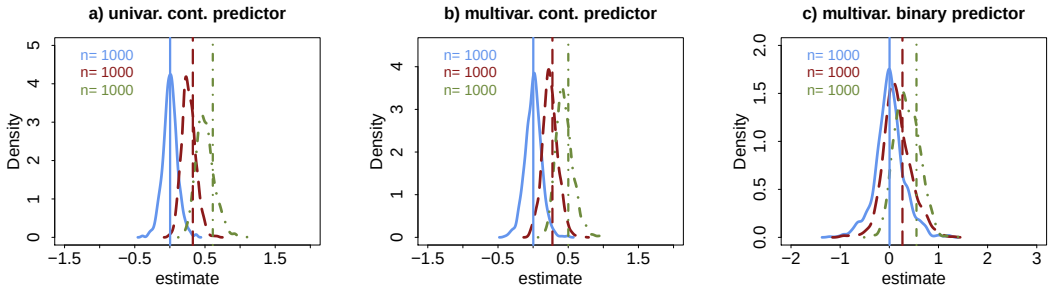


Figure 3.33: **Estimates from ordinal response models - PGEE.** Distributions of the effect sizes based on the simulated data with an input correlation coefficients of $r = 0.01$ (blue solid line), $r = 0.5$ (red dashed line) and $r = 0.9$ (green dashed-dotted line): a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model binary predictor including the data of tree 1 and 50 species. The vertical lines indicate the means overall methods. The bandwidth of the densities was set to the default (`"nrd0"`).

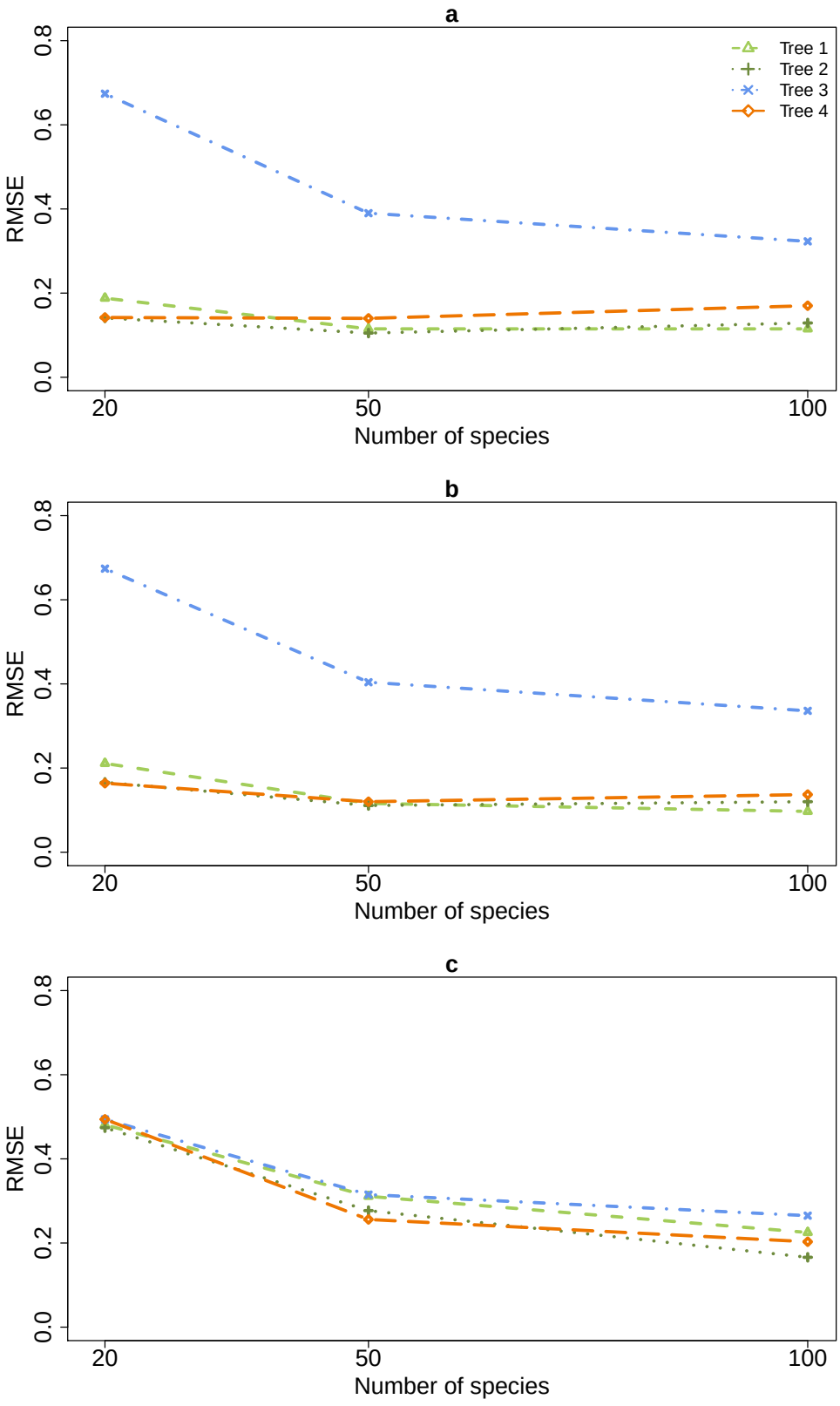


Figure 3.34: **Rooted mean squared errors from ordinal response models - PGEE.** The rooted mean squared errors (RMSE) are shown for the data simulated with a correlation coefficient of $r = 0.5$. a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

3.2.4 PGLMM

One special feature of the phylogenetic mixed model based on MCMC is the multinomial logit model for nominal variables with more than two levels. Instead of treating the ordinal scaled response variable as pseudo-continuous as for the non-phylogenetic, PGLS and PGEE approach, this method measures the effects on the probabilities of being in a certain level compared to the reference level. The simulated ordinal variable contains four levels, thus, the PGLMM estimates the effect on the probabilities of being in category 2, 3 and 4 compared to the reference category 1.
The following analyses on hypothesis testing and estimation abilities of the PGLMM method are based on the effects on the probability of category 4 compared to the reference category 1. In case of a correlation between the ordinal scaled response and the predictor variables, category 4 vs. category 1 should show the strongest effect.

Type I Error and Power

On the one hand, in case of all parameter combinations (different tree types and numbers of species), *type I error rates* are significantly elevated. The analyses based on tree structure 3, with the highest phylogenetic signal among the four trees, show the highest rates of type I error (Table 3.11 and Figure 3.35). On the other hand, the *power* analysis shows a better performance mostly reaching 60% and higher, but in case of the binary predictor of the multivariate model these percentages are decreased again. As usual, the power generally increases with increasing sample size (Table 3.11 and Figure 3.36).

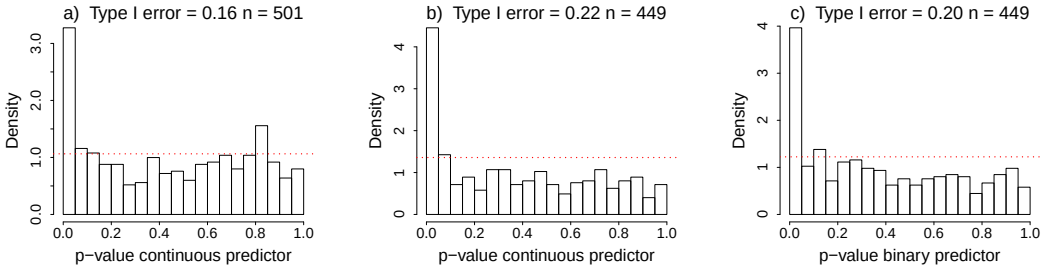


Figure 3.35: **Type I error rates from ordinal response models - PGLMM.** Distributions of p-values from data simulated with $r = 0.01$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The horizontal dashed line indicates the density at the critical value of 0.05.

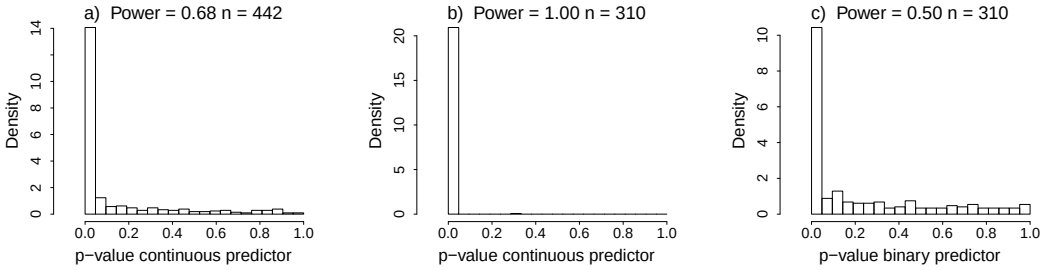


Figure 3.36: **Power from ordinal response models - PGLMM.** Distributions of p-values from data simulated with $r = 0.9$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

Table 3.11: **Type I error rates (TIE) and Power from ordinal response models - PGLMM**: Fisher's exact test was used to compare the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values to the expected ratio (0.05/0.95) under the assumption of a type I error of $\alpha = 5\%$. The different significance levels of the Fisher's exact test are indicated by bold font for $p < 0.001$ and in italics for $p < 0.05$. a) univariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.15**	0.66	**0.16**	0.68	**0.21**	0.80
Tree 2	**0.14**	0.72	**0.30**	0.87	**0.31**	0.94
Tree 3	**0.25**	0.66	**0.26**	0.78	**0.31**	0.81
Tree 4	**0.16**	0.57	**0.15**	0.72	**0.26**	0.78

b) multivariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.19**	0.89	**0.22**	1.00	**0.17**	1.00
Tree 2	**0.17**	0.95	**0.13**	1.00	**0.14**	1.00
Tree 3	**0.32**	0.78	**0.26**	0.97	**0.42**	0.97
Tree 4	**0.24**	0.94	**0.19**	1.00	**0.15**	1.00

c) multivariate model - binary predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.19**	0.33	**0.20**	0.50	**0.18**	0.62
Tree 2	**0.14**	0.34	**0.14**	0.60	**0.13**	0.79
Tree 3	**0.21**	0.31	**0.20**	0.45	**0.18**	0.44
Tree 4	**0.23**	0.41	**0.18**	0.56	**0.14**	0.61

Mean estimates (M), Mean Error (ME) and Rooted Mean Squared Error (RMSE)

Not only concerning hypothesis testing but also the estimation abilities don‘t show a very good performance using PGLMM for the multinomial model. The *mean error* as well as the *rooted mean squared error* are strongly elevated compared to the other methods. Again, especially tree 3 shows strongly biased and inaccurate estimates in case of continuous predictors. Moreover, the binary predictor of the multivariate model shows higher errors and increasing sample size leads to less biased and more accurate parameter estimates. Further, it is important to note that the number of simulations after excluding non-converged models, are remarkably small. In other words, in about 50% of the cases, multinomial logit models with PGLMM do not converge, thus, the results are not meaningful. The corresponding results are found in Tables C.22 to C.24 in Appendix C.2.4 and Figures 3.37, 3.38.

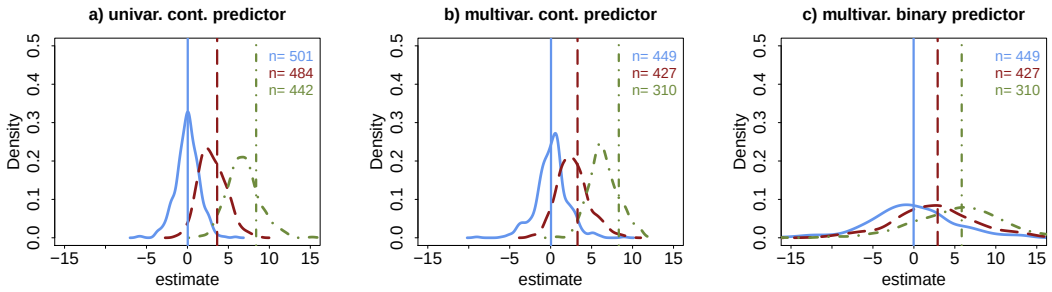


Figure 3.37: **Estimates from nominal response models - PGLMM.** Distributions of the effect sizes based on the simulated data with an input correlation coefficients of $r = 0.01$ (blue solid line), $r = 0.5$ (red dashed line) and $r = 0.9$ (green dashed-dotted line): a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model binary predictor including the data of tree 1 and 50 species. The vertical lines indicate the means overall methods.

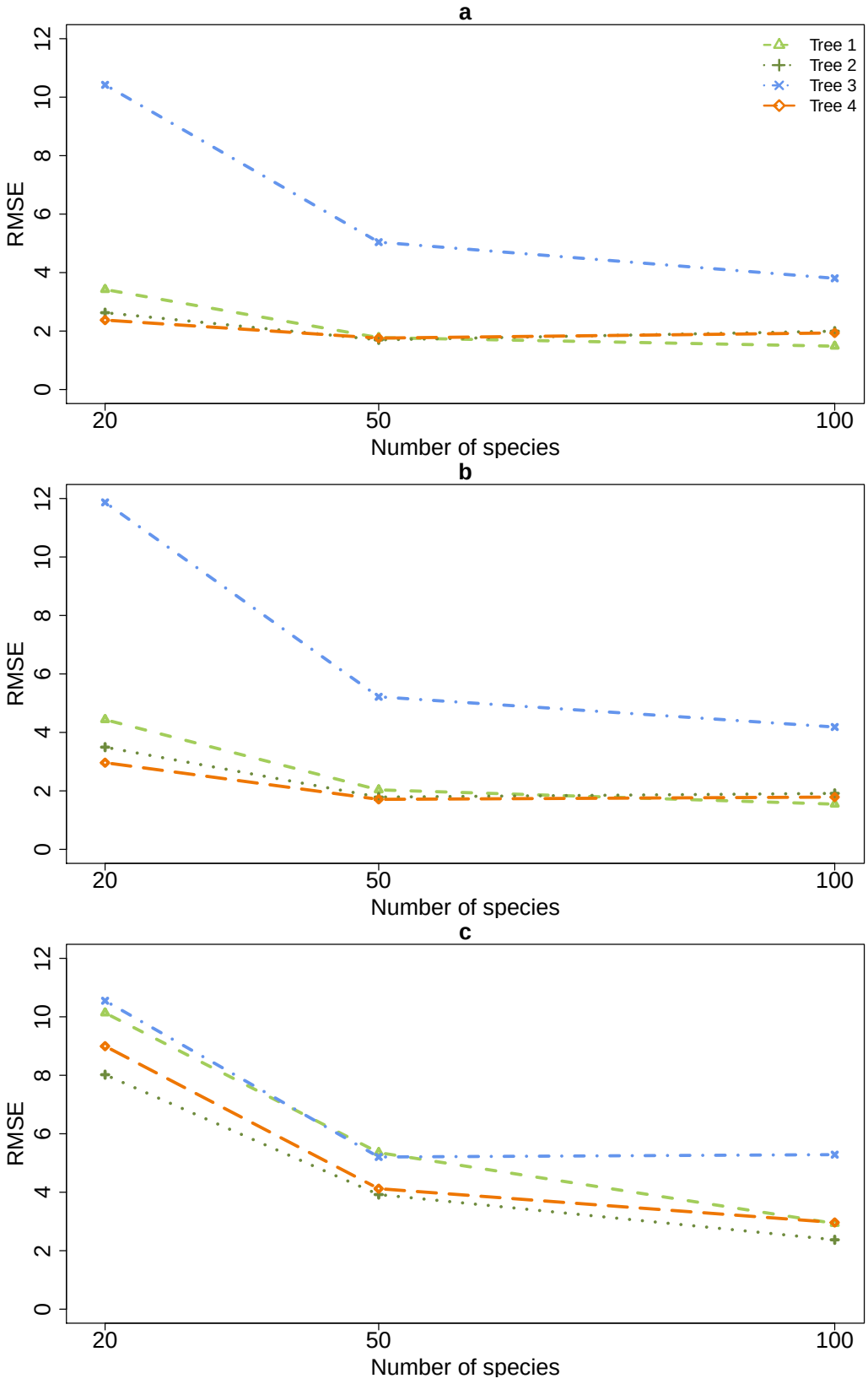


Figure 3.38: **Rooted mean squared errors from ordinal response models - PGLMM.** The rooted mean squared errors (RMSE) are shown for the data simulated with a correlation coefficient of $r = 0.5$. a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

3.2.5 The four methods in comparison

Simulated data

The four methods in direct comparison (Figure 3.39) shows that the mean parameter estimates are very similar for the non-phylogenetic, the PGLS and the PGEE method for an ordinal (i.e. pseudo-continuous) response variable. Also the 95% confidence intervals are very narrow indicating a small variability in the parameter estimates. The PGLMM method, modeling multinomial logit models, measures a different effect, in fact the effect on the probabilities of being in a certain category compared to a reference category. Therefore, the mean estimates of these Bayesian models are different from the other approaches. However, they show a much higher variability in the parameter estimates, and as already seen in the former section, show a much larger estimation bias and a higher inaccuracy in the parameter estimates.
The parameter estimates in pairwise comparison between the four methods (Figure 3.40) shows that the variability in the non-phylogenetic method is much higher compared to PGLS and PGEE. Further, the estimates from the PGLMM method are generally much higher, keeping in mind that they measure rather the effect on probabilities of one category compared to a reference category than the effect on a single unit increase in the dependent variable. Correspondingly, also the variability in the PGLMM estimates is higher.

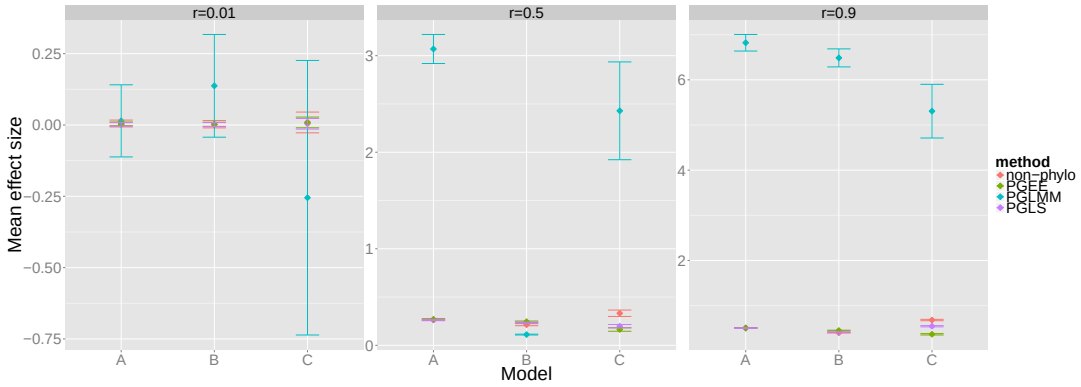


Figure 3.39: **Mean estimates and 95%-confidence intervals from ordinal response models.** Comparison of the mean estimates and corresponding 95%-confidence intervals from the four methods for the three input correlation coefficients including the simulated data for tree 1 and 50 species (number of simulations = 1000). A: univariate model - continuous predictor; B: multivariate model - continuous predictor; C: multivariate model - binary predictor.

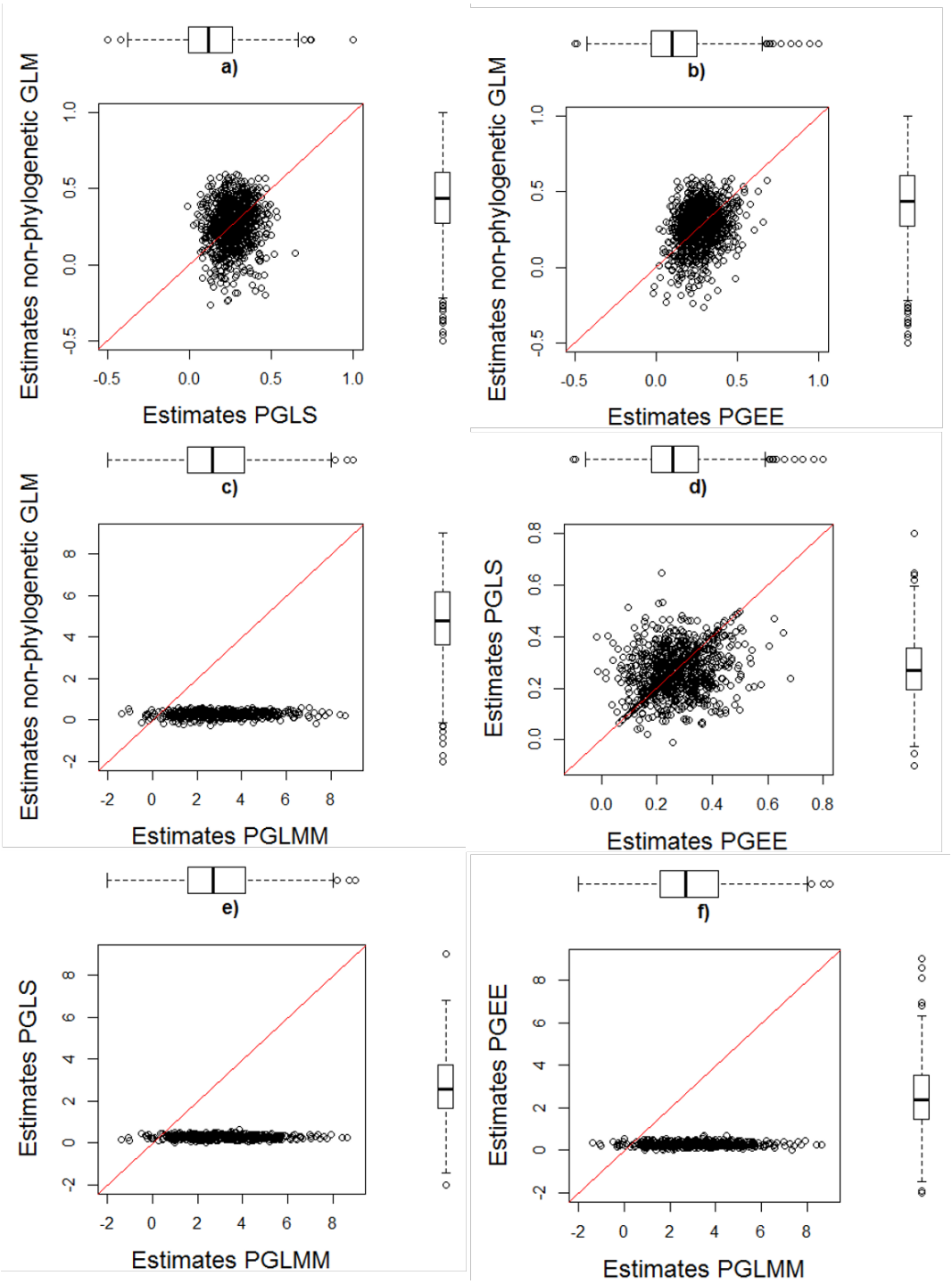


Figure 3.40: **Parameter estimates compared.** The parameter estimates based on the simulated data with $r = 0.5$ and 50 species from the non-phylogenetic and three phylogenetic methods (ordinal response models) plotted against each other pairwise with the corresponding boxplots aside: a) Non-phylogenetic GLM vs. PGLS; b) Non-phylogenetic vs. PGEE; c) Non-phylogenetic vs. PGLMM; d) PGLS vs. PGEE; e) PGLS vs. PGLLMM; f) PGEE vs. PGLMM.

Application on a real data set

In order to compare the four methods directly, they were again applied on a real evolutionary question based on the primate data set including 78 species. The application on an ordinal response variable is based on the second hypothesis described in Material & Methods (Section 2.2). Smarter species with larger relative brain sizes are expected to live in more complex foraging niches compared their smaller brained relatives (Schuppli et al. 2012, Schuppli et al. in prep., Graber et al. in prep.). Thus, the hypothesis predicts a positive correlation between relative brain size (corrected for body size) and the foraging niche complexity, representing the ordinal response variable.
The non-phylogenetic approach and the two phylogenetic methods (i.e. PGLS and PGEE) treating the ordinal scaled response variable (foraging niche complexity) as pseudo-continuous show very similar parameter estimates as well as p-values for both continuous predictor variables, brain size and body size (Table 3.12). The PGLMM method, on the other hand, shows a different parameter estimate with a different meaning (i.e. effect on probability of on category vs. reference category), but the p-values lead in the end to the same conclusion as for the other three methods. In fact, there is a strong positive effect of relative brain size on niche complexity, meaning that larger brained species managed to live in more complex foraging niches.
The corresponding diagnostic plots are shown in Figures 3.41, 3.42, 3.43 and 3.44. The plots showing the residuals vs. the fitted values show a structure according to the four levels of the ordinal response variable, which probably indicates a violation of the model assumptions. The qq-plots show that the residuals are probably not perfectly distributed according to normality, however, are also not too bad. Furthermore, considering the MCMC traces to check for convergence of the Bayesian approach, it is obvious that the model has not converged, which is also confirmed by the Geweke diagnostic test ($p<0.05$). In other words, the results by the PGLMM are probably not meaningful, as the model, even after 50,000 iterations, has not reached convergence. However, further studies on the prior settings in `MCMCglmm` are needed in order to properly judge its statistical performance.

Table 3.12: **Application of the four methods on a real data set - ordinal response model.** Testing the effect of brain size (continuous predictor) on niche complexity (ordinal response) correcting for body size (continuous predictor). Given are sample size (N), phylogenetic signal lambda (λ) in case of PGLS, and the estimate, standard error and p-value (bold if significant) of the explanatory variables brain size and body size. For PGLMM, the number of iterations is set to 50,000 and the estimates represent the posterior mean of the effects of brain size and body size on the probability of the highest niche complexity level versus the lowest niche complexity level. Important to note is that the PGLMM model seems not to have converged (Geweke diagnostic test: $p<0.05$).

			brain mass			**body mass**		
Method	N	λ	estimate	std. error	p-value	estimate	std. error	p-value
Non-phylogenetic GLM	78		2.600	0.543	**0.000**	-1.576	0.447	**0.001**
PGLS	78	0.864	2.389	0.760	**0.002**	-1.446	0.564	**0.012**
PGEE	78		2.028	0.163	**0.000**	-1.266	0.136	**0.000**
PGLMM	78		59.699		**0.000**	-34.078		**0.005**

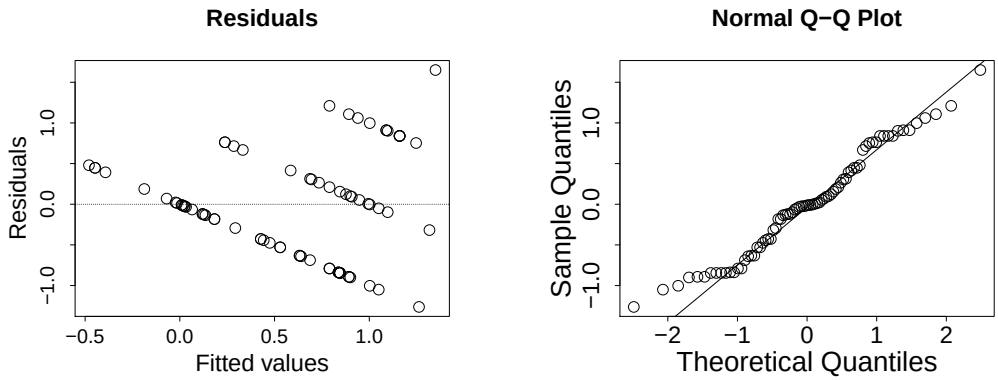


Figure 3.41: **Diagnostic plots from ordinal response model - non-phylogenetic GLM.** The diagnostic plots serve to check the assumptions about the normal and homoscedastic distribution of the residuals. The plot on the left shows the residuals as a function of the fitted values. The plot on the right shows a normal quantile plot.

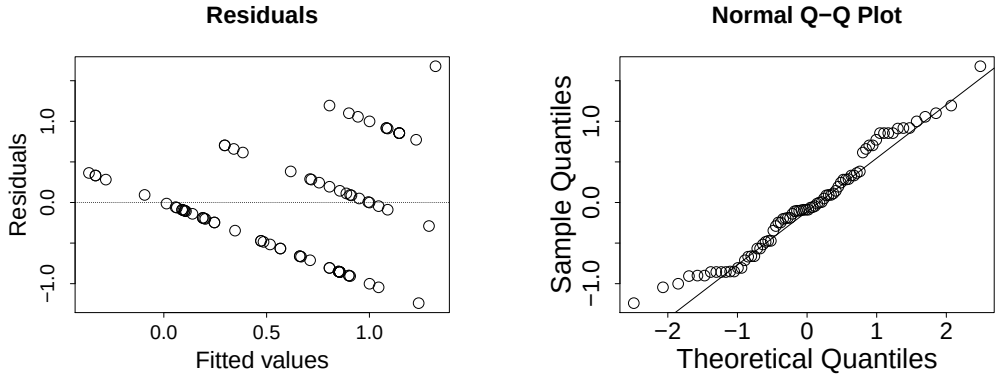


Figure 3.42: **Diagnostic plots from ordinal response model - PGLS.** The diagnostic plots serve to check the assumptions about the normal and homoscedastic distribution of the residuals. The plot on the left shows the residuals as a function of the fitted values. The plot on the right shows a normal quantile plot.

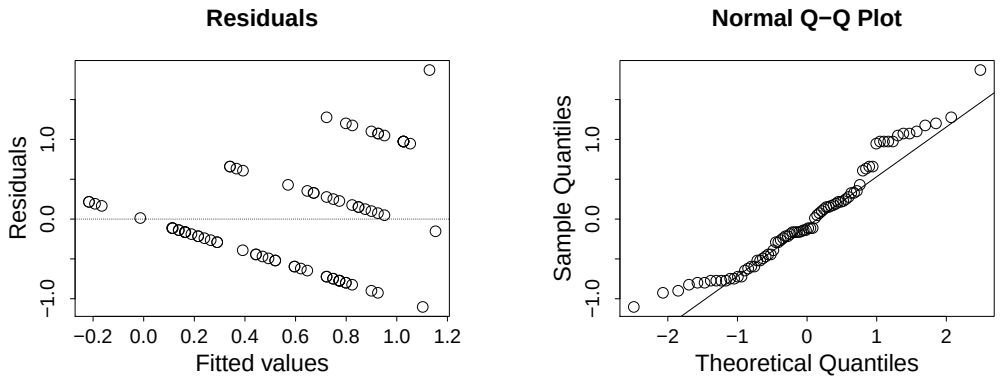


Figure 3.43: **Diagnostic plots from ordinal response model - PGEE.** The diagnostic plots serve to check the assumptions about the normal and homoscedastic distribution of the residuals. The plot on the left shows the residuals as a function of the fitted values. The plot on the right shows a normal quantile plot.

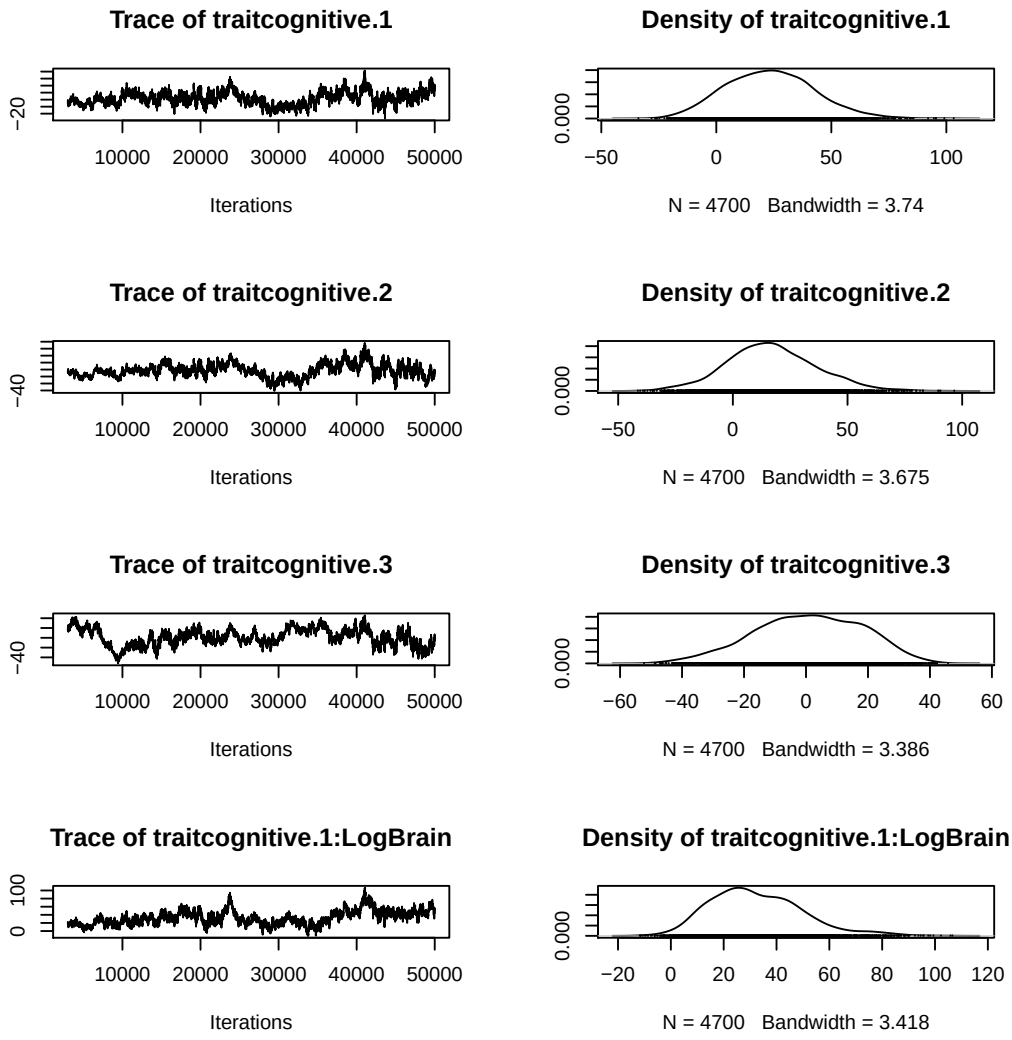


Figure 3.44: **Diagnostic plots from ordinal response model - PGLMM.** The diagnostic plots of the Markov chains serve to check for convergence. The plots on the left show the traces of the sampled posteriors along the iterations. The plots on the right show the distributions of the sampled posteriors.

3.3 Binary response variable

The simulated binary response was modeled using three different approaches: the non-phylogenetic GLM in form of a simple logistic regression, the Bayesian approach with PGLMM and the phylogenetic logistic regression (PLR). At this point, it is important to note two issues about the following results. First, the method by Paradis and Claude (2002) using generalised estimating equations (PGEE) is also implemented for binary responses, however, certain data constellations in form of phylogenetic relatedness structures lead to freezing of `R`, which prevents using this method in a simulation study. Thus, theoretically, this method can be used to analyse binary data, however, in some cases might overburden the statistical program. Second, the analyses of the PLR are only based on maximal 100 simulations, due to two reasons. First, the implementation of this method in `R` is very time consuming and second, several crashes of the server disrupted the simulation loops.

3.3.1 Non-phylogenetic GLM

Type I Error and Power

Not surprisingly based on the analyses of the continuous and ordinal response using a non-phylogenetic approach, modeling binary data with a simple logistic regression model also leads in most cases to elevated *type I error rates*, with the analyses based on tree 3 showing the highest error rates (Table 3.13 and Figure 3.45). However, tree 2 makes again an exception, where most of the species diversification was simulated to occur early in the history of the evolution of a trait. Furthermore, the type I error rates tend to increase for larger samples sizes, as well as the *power*. The power with an average of ca. 80% shows a rather good performance at least for sample sizes of 50 and more, however, not so for a sample size of 20 species. There, especially for the multivariate model, power is strongly decreased (Table 3.13 and Figure 3.46).

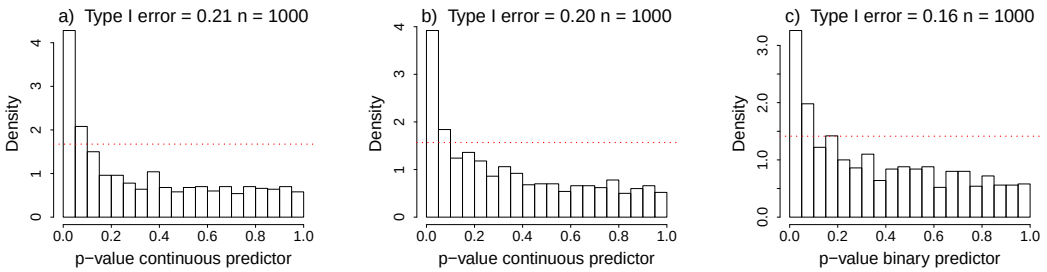


Figure 3.45: **Type I error rates from binary response models - non-phylogenetic GLM.** Distributions of p-values from data simulated with $r = 0.01$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The horizontal dashed line indicates the density at the critical value of 0.05.

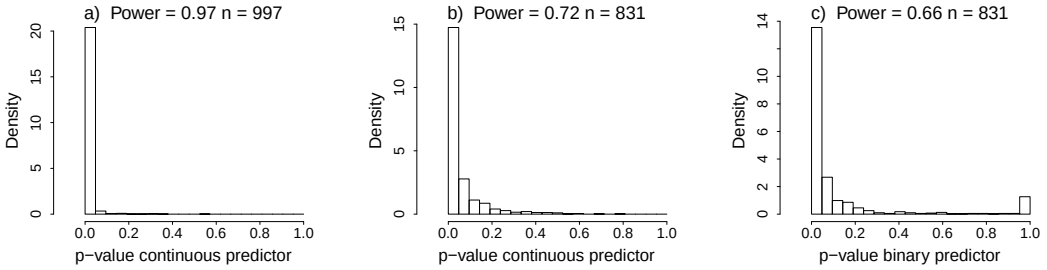


Figure 3.46: **Power from binary response models - non-phylogenetic GLM.** Distributions of p-values from data simulated with $r = 0.9$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

Table 3.13: **Type I error rates (TIE) and Power from binary response models - Non-phylogenetic GLM**: Fisher's exact test was used to compare the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values to the expected ratio (0.05/0.95) under the assumption of a type I error of $\alpha = 5\%$. The different significance levels of the Fisher's exact test are indicated by bold font for $p < 0.001$ and in italics for $p < 0.05$. a) univariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.13**	0.56	**0.21**	0.97	**0.33**	1.00
Tree 2	0.02	0.60	0.04	1.00	0.06	1.00
Tree 3	**0.39**	0.31	**0.34**	0.96	**0.77**	0.90
Tree 4	**0.12**	0.60	**0.28**	0.98	**0.32**	1.00

b) multivariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	*0.08*	0.07	**0.20**	0.72	**0.31**	0.97
Tree 2	0.02	0.09	0.05	0.78	0.06	0.99
Tree 3	0.07	0.04	**0.27**	0.69	**0.56**	0.85
Tree 4	*0.07*	0.07	**0.23**	0.72	**0.29**	0.98

c) multivariate model - binary predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	*0.09*	0.05	**0.16**	0.66	**0.23**	0.96
Tree 2	0.03	0.05	0.05	0.73	0.05	0.99
Tree 3	*0.09*	0.03	**0.25**	0.58	**0.50**	0.73
Tree 4	*0.08*	0.06	**0.18**	0.64	**0.23**	0.97

Mean estimates (M), Mean Error (ME) and Rooted Mean Squared Error (RMSE)

First, the *mean estimates* for the different simulation conditions (tree structure, sample size, strength of correlation) show much higher variation compared to the continuous and ordinal response models.
The *mean error* as an indication for estimation bias is mainly highest for tree 3. Moreover, estimation bias is generally highest for strong correlations (i.e. input correlation $r = 0.9$).
The *rooted mean squared error* is highly elevated for tree structure 3, meaning that models with a phylogeny where most diversification occurs late in the evolutionary history lead to inaccurate parameter estimates. On the other hand, tree 2 shows rather lower mean squared errors. These differences are especially pronounced for a small sample size ($n = 20$), whereas for larger sample sizes, these differences disappear. Furthermore, the multivariate models in case of strong correlations (input correlation $r = 0.9$) lead to extreme high parameter estimates in a about 5% of the cases. This also explains the extreme high mean errors and rooted mean squared errors for the multivariate models with an input correlation of 0.9 for all methods, as the mean over all methods serves as reference for the error calculations. The corresponding results are shown in Tables C.25 to C.27 in Appendix C.3.1 and Figures 3.47, 3.48.

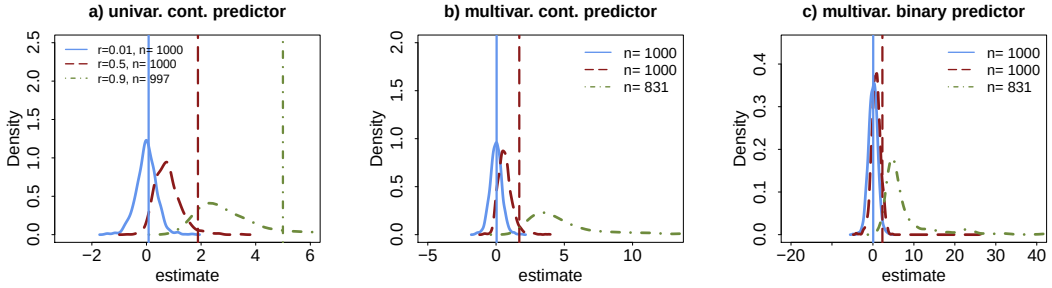


Figure 3.47: **Estimates from binary response models - Non-phylogenetic GLM.** Distributions of the effect sizes based on the simulated data with an input correlation coefficients of $r = 0.01$ (blue solid line), $r = 0.5$ (red dashed line) and $r = 0.9$ (green dashed-dotted line): a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model binary predictor including the data of tree 1 and 50 species. The vertical lines indicate the means overall methods. The overall means for the continuous and binary predictor of the multivariate model with an input correlation of $r = 0.9$ is very large ($> 10^{11}$), and thus, not shown in the graph. The bandwidth of the densities was set to the default (`"nrd0"`).

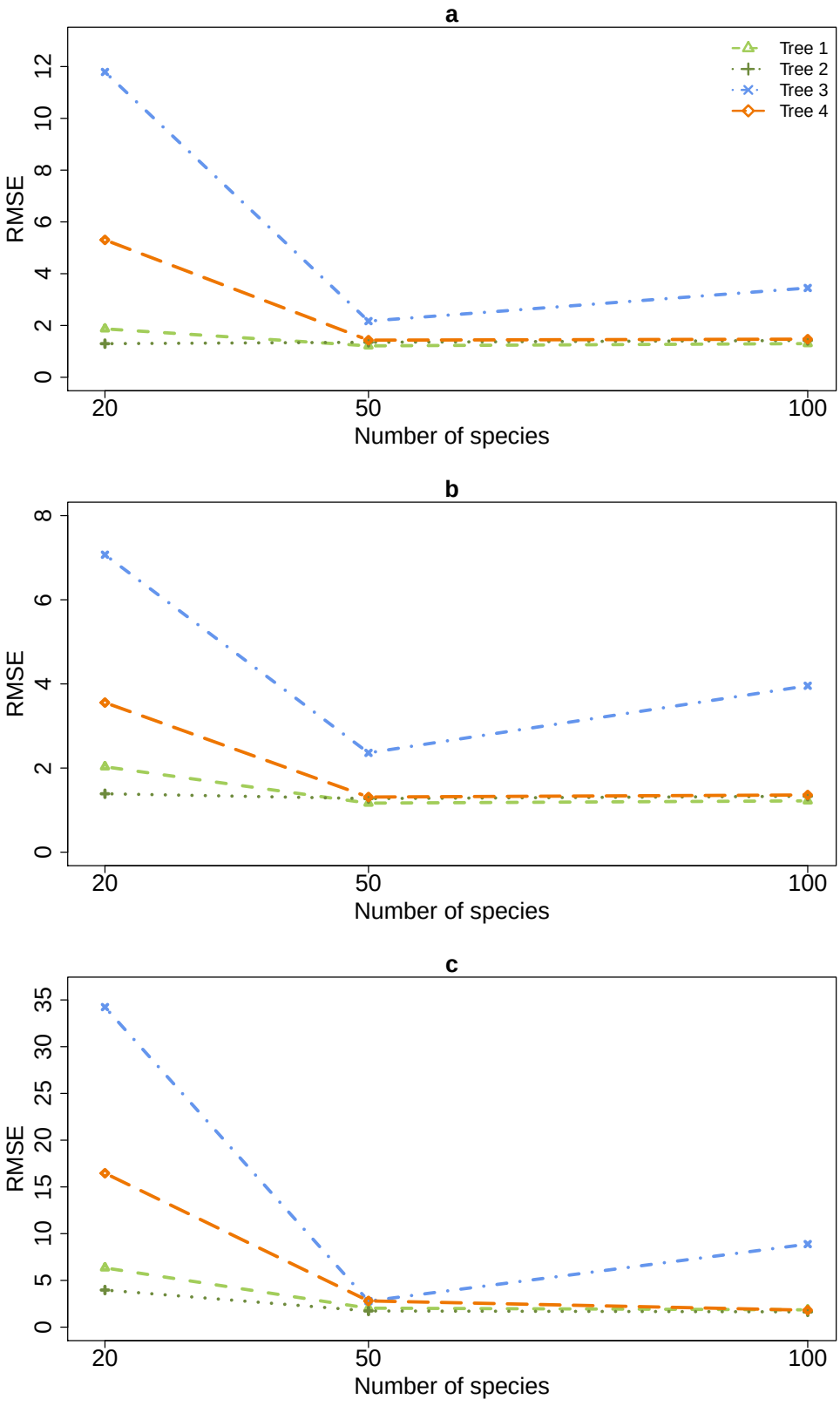


Figure 3.48: **Rooted mean squared errors from binary response models - Non-phylogenetic GLM.** The rooted mean squared errors (RMSE) are shown for the data simulated with a correlation coefficient of $r = 0.5$. a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

3.3.2 PGEE

The simulation loop for the PGEE method was not run for the binary response variable due to freezing problems of `R`. This problem is claimed to come from very strong correlations among some observations, probably leading to numerical problems (personal communication, Emmanuel Paradis).

3.3.3 PGLMM

Type I Error and Power

Most parameter combinations of tree structure and number of species, except tree 2, lead to significantly elevated *type I error rates*, comparable to the non-phylogenetic approach. In other words, the Bayesian approach for binary data misleadingly declares significance too often (Table 3.14 and Figure 3.49). The type I error rates are especially high for tree 3 and small sample sizes ($n = 20$). The *power*, on the other hand, shows a good performance with mostly reaching 80% and higher, with a increasing tendency for larger samples sizes. The binary predictor of the multivariate model shows a slightly reduced power (Table 3.14 and Figure 3.50).

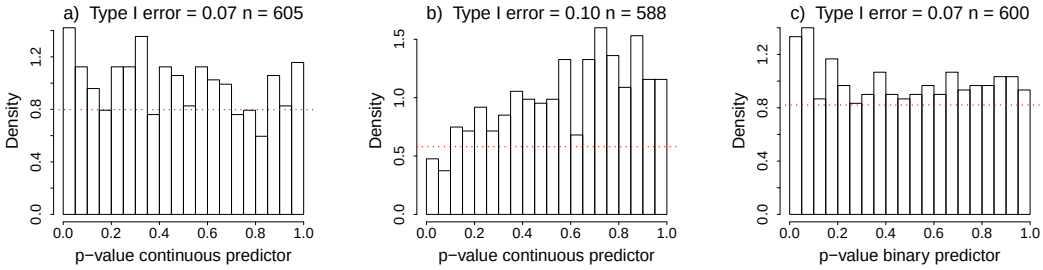


Figure 3.49: **Type I error rates from binary response models - PGLMM.** Distributions of p-values from data simulated with $r = 0.01$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The horizontal dashed line indicates the density at the critical value of 0.05.

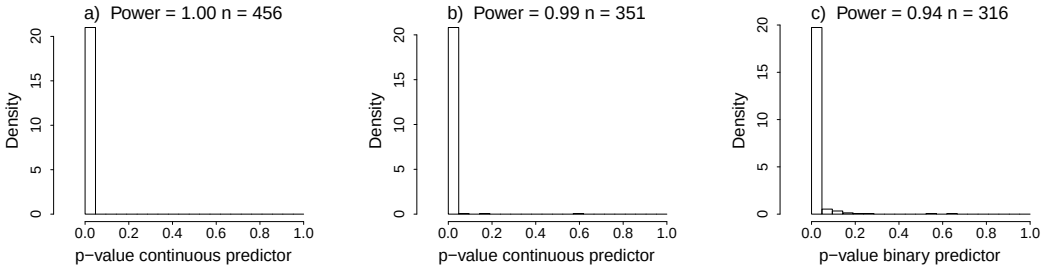


Figure 3.50: **Power from binary response models - PGLMM.** Distributions of p-values from data simulated with $r = 0.9$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

Table 3.14: **Type I error rates (TIE) and Power from binary response models - PGLMM**: Fisher's exact test was used to compare the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values to the expected ratio (0.05/0.95) under the assumption of a type I error of $\alpha = 5\%$. The different significance levels of the Fisher's exact test are indicated by bold font for $p < 0.001$ and in italics for $p < 0.05$. a) univariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.11**	0.97	0.07	1.00	*0.09*	1.00
Tree 2	0.05	1.00	0.05	1.00	0.06	1.00
Tree 3	**0.37**	0.83	*0.11*	1.00	**0.28**	0.89
Tree 4	**0.11**	0.97	*0.08*	1.00	*0.10*	1.00

b) multivariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.11**	0.82	*0.10*	0.99	**0.09**	1.00
Tree 2	0.06	0.89	0.07	1.00	0.06	1.00
Tree 3	**0.31**	0.47	*0.11*	0.98	**0.22**	0.73
Tree 4	**0.11**	0.88	0.06	0.99	0.08	1.00

c) multivariate model - binary predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	**0.13**	0.75	0.07	0.94	*0.08*	1.00
Tree 2	0.05	0.76	0.06	0.99	0.04	1.00
Tree 3	**0.31**	0.44	*0.09*	0.90	**0.27**	0.64
Tree 4	**0.14**	0.78	*0.09*	0.92	0.06	1.00

Mean estimates (M), Mean Error (ME) and Rooted Mean Squared Error (RMSE)

Again, the *mean parameter estimates* show a rather high variability with large standard deviations. For tree 3 the mean errors tend to be higher than for the other tree structures, whereas for tree 2 the opposite pattern is observed. Tree structure 3 shows also the least accurate parameter estimates, and generally, for larger samples sizes the accuracy increases (indicated by the RMSE). The multivariate model with a high correlation between the traits shows extreme high estimation errors which is based on the very high parameter estimates from the non-phylogenetic GLM included in the overall mean. Compared to the non-phylogenetic approach, the estimation performance is worse. The parameter estimates from PGLMM are generally much more biased and less accurate. The corresponding results are found in Tables C.28 to C.30 in Appendix C.3.2 and Figures 3.51, 3.52.

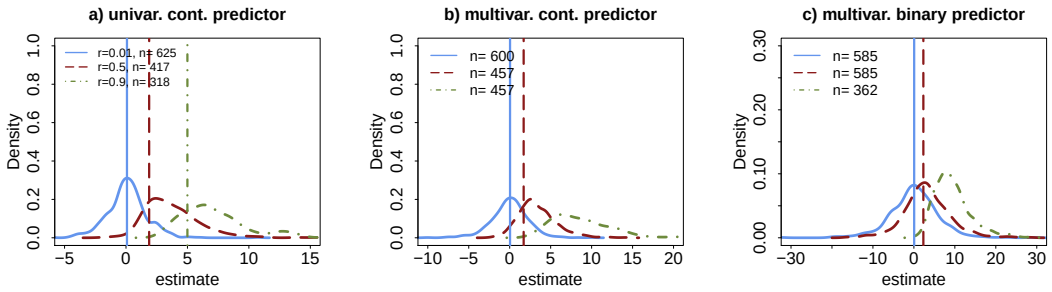


Figure 3.51: **Estimates from binary response models - PGLMM.** Distributions of the effect sizes based on the simulated data with an input correlation coefficients of $r = 0.01$ (blue solid line), $r = 0.5$ (red dashed line) and $r = 0.9$ (green dashed-dotted line): a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model binary predictor including the data of tree 1 and 50 species. The vertical lines indicate the means overall methods. The overall means for the continuous and binary predictor of the multivariate model with an input correlation of $r = 0.9$ is very large ($> 10^{11}$), and thus, not shown in the graph. The bandwidth of the densities was set to the default (`"nrd0"`).

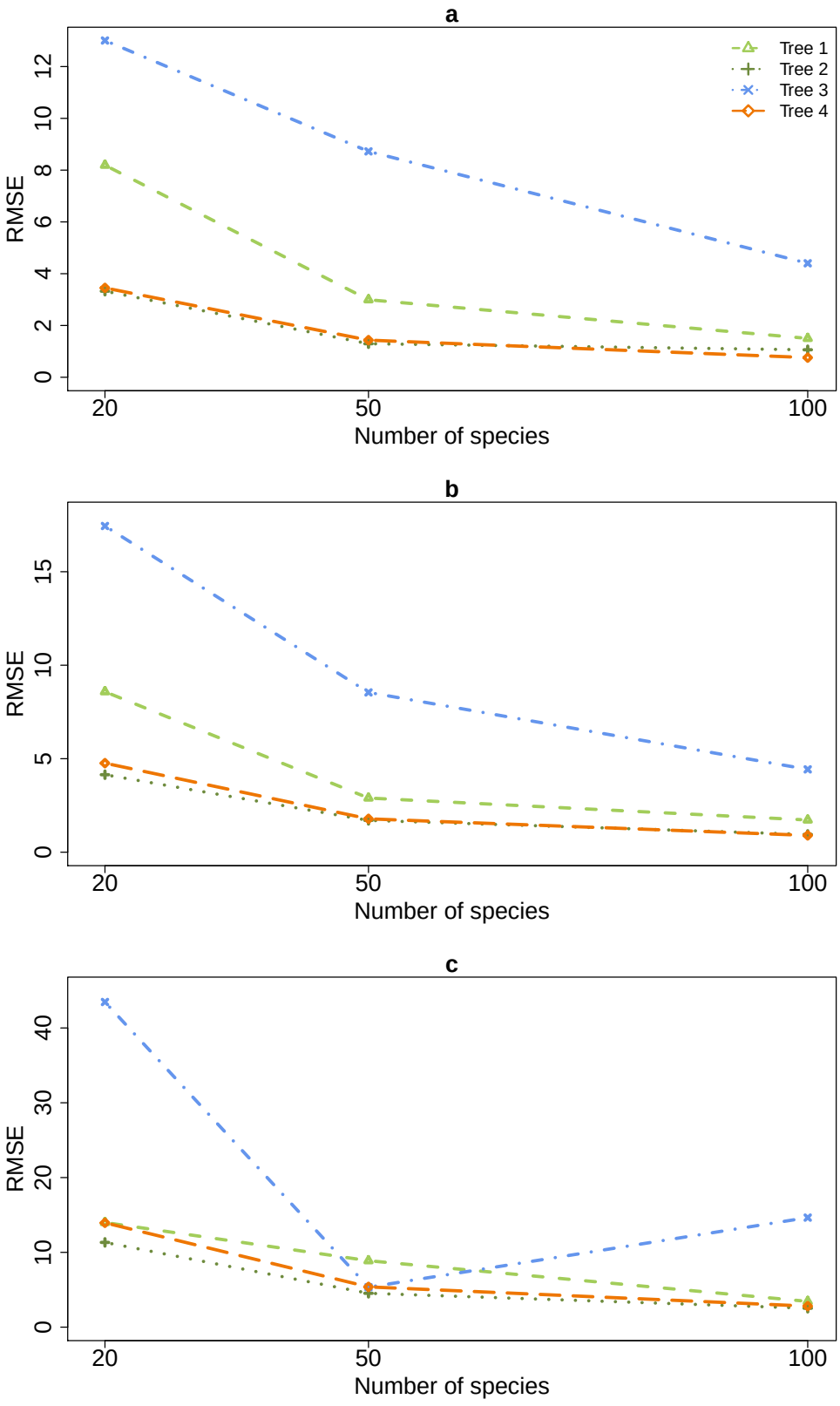


Figure 3.52: **Rooted mean squared errors from binary response models - PGLMM.** The rooted mean squared errors (RMSE) are shown for the data simulated with a correlation coefficient of $r = 0.5$. a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

3.3.4 PLR

Considering the results from PLR it is important to keep it mind that they are based on maximum of 100 simulations, whereas the other methods are as usual based on 1000 simulations. This is due to the fact that the function `phyloglm()` is very slow and further, several crashes of the server led to an early stop of the simulation loop.

Type I Error and Power

In relation to hypothesis testing, the PLR shows a much better performance than the non-phylogenetic approach and PGLMM. Only tree 3 shows significantly elevated *type I error rates* specifically for larger samples sizes ($n = 50$, $n = 100$) (Table 3.15, Figure 3.53). The *power* analysis on the other hand, does not show such a good performance, also compared to the other methods. In particular, the models based on tree 3 and small sample sizes ($n = 20$) show very low power. However, for larger sample size and tree structures other than 3 show very high power (Table 3.15, Figure 3.54).

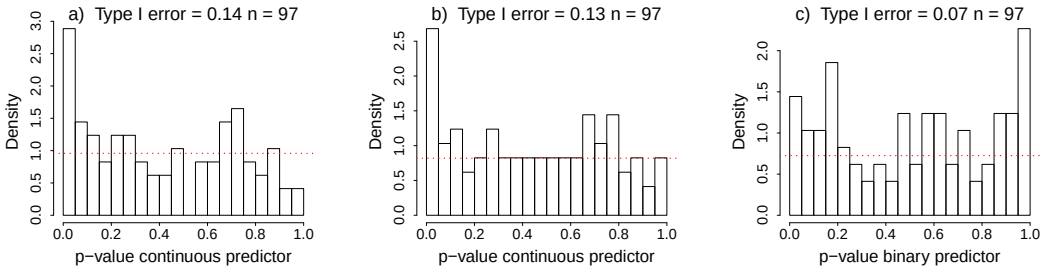


Figure 3.53: **Type I error rates from binary response models - PLR.** Distributions of p-values from data simulated with $r = 0.01$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor. The horizontal dashed line indicates the density at the critical value of 0.05.

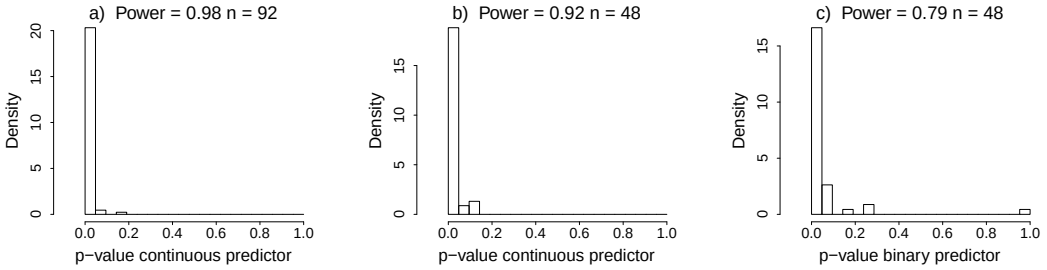


Figure 3.54: **Power from binary response models - PLR.** Distributions of p-values from data simulated with $r = 0.9$ including the data for tree 1 and 50 species: a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

Table 3.15: **Type I error rates (TIE) and Power from binary response models - PLR**: Fisher's exact test was used to compare the observed ratio of significant ($p < 0.05$) vs. non-significant ($p > 0.05$) p-values to the expected ratio (0.05/0.95) under the assumption of a type I error of $\alpha = 5\%$. The different significance levels of the Fisher's exact test are indicated by bold font for $p < 0.001$ and in italics for $p < 0.05$. a) univariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	0.02	0.75	0.14	0.98	**0.18**	1.00
Tree 2	0.02	0.86	0.01	1.00	0.05	1.00
Tree 3	0.04	0.37	**0.33**	0.92	**0.42**	0.97
Tree 4	0.04	0.88	0.08	1.00	0.11	1.00

b) multivariate model - continuous predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	0.04	0.11	0.13	0.92	**0.18**	1.00
Tree 2	0.01	0.15	0.00	0.95	0.06	1.00
Tree 3	0.04	0.06	**0.28**	0.64	**0.29**	0.94
Tree 4	0.03	0.21	0.09	0.93	0.09	0.99

c) multivariate model - binary predictor.

	Species 20		**Species 50**		**Species 100**	
	TIE	Power	TIE	Power	TIE	Power
Tree 1	0.04	0.11	0.07	0.79	0.05	0.99
Tree 2	0.01	0.00	0.03	0.89	0.07	1.00
Tree 3	0.02	0.20	0.13	0.38	0.14	0.84
Tree 4	0.02	0.00	0.09	0.82	0.07	1.00

Mean estimates (M), Mean Error (ME) and Rooted Mean Squared Error (RMSE)

The *mean parameter estimates* of PLR show not such a high variability in contrast to the non-phylogenetic approach and PGLMM. The *mean errors* as an indication for the estimation bias as well as the *rooted mean squared errors* are moderate except for models based on tree 3 showing poor estimation abilities. Moreover, the strong correlations ($r = 0.9$) show more biased and less accurate estimates as well as the binary predictor of the multivariate model. The corresponding results are found in Tables C.31 to C.33 in Appendix C.3.3 and Figures 3.55, 3.56.

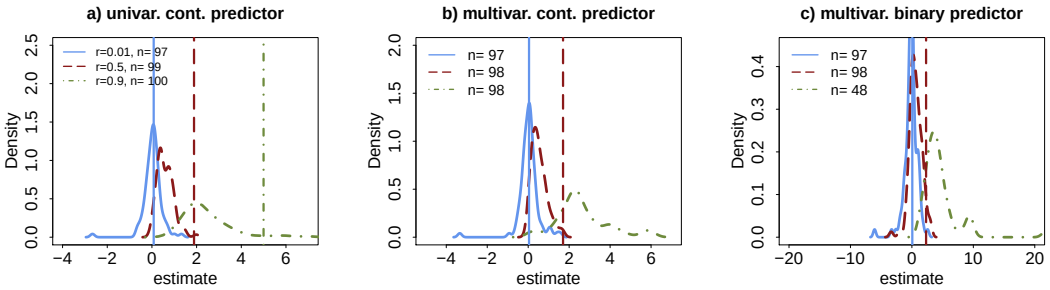


Figure 3.55: **Estimates from binary response models - PLR.** Distributions of the effect sizes based on the simulated data with an input correlation coefficients of $r = 0.01$ (blue solid line), $r = 0.5$ (red dashed line) and $r = 0.9$ (green dashed-dotted line): a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model binary predictor including the data of tree 1 and 50 species. The vertical lines indicate the means overall methods. The overall means for the continuous and binary predictor of the multivariate model with an input correlation of $r = 0.9$ is very large ($> 10^{11}$), and thus, not shown in the graph. The bandwidth of the densities was set to the default (`"nrd0"`).

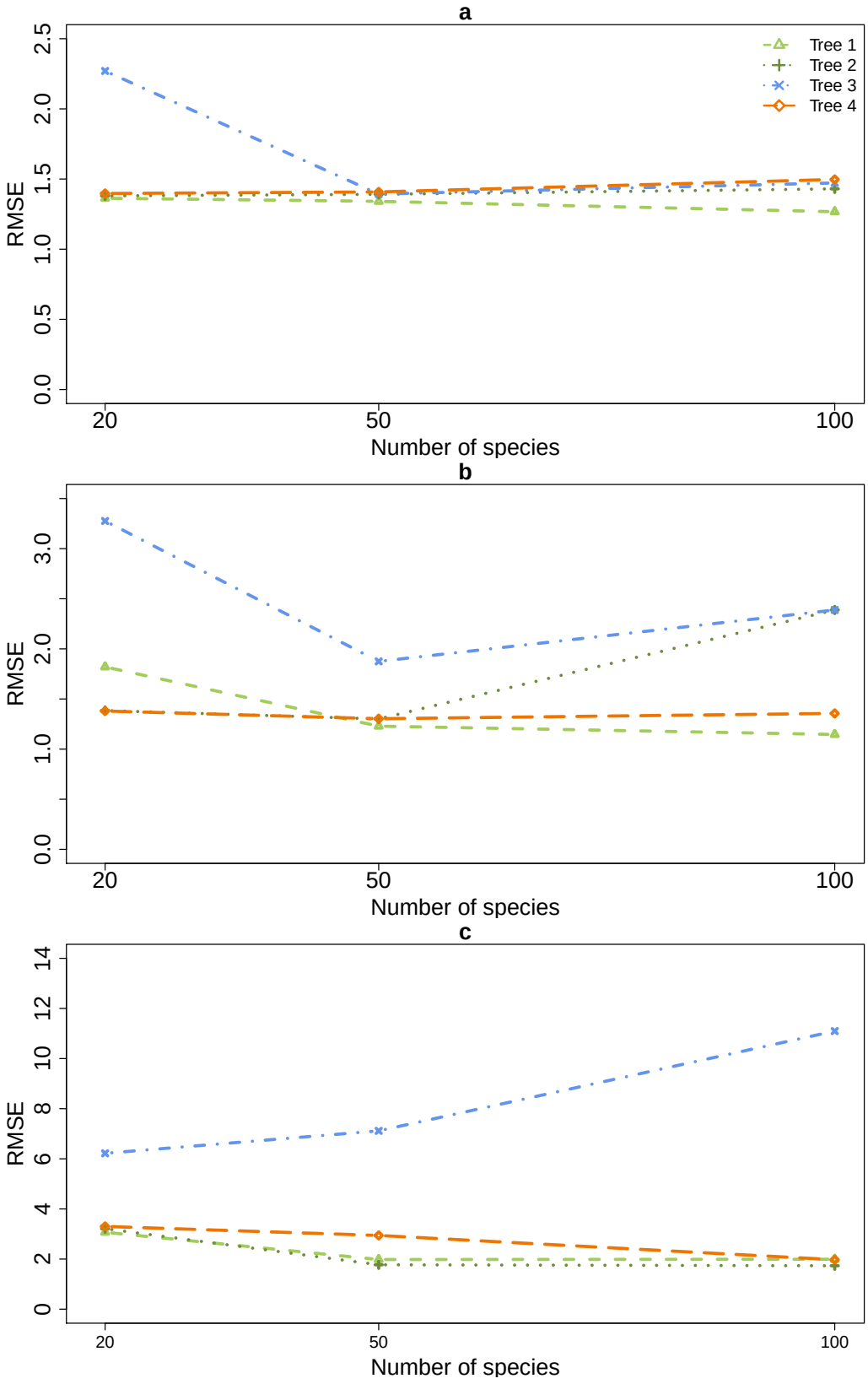


Figure 3.56: **Rooted mean squared errors from binary response models - PLR.** The rooted mean squared errors (RMSE) are shown for the data simulated with a correlation coefficient of $r = 0.5$. a) univariate model - continuous predictor; b) multivariate model - continuous predictor; c) multivariate model - binary predictor.

3.3.5 Phylogenetic signal α

According to Ives and Garland (2010) and Ho and Ané (in review), the phylogenetic signal for a binary variable is measured by α. This phylogenetic signal is based on the transition rates between 0 and 1: the higher the transition rates, the lower the phylogenetic correlation and the lower the phylogenetic signal. However, the interpretation of α is counterintuitive, as a lower value stands for a high phylogenetic signal (further details are found in Section 1.3.3).
As already observed for the phylogenetic signal of continuous traits (λ), the tree structures accord roughly to the estimated phylogenetic signals. Tree 2, with most diversifications occurring early in the phylogenetic history, shows the lowest phylogenetic signal among the four trees (high values of α). However, not tree 3 as expected shows the highest phylogenetic signals (low values of α) but rather tree 4 where all branch lengths were set to 1 illustrating the case where the true branch lengths are unknown. Moreover, the phylogenetic signal increases with an increasing number of species (Table 3.16).

Table 3.16: **Mean phylogenetic signal α with the standard deviation in brackets from binary response models - PLR**:
a) univariate model.

	Species 20	**Species 50**	**Species 100**
Tree 1	3.78(6.54)	0.94(2.24)	1.02(3.12)
Tree 2	10.38(3.86)	6.25(2.13)	4.92(2.37)
Tree 3	7.58(8.63)	2.04(2.45)	1.6(2.22)
Tree 4	1.47(2.68)	0.4(0.94)	0.2(0.46)

b) multivariate model.

	Species 20	**Species 50**	**Species 100**
Tree 1	4.16(6.75)	1.08(2.45)	1.18(2.63)
Tree 2	10.24(3.89)	6.23(2.12)	4.90(2.24)
Tree 3	9.42(9.58)	2.35(2.76)	2.67(4.09)
Tree 4	1.76(2.99)	0.85(1.79)	0.46(1.16)

3.3.6 The four methods in comparison

Simulated data

The direct comparison shows that the estimates of the three methods roughly agree, however, for higher correlations ($r = 0.5$, $r = 0.9$) PGLMM results in higher estimates with higher variabilities. Further, the non-phylogenetic approach in case of the multivariate model shows a higher mean estimate in case of a low correlation ($r = 0.01$). The 95% confidence-intervals of the parameter estimates show that the estimate variabilities in general is very low. For graphical illustration see Figure 3.57.
Figure 3.58 shows the estimates of the three methods plotted against each other. This illustrates that the estimates of PGLMM generally show a much higher variability compared to the estimates of the non-phylogenetic approach and PLR.

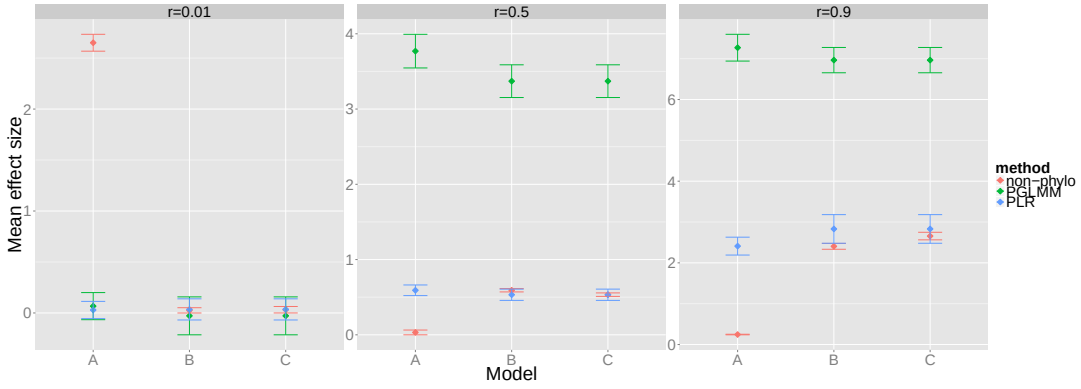


Figure 3.57: **Mean estimates and 95%-confidence intervals from binary response models.** Comparison of the mean estimates and corresponding 95%-confidence intervals from the four methods for the three input correlation coefficients including the simulated data for tree 1 and 50 species (number of simulations = 1000). A: univariate model - continuous predictor; B: multivariate model - continuous predictor; C: multivariate model - binary predictor.

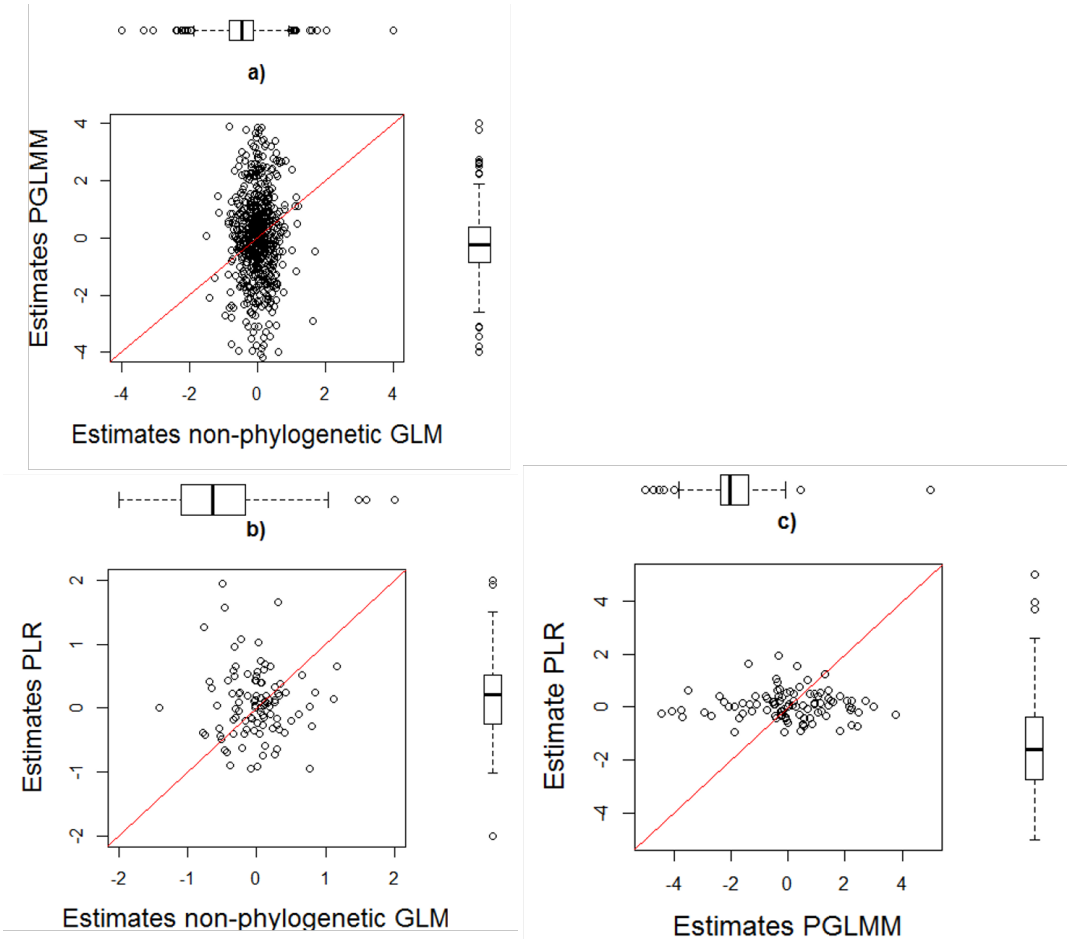


Figure 3.58: **Parameter estimates compared.** The parameter estimates based on the simulated data with $r = 0.5$ and 50 species from the non-phylogenetic and two phylogenetic methods (binary response models) plotted against each other pairwise with the corresponding boxplots aside: a) Non-phylogenetic GLM vs. PGLMM; b) Non-phylogenetic vs. PLR; c) PGLMM vs. PLR.

Application on a real data set

To compare the methods for a binary response variable using a real evolutionary questions, the effect of relative brain size (corrected for body size) (continuous predictor variable) on whether a species shows extractive foraging or not (binary response variable) was tested. The application of the four methods (for the application on a real data set PGEE method worked, whereas for the simulations this method could not be tested due to freezing of `R`) on the same real primate data set shows that the parameter estimates as well as the p-values are very similar for the non-phylogenetic GLM, the PGEE and PLR (Table 3.17). For PGLMM the estimates are much higher and more important, according to the diagnostic plots and the Geweke test, the model seems not to have converged (Figure 3.59). In other words, the results obtained from the PGLMM modeling the binary response variable are not reliable. However, further studies on the prior settings in `MCMCglmm` are needed in order to properly judge its statistical performance.

But all in all, concerning the direction of the effects and p-values, the four methods lead to the same conclusion: larger brained primates more likely use extractive foraging compared to their smaller brained relatives. In other words, smarter primates rely more often on difficult to access food resources.
However, also the underlying model assumptions need to be met in order to trust the results of an analysis. In case of the non-phylogenetic approach, this is not the case as the data is not independent due to phylogenetic relationships between species. Moreover, for continuously distributed data, diagnostic plots help to check whether a model fits the data or not in terms of homoscedasticity and normality of the residuals. But for binary data, these plots do not make sense anymore. Thus, only the diagnostic plot for the convergence of the MCMC chains of the PGLMM are shown (Figure 3.59).

Table 3.17: **Application of the four methods on a real data set - binary response model.** Testing the effect of brain size (continuous predictor) on extractive foraging (binary response) correcting for body size (continuous predictor). Given are sample size (N), phylogenetic signal alpha (α) in case of PLR, and the estimate, standard error and p-value (bold if significant) of the explanatory variables brain size and body size. For PGLMM, the number of iterations is set to 50,000 and the estimates represent the posterior mean of the effects of brain size and body size on extractive foraging. Important to note is that the PGLMM model seems not to have converged (Geweke diagnostic test: $p<0.001$).

			brain mass			**body mass**		
Method	N	α	estimate	std. error	p-value	estimate	std. error	p-value
Non-phylogenetic GLM	78		10.638	2.907	**0.000**	-7.038	2.242	**0.002**
PGEE	78		9.724	2.794	**0.004**	-6.750	2.155	**0.008**
PLR	78	0.076	9.492	2.667	**0.000**	-6.307	2.073	**0.002**
PGLMM	78		80.718		**0.003**	-53.958		**0.011**

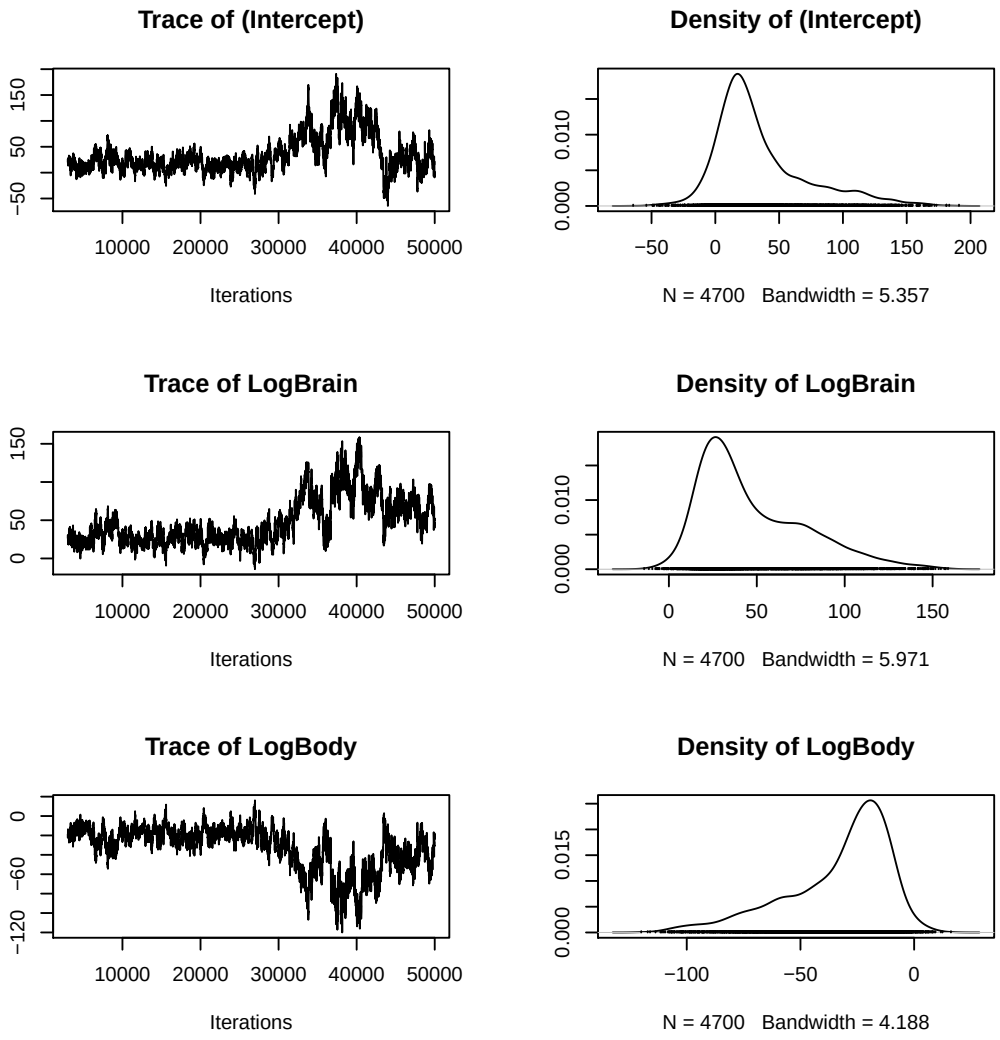


Figure 3.59: **Diagnostic plots from binary response model - PGLMM.** The diagnostic plots of the Markov chains serve to check for convergence. The plots on the left show the traces of the sampled posteriors along the iterations. The plots on the right show the distributions of the sampled posteriors.

4 Discussion

First, a comment on the setup should give an overview of the main parts of the following discussion. The setup of the discussion of the statistical performances of the phylogenetic methods is different from the structure in the results section. In fact, the methods are not discussed separately but rather as a whole trying to pool and compare the important findings and conclusions of the extensive results from the former chapter. In the following three sections, the results for the three types of response variables, continuous, ordinal and binary, are summarized and subsequently discussed in terms of their statistical abilities and practical applicabilities with respect to various conditions (tree structure, sample size, strength of correlation and predictor variable). Three qualitative summaries, one for each type of response, give an overview of the statistical performances of the four compared methods (Tables 4.1, 4.2, 4.3).

4.1 Continuous response

4.1.1 General comparison of the statistical performances

Not surprisingly, the non-phylogenetic approach testing the correlated evolution between two traits shows highly elevated type I error rates, error rates which are unacceptable in terms of hypothesis testing. In other words, if ignoring the phylogenetic dependencies between species, the statistical significance of an evolutionary correlation is mostly overestimated. These results are confirmed by the findings of several other studies (e.g. Martins and Garland 1991, Martins et al. 2002). The comparative methods PGLS, PGEE and PGLMM, taking into account the phylogenetic dependencies between species perform much better in that respect. Generally, all the three methods wrongly declare statistical significance only in about 5% of the cases, thus, showing acceptable type I error rates. However, the PGLMM stands out among those methods as it never shows any false positive results. The power analyses based on the simulations with an input correlation of $r = 0.9$ generally show very good performances for all methods. In relation to the estimation abilities, the non-phylogenetic approach shows more biased and less accurate parameter estimates compared to the phylogenetic methods, but within those the estimation abilities are comparable. In conclusion, ignoring phylogenetic dependencies between species is statistically not tolerable not only in terms of assumptions violation (i.e. dependency between observations ignored) but also due to a poor performance confirming the findings of former studies (e.g. Martins and Garland 1991, Martins et al. 2002).

4.1.2 Statistical performances with respect to varying evolutionary and empirical conditions

It is of special interest to discuss and verify these findings under different evolutionary scenarios and data conditions. Starting with the structure of the phylogenetic tree, striking are the results based on tree 2 and 3. In combination with tree 2, the methods and in particular the non-phylogenetic approach, shows rather good performances whereas with tree 3 rather bad performances. This can be explained by considering the biological meaning of the tree structure: Tree structure 2 shows most diversifications between species at a early point in the evolutionary history which goes along with short shared evolutionary paths and long independent trait evolution between species. Tree 3, on the other hand, shows the opposite evolutionary scenario, where most diversification occurs at the tips of the tree. In terms of the phylogenetic signal, tree 2 has a rather low phylogenetic signal and tree 3 a rather higher phylogenetic signal, which is also observed in the λ estimates of the PGLS regression (Table 3.3). To sum up, if species show short common evolutionary histories (low phylogenetic signal –> tree 2) implying a weaker dependency structure, results in better statistical performances. On the other hand, the opposite is true for an evolutionary scenario where species show a long common phylogenetic history (tree 3). These patterns are especially pronounced in the non-phylogenetic analysis. However, also the PGLS regression which accounts for phylogenetic dependencies shows in combination with a small sample size a tendency in that direction. Therefore, although this has not explicitly been part of simulation setup, the results illustrate certain comparative methods are more sensitive (i.e PGLS) to the strength of the phylogenetic signal than others (i.e. PGLMM).
Not surprisingly, also sample size has an impact on hypothesis testing and estimation abilities of these methods. Increasing sample sizes result in lower false positive (i.e. type I error) and higher true positive rates (i.e. power) as well as increased estimation accuracy. Intuitively, this can be explained by the fact that increasing samples sizes more and more approximate the whole sample/population containing the true correlation coefficient.
Furthermore, how well a method performs in terms of estimation accuracy also seems to be affected by the strength of an evolutionary correlation. In case of all methods, stronger correlations tend to show more accurate estimates.
As a last factor, the type of the predictor variable leads generally to variation in the estimation performances. In fact, in case of the binary predictor of the multivariate model, all methods show remarkably worse statistical abilities (i.e. power and estimation abilities) compared to the continuous predictors.

In conclusion, the statistical performances with respect to different conditions might vary among the three comparative methods. Although, all methods are comparably sensitive to sample size, strength of correlation and type of the predictor variable, PGLS seems to be especially sensitive in relation to the tree structure or rather the strength of the phylogenetic signal, whereas on the other extreme, PGLMM constantly shows a very good statistical performance. Therefore, based on these findings, the phylogenetic mixed model is most recommendable for analyzing the effects on a continuous response variable, however, in case of non-extreme or known evolutionary scenarios

(i.e. tree 1 in contrast to tree 2,3 and 4), also PGLS and PGEE reliably measure the correlated evolution among traits. This final conclusion is further supported by the direct comparison of the three methods analyzing the same data set (Section 3.1.5): the parameter estimates as well as the p-values are strongly consistent among the three methods resulting in the same evolutionary interpretations.

Table 4.1: **Overview statistical performances for continuous response models.** Qualitative summary of the statistical performances in terms of type I error rate, power, bias and estimation accuracy of the four methods (non-phylogenetic GLM, PGLS, PGEE and PGLMM) with respect to different varying simulation parameters (i.e. tree structure, number of species, strength of input correlation and type of predictor variable). In particular, clear identifiable tendencies are indicated by down- and upwards arrows. A minus sign indicates that no tendency is observable. Additionally to each of the simulation parameters, an overall rating of the corresponding performances is given. Upward arrows indicate rather high values, downward arrows indicate lower values. Green colored arrows stand for a good performance, red colored arrows represent rather bad performances.

		Hypothesis testing			Estimation ability
Method	**Parameter**	**Type I error**	**Power**	**Bias**	**Accuracy**
Non-phyl. GLM	Overall	↑	↑	↑	↓
	Tree structure	↓ T2, ↑ T3	↓ T3	-	↓ T3, ↑ T2
	Sample size	↓ $n = 20$, ↑ $n = 100$	↑ $n = 100$	-	↓ $n = 20$, ↑ n=100
	Strength of correlation	-	-	-	↓ $r = 0.01$, ↑ $r = 0.9$
	Predictor	↓ binary, ↑ univ. cont.	↓ binary	↑ binary	↓ binary
PGLS	Overall	↓	↑	↓	↑
	Tree structure	↓ T2, ↑ T3	↓ T3	-	-
	Sample size	-	↓ $n = 20$, ↑ $n = 100$	-	↓ $n = 20$, ↑ $n = 100$
	Strength of correlation	-	-	-	↓ $r = 0.01$, ↑ $r = 0.9$
	Predictor	-	↓ binary	↑ binary	↓ binary
PGEE	Overall	↓	↑	↓	↑
	Tree structure	↑ T4	-	-	-
	Sample size	↑ $n = 100$	↓ $n = 20$, ↑ $n = 100$	-	↓ $n = 20$, ↑ $n = 100$
	Strength of correlation	-	-	↑ $r = 0.9$	↓ $r = 0.01$, ↑ $r = 0.9$
	Predictor	↑ univ. cont.	↓ binary	↑ binary	↓ binary
PGLMM	Overall	↓	↑	↓	↑
	Tree structure	-	↓ T3	-	-
	Sample size	-	↓ $n = 20$, ↑ $n = 100$	-	↓ $n = 20$, ↑ $n = 100$
	Strength of correlation	-	-	↑ $r = 0.9$	↓ $r = 0.01$, ↑ $r = 0.9$
	Predictor	-	↓ binary	↑ binary	↓ binary

4.2 Ordinal response

Before assessing any results, it is important to remember the assumption of pseudo-continuity for the ordinal scaled response variable applying three of the four methods, the non-phylogenetic GLM, PGLS and PGEE. This assumption is often used in comparative studies either due to a lack of knowledge and or a lack of suitable phylogenetic methods. Modeling this kind of pseudo-continuous data, however, might lead to elevated type I error rates, especially if the distances between the levels are unequal (Matthews et al. 2011, Stevens 1946, Purvis et al. 2005). Therefore, the aim of this study was to test whether assuming pseudo-continuity of ordinal scaled variables still lead to reliable results and compare them to the very new approach, the phylogenetic generalised mixed model (PGLMM), explicitly modeling multinomial logit models.

4.2.1 General comparison of the statistical performances

Generally, the statistical performances for the ordinal response models are very similar to the continuous response models, except for the phylogenetic mixed model. The qualitative summary of the four methods is found in Table 4.2.
First, considering the non-phylogenetic approach for analyzing ordinal data, using a simple GLM with the identity link, the performances are rather poor. Especially the highly elevated type I error rates make this approach statistically unacceptable. The estimation performance, however, is only slightly decreased in comparison to the phylogenetic approaches PGLS and PGEE. But the two phylogenetic methods assuming pseudo-continuity of the response variable show much better performances in significance testing. PGLS shows only significantly elevated type I error rates for small samples sizes ($n = 20$), whereas PGEE counterintuitively shows dominantly elevated type I error rates for larger sample sizes ($n = 50$ and $n = 100$) as well as generally more elevated type I error rates. Furthermore, estimation abilities based on the simulations and the direct comparison of the parameter estimates on a real evolutionary question (Section 3.2.5) shows that the estimates and their errors are very similar with the p-values leading to the exact same conclusions. In sum, whereas PGLS is slightly better performing in hypothesis testing, the estimation abilities are equally good for the PGLS and PGEE. Therefore, although assuming pseudo-continuity which basically violates the model assumptions, the performances of the two phylogenetic approaches are statistically acceptable. In concordance with this finding, a study by Matthews et al. (2011) showed that treating ordinal scaled data as continuous is fully acceptable in terms of type I error rates and thus, also verifies the application of linear models on ordinal scaled response variables. All in all, in particular due to slightly lower type I error rates for PGLS compared to PGEE, the former method is more recommendable.
After the comparison of the methods treating ordinal data as continuous, it is of special interest whether the phylogenetic mixed model explicitly modeling a multinomial model shows a better performance.
First of all, the phylogenetic mixed model seems to be very flexible and seems to have many advantages over other comparative methods. Starting with a rather philosophical argument, the phylogenetic mixed model aims to explain part of the variance using phylogeny, whereas other methods such as PGLS, incorporate it into the error term

and transform the data in order to met the assumptions about independency of the observations. Furthermore, it allows to model multiple response variables simultaneously and to account for additional random effects other than the phylogeny. Besides modeling phylogenetic comparative analysis, the package allows for the application of many other types of models, such as the animal model and meta-analyses (Hadfield 2012). Therefore, the phylogenetic mixed model offers a wide range of applicabilities in the field of evolutionary biology. Based on the fact that it claims to model explicitly nominal data, the statistical performance is expected to be better compared to models assuming pseudo-continuity. However, the results of the simulations using PGLMM do not support that presumption. For every combination of the simulated conditions (i.e. tree structure and sample size), the type I error rates are significantly elevated. Moreover, the estimation abilities in terms of bias and accuracy, show a very bad performance. All in all, based on these simulations, the `MCMCglmm` package does not produce reliable results for categorical data with more than two levels. But why is that?

Before trying to give an answer to that question, some further general comments about the `MCMCglmm` package are necessary. One main disadvantage of the `MCMCglmm()` function is its inconvenient application. Using a Bayesian approach, the function allows to specify the priors for the variance components of the fixed effects, the random effects and the residuals. These prior settings allow to specify the expected value as well as the strength of belief in that expected value. However, besides the fact that the implementation of the function and prior settings are not intuitive at all, most biologists are probably overstrained with the question of what is an appropriate prior for a specific evolutionary question. Inappropriate priors probably also explain why these models very often don‘t reach convergence, despite high numbers of iterations (example see Figure 3.44). Further, this also explains the strongly decreased number of simulations for the results of the PGLMM in case of the ordinal response models, as all non-converged models are excluded before analyses. In sum, the `MCMCglmm()` function is not at all user-friendly for a biologists with a standard statistical background.

These issues are probably also the key to the very bad performances of the multinomial logit models in case of the ordinal scaled response variable. Unfortunately, understanding the Bayesian approach and with that the extensive details of the prior settings in `MCMCglmm` was beyond the scope of this thesis, therefore, the default priors were used in the simulations of this study. These default priors are rather flat and uninformative which in the end probably caused the bad statistical performance. Although, they seemed to have worked in case of continuous response variable, they ended up in bad results for the multinomial model. Moreover, for every parameter condition in terms of phylogenetic tree structure, sample size and strength of correlation, the simulated analyses used the same prior (i.e. default prior) which probably is also not appropriate. Because the priors in some cases, even though they seem to be harmless, have strong impacts on the posterior parameter distributions (personal communication, Jarrod Hadfield), they should be specified specifically for each type of data in terms of phylogenetic structure, sample size and strength of correlation. Thus, using a global prior, as it has been done in this simulation study, leads to unreliable results. Therefore, future simulation studies testing the PGLMM should probably adjust the priors

properly for each parameter condition. In conclusion, in case of non-continuous data, it is very dangerous to use this method without properly understanding and justifying the settings of the priors. Although, also based on the fact that for the continuous response the PGLMM performed really well, it would probably be premature to dismiss the multinomial model implemented in `MCMCglmm`, it is certainly not ready to be used. Besides that further studies are needed to investigate the statistical abilities of modeling nominal data, the practical applicabilities of this package need to be substantially improved. Therefore, `MCMCglmm` is not recommended to be used for non-continuous data without properly understanding its theoretical and practical implementation.

4.2.2 Statistical performances with respect to varying evolutionary and empirical conditions

The statistical performances with respect to the phylogenetic tree structure, sample size, strength of correlation and type of response variable are comparable to the continuous case (Section 4.1.2) and thus, are not discussed in detail here (qualitative summary given in Table 4.2). Lower phylogenetic signals (i.e. tree 2) generally lead to better performances whereas in case of higher phylogenetic signals (i.e. tree 3), the methods perform rather poor. Furthermore, larger sample sizes lead to higher power and more accurate estimates and the binary predictor variable shows generally a worse performance compared to the continuous predictors. The detailed discussion about these issues can be found in Section 4.1.2.

Table 4.2: **Overview statistical performances for ordinal response models.** Qualitative summary of the statistical performances in terms of type I error rate, power, bias and estimation accuracy of the four methods (non-phylogenetic GLM, PGLS, PGEE and PGLMM) with respect to different varying simulation parameters (i.e. tree structure, number of species, strength of input correlation and type of predictor variable). In particular, clear identifiable tendencies are indicated by down- and upwards arrows. A minus sign indicates that no tendency is observable. Additionally to each of the simulation parameters, an overall rating of the corresponding performances is given. Upward arrows indicate rather high values, downward arrows indicate lower values. Green colored arrows stand for a good performance, red colored arrows represent rather bad performances and arrows in brackets represent rather weak trends in the corresponding directions.

		Hypothesis testing			Estimation ability
Method	**Parameter**	**Type I error**	**Power**	**Bias**	**Accuracy**
Non-phylogenetic GLM	Overall	$\uparrow$	$\uparrow$	($\uparrow$)	($\downarrow$)
	Tree structure	$\downarrow$ T2, $\uparrow$ T3	-	-	$\downarrow$ T3, $\uparrow$ T2
	Sample size	$\downarrow$ $n = 100$, $\uparrow$ $n = 20$	$\downarrow$ $n = 20$, $\uparrow$ $n = 100$	-	$\downarrow$ $n = 20$, $\uparrow$ $n = 100$
	Strength of correlation	-	-	$\downarrow$ $r = 0.01$, $\uparrow$ $r = 0.9$	-
	Predictor	$\downarrow$ binary, $\uparrow$ univ. Cont.	$\downarrow$ binary	-	-
PGLS	Overall	$\downarrow$	$\uparrow$	$\downarrow$	$\uparrow$
	Tree structure	$\downarrow$ T2, $\uparrow$ T3	$\downarrow$ T3	$\uparrow$ T3	$\downarrow$ T3
	Sample size	$\downarrow$ $n = 20$, $\uparrow$ $n = 100$	$\downarrow$ $n = 20$, $\uparrow$ $n = 100$	-	$\downarrow$ $n = 20$, $\uparrow$ $n = 100$
	Strength of correlation	-	-	$\downarrow$ $r = 0.01$, $\uparrow$ $r = 0.9$	$\downarrow$ $r = 0.9$, $\uparrow$ $r = 0.01$
	Predictor	-	$\downarrow$ binary	-	-
PGEE	Overall	$\downarrow$	$\uparrow$	$\downarrow$	$\uparrow$
	Tree structure	$\uparrow$ T2, $\uparrow$ T1, $\uparrow$ T3, $\uparrow$ T4	$\downarrow$ T3	$\uparrow$ T3	$\downarrow$ T3
	Sample size	$\downarrow$ $n = 20$, $\uparrow$ $n = 100$	$\downarrow$ $n = 20$, $\uparrow$ $n = 100$	-	$\downarrow$ $n = 20$, $\uparrow$ $n = 100$
	Strength of correlation	-	-	$\downarrow$ $r = 0.01$, $\uparrow$ $r = 0.9$	-
	Predictor	-	$\downarrow$ binary	-	-
PGLMM	Overall	$\uparrow$	($\uparrow$)	$\uparrow$	$\downarrow$
	Tree structure	$\uparrow$ T3	$\downarrow$ T3	$\uparrow$ T3	$\downarrow$ T3
	Sample size	-	$\downarrow$ $n = 20$, $\uparrow$ n=100	-	$\downarrow$ $n = 20$, $\uparrow$ $n = 100$
	Strength of correlation	-	-	$\downarrow$ $r = 0.01$, $\uparrow$ $r = 0.9$	$\downarrow$ $r = 0.9$, $\uparrow$ $r = 0.01$
	Predictor	-	$\downarrow$ binary	-	-

4.3 Binary response

For the binary response variable, only two phylogenetic comparative methods, PGLMMM and PLR, and the non-phylogenetic approach were compared based on their hypothesis testing and estimation abilities. Unfortunately, the PGEE method which is also implemented for binary data, could not be tested with simulations due to freezing of R for certain data constellations. Moreover, several crashes of the server led to an early stop of the PLR simulation loop ending up in maximal 100 simulations.

4.3.1 General comparison of the statistical performances

Starting with the non-phylogenetic approach, the rather poor performance is not surprising. As already in case of the continuous and ordinal response, for most conditions in terms of tree structure and sample size, the type I error rates are significantly elevated. Also the power analyses don‘t show such a good performance, especially for small sample sizes and the multivariate model. Moreover, the estimates are rather inaccurate. In sum, a non-phylogenetic approach is also for a binary response variable statistically not reliable.
Although, the PGLMM accounting for the phylogenetic dependencies in form of a random effect would be expected to show a better performance, shows a similar or even worse performance compared to the non-phylogenetic GLM. Almost for any parameter condition, the type I error rates are significantly elevated, whereas the probability to declare significance if the alternative is indeed true (i.e. power), shows a better performance with on average 80% and higher. In relation to the estimation abilities, the statistical performance is even worse compared to the non-phylogenetic approach with a slightly higher bias and less inaccurate parameter estimates. This can probably also be explained by the prior settings. As already discussed in the former Section 4.2.1, the prior specification might have a large impact on the the posterior distribution of the parameters. As in case for the ordinal response variable, also the settings suggested by the author of the package was used for the simulations. However, this is probably not an appropriate prior for each parameter condition leading unreliable results and the problem of non-convergence. And in fact, in about 50% of the cases the models in the simulations did not converge (see number of simulations for the PGLMM). In sum, it would probably be premature to dismiss the phylogenetic mixed model based on these finding, however, its implementation needs to be substantially improved and in future studies, more reasonable prior settings are needed to further investigate the performance of PGLMM. Therefore, the `MCMCglmm` is not recommendable for modeling binary data without properly understanding and justifying the settings of the priors.
The last method modeling binary data is the phylogenetic logistic regression by Ives and Garland (2010), which has very recently been implemented in `R` by Ho and Ané (in review). Compared to the non-phylogenetic approach and the phylogenetic mixed model, this method performed much better in terms of both, hypothesis testing and estimation abilities. Only with a few exceptions, the type I error rates never significantly exceeded the expected 5%. Power generally also shows a good performance, however, small sample sizes ($n = 20$) are problematic. Moreover, the estimates are less biased and more accurate for the PLR compared to the other methods. In conclu-

sion, based on the simulations of this study, the phylogenetic logistic regression gives the most reliable results. For future studies it would be interesting to further look at the phylogenetic mixed model and in particular to understand the prior setting and its implications on the posterior distributions. Additionally, although PGEE showed comparable results to PLR in the application on a real data set, this method should be included in a simulation study in order to compare its abilities with the ones of other methods. Without claiming anything about the statistical performance of the PGEE, the PLR makes based on the correlation structure intuitively more sense. In fact, in PGEE this correlation structure assumes a Brownian motion model of evolution which is not appropriate for a discrete variable, whereas in PLR the matrix is adjusted based on a Markov process.

4.3.2 Statistical performances with respect to varying evolutionary and empirical conditions

With respect to different parameter conditions, the statistical behviour of the binary response model is comparable to the continuous and ordinal response models (Sections 4.1.2, 4.2.2) (qualitative summary given in Table 4.3). In fact, in case of tree 3 with a high phylogenetic signal, all methods show rather unreliable results. Moreover, increasing sample size positively affects the statistical ability of each method. For a binary predictor variable as well as for strong correlations, the power and the estimation abilities are generally not that good. All in all, the decision among different phylogenetic methods does not only depend on the overall performance of a phylogenetic method, but also on the evolutionary and data conditions.

Table 4.3: **Overview statistical performances for binary response models.** Qualitative summary of the statistical performances in terms of type I error rate, power, bias and estimation accuracy of the four methods (non-phylogenetic GLM, PGLMM and PLR) with respect to different varying simulation parameters (i.e. tree structure, number of species, strength of input correlation and type of predictor variable). In particular, clear identifiable tendencies are indicated by down- and upwards arrows. A minus sign indicates that no tendency is observable. Additionally to each of the simulation parameters, an overall rating of the corresponding performances is given. Upward arrows indicate rather high values, downward arrows indicate lower values. Green colored arrows stand for a good performance, red colored arrows represent rather bad performances and arrows in brackets represent rather weak trends in the corresponding directions.

		Hypothesis testing		Estimation ability	
Method	**Parameter**	**Type I error**	**Power**	**Bias**	**Accuracy**
Non-phylogenetic GLM	Overall	↑	(↑)	↑	↓
	Tree structure	↓ T2, ↑ T3	↓ T3, ↑ T2	↑ T3	↓ T3, ↑ T2
	Sample size	↓ $n = 20$, ↑ $n = 100$	↓ $n = 20$, ↑ $n = 100$	-	↓ $n = 20$, ↑ $n = 100$
	Strength of correlation	-	-	↓ $r = 0.01$, ↑ $r = 0.9$	↓ $r = 0.9$, ↑ $r = 0.01$
	Predictor	↓ binary	↓binary	-	↓ binary
PGLMM	Overall	↑	↑	↑	↓
	Tree structure	↓ T2, ↑ T3	↓ T3, ↑ T2	↑ T3	↓ T3
	Sample size	-	↓ $n = 20$, ↑ $n = 100$	-	↓ $n = 20$, ↑ $n = 100$
	Strength of correlation	-	-	↓ $r = 0.01$, ↑ $r = 0.9$	↓ $r = 0.9$
	Predictor	-	↓ binary	-	↓ binary
PLR	Overall	↓	(↑)	↓	↑
	Tree structure	↓ T2, ↑ T3	↑ T2	-	↓ T3
	Sample size	↓ $n = 20$, ↑ $n = 100$	↓ $n = 20$, ↑ $n = 100$	-	↓ $n = 20$
	Strength of correlation	-	-	↑ $r = 0.9$	↓ $r = 0.9$
	Predictor	-	↓ binary	-	↓ binary

4.4 Conclusions

The statistical performances of different phylogenetic methods were shown to vary among different types of responses. Whereas PGLMM showed the overall best performance for the continuous response variable, PGLS gives the most reliable results treating an ordinal scaled variable as continuous. This shows that, although the model assumptions may not be fully met, methods originally designed for continuous data also produce reliable results for ordinal scaled variables. Moreover, the phylogenetic logistic regression implemented in the package `phylolm` shows the best performance modeling binary data. Although, it is probably premature to dismiss its statistical abilities, the phylogenetic mixed model is not recommended for nominal and binary responses without properly understanding the prior settings and its implications. Moreover, the practical applicabilities and implementations of PGLS, PGEE and PLR in `R` are much more user-friendly compared to PGLMM.
Furthermore, all of the investigated phylogenetic methods were shown to be sensitive with respect to the phylogenetic tree structure as well as to sample size, however, the grade of sensitivity for specific conditions varies among the methods. For example, in case of a continuous response, PGLS shows a very good performance for larger sample sizes, whereas PGEE shows better performances for smaller sample sizes. What method to use for what kind of data in terms of tree structure and sample size is summarized in Table 4.4. It shows that for a continuous response variable, there is an appropriate approach for each combination of tree structure and sample size, whereas in case of an ordinal and a binary response, small sample sizes and high phylogenetic signals, respectively, should be avoided if possible because even the method with the best performance shows rather weak statistical abilities (indicated by brackets in Table 4.4).
All in all, these findings show that the decision of what approach to use does not only depend on the overall statistical performance of a phylogenetic method, but also on the type of data in terms of tree structure, sample size and in the broadest sense also on the type of evolutionary question.

Table 4.4: **Decision table.** Summary of what method to use for what kind of data in terms of the type of response, tree structure and sample size. Brackets indicate data constellations which sould be avoided if possible, but the given method, alhtough it shows rather weak statistical abilities, performs best among the four methods.

Continuous response	20 Species	50 Species	100 Species
Tree 1	PGEE, PGLMM	PGLS, PGLMM	PGLS, PGEE, PGLMM
Tree 2	non-phylog. GLM, PGEE, PGLMM	non-phylog. GLM, PGLS, PGEE, PGLMM	non-phylog. GLM, PGLS, PGEE, PGLMM
Tree 3	PGEE	PGLS, PGLMM	PGLS, PGLMM
Tree 4	PGEE, PGLMM	PGLMM	PGLMM
Ordinal response			
Tree 1	(PGEE)	PGLS	PGLS
Tree 2	non-phylog. GLM, PGEE, PGLS	non-phylog. GLM, PGLS, PGEE	non-phylog. GLM, PGLS, PGEE
Tree 3	(PGEE)	PGLS	PGLS
Tree 4	(PGEE)	PGLS	PGLS
Binary response			
Tree 1	PLR	PLR	(PGLMM)
Tree 2	PGLMM	non-phylog. GLM, PGLMM	non-phylog. GLM, PGLMM, PLR
Tree 3	(PLR)	(PGLMM)	(PGLMM)
Tree 4	(PLR)	(PLR)	PLR

Acknowledgements

I am very grateful for the support of Reinhard Furrer as a supervisor and the opportunity for doing this interdisciplinary thesis in the framework of the master program in biostatistics. This master program offers me a great opportunity to broaden and deepen my statistical knowledge besides my research in evolutionary biology.
Further, I want to thank Karin Isler for her support as a supervisor and for the many inspiring and fruitful discussions. Special thanks also go to the following communities and people, who contributed a lot to this thesis: AnthroTree Facebook group, AnthroTree Workshop, AnthroTree Website, R-sig-phylo group, Charly Nunn, Luke Matthews, Emmanuel Paradis, Vincent Carey, Jarrod Hadfield, Lam si Tung Ho, Luke McNally, Markus Gisi, Sandra Heldstab, Caroline Schuppli, Livia Gerber.

References

Abouheif, E. (1999). A method for testing the assumption of phylogenetic independence in comparative data. *Evolutionary Ecology Research*, 1:895–909.

Blomberg, S. P. and Garland, T. J. (2002). Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. *Journal of evolutionary biology*, 15:899–910.

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

Blombgerg, S. P., Lefevre, J. G., Wells, J. A., and Waterhouse, M. (2011). Independent contrasts and pgls regression estimators are equivalent. *Syst. Biol.*, 61(3):382–391.

Cheverud, J. M., Dow, M. M., and Leutenegger, W. (1985). The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution*, 39:1335–1351.

Diaz-Uriarte, R. and Garland, T. (1996). Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Systematic Biology*, 45(1):27–47.

Edwards, A. W. F. and Cavalli-Sforza, L. L. (1964). Reconstruction of evolutionary trees. *Phenetic and Phylogenetic Classification*, (6):67–76.

Fahrmeir, L., Kneib, T., and Lang, S. (2007). *Regression - Modelle, Methoden und Anwendungen*. Springer-Verlag Berlin Heidelberg.

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

Felsenstein, J. (1988). Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics*, 19:445–471.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.

Freckleton, R. P., Harvey, P. H., and Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of the evidence. *The American Naturalist*, 160:712–726.

Fritz, S. A. and Purvis, A. (2010). Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, 24(4):1042–1051.

Garland, T. J. and R, I. A. (2000). Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *The American Naturalist*, 155:346–364.

Ghisletta, P. and Spini, D. (2004). An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational and Behavioural Statistics*, 29(4):421–437.

Gilmour, A. R., Gogel, B., Cullis, B., and Thompson, R. (2009). ASReml user guide release 3.0. *VSN International Ltd, Hemel Hempstead, UK*.

Gittleman, J. L. and Kot, M. (1999). Adaptation: statistics and a null model for estimating phylogenetic effects. *Syst.Zool.*, 39:227–241.

Graber, S., Schuppli, C., Isler, K., and van Schaik, C. (2013). Do larger brained species live in more complex niches? *in preparation*.

Grafen, A. (1989). The phylogenetic regression. *Trans. R. Soc. Lond. B, Biol. Sci.*, 326:119– 157.

Hadfield, J. (2012). MCMCglmm course notes.

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.

Hadfield, J. D. and Nakagawa, S. (2010). General quantitative methods for comparative biology: phylogenies, taxanomies and multi-trait models for continuous and categorical characters. *J. Evol. Biol.*, 23:494–508.

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., and Challenger, W. (2008). Geiger: investigating evolutionary radiations. *Bioinformatics*, 24:129–131.

Heinze, G. and Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in medicine*, 21(16):2409–2419.

Held, L. and Sabanes Bove, D. (2013). *Applied Statistical Inference: Likelihood and Bayes*. Springer Verlag.

Hernández, C. E., Rodríguez-Serrano, E., Avaria-Llautureo, J., Inostroza-Michael, O., Morales-Pallero, B., Boric-Bargetto, D., Canales-Aguirre, C. B., Marquet, P. A., and Meade, A. (2013). Using phylogenetic information and the comparative method to evaluate hypotheses in macroecology. *Methods in Ecology and Evolution*, 4:401–415.

Ho, L. S. and Cecile, A. (2013a). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *in review*.

Ho, L. S. and Cecile, A. (2013b). Package phylolm.

Horton, N. J. and Lipsitz, S. R. (1999). Review of software to fit generalized estimating equation regression models. *The American Statistician*, 53(2):160–169.

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

Ives, A. R. and Garland, T. (2010). Phylogenetic logistic regression for binary dependent variables. *Syst. Biol.*, 59(1):9–26.

Ives, A. R., Midford, P. E., and Garland, T. J. (2007). Within-species variation and measurement error in phylogenetic comparative methods. *Syst. Biol.*, 56:252–270.

Janson, C. H. and van Schaik, C. (1993). *Ecological Risk Aversion in Juvenile Primates: Slow and Steady Wins the Race*. Oxford University Press, New York.

Lavin, S. R., Karasov, W. H., Ives, A. R., Middleton, K. M., and Garland, T. J. (2008). Morphometrics of the avian small intestine compared with that of nonflying mammals: A phylogenetic approach. *Physiological and Biochemical Zoology*, 81(5):526–550.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22.

Lynch, M. (1991). Method for the analysis of comparative data in evolutionary biology. *Evolution*, 50:1065–1080.

Maddison, D. R. and Maddison, W. P. (2000). Macclade 4 manual.

Mancl, L. A. and DeRouen, T. A. (2001). A covariance estimator for gee with improved small-sample properties. *Biometrics*, 57(1):126–134.

Martins, E. P. (1999). Estimation of ancestral states of continuous characters: a computer simulation study. *Systematic Biology*, 48(3):642–650.

Martins, E. P., Diniz-Filho, J. A., and Housworth, E. A. (2002). Adaptive constraints and the phylogenetic comparative method: a computer simulation test. *Evolution*, 56(1):1–13.

Martins, E. P. and Garland, T. J. (1991). Phylogenetic analyses of the correlated evolution of continuous characters: A simulation study. *Evolution*, 45(3):534–557.

Martins, E. P. and Hansen, T. F. (1996). *The statistical analysis of interspecific data: a reviewand evaluation of phylogenetic comparative methods*, pages 22–75. Oxford University press.

Martins, E. P. and Hansen, T. F. (1997). Phylogenies and the comparative method: A general approach to incorporating phylgoenetic information into the analysis of interspecific data. *The American Naturalist*, 149(4):646–667.

Martins, E. P., Pienaar, J., and Hecht Orzack, S. (2008). A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, 62(8):1965–1977.

McCullagh, P. and Nelder, J. A. (1983). Quasi-likelihood functions. *Annals of Statistics*, 11:59–67.

McCullagh, P. and Nelder, J. A. (1989a). *Generalized linear models*. London: Chapman and Hall.

McCullagh, P. and Nelder, J. A. (1989b). Generalized linear models (monographs on statistics and applied probability 37). *Chapman Hall, London*.

Nelder, J. A. and Wedderburn, R. W. A. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series B*, 54:3–40.

Norton, E. C., Bieler, G. S., Ennett, S. T., and Zarkin, G. A. (1996). Analysis of prevention program effectiveness with clustered data using generalized estimating equations. *Journal of Consulting and Clinical Psychology*, 64(5):919–926.

Nunn, C. L. (2011). *The Comparative Approach in Evolutionary Anthropology and Biology*. The University of Chicago Press.

Orme, C., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., and Pearse, W. (2012). Caper: comparative analyses of phylogenetics and evolution in R. *R package version 0.5*.

Orme, D. (2012). The caper package: comparative analysis of phylogenetics and evolution in R.

Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings: Biological Sciences*, 255:37–45.

Pagel, M. (1997). Inferring evolutionary processes from phylogenies. *Zoologica Scripta*, 26(4):331–348.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401:877–884.

Paradis, E., Bolker, B., Claude, J., Sien Cuong, H., Desper, R., Durand, B., Dutheil, J., Gascuel, O., Jobb, G., Heibl, C., Lefort, V., Lemon, J., Noel, Y., Nylander, J., Opgen-Rhein, R., Strimmer, K., and de Viennem, D. (2009). The ape package.

Paradis, E. and Claude, J. (2002). Analysis of comparative data using generalized estimating equations. *J. Theor. Biol.*, 218:175–185.

Paradis, E., Claude, J., and Strimmer, K. (2004). Ape: Analyses of phylogenetics and evolution in r language. *Bioinformatics*, 20(2):289–290.

Plummer, M., Best, N., and Cowles, K. (2012). Package coda.

Postma, E. and Charmantier, A. (2007). What animal models can and cannot tell ornithologists about the genetics of wild populations. *J. Ornithol.*, 148:633–642.

Purvis, A., Cardillo, M., Grenyer, R., and Collen, B. (2005). Correlates of extinction risk: phylogeny, biology, threat and scale. *Phylogeny and conservation (eds A. Purvis, JL Gittleman & T. Brooks)*, pages 295–316.

R-Development-Core-Team (2011). R: A language and environment for statistical computing.

Rao, C. R. and Toutenberg, H. (1995). *Linear Models. Least-squares and Alternatives*. Berlin: Springer Series in Statistics Springer.

Revell, L. J. (2010). Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution*, 1:319–329.

Rohlf, F. J. (2006). A comment on phylogenetic correction. *Evolution*, 60:1509–15.

Schluter, D., Price, T., Mooers, A. O., and Ludwid, D. (1997). Likelihood of ancestor states in adaptive radiation. *Evolution*, 51:1699–1711.

Schuppli, C., Graber, S., Isler, K., and van Schaik, C. (2013). How can species evolve into complex niches. *in preparation*.

Schuppli, C., Isler, K., and van Schaik, C. P. (2012). How to explain the unusually late age at skill competence among humans. *Journal of Human Evolution*, 63(6):843–850.

Stevens, S. S. (1946). On the theory of scales of measurement.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical review*, 36(5):823.

Wedderburn, R. W. M. (1974). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*, 61:439–447.

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrika*, 42:121–130.

Zorn, C. J. W. (2001). Generalized estimating equation models for correlated data: A review with applications. *American Journal of Political Science*, 45(2):470–490.

Appendices

A Comparative methods

A.1 Sampling algorithms used in PGLMM

Metropolis-Hastings-algorithm (Fahrmeir et al. 2007, p. 484):
The aim is to draw random numbers from the density function of a certain posterior distribution is $p(\boldsymbol{\theta}|\boldsymbol{y})$ where direct sampling is difficult:

1. Choose an initial value $\boldsymbol{\theta}^{(0)}$ and the number of total iterations T.

2. Draw a random value $\boldsymbol{\theta}^*$ out of a so called proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$ and accept this as a new $\boldsymbol{\theta}^{(t)}$ with probability of $\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$, otherwise use $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$.
   Probability α is basically the quotient of the posterior distribution and the proposal distribution at the current state of $\boldsymbol{\theta}^{(t-1)}$ and the proposed value $\boldsymbol{\theta}^*$:

$$\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)}) = min\left\{\frac{p(\boldsymbol{\theta}^*|\boldsymbol{y}q(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^*}{p(\boldsymbol{\theta}^{(t-1}|\boldsymbol{y})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})}, 1\right\}.$$

3. If $t = T$ stop the algorithm, otherwise continue with $t = t + 1$ in 2..

After a certain convergence time t_0 the random numbers $\boldsymbol{\theta}^{(t_0+1)}, \ldots, \boldsymbol{\theta}^{(T)}$ represent a random sample from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$.

Gibbs-sampler (Fahrmeir et al. 2007, p. 487):
The aim is to draw random numbers from the density function of a certain posterior distribution is $p(\boldsymbol{\theta}|\boldsymbol{y})$. Assume the parameter vector contains P parameters $\boldsymbol{\theta} = \theta_1, \ldots, \theta_P$:

1. Choose initial values $\theta_1^{(0)}, \ldots, \theta_P^{(0)}$ and the number of total iterations T.

2. For $p = 1, \ldots, P$: draw random numbers $\theta_p^{(t)}$ out of the conditional density: $p(\theta_p|\theta_1^{(t)}, \ldots, \theta_{p-1}^{(t)}, \theta_{p+1}^{(t-1)}, \ldots, \theta_P^{(t-1)}, \boldsymbol{y})$

3. If $t = T$ stop the algorithm, otherwise continue with $t = t + 1$ in 2.

After a certain convergence time t_0 the random numbers $\boldsymbol{\theta}_p^{(t_0+1)}, \ldots, \boldsymbol{\theta}_p^{(T)}$ represent a random sample from the marginal distribution $\boldsymbol{\theta}_p|\boldsymbol{y}$.

B Simulation loops in R

B.1 Simulation loop for continuous response

```
################################################################################
#packages needed:
################################################################################

library(gee)
library(phytools)
library(geiger)
library(ape)
library(caper)
library(MCMCglmm)
library(phylolm)


#################################################################################################################
#Simulation loop continuous response
#################################################################################################################

foo <- function(x) #function to extract all coefficients from the compar.gee output (by Emmanuel Paradis)
{
  nas <- is.na(x$coef)
  coef <- x$coef[!nas]
  cnames <- names(coef)
  coef <- matrix(rep(coef, 4), ncol = 4)
  dimnames(coef) <- list(cnames,
                         c("Estimate", "S.E.", "t", "Pr(T > |t|)"))
  df <- x$dfP - dim(coef)[1]
  coef[, 2] <- sqrt(diag(x$W))
  coef[, 3] <- coef[, 1]/coef[, 2]
  if (df < 0) {
    warning("not enough degrees of freedom to compute P-values.")
    coef[, 4] <- NA
  } else coef[, 4] <- 2 * (1 - pt(abs(coef[, 3]), df))
  coef
}



#set seed, number of simulations and the empty array for saving the outputs of the analyses:
set.seed(1234)

n_sim <- 1000#number of simulations


#empty arrays to save the outputs fo continuous response
results.c.cont <- array(NA, c(4,3,n_sim, 4,  21)) #array for coefficients (1 continuous predictor)
results.c2.cont <- array(NA, c(4,3,n_sim, 4, 31)) #array for coefficients (1 continuous and 1 binary predictor)

results.coefficients.cont <- NULL #empty list (for all 3 number of species) for output with 1 continuous predictor
results.res.cont <- NULL #empty list (for all 3 number of species) for residuals with 1 continuous predictor
results.coefficients2.cont <- NULL #empty list (for all 3 number of species) for output with 1 continuous and 1 binary predictor
results.res2.cont <- NULL #empty list (for all 3 number of species) for residuals with 1 continuous and 1 binary predictor


#start of simulation loop:
for (species in c(20,50,100)){#three different number of species

  res.cont <- array(NA, c(4,3,n_sim,4,species))#array for residuals continuous response
  res2.cont <- array(NA, c(4,3,n_sim,4,species))#array for residuals continuous response


  for (t in 1:4) {#the four different trees:
    tree1 <- sim.bdtree(b=1, d=0, n=species, extinct=T)#random ultrametric tree
    if (t==1) tree <- tree1
    if (t==2) tree <- transform(tree1, "lambda", 0.1) #tree similar to star phylogeny (diversification at base);
    if (t==3) tree <- transform(tree1, "delta", 0.1)#tree opposite to a star phylogeny (diversification at tips)
    tree$edge.length <- replace(tree$edge.length, tree$edge.length=="NaN", 0.5)#if producing NaN in tree3$edge.length
    #-->replace with 0.5, ending up in probably non-ultrametric tree anymore!
    if (t==4) tree <- transform(tree1, "kappa", 0) #tree with all branch lengths equal to 1

    # for(corr in 1:3) {#three different correlation coefficients
    #simulate 2 correlated continuous traits:
    #create variance-covariance matrix for characters
    #q <- cbind (c(1,corrs[corr]),c(corrs[corr],1));
    #q2 <- cbind (c(1,corrs[corr], corrs[corr]),c(corrs[corr],1, corrs[corr]), c(corrs[corr], corrs[corr],1));
```

```
for(corr in 1:3) {#three different correlation coefficients
  #simulate 3 correlated continuous traits:
  #create variance-covariance matrix for characters
  #q <- cbind(c(1,corrs[corr]),c(corrs[corr],1)); #if only simulating two correlated traits
  if(corr==1) corrs=0.01
  q2 <- cbind(c(1,corrs, corrs),c(corrs,1, 0), c(corrs, 0,1)); #if simulating 3 correlated traits
  if(corr==2) corrs=0.5
  q2 <- cbind(c(1,corrs, corrs),c(corrs,1, 0), c(corrs, 0,1))
  if(corr==3) corrs=0.9
  q2 <- cbind(c(1,corrs, corrs),c(corrs,1, 0.65), c(corrs, 0.65,1))



  for(i in 1:n_sim){#number of simulations

    #simulate character evolution along a tree
    sims <- sim.char(tree, par=q2, model="BM", nsim=1);

    ##############prepare data sets: ##################################################################3

    #make binary data out of X3.1 (independent variable-->second covariate):
    trait.x3 = sims[,3,1]
    binary.X3 <- rep(NA,species)
    for(j in 1:species)
      ifelse (trait.x3[j] < mean(trait.x3), binary.X3[j] <- 0, binary.X3[j] <- 1)
    #binary.X3

    #make ordinal data out of X1.1 (dependent variable):
    nominal.y <- rep(NA,species)
    trait.x1 = sims[,1,1]
    nominal.y <- cut(trait.x1, 4, labels=c("1", "2", "3", "4"))

    #make binary data out of X1.1 (dependent variable):
    binary.y <- rep(NA,species)
    for(j in 1:species)
      ifelse (trait.x1[j] < mean(trait.x1), binary.y[j] <- 0, binary.y[j] <- 1)


    animal <- paste("s",c(1:species), sep="")#give species names to each species for creating the comparative data for pgls
    data <- data.frame(animal,sims, binary.X3, nominal.y, binary.y);
    data$nominal.ynumeric <- as.numeric(data$nominal.y)
    data$binary.y <- as.factor(as.character(data[,"binary.y"]))
    #print(tail(data))
    #all(rownames(data) %in% tree$tip.label) #to check wheter tree tip labels and species names in data match


    ########################################
    #perfect fit:
    #1 explanatory variable:
    if(min(data[data[,"binary.y"]==1,"X2.1"]) > max(data[data[,"binary.y"]==0,"X2.1"])){

      results.c.cont[t, corr, i, 1, 17] <- 333
      results.c.cont[t, corr, i, 2, 17] <- 333
      results.c.cont[t, corr, i, 3, 17] <- 333
      results.c.cont[t, corr, i, 4, 17] <- 333

    }

    if (F){
    #two explanatory variables
    if((min(data[data[,"binary.y"]==1,"X2.1"]) > max(data[data[,"binary.y"]==0,"X2.1"])) & (isTRUE(data[,"binary.y"] == data[,"binary.X3"]))){

      results.c2.cont[t, corr, i, 1, 25] <- 333
      results.c2.cont[t, corr, i, 2, 25] <- 333
      results.c2.cont[t, corr, i, 3, 25] <- 333
      results.c2.cont[t, corr, i, 4, 25] <- 333

    }

    }#end if (F)
    ########### continuous response variable ###########################################################


      ######################
      #non-phylogenetic glm:

      #1 continuous predictor:
      lm1 <- try(lm(X1.1 ~ X2.1, data), silent=T)
      #print(summary.lm(lm1))

      if (class(lm1) != "try-error"){
        results.c.cont[t, corr, i, 1, c(1:8)] <- matrix(summary(lm1)$coefficients, 1, 8)
        res.cont[t,corr,i,1,c(1:species)] <- matrix(lm1$residual, 1, species)
      }
      else{
        results.c.cont[t, corr, i, 1, c(1:8)] <- matrix(99, 1, 8)
        res.cont[t,corr,i,1,c(1:species)] <- matrix(99, 1, species)
```

```
}

#1 continuous and 1 binary predictor:
lm2 <- try(lm(X1.1 ~ X2.1 + binary.X3, data), silent=T)
#print(summary.lm(lm2))

if (class(lm2) != "try-error"){
  results.c2.cont[t, corr, i, 1, c(1:12)] <- matrix(summary(lm2)$coefficients, 1, 12)
  res2.cont[t,corr,i,1,c(1:species)] <- matrix(lm2$residual, 1, species)
}

else {
  results.c2.cont[t, corr, i, 1, c(1:12)] <- matrix(99, 1, 12)
  res2.cont[t,corr,i,1,c(1:species)] <- matrix(99, 1, species)
}



########################
#phylogenetic glm = pgls:
comp.data <- comparative.data(phy = tree, data = data, names.col = animal, vcv = TRUE)
#print(comp.data)

#1 continuous predictor:
pgls1 <- try(pgls(X1.1 ~ X2.1, data = comp.data, lambda="ML", bounds = list(lambda=c(0.001,1), kappa=c(1e-6,3), delta=c(1e-6,3))), silent=T)
#print(summary(pgls1))

if (class(pgls1) != "try-error"){
  results.c.cont[t, corr, i, 2, c(1:8)] <- matrix(summary(pgls1)$coefficients, 1, 8)
  results.c.cont[t, corr, i, 2, 9] <- matrix(as.numeric(summary(pgls1)$param[2]), 1, 1)
  res.cont[t,corr,i,2,c(1:species)] <- matrix(pgls1$residual, 1, species)
}

else {
  results.c.cont[t, corr, i, 2, c(1:9)] <- matrix(99, 1, 9)
  res.cont[t,corr,i,2,c(1:species)] <- matrix(99, 1, species)
}


#1 continuous and 1 binary predictor:
pgls2 <- try(pgls(X1.1 ~ X2.1 + binary.X3, data = comp.data, lambda="ML", bounds = list(lambda=c(0.001,1), kappa=c(1e-6,3), delta=c(1e-6,3))), sile
#print(summary(pgls2))

if (class(pgls2) != "try-error"){
  results.c2.cont[t, corr, i, 2, c(1:12)] <- matrix(summary(pgls2)$coefficients, 1, 12)
  results.c2.cont[t, corr, i, 2, 13] <- matrix(as.numeric(summary(pgls2)$param[2]), 1, 1)
  res2.cont[t,corr,i,2,c(1:species)] <- matrix(pgls2$residual, 1, species)
}

else {
  results.c2.cont[t, corr, i, 2, c(1:13)] <- matrix(99, 1, 13)
  res2.cont[t,corr,i,2,c(1:species)] <- matrix(99, 1, species)
}


####################
#phylogenetic GEE:

#1 continuous predictor:
pgee <- try(compar.gee(X1.1 ~ X2.1,  data = data, phy=tree), silent=T)
#print(pgee)

if (class(pgee) != "try-error"){
  results.c.cont[t, corr, i, 3, c(1:8)] <- matrix(foo(pgee), 1, 8)
  res.cont[t,corr,i,3,c(1:species)] <- matrix(pgee$residual, 1, species)
}

else {
  results.c.cont[t, corr, i, 3, c(1:8)] <- matrix(99, 1, 8)
  res.cont[t,corr,i,3,c(1:species)] <- matrix(99, 1, species)
}

#1 continuous and 1 binary predictor:
pgee2 <- try(compar.gee(X1.1 ~ X2.1 + binary.X3,  data = data, phy=tree), silent=T)
#print(pgee2)

if (class(pgee2) != "try-error"){
  results.c2.cont[t, corr, i, 3, c(1:12)] <- matrix(foo(pgee2), 1, 12)
  res2.cont[t,corr,i,3,c(1:species)] <- matrix(pgee2$residual, 1, species)
}

else {
  results.c2.cont[t, corr, i, 3, c(1:12)] <- matrix(99, 1, 12)
  res2.cont[t,corr,i,3,c(1:species)] <- matrix(99, 1, species)
}
```

```
          ##########################
          #Phylogenetic mixed model:


          #1 continuous predictor:
          prior <- list(R=list(V=1,nu=0.002), G=list(G1=list(V=1, nu=0.002)))
          m1 <- try(MCMCglmm(X1.1 ~ X2.1, random=~animal, data=data, pedigree=tree, prior=prior, pr=F, scale=F,saveX=F, nitt=30000), silent=T)#iterations!
          #print(summary(m1))

          #convergence diagnostics:
          diag1=geweke.diag(m1$Sol)
          results.c.cont[t, corr, i, 4, 18] <- matrix(diag1$z[1],1,1)#save z-value from geweke diagnostic test
          results.c.cont[t, corr, i, 4, 19] <- matrix(diag1$z[2],1,1)#save z-value from geweke diagnostic test
          results.c.cont[t, corr, i, 4, 20] <- matrix(2*pnorm(-abs(diag1$z[1])),1,1)#save p-value from geweke diagnostic test
          results.c.cont[t, corr, i, 4, 21] <- matrix(2*pnorm(-abs(diag1$z[2])),1,1)#save p-value from geweke diagnostic test


          if (class(m1) != "try-error"){
            results.c.cont[t, corr, i, 4, c(1:10)] <- matrix(summary(m1)$solutions[,], 1, 10)
            results.c.cont[t, corr, i, 4, c(11:16)] <- matrix(summary(m1$Sol)$statistics[,c(2:4)], 1, 6)
          }

          else {
            results.c.cont[t, corr, i, 4, c(1:10)] <- matrix(99, 1, 10)
            results.c.cont[t, corr, i, 4, c(11:16)] <- matrix(99, 1, 6)
          }

          #1 continuous and 1 binary predictor:
          #r0=0.5
          #var <- cbind(c(1e+08,r0,r0), c(r0,1e+08,r0),c(r0,r0,1e+08))#vcv for prior of fixed effects (B)
          #B=list(mu=rep(0,3), V=var)
          prior2 <- list(R=list(V=1,nu=0.002), G=list(G1=list(V=1, nu=0.002)))
          m2 <- try(MCMCglmm(X1.1 ~ X2.1 + binary.X3, random=~animal, data=data, scale=F,pedigree=tree, prior=prior2, pr=F, saveX=F, nitt=30000), silent=T)#i
          #print(summary(m2))

          #convergence diagnostics:
          diag2=geweke.diag(m2$Sol)
          results.c2.cont[t, corr, i, 4, 26] <- matrix(diag2$z[1],1,1)#save z-value from geweke diagnostic test
          results.c2.cont[t, corr, i, 4, 27] <- matrix(diag2$z[2],1,1)#save z-value from geweke diagnostic test
          results.c2.cont[t, corr, i, 4, 28] <- matrix(diag2$z[3],1,1)#save z-value from geweke diagnostic test
          results.c2.cont[t, corr, i, 4, 29] <- matrix(2*pnorm(-abs(diag2$z[1])),1,1)#save p-value from geweke diagnostic test
          results.c2.cont[t, corr, i, 4, 30] <- matrix(2*pnorm(-abs(diag2$z[2])),1,1)#save p-value from geweke diagnostic test
          results.c2.cont[t, corr, i, 4, 31] <- matrix(2*pnorm(-abs(diag2$z[3])),1,1)#save p-value from geweke diagnostic test

          if (class(m1) != "try-error"){
            results.c2.cont[t, corr, i, 4, c(1:15)] <- matrix(summary(m2)$solutions[,], 1, 15)
            results.c2.cont[t, corr, i, 4, c(16:24)] <- matrix(summary(m2$Sol)$statistics[,c(2:4)], 1, 9)
          }

          else {
            results.c2.cont[t, corr, i, 4, c(1:15)] <- matrix(99, 1, 15)
            results.c2.cont[t, corr, i, 4, c(16:24)] <- matrix(99, 1, 9)
          }



       }#end of n_sim loop


     }#end of corr loop

   } #end of tree loop


   #results continuous response
   results.coefficients.cont <- list(results.coefficients.cont, results.c.cont)
   results.coefficients2.cont <- list(results.coefficients2.cont, results.c2.cont)


   #residuals continuous response
   results.res.cont <- list(results.res.cont, res.cont)
   results.res2.cont <- list(results.res2.cont, res2.cont)


}# end of species loop
```

B.2 Simulation loop for ordinal response

```
################################################################################
#packages needed:
################################################################################

library(gee)
library(phytools)
library(geiger)
library(ape)
library(caper)
library(MCMCglmm)
library(phylolm)



#################################################################################################################
#Simulation loop ordinal response
#################################################################################################################
foo <- function(x) #function to extract all coefficients from the compar.gee output (by Emmanuel Paradis)
{
  nas <- is.na(x$coef)
  coef <- x$coef[!nas]
  cnames <- names(coef)
  coef <- matrix(rep(coef, 4), ncol = 4)
  dimnames(coef) <- list(cnames,
                         c("Estimate", "S.E.", "t", "Pr(T > |t|)"))
  df <- x$dfP - dim(coef)[1]
  coef[, 2] <- sqrt(diag(x$W))
  coef[, 3] <- coef[, 1]/coef[, 2]
  if (df < 0) {
    warning("not enough degrees of freedom to compute P-values.")
    coef[, 4] <- NA
  } else coef[, 4] <- 2 * (1 - pt(abs(coef[, 3]), df))
  coef
}



#set seed, number of simulations and the empty array for saving the outputs of the analyses:
set.seed(1234)

n_sim <- 1000 #number of simulations


#empty arrays to save the outputs fo ordinal response
results.c.nominal <- array(NA, c(4,3,n_sim, 4, 53)) #array for coefficients (1 continuous predictor)
results.c2.nominal <- array(NA, c(4,3,n_sim, 4, 79)) #array for coefficients (1 continuous and 1 binary predictor)

results.coefficients.nominal <- NULL #empty list (for all 3 number of species) for output with 1 continuous predictor
results.res.nominal <- NULL #empty list (for all 3 number of species) for residuals with 1 continuous predictor
results.coefficients2.nominal <- NULL #empty list (for all 3 number of species) for output with 1 continuous and 1 binary predictor
results.res2.nominal <- NULL #empty list (for all 3 number of species) for residuals with 1 continuous and 1 binary predictor



#start of simulation loop:
for (species in c(20,50,100)){#three different number of species

  res.nominal <- array(NA, c(4,3,n_sim,4,species))#array for residuals nominal respsonse
  res2.nominal <- array(NA, c(4,3,n_sim,4,species))#array for residuals nominal response

  for (t in 1:4) {#the four different trees:
    tree1 <- sim.bdtree(b=1, d=0, n=species, extinct=T)#random ultrametric tree
    if (t==1) tree <- tree1
    if (t==2) tree <- transform(tree1, "lambda", 0.1) #tree similar to star phylogeny (diversification at base);
    if (t==3) tree <- transform(tree1, "delta", 0.1)#tree opposite to a star phylogeny (diversification at tips)
    tree$edge.length <- replace(tree$edge.length, tree$edge.length=="NaN", 0.5)#if producing NaN in tree3$edge.length
    #-->replace with 0.5, ending up in probably non-ultrametric tree anymore!
    if (t==4) tree <- transform(tree1, "kappa", 0) #tree with all branch lengths equal to 1

    # for(corr in 1:3) {#three different correlation coefficients
    #simulate 2 correlated continuous traits:
    #create variance-covariance matrix for characters
    #q <- cbind (c(1,corrs[corr]),c(corrs[corr],1));
    #q2 <- cbind (c(1,corrs[corr], corrs[corr]),c(corrs[corr],1, corrs[corr]), c(corrs[corr], corrs[corr],1));


    for(corr in 1:3) {#three different correlation coefficients
      #simulate 3 correlated continuous traits:
      #create variance-covariance matrix for characters
      #q <- cbind(c(1,corrs[corr]),c(corrs[corr],1)); #if only simulating two correlated traits
      if(corr==1) corrs=0.01
      q2 <- cbind(c(1,corrs, corrs),c(corrs,1, 0), c(corrs, 0,1)); #if simulating 3 correlated traits
      if(corr==2) corrs=0.5
      q2 <- cbind(c(1,corrs, corrs),c(corrs,1, 0), c(corrs, 0,1))
      if(corr==3) corrs=0.9
```

```
q2 <- cbind(c(1,corrs, corrs),c(corrs,1, 0.65), c(corrs, 0.65,1))



for(i in 1:n_sim){#number of simulations

  #simulate character evolution along a tree
  sims <- sim.char(tree, par=q2, model="BM", nsim=1);

  ###############prepare data sets: ##################################################################3

  #make binary data out of X3.1 (independent variable-->second covariate):
  trait.x3 = sims[,3,1]
  binary.X3 <- rep(NA,species)
  for(j in 1:species)
    ifelse (trait.x3[j] < mean(trait.x3), binary.X3[j] <- 0, binary.X3[j] <- 1)
  #binary.X3

  #make ordinal data out of X1.1 (dependent variable):
  nominal.y <- rep(NA,species)
  trait.x1 = sims[,1,1]
  nominal.y <- cut(trait.x1, 4, labels=c("1", "2", "3", "4"))

  #make binary data out of X1.1 (dependent variable):
  binary.y <- rep(NA,species)
  for(j in 1:species)
    ifelse (trait.x1[j] < mean(trait.x1), binary.y[j] <- 0, binary.y[j] <- 1)


  animal <- paste("s",c(1:species), sep="")#give species names to each species for creating the comparative data for pgls
  data <- data.frame(animal,sims, binary.X3, nominal.y, binary.y);
  data$nominal.ynumeric <- as.numeric(data$nominal.y)
  data$binary.y <- as.factor(as.character(data[,"binary.y"]))
  #print(tail(data))
  #all(rownames(data) %in% tree$tip.label) #to check wheter tree tip labels and species names in data match


  ########################################
  #perfect fit:
  #1 explanatory variable:
  if(min(data[data[,"binary.y"]==1,"X2.1"]) > max(data[data[,"binary.y"]==0,"X2.1"])){

    results.c.nominal[t, corr, i, 1, 49] <- 333
    results.c.nominal[t, corr, i, 2, 49] <- 333
    results.c.nominal[t, corr, i, 3, 49] <- 333
    results.c.nominal[t, corr, i, 4, 49] <- 333

  }

  if(F){
  #two explanatory variables
  if((min(data[data[,"binary.y"]==1,"X2.1"]) > max(data[data[,"binary.y"]==0,"X2.1"])) & (isTRUE(data[,"binary.y"] == data[,"binary.X3"]))){


    results.c2.nominal[t, corr, i, 1, 73] <- 333
    results.c2.nominal[t, corr, i, 2, 73] <- 333
    results.c2.nominal[t, corr, i, 3, 73] <- 333
    results.c2.nominal[t, corr, i, 4, 73] <- 333

  }

  }#end if(F)

    ############ nominal response variable ###############################################


    ######################
    #non-phylogenetic glm:

    #1 continuous predictor:
    lm21 <- try(lm(nominal.ynumeric ~ X2.1, data), silent=T)
    #print(summary(lm21))

    if (class(lm21) != "try-error"){
      results.c.nominal[t, corr, i, 1, c(1:8)] <- matrix(summary(lm21)$coefficients, 1, 8)
      res.nominal[t,corr,i,1,c(1:species)] <- matrix(lm21$residual, 1, species)
    }

    else {
      results.c.nominal[t, corr, i, 1, c(1:8)] <- matrix(99, 1, 8)
      res.nominal[t,corr,i,1,c(1:species)] <- matrix(99, 1, species)
    }


    #1 continuous and 1 binary predictor:
    lm22 <- try(lm(nominal.ynumeric ~ X2.1 + binary.X3, data), silent=T)
    #print(summary(lm22))
```

```
    if (class(lm22) != "try-error"){
      results.c2.nominal[t, corr, i, 1, c(1:12)] <- matrix(summary(lm22)$coefficients, 1, 12)
      res2.nominal[t,corr,i,1,c(1:species)] <- matrix(lm22$residual, 1, species)
    }

    else {
      results.c2.nominal[t, corr, i, 1, c(1:12)] <- matrix(99, 1, 12)
      res2.nominal[t,corr,i,1,c(1:species)] <- matrix(99, 1, species)
    }



    ########################
    #phylogenetic glm = pgls:
    comp.data <- comparative.data(phy = tree, data = data, names.col = animal, vcv = TRUE)
    #print(comp.data)

    #1 continuous predictor:
    pgls21 <- try(pgls(nominal.ynumeric ~ X2.1, data = comp.data, lambda="ML"), silent=T)
    #print(summary(pgls21))

    if (class(pgls21) != "try-error"){
      results.c.nominal[t, corr, i, 2, c(1:8)] <- matrix(summary(pgls21)$coefficients, 1, 8)
      results.c.nominal[t, corr, i, 2, 9] <- matrix(as.numeric(summary(pgls21)$param[2]), 1, 1)
      res.nominal[t,corr,i,2,c(1:species)] <- matrix(pgls21$residual, 1, species)
    }

    else {
      results.c.nominal[t, corr, i, 2, c(1:9)] <- matrix(99, 1, 9)
      res.nominal[t,corr,i,2,c(1:species)] <- matrix(99, 1, species)
    }

    #1 continuous and 1 binary predictor:
    pgls22 <- try(pgls(nominal.ynumeric ~ X2.1 + binary.X3, data = comp.data, lambda="ML"), silent=T)
    #print(summary(pgls22))

    if (class(pgls22) != "try-error"){
      results.c2.nominal[t, corr, i, 2, c(1:12)] <- matrix(summary(pgls22)$coefficients, 1, 12)
      results.c2.nominal[t, corr, i, 2, 13] <- matrix(as.numeric(summary(pgls22)$param[2]), 1, 1)
      res2.nominal[t,corr,i,2,c(1:species)] <- matrix(pgls22$residual, 1, species)
    }

    else {
      results.c2.nominal[t, corr, i, 2, c(1:13)] <- matrix(99, 1, 13)
      res2.nominal[t,corr,i,2,c(1:species)] <- matrix(99, 1, species)
    }


    ####################
    #phylogenetic GEE:

    #1 continuous predictor:
    pgee2 <- try(compar.gee(nominal.ynumeric ~ X2.1,  data = data, phy=tree), silent=T)
    #print(pgee2)

    if (class(pgee2) != "try-error"){
      results.c.nominal[t, corr, i, 3, c(1:8)] <- matrix(foo(pgee2), 1, 8)
      res.nominal[t,corr,i,3,c(1:species)] <- matrix(pgee2$residual, 1, species)
    }

    else {
      results.c.nominal[t, corr, i, 3, c(1:8)] <- matrix(99, 1, 8)
      res.nominal[t,corr,i,3,c(1:species)] <- matrix(99, 1, species)
    }



    #1 continuous and 1 binary predictor:
    pgee22 <- try(compar.gee(nominal.ynumeric ~ X2.1 + binary.X3,  data = data, phy=tree), silent=T)
    #print(pgee22)

    if (class(pgee22) != "try-error"){
      results.c2.nominal[t, corr, i, 3, c(1:12)] <- matrix(foo(pgee22), 1, 12)
      res2.nominal[t,corr,i,3,c(1:species)] <- matrix(pgee22$residual, 1, species)
    }
    else {
      results.c2.nominal[t, corr, i, 3, c(1:12)] <- matrix(99, 1, 12)
      res2.nominal[t,corr,i,3,c(1:species)] <- matrix(99, 1, species)
    }



########################
#Phylogenetic mixed model:

#1 continuous predictor:
k <- length(levels(data$nominal.y)); k
I <- diag(k-1); I
```

```
      J <- matrix(rep(1, (k-1)^2), c(k-1, k-1)); J
      IJ <- (1/k) * (I + J); IJ #this prior implies that the variance in each probability is constant and that probabilities of
      #different outcomes are mutually independent, conditional on the constraint that they must sum to one.
      prior22 <- list(R = list(V = IJ, fix=1),G = list(G1 = list(V = IJ,nu=0.002)))
      m21 <- try(MCMCglmm(nominal.y ~ trait-1 + trait:X2.1, random=~us(trait):animal,rcov = ~us(trait):units, scale=F,family="categorical", data=data, pedigr
      #print(summary(m21))

    #convergence diagnostics:
    diag11=geweke.diag(m21$Sol)
    results.c.nominal[t, corr, i, 4, 50] <- matrix(diag11$z[1],1,1)#save z-value from geweke diagnostic test
    results.c.nominal[t, corr, i, 4, 51] <- matrix(diag11$z[2],1,1)#save z-value from geweke diagnostic test
    results.c.nominal[t, corr, i, 4, 52] <- matrix(2*pnorm(-abs(diag11$z[1])),1,1)#save p-value from geweke diagnostic test
   results.c.nominal[t, corr, i, 4, 53] <- matrix(2*pnorm(-abs(diag11$z[2])),1,1)#save p-value from geweke diagnostic test


          if (class(m21) != "try-error"){
            results.c.nominal[t, corr, i, 4, c(1:30)] <- matrix(summary(m21)$solutions[,], 1, 30)
            results.c.nominal[t, corr, i, 4, c(31:48)] <- matrix(summary(m21$Sol)$statistics[,c(2:4)], 1, 18)
          }

          else {
            results.c.nominal[t, corr, i, 4, c(1:30)] <- matrix(99, 1, 30)
            results.c.nominal[t, corr, i, 4, c(31:48)] <- matrix(99, 1, 18)
          }

          #1 continuous and 1 binary predictor:
          #r=0.5
          Prior.phyl6 = list(R = list(V = IJ, fix=1),G = list(G1 = list(V = IJ,nu=0.002)))
          m22 <- try(MCMCglmm(nominal.y ~ trait-1+ trait:X2.1 + trait:binary.X3, random=~us(trait):animal,scale=F,rcov = ~us(trait):units, pedigree=tree,
                              data = data, family="categorical", nitt=12000, pl=T,
                              prior=Prior.phyl6), silent=T)
          #print(summary(m22))

   #convergence diagnostics:
   diag12=geweke.diag(m22$Sol)
   results.c2.nominal[t, corr, i, 4, 74] <- matrix(diag12$z[1],1,1)#save z-value from geweke diagnostic test
   results.c2.nominal[t, corr, i, 4, 75] <- matrix(diag12$z[2],1,1)#save z-value from geweke diagnostic test
   results.c2.nominal[t, corr, i, 4, 76] <- matrix(diag12$z[3],1,1)#save z-value from geweke diagnostic test
   results.c2.nominal[t, corr, i, 4, 77] <- matrix(2*pnorm(-abs(diag12$z[1])),1,1)#save p-value from geweke diagnostic test
  results.c2.nominal[t, corr, i, 4, 78] <- matrix(2*pnorm(-abs(diag12$z[2])),1,1)#save p-value from geweke diagnostic test
   results.c2.nominal[t, corr, i, 4, 78] <- matrix(2*pnorm(-abs(diag12$z[3])),1,1)#save p-value from geweke diagnostic test


          if (class(m22) != "try-error"){
            results.c2.nominal[t, corr, i, 4, c(1:45)] <- matrix(summary(m22)$solutions[,], 1, 45)
            results.c2.nominal[t, corr, i, 4, c(46:72)] <- matrix(summary(m22$Sol)$statistics[,c(2:4)], 1, 27)
          }

          else {
            results.c2.nominal[t, corr, i, 4, c(1:45)] <- matrix(99, 1, 45)
            results.c2.nominal[t, corr, i, 4, c(46:72)] <- matrix(99, 1, 27)
          }




      }#end of n_sim loop


    }#end of corr loop

   } #end of tree loop


   #results nominal response
   results.coefficients.nominal <- list(results.coefficients.nominal, results.c.nominal)
   results.coefficients2.nominal <- list(results.coefficients2.nominal, results.c2.nominal)

   #residuals nominal response
   results.res.nominal <- list(results.res.nominal, res.nominal)
   results.res2.nominal <- list(results.res2.nominal, res2.nominal)

}# end of species loop
```

B.3 Simulation loop for binary response

```
################################################################################
#packages needed:
################################################################################

library(gee)
library(phytools)
library(geiger)
library(ape)
library(caper)
library(MCMCglmm)
library(phylolm)


################################################################################
#Simulation loop binary response
################################################################################


set.seed(123)

foo <- function(x) #function to extract all coefficients from the compar.gee output (by Emmanuel Paradis)
{
  nas <- is.na(x$coef)
  coef <- x$coef[!nas]
  cnames <- names(coef)
  coef <- matrix(rep(coef, 4), ncol = 4)
  dimnames(coef) <- list(cnames,
                         c("Estimate", "S.E.", "t", "Pr(T > |t|)"))
  df <- x$dfP - dim(coef)[1]
  coef[, 2] <- sqrt(diag(x$W))
  coef[, 3] <- coef[, 1]/coef[, 2]
  if (df < 0) {
    warning("not enough degrees of freedom to compute P-values.")
    coef[, 4] <- NA
  } else coef[, 4] <- 2 * (1 - pt(abs(coef[, 3]), df))
  coef
}


n_sim <- 1000 #number of simulations

#empty arrays to save the outputs fo binary response
results.c.binary <- array(NA, c(4,3,n_sim, 4, 27)) #array for coefficients (1 continuous predictor)
results.c2.binary <- array(NA, c(4,3,n_sim, 4, 32)) #array for coefficients (1 continuous and 1 binary predictor)

results.coefficients.partII.binary <- NULL #empty list (for all 3 number of species) for output with 1 continuous predictor
results.res.partII.binary <- NULL #empty list (for all 3 number of species) for residuals with 1 continuous predictor
results.coefficients2.partII.binary <- NULL #empty list (for all 3 number of species) for output with 1 continuous and 1 binary predictor
results.res2.partII.binary <- NULL #empty list (for all 3 number of species) for residuals with 1 continuous and 1 binary predictor



#start of simulation loop:
for (species in c(20,50,100)){#three different number of species

  res.binary <- array(NA, c(4,3,n_sim,4,species))#array for residuals binary response
  res2.binary <- array(NA, c(4,3,n_sim,4,species))#array for residuals binary response

  for (t in 1:4) {#the four different trees:
    tree1 <- sim.bdtree(b=1, d=0, n=species, extinct=T)#random ultrametric tree
    if (t==1) tree <- tree1
    if (t==2) tree <- transform(tree1, "lambda", 0.1) #tree similar to star phylogeny (diversification at base);
    if (t==3) tree <- transform(tree1, "delta", 0.1)#tree opposite to a star phylogeny (diversification at tips)
    tree$edge.length <- replace(tree$edge.length, tree$edge.length=="NaN", 0.5)#if producing NaN in tree3$edge.length
    #-->replace with 0.5, ending up in probably non-ultrametric tree anymore!
    if (t==4) tree <- transform(tree1, "kappa", 0) #tree with all branch lengths equal to 1

    # for(corr in 1:3) {#three different correlation coefficients
    #simulate 2 correlated continuous traits:
    #create variance-covariance matrix for characters
    #q <- cbind (c(1,corrs[corr]),c(corrs[corr],1));
    #q2 <- cbind (c(1,corrs[corr], corrs[corr]),c(corrs[corr],1, corrs[corr]), c(corrs[corr], corrs[corr],1));


    for(corr in 1:3) {#three different correlation coefficients
      #simulate 3 correlated continuous traits:
      #create variance-covariance matrix for characters
      #q <- cbind(c(1,corrs[corr]),c(corrs[corr],1)); #if only simulating two correlated traits
      if(corr==1) corrs=0.01
      q2 <- cbind(c(1,corrs, corrs),c(corrs,1, 0), c(corrs, 0,1)); #if simulating 3 correlated traits
      if(corr==2) corrs=0.5
      q2 <- cbind(c(1,corrs, corrs),c(corrs,1, 0), c(corrs, 0,1))
      if(corr==3) corrs=0.9
      q2 <- cbind(c(1,corrs, corrs),c(corrs,1, 0.65), c(corrs, 0.65,1))
```

```
for(i in 1:n_sim){#number of simulations

  #simulate character evolution along a tree
  sims <- sim.char(tree, par=q2, model="BM", nsim=1);

  ###############prepare data sets: ##################################################################3

  #make binary data out of X3.1 (independent variable-->second covariate):
  trait.x3 = sims[,3,1]
  binary.X3 <- rep(NA,species)
  for(j in 1:species)
    ifelse (trait.x3[j] < mean(trait.x3), binary.X3[j] <- 0, binary.X3[j] <- 1)
  #binary.X3

  #make ordinal data out of X1.1 (dependent variable):
  nominal.y <- rep(NA,species)
  trait.x1 = sims[,1,1]
  nominal.y <- cut(trait.x1, 4, labels=c("1", "2", "3", "4"))

  #make binary data out of X1.1 (dependent variable):
  binary.y <- rep(NA,species)
  for(j in 1:species)
    ifelse (trait.x1[j] < mean(trait.x1), binary.y[j] <- 0, binary.y[j] <- 1)


  animal <- paste("s",c(1:species), sep="")#give species names to each species for creating the comparative data for pgls
  data <- data.frame(animal,sims, binary.X3, nominal.y, binary.y);
  data$nominal.ynumeric <- as.numeric(data$nominal.y)
  data$binary.y <- as.factor(as.character(data[,"binary.y"]))
  #print(tail(data))


  ########################################
    #perfect fit:
    #1 explanatory variable:
    if(min(data[data[,"binary.y"]==1,"X2.1"]) > max(data[data[,"binary.y"]==0,"X2.1"])){

      results.c.binary[t, corr, i, 1, 17] <- 333
      results.c.binary[t, corr, i, 2, 17] <- 333
      results.c.binary[t, corr, i, 3, 17] <- 333
      results.c.binary[t, corr, i, 4, 17] <- 333

    }

  if(F){
    #two explanatory variables
    if((min(data[data[,"binary.y"]==1,"X2.1"]) > max(data[data[,"binary.y"]==0,"X2.1"])) & (isTRUE(data[,"binary.y"] == data[,"binary.X3"]))){

      results.c2.binary[t, corr, i, 1, 25] <- 333
      results.c2.binary[t, corr, i, 2, 25] <- 333
      results.c2.binary[t, corr, i, 3, 25] <- 333
      results.c2.binary[t, corr, i, 4, 25] <- 333
    }
  }#if F end


  ##################### binary response variable ############################################################

  ######################
  #non-phylogenetic glm:

  #1 continuous predictor:
  lm31 <- try(glm(binary.y ~ X2.1, data, family=binomial(link = "logit")), silent=t)
  #print(summary(lm31))

  if (class(lm31) != "try-error"){
    results.c.binary[t, corr, i, 1, c(1:8)] <- matrix(summary(lm31)$coefficients, 1, 8)
    results.c.binary[t, corr, i, 1, 13] <- matrix(summary(lm31)$iter, 1, 1)
    res.binary[t,corr,i,1,c(1:species)] <- matrix(lm31$residual, 1, species)

  }

  else {
    results.c.binary[t, corr, i, 1, c(1:8)] <- matrix(99, 1, 8)
    res.binary[t,corr,i,1,c(1:species)] <- matrix(99, 1, species)
  }


  #1 continuous and 1 binary predictor:
  lm32 <- try(glm(binary.y ~ X2.1 + binary.X3, data, family=binomial(link = "logit")), silent=T)
  #print(summary(lm32))

  if (class(lm32) != "try-error"){
    results.c2.binary[t, corr, i, 1, c(1:12)] <- matrix(summary(lm32)$coefficients, 1, 12)
    results.c2.binary[t, corr, i, 1, 13] <- matrix(summary(lm32)$iter, 1, 1)
    res2.binary[t,corr,i,1,c(1:species)] <- matrix(lm32$residual, 1, species)
```

```
  }

  else {
    results.c2.binary[t, corr, i, 1, c(1:12)] <- matrix(99, 1, 12)
    res2.binary[t,corr,i,1,c(1:species)] <- matrix(99, 1, species)
  }


######################
#phylogenetic GEE:
#freezes sometimes, not possible to run simulation loop

  if(F){
  #1 continuous predictor:
  pgee3 <- try(compar.gee(binary.y ~ X2.1,  data = data, family="binomial",phy=tree), silent=T)
  print(pgee3)

  if (class(pgee3) != "try-error"){
    results.c.binary[t, corr, i, 2, c(1:8)] <- matrix(foo(pgee3), 1, 8)
    res.binary[t,corr,i,2,c(1:species)] <- matrix(pgee3$residual, 1, species)
  }

  else {
    results.c.binary[t, corr, i, 2, c(1:8)] <- matrix(99, 1, 8)
    res.binary[t,corr,i,2,c(1:species)] <- matrix(99, 1, species)
  }



  #1 continuous and 1 binary predictor:
  pgee32 <- try(compar.gee(binary.y ~ X2.1 + binary.X3,  data = data, family="binomial", phy=tree), silent=T)
  print(pgee32)

  if (class(pgee32) != "try-error"){
    results.c2.binary[t, corr, i, 2, c(1:12)] <- matrix(foo(pgee32), 1, 12)
    res2.binary[t,corr,i,2,c(1:species)] <- matrix(pgee32$residual, 1, species)
  }

  else {
    results.c2.binary[t, corr, i, 2, c(1:12)] <- matrix(99, 1, 12)
    res2.binary[t,corr,i,2,c(1:species)] <- matrix(99, 1, species)
  }

  }


########################
#Phylogenetic mixed model:

  #1 continuous predictor:
  Prior.phyl51 = list(R = list(V = 1, fix=1),G = list(G1 = list(V = 1,nu=0.002)))
  m31 <- try(MCMCglmm(binary.y ~ X2.1, random=~animal, rcov = ~units, pedigree=tree,
                      data = data, family="categorical", nitt=12000, scale=F,
                      prior=Prior.phyl51), silent=T)#iterations!

  #print(summary(m31))



  if (class(m31) != "try-error"){
    #convergence diagnostics:
    diag1=geweke.diag(m31$Sol)
    results.c.binary[t, corr, i, 3, 18] <- matrix(diag1$z[1],1,1)#save z-value from geweke diagnostic test
    results.c.binary[t, corr, i, 3, 19] <- matrix(diag1$z[2],1,1)#save z-value from geweke diagnostic test
    results.c.binary[t, corr, i, 3, 20] <- matrix(2*pnorm(-abs(diag1$z[1])),1,1)#save p-value from geweke diagnostic test
    results.c.binary[t, corr, i, 3, 21] <- matrix(2*pnorm(-abs(diag1$z[2])),1,1)#save p-value from geweke diagnostic test

    results.c.binary[t, corr, i, 3, c(1:10)] <- matrix(summary(m31)$solutions[,], 1, 10)
    results.c.binary[t, corr, i, 3, c(11:16)] <- matrix(summary(m31$Sol)$statistics[,c(2:4)], 1, 6)
  }

  else {
    results.c.binary[t, corr, i, 3, c(1:16)] <- matrix(99, 1, 16)
  }


  #1 continuous and 1 binary predictor:
  #r1=0.5
  #var2 <- cbind(c(1e+08,r1,r1), c(r1,1e+08,r1),c(r1,r1,1e+08)); var1#vcv for prior of fixed effects (B)
  #B=list(mu=rep(0,3), V=var2)
  Prior.phyl52 = list(R = list(V = 1, fix=1),G = list(G1 = list(V = 1,nu=0.002)))
  m32 <- try(MCMCglmm(binary.y ~ X2.1 + binary.X3, random=~animal, rcov = ~units, pedigree=tree,
                      data = data, family="categorical", nitt=12000, scale=F,
                      prior=Prior.phyl52), silent=T)#iterations!

  #print(summary(m32))
```

```
    if (class(m32) != "try-error"){
      #convergence diagnostics:
      diag2=geweke.diag(m32$Sol)
      results.c2.binary[t, corr, i, 3, 26] <- matrix(diag2$z[1],1,1)#save z-value from geweke diagnostic test
      results.c2.binary[t, corr, i, 3, 27] <- matrix(diag2$z[2],1,1)#save z-value from geweke diagnostic test
      results.c2.binary[t, corr, i, 3, 28] <- matrix(diag2$z[3],1,1)#save z-value from geweke diagnostic test
      results.c2.binary[t, corr, i, 3, 29] <- matrix(2*pnorm(-abs(diag2$z[1])),1,1)#save p-value from geweke diagnostic test
      results.c2.binary[t, corr, i, 3, 30] <- matrix(2*pnorm(-abs(diag2$z[2])),1,1)#save p-value from geweke diagnostic test
      results.c2.binary[t, corr, i, 3, 31] <- matrix(2*pnorm(-abs(diag2$z[3])),1,1)#save p-value from geweke diagnostic test

      results.c2.binary[t, corr, i, 3, c(1:15)] <- matrix(summary(m32)$solutions[,], 1, 15)
      results.c2.binary[t, corr, i, 3, c(16:24)] <- matrix(summary(m32$Sol)$statistics[,c(2:4)], 1, 9)
    }

    else {
      results.c2.binary[t, corr, i, 3, c(1:24)] <- matrix(99, 1, 24)
    }




    ##########################################
    #phylogenetic logistic regression by Lam Ho and Cecile Ane using the phylolm package:

    #1 continuous predictor:
    data$binaryylogreg=as.numeric(as.character(data[,7]));
    data$binaryX3logreg=as.numeric(as.character(data[,5]));

    plogreg1 <- try(phyloglm(binaryylogreg ~ X2.1, phy=tree, data=data, btol=30), silent=T)

    #print(summary(plogreg1))

    if (class(plogreg1) != "try-error"){
      #res_matrix_phylolm=as.matrix(summary(plogreg1)$coefficients); res_matrix_phylolm
      #res1_phylolm <- as.numeric(res_matrix_phylolm); #res1_phylolm
      #tvalue <- (res1_phylolm[c(1:2)]/res1_phylolm[c(3:4)])
      tvalue <- (summary(plogreg1)$coefficients[,1])/(summary(plogreg1)$coefficients[,2])

      p11 = 2 * pnorm(-abs(tvalue[1])); #p11
      p21 = 2 * pnorm(-abs(tvalue[2])); #p21

      results.c.binary[t, corr, i, 4, c(1:4)] <- matrix(summary(plogreg1)$coefficients, 1, 4)
      results.c.binary[t, corr, i, 4, c(5:6)] <- matrix(tvalue, 1, 2)
      results.c.binary[t, corr, i, 4, c(7:8)] <- matrix(c(p11,p21), 1, 2)
      results.c.binary[t, corr, i, 4, 18] <- matrix(summary(plogreg1)$alpha, 1, 1)
      results.c.binary[t, corr, i, 4, 27] <- matrix(plogreg1$convergeflag, 1, 1)
      res.binary[t,corr,i,4,c(1:species)] <- matrix(summary(plogreg1)$residuals, 1, species)
    }

    else {
      results.c.binary[t, corr, i, 4, c(1:27)] <- matrix(99, 1, 27)
      res.binary[t,corr,i,4,c(1:species)] <- matrix(99, 1, species)
    }



    #1 continuous and 1 binary predictor:
    plogreg2 <- try(phyloglm(binaryylogreg ~ X2.1 + binaryX3logreg, phy=tree, data=data, btol=30), silent=T)
    print(summary(plogreg2))

    if (class(plogreg2) != "try-error"){

      results.c2.binary[t, corr, i, 4, c(1:6)] <- matrix(summary(plogreg2)$coefficients, 1, 6)
      t2value <- (summary(plogreg2)$coefficients[,1])/(summary(plogreg2)$coefficients[,2])

      p12=2 * pnorm(-abs(t2value[1])); #p12
      p22=2 * pnorm(-abs(t2value[2])); #p22
      p32=2 * pnorm(-abs(t2value[3])); #p32

      results.c2.binary[t, corr, i, 4, c(7:9)] <- matrix(t2value, 1, 3)
      results.c2.binary[t, corr, i, 4, c(10:12)] <- matrix(c(p12,p22,p32), 1, 3)
      results.c2.binary[t, corr, i, 4, 26] <- matrix(summary(plogreg2)$alpha, 1, 1)
      results.c2.binary[t, corr, i, 4, 32] <- matrix(plogreg2$convergeflag, 1, 1)
      res2.binary[t,corr,i,4,c(1:species)] <- matrix(summary(plogreg2)$residuals, 1, species)
    }

    else {
      results.c2.binary[t, corr, i, 4, c(1:32)] <- matrix(99, 1, 32)
      res2.binary[t,corr,i,4,c(1:species)] <- matrix(99, 1, species)
    }



  }#end of n_sim loop

}#end of corr loop
```

```
  } #end of tree loop



  #results binary response
  results.coefficients.binary <- list(results.coefficients.binary, results.c.binary)
  results.coefficients2.binary <- list(results.coefficients2.binary, results.c2.binary)

  #residuals binary response
  results.res.partII.binary <- list(results.res.partII.binary, res.binary)
  results.res2.partII.binary <- list(results.res2.partII.binary, res2.binary)

}# end of species loop
```

C Results - Mean, Mean error and Rooted mean squared error

C.1 Continuous response

C.1.1 Non-phylogenetic GLM

Table C.1: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - Non-phylogenetic GLM (univariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.006 (0.398)	0.001	0.398	1000	-0.005 (0.345)	-0.010	0.345	1000	-0.001 (0.229)	-0.006	0.229	1000
Tree 2	-0.005 (0.244)	-0.010	0.244	1000	0.005 (0.149)	0.000	0.149	1000	0.007 (0.104)	0.002	0.104	1000
Tree 3	-0.036 (0.796)	-0.041	0.796	1000	0.027 (0.451)	0.022	0.452	1000	0.006 (0.615)	0.001	0.615	1000
Tree 4	-0.007 (0.418)	-0.012	0.418	1000	0.017 (0.299)	0.012	0.299	1000	0.008 (0.274)	0.003	0.274	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.503 (0.341)	0.002	0.341	1000	0.487 (0.297)	-0.014	0.297	1000	0.502 (0.216)	0.001	0.216	1000
Tree 2	0.500 (0.215)	-0.001	0.215	1000	0.496 (0.133)	-0.005	0.133	1000	0.502 (0.089)	0.001	0.089	1000
Tree 3	0.518 (0.708)	0.017	0.708	1000	0.479 (0.389)	-0.022	0.390	1000	0.527 (0.541)	0.026	0.541	1000
Tree 4	0.510 (0.324)	0.009	0.439	1000	0.501 (0.286)	0.000	0.286	1000	0.495 (0.225)	-0.006	0.225	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.899 (0.18)	0.000	0.18	1000	0.909 (0.150)	0.01	0.150	1000	0.902 (0.105)	0.003	0.105	1000
Tree 2	0.892 (0.109)	-0.007	0.109	1000	0.902 (0.064)	0.003	0.064	1000	0.901 (0.045)	0.002	0.045	1000
Tree 3	0.900 (0.371)	0.001	0.371	1000	0.890 (0.203)	-0.009	0.203	1000	0.895 (0.270)	-0.004	0.270	1000
Tree 4	0.902 (0.177)	0.003	0.231	1000	0.895 (0.137)	-0.004	0.137	1000	0.895 (0.119)	-0.004	0.119	1000

Table C.2: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - Non-phylogenetic GLM (multivariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.003 (0.46)	-0.007	0.459	1000	-0.008 (0.362)	-0.011	0.362	1000	-0.006 (0.254)	-0.01	0.254	1000
Tree 2	-0.017 (0.291)	-0.02	0.292	1000	0.003 (0.172)	-0.001	0.172	1000	0.007 (0.123)	0.004	0.123	1000
Tree 3	0.001 (0.778)	-0.003	0.778	1000	0.036 (0.517)	0.033	0.518	1000	0.002 (0.611)	-0.001	0.611	1000
Tree 4	-0.008 0.459)	-0.011	0.527	1000	0.013 (0.328)	0.010	0.328	1000	0.000 (0.295)	-0.003	0.295	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.393 (0.381)	-0.039	0.383	1000	0.382 (0.309)	-0.050	0.313	1000	0.403 (0.231)	-0.028	0.232	1000
Tree 2	0.400 (0.256)	-0.032	0.258	1000	0.399 (0.156)	-0.032	0.159	1000	0.402 (0.104)	-0.030	0.109	1000
Tree 3	0.436 (0.634)	0.005	0.634	1000	0.408 (0.444)	-0.023	0.444	1000	0.435 (0.522)	0.003	0.522	1000
Tree 4	0.406 (0.361)	-0.026	0.432	1000	0.409 (0.302)	-0.023	0.303	1000	0.398 (0.233)	-0.034	0.235	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.722 (0.142)	-0.051	0.150	1000	0.731 (0.110)	-0.042	0.117	1000	0.725 (0.082)	-0.047	0.095	1000
Tree 2	0.708 (0.100)	-0.065	0.119	1000	0.721 (0.059)	-0.052	0.079	1000	0.722 (0.041)	-0.051	0.066	1000
Tree 3	0.747 (0.206)	-0.026	0.208	1000	0.745 (0.195)	-0.028	0.197	1000	0.743 (0.187)	-0.030	0.189	1000
Tree 4	0.725 (0.132)	-0.048	0.158	1000	0.724 (0.103)	-0.049	0.114	1000	0.724 (0.092)	-0.049	0.104	1000

Table C.3: **Mean effect size (M) with standard deviation in brackets, mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - Non-phylogenetic GLM (multivariate model - binary predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.058 (1.118)	0.042	1.118	1000	0.013 (1.15)	-0.003	1.149	1000	0.031 (0.864)	0.015	0.864	1000
Tree 2	0.048 (0.710)	0.032	0.710	1000	0.022 (0.544)	0.006	0.544	1000	0.006 (0.518)	-0.01	0.518	1000
Tree 3	-0.040 (1.476)	-0.056	1.477	1000	-0.015 (0.457)	-0.031	0.458	1000	0.041 (1.031)	0.026	1.031	1000
Tree 4	0.010 (1.399)	-0.005	1.213	1000	0.045 (1.305)	0.029	1.305	1000	0.067 (1.203)	0.051	1.204	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.549 (0.928)	0.089	0.932	1000	0.721 (1.017)	0.261	1.049	1000	0.694 (0.726)	0.234	0.763	1000
Tree 2	0.480 (0.603)	0.020	0.603	1000	0.611 (0.482)	0.151	0.505	1000	0.836 (0.454)	0.376	0.589	1000
Tree 3	0.394 (1.298)	-0.066	1.299	1000	0.203 (0.394)	-0.257	0.470	1000	0.429 (0.939)	-0.031	0.939	1000
Tree 4	0.727 (1.229)	0.267	1.061	1000	0.793 (1.113)	0.333	1.161	1000	0.945 (1.034)	0.485	1.142	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.915 (0.337)	0.073	0.345	1000	1.198 (0.365)	0.357	0.51	1000	1.237 (0.256)	0.395	0.470	1000
Tree 2	0.881 (0.203)	0.039	0.207	1000	1.124 (0.166)	0.282	0.327	1000	1.512 (0.149)	0.670	0.687	1000
Tree 3	0.692 (0.552)	-0.150	0.572	1000	0.412 (0.162)	-0.430	0.459	1000	0.711 (0.390)	-0.130	0.411	1000
Tree 4	1.274 (0.438)	0.432	0.466	1000	1.521 (0.407)	0.679	0.791	1000	1.697 (0.377)	0.855	0.935	1000

C.1.2 PGLS

Table C.4: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - PGLS (univariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.005 (0.253)	0	0.253	993	0.014 (0.151)	0.009	0.151	912	0.005 (0.098)	0.000	0.098	906
Tree 2	-0.006 (0.244)	-0.011	0.244	996	0.005 (0.148)	0.000	0.148	996	0.007 (0.102)	0.002	0.102	1000
Tree 3	0.003 (0.283)	-0.002	0.283	934	0.005 (0.143)	0.000	0.143	915	0.012 (0.104)	0.007	0.104	957
Tree 4	0.000 (0.27)	-0.005	0.27	1000	0.010 (0.147)	0.005	0.147	997	0.008 (0.105)	0.003	0.105	996

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.502 (0.226)	0.001	0.226	996	0.5 (0.123)	-0.001	0.123	910	0.503 (0.088)	0.003	0.088	911
Tree 2	0.499 (0.213)	-0.002	0.213	992	0.495 (0.131)	-0.006	0.131	998	0.502 (0.086)	0.001	0.086	999
Tree 3	0.514 (0.235)	0.013	0.235	928	0.496 (0.125)	-0.005	0.125	905	0.500 (0.089)	-0.001	0.089	952
Tree 4	0.509 (0.225)	0.008	0.225	998	0.495 (0.133)	-0.006	0.133	1000	0.500 (0.091)	-0.001	0.091	996

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.9 (0.113)	0.001	0.113	978	0.9 (0.062)	0.001	0.062	915	0.9 (0.044)	0.001	0.044	901
Tree 2	0.892 (0.109)	-0.007	0.109	994	0.902 (0.063)	0.003	0.063	995	0.901 (0.044)	0.002	0.044	997
Tree 3	0.896 (0.119)	-0.003	0.119	944	0.9 (0.065)	0.001	0.065	916	0.901 (0.045)	0.001	0.045	953
Tree 4	0.904 (0.115)	0.004	0.115	999	0.898 (0.065)	-0.001	0.064	995	0.9 (0.046)	0.001	0.046	997

Table C.5: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - PGLS (multivariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.003 (0.284)	0.000	0.284	993	0.014 (0.163)	0.011	0.164	909	0.004 (0.110)	0.001	0.110	910
Tree 2	-0.017 (0.291)	-0.020	0.291	991	0.003 (0.169)	0.000	0.169	999	0.007 (0.120)	0.004	0.12	999
Tree 3	-0.003 (0.332)	-0.006	0.332	945	0.008 (0.157)	0.005	0.157	914	0.010 (0.107)	0.007	0.107	952
Tree 4	-0.004 (0.304)	-0.008	0.304	999	0.007 (0.163)	0.004	0.163	997	0.009 (0.115)	0.006	0.115	997

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.441 (0.245)	0.01	0.245	990	0.456 (0.134)	0.024	0.136	897	0.459 (0.095)	0.028	0.099	911
Tree 2	0.399 (0.254)	-0.032	0.256	993	0.399 (0.153)	-0.033	0.157	999	0.404 (0.102)	-0.028	0.106	999
Tree 3	0.477 (0.267)	0.045	0.271	942	0.460 (0.131)	0.028	0.134	909	0.477 (0.092)	0.046	0.103	956
Tree 4	0.431 (0.261)	-0.001	0.261	999	0.439 (0.150)	0.007	0.15	998	0.448 (0.101)	0.016	0.102	999

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.784 (0.126)	0.011	0.126	986	0.819 (0.07)	0.047	0.084	904	0.82 (0.048)	0.047	0.068	884
Tree 2	0.709 (0.1)	-0.064	0.119	997	0.722 (0.058)	-0.051	0.078	998	0.723 (0.041)	-0.05	0.065	999
Tree 3	0.819 (0.142)	0.047	0.149	958	0.831 (0.074)	0.058	0.094	844	0.861 (0.048)	0.088	0.100	921
Tree 4	0.762 (0.121)	-0.011	0.122	996	0.785 (0.073)	0.012	0.073	993	0.804 (0.051)	0.031	0.060	994

Table C.6: **Mean effect size (M) with standard deviation in brackets, mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - PGLS (multivariate model - binary predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.03 (0.714)	0.014	0.714	993	0.005 (0.478)	-0.011	0.477	909	0.005 (0.306)	-0.01	0.306	910
Tree 2	0.049 (0.705)	0.033	0.705	991	0.022 (0.538)	0.007	0.537	999	0.005 (0.508)	-0.011	0.508	999
Tree 3	-0.029 (1.137)	-0.045	1.138	945	-0.004 (0.142)	-0.02	0.143	914	0.012 (0.266)	-0.004	0.266	952
Tree 4	0.043 (0.931)	0.028	0.931	999	0.027 (0.604)	0.012	0.604	997	-0.002 (0.441)	-0.018	0.441	997

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.332 (0.597)	-0.128	0.61	990	0.326 (0.441)	-0.134	0.46	897	0.298 (0.270)	-0.162	0.315	911
Tree 2	0.478 (0.600)	0.018	0.6	993	0.608 (0.474)	0.148	0.496	999	0.827 (0.449)	0.367	0.58	999
Tree 3	0.238 (1.036)	-0.222	1.059	942	0.088 (0.133)	-0.372	0.396	909	0.122 (0.371)	-0.338	0.502	956
Tree 4	0.557 (0.892)	0.097	0.897	999	0.485 (0.534)	0.025	0.534	998	0.498 (0.359)	0.038	0.361	999

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.619 (0.324)	-0.223	0.393	986	0.605 (0.249)	-0.237	0.343	904	0.579 (0.157)	-0.263	0.306	884
Tree 2	0.88 (0.203)	0.038	0.207	997	1.118 (0.163)	0.276	0.321	998	1.502 (0.148)	0.66	0.677	999
Tree 3	0.515 (0.538)	-0.327	0.629	958	0.172 (0.091)	-0.67	0.676	844	0.198 (0.239)	-0.644	0.687	921
Tree 4	1.011 (0.361)	0.169	0.399	996	0.961 (0.226)	0.119	0.256	993	0.912 (0.165)	0.07	0.179	994

C.1.3 PGEE

Table C.7: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - PGEE (univariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.008 (0.241)	0.003	0.241	1000	0.012 (0.15)	0.007	0.15	1000	0.006 (0.098)	0.001	0.098	1000
Tree 2	-0.005 (0.243)	-0.01	0.243	1000	0.006 (0.147)	0.001	0.147	1000	0.006 (0.101)	0.001	0.101	1000
Tree 3	0.007 (0.249)	0.002	0.249	1000	0.008 (0.142)	0.003	0.142	1000	0.011 (0.103)	0.006	0.103	1000
Tree 4	-0.002 (0.269)	-0.006	0.269	1000	0.009 (0.151)	0.004	0.15	1000	0.007 (0.11)	0.002	0.11	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.499 (0.214)	-0.002	0.214	1000	0.5 (0.123)	-0.001	0.123	1000	0.502 (0.088)	0.001	0.088	1000
Tree 2	0.499 (0.213)	-0.002	0.213	1000	0.496 (0.13)	-0.005	0.13	1000	0.502 (0.086)	0.001	0.086	1000
Tree 3	0.51 (0.208)	0.009	0.208	1000	0.496 (0.123)	-0.005	0.123	1000	0.500 (0.09)	-0.001	0.09	1000
Tree 4	0.511 (0.23)	0.01	0.230	1000	0.494 (0.135)	-0.006	0.135	1000	0.498 (0.096)	-0.003	0.096	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.901 (0.106)	0.002	0.106	1000	0.900 (0.062)	0.001	0.062	1000	0.901 (0.044)	0.001	0.044	1000
Tree 2	0.892 (0.109)	-0.007	0.109	1000	0.902 (0.062)	0.003	0.062	1000	0.901 (0.044)	0.001	0.044	1000
Tree 3	0.899 (0.105)	-0.001	0.105	1000	0.901 (0.065)	0.002	0.065	1000	0.901 (0.045)	0.002	0.045	1000
Tree 4	0.905 (0.114)	0.006	0.114	1000	0.898 (0.066)	-0.001	0.066	1000	0.901 (0.048)	0.002	0.048	1000

Table C.8: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - PGEE (multivariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.008 (0.271)	0.005	0.271	1000	0.012 (0.162)	0.009	0.162	1000	0.005 (0.109)	0.001	0.109	1000
Tree 2	-0.017 (0.29)	-0.02	0.291	1000	0.004 (0.169)	0.001	0.169	1000	0.006 (0.12)	0.003	0.12	1000
Tree 3	0.007 (0.271)	0.003	0.270	1000	0.009 (0.155)	0.006	0.155	1000	0.011 (0.107)	0.008	0.108	1000
Tree 4	0.001 (0.305)	-0.003	0.305	1000	0.007 (0.165)	0.004	0.165	1000	0.008 (0.122)	0.005	0.122	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.444 (0.23)	0.012	0.231	1000	0.458 (0.133)	0.027	0.136	1000	0.462 (0.095)	0.03	0.099	1000
Tree 2	0.399 (0.254)	-0.033	0.256	1000	0.399 (0.152)	-0.032	0.156	1000	0.404 (0.102)	-0.028	0.106	1000
Tree 3	0.485 (0.224)	0.053	0.230	1000	0.462 (0.13)	0.030	0.133	1000	0.478 (0.093)	0.046	0.104	1000
Tree 4	0.44 (0.257)	0.008	0.257	1000	0.441 (0.152)	0.009	0.152	1000	0.449 (0.106)	0.017	0.107	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.811 (0.112)	0.038	0.119	1000	0.825 (0.066)	0.052	0.084	1000	0.826 (0.047)	0.053	0.071	1000
Tree 2	0.709 (0.100)	-0.064	0.119	1000	0.722 (0.058)	-0.05	0.077	1000	0.724 (0.041)	-0.049	0.064	1000
Tree 3	0.847 (0.113)	0.074	0.135	1000	0.839 (0.068)	0.066	0.095	1000	0.862 (0.048)	0.089	0.101	1000
Tree 4	0.782 (0.119)	0.009	0.119	1000	0.798 (0.07)	0.025	0.075	1000	0.814 (0.052)	0.041	0.066	1000

Table C.9: **Mean effect size (M) with standard deviation in brackets, mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - PGEE (multivariate model - binary predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.022 (0.688)	0.007	0.688	1000	0.007 (0.473)	-0.009	0.473	1000	0.008 (0.303)	-0.008	0.303	1000
Tree 2	0.049 (0.702)	0.033	0.702	1000	0.022 (0.536)	0.007	0.536	1000	0.006 (0.507)	-0.01	0.507	1000
Tree 3	-0.024 (1.084)	-0.039	1.085	1000	-0.002 (0.142)	-0.018	0.143	1000	0.01 (0.279)	-0.006	0.279	1000
Tree 4	0.008 (0.917)	-0.008	0.916	1000	0.021 (0.608)	0.005	0.608	1000	0.000 (0.467)	-0.016	0.467	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.306 (0.581)	-0.154	0.601	1000	0.32 (0.435)	-0.14	0.456	1000	0.296 (0.271)	-0.164	0.317	1000
Tree 2	0.479 (0.598)	0.019	0.598	1000	0.609 (0.474)	0.149	0.496	1000	0.824 (0.448)	0.364	0.578	1000
Tree 3	0.239 (1.019)	-0.221	1.042	1000	0.087 (0.13)	-0.373	0.395	1000	0.125 (0.377)	-0.335	0.504	1000
Tree 4	0.516 (0.844)	0.056	0.846	1000	0.475 (0.538)	0.015	0.537	1000	0.485 (0.37)	0.025	0.371	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.526 (0.305)	-0.316	0.439	1000	0.58 (0.245)	-0.262	0.359	1000	0.549 (0.175)	-0.292	0.341	1000
Tree 2	0.877 (0.203)	0.036	0.206	1000	1.114 (0.164)	0.272	0.317	1000	1.495 (0.147)	0.653	0.669	1000
Tree 3	0.467 (0.55)	-0.375	0.666	1000	0.164 (0.088)	-0.678	0.684	1000	0.196 (0.231)	-0.646	0.686	1000
Tree 4	0.917 (0.346)	0.076	0.354	1000	0.893 (0.213)	0.051	0.219	1000	0.868 (0.165)	0.026	0.167	1000

C.1.4 PGLMM

Table C.10: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - PGLMM (univariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.006 (0.247)	0.001	0.247	896	0.012 (0.151)	0.007	0.151	820	0.005 (0.101)	0.000	0.101	773
Tree 2	-0.005 (0.243)	-0.01	0.243	935	0.005 (0.148)	0.000	0.148	879	0.006 (0.103)	0.001	0.103	825
Tree 3	0.009 (0.260)	0.004	0.260	898	0.01 (0.145)	0.005	0.145	862	0.011 (0.105)	0.006	0.105	751
Tree 4	0.000 (0.264)	-0.004	0.263	860	0.011 (0.149)	0.006	0.149	826	0.009 (0.104)	0.004	0.104	792

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.504 (0.216)	0.003	0.216	904	0.498 (0.125)	-0.003	0.125	819	0.501 (0.089)	0.000	0.089	801
Tree 2	0.499 (0.215)	-0.002	0.215	937	0.495 (0.132)	-0.006	0.132	903	0.502 (0.085)	0.001	0.085	841
Tree 3	0.511 (0.215)	0.011	0.215	870	0.49 (0.125)	-0.011	0.126	864	0.5 (0.093)	-0.001	0.093	791
Tree 4	0.513 (0.213)	0.012	0.213	858	0.495 (0.134)	-0.006	0.134	857	0.497 (0.089)	-0.004	0.089	802

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.899 (0.112)	0.000	0.112	914	0.9 (0.063)	0.001	0.063	876	0.9 (0.046)	0.001	0.046	815
Tree 2	0.892 (0.109)	-0.007	0.109	958	0.902 (0.062)	0.003	0.062	911	0.9 (0.044)	0.001	0.044	893
Tree 3	0.896 (0.112)	-0.003	0.112	905	0.896 (0.068)	-0.003	0.068	783	0.898 (0.047)	-0.002	0.047	840
Tree 4	0.902 (0.11)	0.003	0.110	900	0.897 (0.063)	-0.002	0.063	873	0.901 (0.045)	0.002	0.045	828

Table C.11: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - PGLMM (multivariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.005 (0.273)	0.002	0.273	879	0.011 (0.165)	0.008	0.165	800	0.000 (0.110)	-0.003	0.110	788
Tree 2	-0.020 (0.289)	-0.024	0.29	930	0.002 (0.168)	-0.001	0.167	910	0.005 (0.118)	0.002	0.118	880
Tree 3	0.004 (0.287)	0.001	0.287	883	0.011 (0.156)	0.008	0.156	828	0.010 (0.111)	0.007	0.111	771
Tree 4	-0.003 (0.294)	-0.006	0.294	857	0.003 (0.162)	-0.001	0.162	834	0.010 (0.117)	0.007	0.117	791

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.436 (0.237)	0.005	0.237	881	0.457 (0.134)	0.026	0.137	805	0.461 (0.096)	0.03	0.1	794
Tree 2	0.397 (0.256)	-0.034	0.258	932	0.397 (0.155)	-0.035	0.159	907	0.404 (0.101)	-0.027	0.105	867
Tree 3	0.481 (0.233)	0.05	0.239	898	0.454 (0.132)	0.022	0.134	843	0.474 (0.096)	0.042	0.105	740
Tree 4	0.43 (0.247)	-0.002	0.247	872	0.438 (0.15)	0.007	0.15	853	0.447 (0.1)	0.016	0.101	781

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.788 (0.116)	0.015	0.117	909	0.815 (0.068)	0.042	0.080	865	0.816 (0.048)	0.043	0.065	837
Tree 2	0.708 (0.099)	-0.065	0.118	946	0.722 (0.058)	-0.051	0.077	927	0.723 (0.041)	-0.050	0.065	900
Tree 3	0.829 (0.124)	0.056	0.136	903	0.822 (0.071)	0.049	0.086	705	0.853 (0.049)	0.080	0.094	832
Tree 4	0.765 (0.113)	-0.008	0.113	900	0.787 (0.07)	0.014	0.071	895	0.805 (0.05)	0.032	0.060	841

Table C.12: **Mean effect size (M) with standard deviation in brackets, mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from continuous response models - PGLMM (multivariate model - binary predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.019 (0.701)	0.003	0.701	884	0.008 (0.497)	-0.008	0.497	831	0.009 (0.309)	-0.006	0.309	813
Tree 2	0.057 (0.7)	0.042	0.701	931	0.026 (0.536)	0.01	0.536	904	0.003 (0.512)	-0.013	0.512	888
Tree 3	-0.016 (1.075)	-0.032	1.075	839	-0.002 (0.142)	-0.018	0.143	894	0.009 (0.184)	-0.007	0.184	810
Tree 4	0.042 (0.896)	0.027	0.896	858	0.016 (0.597)	0.001	0.597	829	-0.011 (0.446)	-0.027	0.447	795

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.325 (0.572)	-0.135	0.588	901	0.327 (0.438)	-0.133	0.458	831	0.298 (0.272)	-0.162	0.316	822
Tree 2	0.475 (0.594)	0.015	0.593	934	0.618 (0.473)	0.158	0.498	901	0.823 (0.452)	0.363	0.58	876
Tree 3	0.272 (0.962)	-0.188	0.979	832	0.088 (0.129)	-0.372	0.394	889	0.115 (0.337)	-0.345	0.482	809
Tree 4	0.556 (0.832)	0.096	0.837	849	0.471 (0.531)	0.011	0.531	846	0.504 (0.367)	0.044	0.369	813

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.607 (0.293)	-0.235	0.376	908	0.621 (0.23)	-0.221	0.318	874	0.6 (0.142)	-0.242	0.28	863
Tree 2	0.875 (0.201)	0.034	0.204	938	1.116 (0.165)	0.274	0.32	935	1.497 (0.146)	0.655	0.671	904
Tree 3	0.472 (0.525)	-0.37	0.642	853	0.186 (0.077)	-0.656	0.661	784	0.191 (0.161)	-0.651	0.671	834
Tree 4	0.993 (0.328)	0.151	0.361	906	0.956 (0.217)	0.114	0.245	881	0.914 (0.165)	0.072	0.18	859

C.2 Ordinal response

C.2.1 Non-phylogenetic GLM

Table C.13: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - Non-phylogenetic GLM (univariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.008 (0.275)	-0.012	0.275	1000	0.005 (0.194)	0.002	0.194	1000	0.007 (0.110)	0.004	0.110	1000
Tree 2	0.005 (0.174)	0.002	0.174	1000	0.006 (0.079)	0.003	0.079	1000	0.003 (0.045)	0.000	0.045	1000
Tree 3	-0.014 (0.696)	-0.017	0.696	1000	0.010 (0.364)	0.007	0.364	1000	-0.009 (0.482)	-0.012	0.482	1000
Tree 4	0.004 (0.196)	0.000	0.195	1000	0.004 (0.111)	0.000	0.111	1000	0.008 (0.087)	0.005	0.087	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.38 (0.223)	0.053	0.229	1000	0.265 (0.146)	-0.062	0.159	1000	0.236 (0.084)	-0.09	0.123	1000
Tree 2	0.337 (0.141)	0.01	0.141	1000	0.242 (0.062)	-0.085	0.105	1000	0.204 (0.038)	-0.123	0.129	1000
Tree 3	0.649 (0.622)	0.322	0.700	1000	0.518 (0.306)	0.191	0.36	1000	0.394 (0.433)	0.067	0.438	1000
Tree 4	0.264 (0.151)	-0.062	0.163	1000	0.191 (0.094)	-0.136	0.165	1000	0.163 (0.066)	-0.164	0.177	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.696 (0.156)	0.085	0.178	1000	0.505 (0.101)	-0.107	0.147	1000	0.44 (0.069)	-0.171	0.184	1000
Tree 2	0.622 (0.117)	0.011	0.117	1000	0.442 (0.063)	-0.169	0.181	1000	0.37 (0.046)	-0.241	0.245	1000
Tree 3	1.345 (0.542)	0.734	0.912	1000	0.982 (0.297)	0.371	0.475	1000	0.816 (0.319)	0.205	0.379	1000
Tree 4	0.498 (0.114)	-0.113	0.160	1000	0.371 (0.069)	-0.240	0.249	1000	0.3 (0.051)	-0.311	0.315	1000

Table C.14: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - Non-phylogenetic GLM (multivariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.01 (0.312)	-0.013	0.312	1000	0.003 (0.207)	0.000	0.207	1000	0.01 (0.12)	0.007	0.121	1000
Tree 2	0.007 (0.208)	0.005	0.208	1000	0.005 (0.09)	0.002	0.09	1000	0.003 (0.053)	0.001	0.053	1000
Tree 3	0 (0.774)	-0.003	0.774	1000	0.008 (0.394)	0.005	0.393	1000	-0.01 (0.503)	-0.013	0.503	1000
Tree 4	0.003 (0.227)	0.000	0.227	1000	0.001 (0.124)	-0.002	0.124	1000	0.006 (0.093)	0.004	0.093	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.3 (0.258)	0.025	0.259	1000	0.214 (0.17)	-0.061	0.18	1000	0.185 (0.097)	-0.09	0.133	1000
Tree 2	0.267 (0.168)	-0.008	0.168	1000	0.19 (0.073)	-0.085	0.112	1000	0.162 (0.044)	-0.113	0.121	1000
Tree 3	0.521 (0.687)	0.246	0.729	1000	0.419 (0.329)	0.143	0.359	1000	0.328 (0.458)	0.052	0.461	1000
Tree 4	0.211 (0.178)	-0.064	0.189	1000	0.15 (0.106)	-0.125	0.164	1000	0.131 (0.074)	-0.144	0.162	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.54 (0.132)	0.039	0.138	1000	0.398 (0.078)	-0.103	0.129	1000	0.348 (0.053)	-0.153	0.162	1000
Tree 2	0.482 (0.1)	-0.019	0.101	1000	0.345 (0.05)	-0.156	0.164	1000	0.291 (0.037)	-0.21	0.213	1000
Tree 3	1.067 (0.424)	0.565	0.706	1000	0.78 (0.22)	0.279	0.355	1000	0.671 (0.25)	0.17	0.303	1000
Tree 4	0.388 (0.092)	-0.113	0.146	1000	0.291 (0.053)	-0.21	0.217	1000	0.239 (0.039)	-0.262	0.265	1000

Table C.15: **Mean effect size (M) with standard deviation in brackets, mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - Non-phylogenetic GLM (multivariate model - binary predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.023 (0.771)	0.018	0.771	1000	0.009 (0.587)	0.004	0.587	1000	-0.015 (0.371)	-0.02	0.371	1000
Tree 2	-0.006 (0.564)	-0.011	0.563	1000	0.009 (0.294)	0.004	0.293	1000	0.001 (0.192)	-0.004	0.192	1000
Tree 3	-0.018 (0.777)	-0.023	0.777	1000	-0.004 (0.55)	-0.009	0.55	1000	0.009 (0.497)	0.005	0.497	1000
Tree 4	0.012 (0.72)	0.007	0.720	1000	0.02 (0.485)	0.015	0.485	1000	0.014 (0.398)	0.009	0.398	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.379 (0.629)	0.112	0.638	1000	0.332 (0.535)	0.065	0.539	1000	0.339 (0.31)	0.072	0.318	1000
Tree 2	0.366 (0.468)	0.099	0.478	1000	0.33 (0.273)	0.063	0.28	1000	0.295 (0.168)	0.028	0.171	1000
Tree 3	0.339 (0.7)	0.072	0.703	1000	0.281 (0.478)	0.014	0.478	1000	0.274 (0.507)	0.007	0.507	1000
Tree 4	0.388 (0.616)	0.12	0.628	1000	0.353 (0.41)	0.086	0.419	1000	0.312 (0.359)	0.045	0.362	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.762 (0.323)	0.209	0.385	1000	0.681 (0.244)	0.128	0.275	1000	0.592 (0.149)	0.039	0.154	1000
Tree 2	0.735 (0.253)	0.182	0.312	1000	0.624 (0.158)	0.071	0.173	1000	0.568 (0.113)	0.015	0.114	1000
Tree 3	0.669 (0.448)	0.116	0.462	1000	0.59 (0.286)	0.037	0.288	1000	0.566 (0.361)	0.013	0.361	1000
Tree 4	0.771 (0.296)	0.218	0.367	1000	0.665 (0.2)	0.112	0.229	1000	0.597 (0.171)	0.044	0.176	1000

C.2.2 PGLS

Table C.16: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - PGLS (univariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.01 (0.206)	-0.013	0.206	988	0.002 (0.099)	-0.001	0.098	995	0.005 (0.061)	0.002	0.061	990
Tree 2	0.002 (0.175)	-0.001	0.175	903	0.007 (0.079)	0.004	0.08	972	0.003 (0.044)	0.000	0.044	989
Tree 3	0.015 (0.485)	0.012	0.485	971	0.009 (0.217)	0.006	0.217	961	0.006 (0.136)	0.003	0.136	864
Tree 4	0.004 (0.147)	0.001	0.147	999	0.003 (0.067)	0.000	0.067	997	0.003 (0.04)	0.000	0.040	999

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.367 (0.179)	0.04	0.183	990	0.263 (0.089)	-0.064	0.109	994	0.231 (0.059)	-0.095	0.112	987
Tree 2	0.338 (0.140)	0.011	0.141	913	0.242 (0.062)	-0.085	0.105	971	0.204 (0.038)	-0.123	0.129	985
Tree 3	0.687 (0.498)	0.36	0.615	974	0.506 (0.246)	0.179	0.304	955	0.421 (0.249)	0.094	0.266	890
Tree 4	0.264 (0.125)	-0.063	0.140	997	0.203 (0.063)	-0.124	0.139	988	0.164 (0.041)	-0.163	0.168	996

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.692 (0.167)	0.081	0.186	994	0.502 (0.113)	-0.109	0.157	987	0.433 (0.075)	-0.178	0.193	986
Tree 2	0.622 (0.117)	0.011	0.117	909	0.441 (0.063)	-0.17	0.181	978	0.37 (0.047)	-0.241	0.245	986
Tree 3	1.354 (0.669)	0.742	0.999	961	0.932 (0.346)	0.32	0.471	956	0.768 (0.403)	0.157	0.432	933
Tree 4	0.498 (0.122)	-0.113	0.166	989	0.369 (0.072)	-0.242	0.253	994	0.303 (0.055)	-0.309	0.313	996

Table C.17: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - PGLS (multivariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.008 (0.243)	-0.011	0.244	992	0.002 (0.109)	-0.001	0.109	992	0.005 (0.066)	0.002	0.066	987
Tree 2	0.005 (0.208)	0.003	0.208	884	0.005 (0.09)	0.003	0.09	972	0.002 (0.052)	0.000	0.052	992
Tree 3	0.003 (0.543)	0.000	0.542	977	0.001 (0.242)	-0.002	0.242	952	0.006 (0.147)	0.003	0.147	877
Tree 4	0.002 (0.178)	-0.001	0.177	990	0.004 (0.076)	0.001	0.076	996	0.003 (0.044)	0.000	0.044	997

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.299 (0.209)	0.024	0.211	990	0.234 (0.099)	-0.041	0.107	994	0.206 (0.062)	-0.069	0.093	993
Tree 2	0.268 (0.167)	-0.007	0.167	897	0.189 (0.072)	-0.086	0.112	972	0.163 (0.043)	-0.113	0.121	985
Tree 3	0.568 (0.544)	0.293	0.618	974	0.459 (0.249)	0.183	0.309	965	0.387 (0.246)	0.112	0.27	881
Tree 4	0.218 (0.152)	-0.057	0.162	990	0.174 (0.069)	-0.101	0.122	994	0.146 (0.041)	-0.129	0.135	992

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.548 (0.14)	0.047	0.148	995	0.414 (0.089)	-0.087	0.125	992	0.365 (0.058)	-0.136	0.148	999
Tree 2	0.484 (0.1)	-0.017	0.101	919	0.345 (0.05)	-0.156	0.164	977	0.291 (0.036)	-0.21	0.213	991
Tree 3	1.062 (0.516)	0.56	0.762	987	0.789 (0.285)	0.288	0.405	966	0.667 (0.351)	0.165	0.388	960
Tree 4	0.397 (0.103)	-0.104	0.147	990	0.301 (0.057)	-0.200	0.208	996	0.255 (0.045)	-0.246	0.25	998

Table C.18: **Mean effect size (M) with standard deviation in brackets, mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - PGLS (multivariate model - binary predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.01 (0.608)	0.005	0.607	992	0.004 (0.298)	-0.001	0.298	992	0.000 (0.187)	-0.005	0.187	987
Tree 2	-0.004 (0.565)	-0.009	0.565	884	0.011 (0.291)	0.006	0.291	972	0.001 (0.190)	-0.004	0.19	992
Tree 3	0.018 (0.576)	0.013	0.576	977	0.009 (0.296)	0.005	0.296	952	0.002 (0.166)	-0.003	0.166	877
Tree 4	0.010 (0.602)	0.005	0.601	990	-0.005 (0.284)	-0.01	0.284	996	0.006 (0.18)	0.001	0.18	997

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.337 (0.501)	0.070	0.506	990	0.198 (0.285)	-0.07	0.294	994	0.173 (0.179)	-0.094	0.202	993
Tree 2	0.362 (0.464)	0.095	0.474	897	0.334 (0.270)	0.067	0.278	972	0.294 (0.166)	0.027	0.168	985
Tree 3	0.268 (0.545)	0	0.544	974	0.147 (0.255)	-0.12	0.282	965	0.104 (0.179)	-0.163	0.242	881
Tree 4	0.315 (0.516)	0.047	0.518	990	0.236 (0.255)	-0.031	0.256	994	0.174 (0.166)	-0.094	0.191	992

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.711 (0.316)	0.158	0.353	995	0.543 (0.224)	-0.009	0.224	992	0.459 (0.14)	-0.094	0.169	999
Tree 2	0.734 (0.252)	0.181	0.311	919	0.623 (0.157)	0.07	0.172	977	0.565 (0.114)	0.012	0.114	991
Tree 3	0.596 (0.414)	0.043	0.416	987	0.430 (0.274)	-0.123	0.300	966	0.313 (0.228)	-0.24	0.331	960
Tree 4	0.687 (0.301)	0.134	0.329	990	0.561 (0.185)	0.008	0.185	996	0.439 (0.137)	-0.114	0.178	998

C.2.3 PGEE

Table C.19: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - PGEE (univariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.007 (0.201)	-0.01	0.201	1000	0.004 (0.105)	0.001	0.105	1000	0.006 (0.067)	0.003	0.067	1000
Tree 2	0.005 (0.173)	0.002	0.173	1000	0.006 (0.079)	0.003	0.079	1000	0.003 (0.044)	-0.001	0.044	1000
Tree 3	0.017 (0.467)	0.014	0.467	1000	0.002 (0.255)	-0.001	0.255	1000	0 (0.171)	-0.003	0.171	1000
Tree 4	0.004 (0.142)	0.001	0.142	1000	0.002 (0.066)	-0.001	0.066	1000	0.003 (0.043)	0.000	0.042	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.375 (0.182)	0.048	0.188	1000	0.269 (0.1)	-0.058	0.115	1000	0.235 (0.068)	-0.092	0.115	1000
Tree 2	0.337 (0.14)	0.011	0.141	1000	0.242 (0.062)	-0.085	0.105	1000	0.204 (0.038)	-0.123	0.129	1000
Tree 3	0.735 (0.536)	0.408	0.674	1000	0.559 (0.314)	0.232	0.39	1000	0.466 (0.292)	0.139	0.323	1000
Tree 4	0.265 (0.129)	-0.061	0.142	1000	0.202 (0.064)	-0.125	0.14	1000	0.162 (0.043)	-0.164	0.17	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.69 (0.189)	0.079	0.205	1000	0.504 (0.131)	-0.107	0.169	1000	0.432 (0.089)	-0.179	0.200	1000
Tree 2	0.622 (0.117)	0.011	0.117	1000	0.442 (0.063)	-0.169	0.181	1000	0.371 (0.046)	-0.241	0.245	1000
Tree 3	1.369 (0.730)	0.757	1.052	1000	0.971 (0.408)	0.36	0.543	1000	0.812 (0.422)	0.201	0.468	1000
Tree 4	0.494 (0.130)	-0.117	0.174	1000	0.364 (0.080)	-0.247	0.26	1000	0.299 (0.058)	-0.312	0.317	1000

Table C.20: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - PGEE (multivariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.004 (0.233)	-0.007	0.233	1000	0.002 (0.117)	0.000	0.117	1000	0.006 (0.073)	0.004	0.074	1000
Tree 2	0.007 (0.208)	0.004	0.208	1000	0.005 (0.09)	0.002	0.09	1000	0.002 (0.052)	0.000	0.052	1000
Tree 3	0.022 (0.532)	0.019	0.532	1000	-0.005 (0.285)	-0.007	0.285	1000	-0.001 (0.188)	-0.004	0.188	1000
Tree 4	0.002 (0.171)	0.000	0.171	1000	0.003 (0.075)	0.000	0.075	1000	0.002 (0.047)	-0.001	0.047	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.313 (0.207)	0.037	0.211	1000	0.245 (0.112)	-0.03	0.116	1000	0.213 (0.074)	-0.062	0.097	1000
Tree 2	0.267 (0.166)	-0.008	0.166	1000	0.19 (0.073)	-0.085	0.111	1000	0.163 (0.043)	-0.112	0.12	1000
Tree 3	0.638 (0.569)	0.362	0.674	1000	0.519 (0.322)	0.244	0.404	1000	0.439 (0.294)	0.164	0.336	1000
Tree 4	0.225 (0.153)	-0.05	0.161	1000	0.176 (0.069)	-0.099	0.12	1000	0.146 (0.045)	-0.129	0.137	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.573 (0.168)	0.072	0.182	1000	0.448 (0.123)	-0.053	0.134	1000	0.39 (0.085)	-0.111	0.139	1000
Tree 2	0.482 (0.1)	-0.019	0.101	1000	0.346 (0.051)	-0.155	0.164	1000	0.292 (0.037)	-0.209	0.213	1000
Tree 3	1.151 (0.667)	0.649	0.931	1000	0.889 (0.405)	0.387	0.560	1000	0.740 (0.412)	0.239	0.476	1000
Tree 4	0.409 (0.116)	-0.092	0.148	1000	0.314 (0.069)	-0.187	0.200	1000	0.265 (0.051)	-0.236	0.242	1000

Table C.21: **Mean effect size (M) with standard deviation in brackets, mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - PGEE (multivariate model - binary predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.004 (0.585)	-0.001	0.585	1000	0.009 (0.302)	0.004	0.302	1000	0.002 (0.194)	-0.003	0.194	1000
Tree 2	-0.006 (0.564)	-0.011	0.564	1000	0.01 (0.29)	0.005	0.29	1000	0.002 (0.189)	-0.003	0.189	1000
Tree 3	0.002 (0.557)	-0.003	0.556	1000	0.02 (0.297)	0.015	0.297	1000	0.003 (0.203)	-0.002	0.203	1000
Tree 4	0.01 (0.572)	0.006	0.571	1000	-0.004 (0.284)	-0.009	0.284	1000	0.008 (0.191)	0.003	0.191	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.305 (0.48)	0.037	0.481	1000	0.164 (0.294)	-0.103	0.311	1000	0.146 (0.19)	-0.121	0.225	1000
Tree 2	0.364 (0.464)	0.097	0.474	1000	0.329 (0.271)	0.061	0.277	1000	0.292 (0.165)	0.025	0.166	1000
Tree 3	0.225 (0.492)	-0.043	0.494	1000	0.115 (0.276)	-0.152	0.315	1000	0.079 (0.187)	-0.188	0.265	1000
Tree 4	0.285 (0.493)	0.018	0.494	1000	0.215 (0.25)	-0.053	0.256	1000	0.161 (0.173)	-0.106	0.203	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.599 (0.323)	0.046	0.326	1000	0.365 (0.249)	-0.188	0.312	1000	0.281 (0.182)	-0.272	0.328	1000
Tree 2	0.733 (0.254)	0.18	0.311	1000	0.619 (0.158)	0.066	0.171	1000	0.562 (0.113)	0.009	0.113	1000
Tree 3	0.443 (0.415)	-0.11	0.429	1000	0.257 (0.303)	-0.295	0.423	1000	0.21 (0.269)	-0.342	0.436	1000
Tree 4	0.599 (0.295)	0.046	0.298	1000	0.432 (0.185)	-0.121	0.221	1000	0.329 (0.134)	-0.224	0.261	1000

C.2.4 PGLMM

Table C.22: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - PGLMM (univariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.012 (3.216)	-0.049	3.213	488	0.015 (1.444)	-0.023	1.442	501	0 (0.819)	-0.037	0.819	495
Tree 2	0.36 (2.226)	0.322	2.246	453	0.096 (0.814)	0.058	0.815	427	0.045 (0.414)	0.007	0.414	451
Tree 3	-0.178 (9.457)	-0.216	9.449	431	0.104 (3.59)	0.066	3.587	428	0.057 (3.191)	0.019	3.187	413
Tree 4	-0.144 (2.244)	-0.181	2.248	433	0.041 (0.93)	0.003	0.929	504	0.068 (0.579)	0.03	0.58	490

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	4.482 (3.316)	0.855	3.421	475	3.069 (1.687)	-0.557	1.775	484	2.559 (1.029)	-1.068	1.483	488
Tree 2	3.874 (2.620)	0.248	2.628	444	2.293 (1.053)	-1.334	1.699	418	1.72 (0.59)	-1.907	1.996	433
Tree 3	8.879 (9.012)	5.253	10.421	379	6.464 (4.173)	2.837	5.041	377	4.905 (3.588)	1.278	3.804	380
Tree 4	3.183 (2.336)	-0.443	2.375	468	2.179 (1.000)	-1.448	1.759	526	1.806 (0.657)	-1.82	1.935	505

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	10.042 (3.671)	1.653	4.022	433	6.821 (1.967)	-1.568	2.514	442	6.055 (1.296)	-2.334	2.669	444
Tree 2	9.006 (3.115)	0.617	3.172	414	6.016 (1.473)	-2.373	2.793	454	4.962 (0.931)	-3.427	3.551	444
Tree 3	19.669 (10.670)	11.28	15.517	367	14.014 (5.178)	5.625	7.64	345	10.841 (4.886)	2.452	5.46	333
Tree 4	7.408 (2.646)	-0.981	2.819	451	5.216 (1.455)	-3.173	3.491	479	4.314 (0.991)	-4.075	4.194	387

Table C.23: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - PGLMM (multivariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.074 (4.433)	-0.116	4.43	472	0.137 (1.947)	0.095	1.947	449	0.042 (0.94)	0.000	0.939	440
Tree 2	-0.062 (3.464)	-0.105	3.462	436	0.062 (1.096)	0.020	1.094	411	-0.001 (0.558)	-0.044	0.559	488
Tree 3	-0.096 (10.938)	-0.139	10.926	433	0.394 (4.372)	0.352	4.381	438	-0.017 (3.844)	-0.06	3.839	400
Tree 4	0.066 (3.153)	0.024	3.149	431	0.023 (1.213)	-0.019	1.211	502	0.044 (0.66)	0.002	0.66	458

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	4.043 (4.375)	0.763	4.436	412	2.743 (1.964)	-0.538	2.034	427	2.195 (1.103)	-1.085	1.546	428
Tree 2	3.5 (3.494)	0.220	3.495	344	1.954 (1.200)	-1.326	1.787	384	1.484 (0.662)	-1.796	1.914	429
Tree 3	7.866 (10.96)	4.586	11.868	392	5.437 (4.758)	2.157	5.218	366	4.391 (4.039)	1.111	4.184	375
Tree 4	2.874 (2.939)	-0.406	2.964	409	2.028 (1.166)	-1.252	1.71	452	1.643 (0.721)	-1.637	1.788	420

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	9.333 (3.834)	1.016	3.96	310	6.486 (1.794)	-1.831	2.561	310	5.449 (1.243)	-2.869	3.125	228
Tree 2	8.641 (3.215)	0.324	3.226	262	5.731 (1.48)	-2.586	2.978	262	4.491 (0.961)	-3.827	3.945	185
Tree 3	18.268 (10.652)	9.950	14.564	317	12.899 (4.882)	4.582	6.689	315	9.800 (4.617)	1.483	4.841	267
Tree 4	7.002 (2.902)	-1.315	3.183	366	4.926 (1.315)	-3.392	3.637	312	3.817 (0.879)	-4.500	4.585	277

Table C.24: **Mean effect size (M) with standard deviation in brackets, mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from ordinal response models - PGLMM (multivariate model - binary predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.476 (10.723)	-0.443	10.72	472	-0.255 (5.202)	-0.223	5.201	449	-0.162 (2.885)	-0.13	2.885	440
Tree 2	0.148 (8.804)	0.18	8.796	436	0.179 (3.446)	0.212	3.449	411	0.042 (1.975)	0.075	1.975	488
Tree 3	0.644 (11.545)	0.676	11.551	433	0.053 (5.775)	0.085	5.769	438	0.065 (3.84)	0.098	3.836	400
Tree 4	-0.337 (10.656)	-0.304	10.648	431	-0.207 (4.899)	-0.175	4.897	502	-0.002 (3.017)	0.031	3.014	458

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	3.171 (10.142)	0.269	10.133	412	2.429 (5.339)	-0.473	5.354	427	2.882 (2.936)	-0.02	2.932	428
Tree 2	2.971 (8.031)	0.069	8.020	344	3.172 (3.916)	0.27	3.92	384	2.346 (2.308)	-0.556	2.372	429
Tree 3	3.681 (10.533)	0.779	10.549	392	2.635 (5.205)	-0.267	5.205	366	1.481 (5.097)	-1.421	5.285	375
Tree 4	4.907 (8.78)	2.005	8.995	409	2.725 (4.122)	-0.177	4.121	452	2.417 (2.929)	-0.485	2.965	420

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	7.604 (8.854)	1.766	9.015	310	5.306 (5.337)	-0.531	5.355	310	4.876 (2.837)	-0.961	2.99	228
Tree 2	7.077 (6.819)	1.239	6.918	262	5.865 (3.542)	0.027	3.535	262	5.147 (2.15)	-0.690	2.252	185
Tree 3	5.963 (10.155)	0.125	10.14	317	4.855 (5.217)	-0.982	5.301	315	3.811 (4.998)	-2.026	5.385	267
Tree 4	7.915 (8.225)	2.078	8.472	366	5.656 (4.254)	-0.181	4.251	312	4.893 (3.387)	-0.944	3.51	277

C.3 Binary response

C.3.1 Non-phylogenetic GLM

Table C.25: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from binary response models - Non-phylogenetic GLM (univariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.021 (1.044)	-0.016	1.044	999	-0.002 (0.38)	-0.039	0.381	1000	0.005 (0.266)	-0.032	0.268	1000
Tree 2	-0.016 (0.351)	-0.053	0.355	1000	0.007 (0.173)	-0.03	0.176	1000	0.006 (0.106)	-0.031	0.111	1000
Tree 3	0.416 (8.635)	0.378	8.637	737	0.068 (1.472)	0.03	1.472	1000	-0.072 (2.925)	-0.109	2.925	865
Tree 4	0.017 (0.534)	-0.02	0.534	999	0.000 (0.285)	-0.038	0.288	1000	0.003 (0.185)	-0.034	0.188	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	1.399 (1.806)	-0.486	1.869	978	0.773 (0.482)	-1.112	1.212	1000	0.626 (0.323)	-1.259	1.299	1000
Tree 2	0.712 (0.553)	-1.173	1.297	999	0.555 (0.238)	-1.329	1.351	1000	0.471 (0.132)	-1.413	1.42	1000
Tree 3	3.337 (11.71)	1.453	11.791	687	2.134 (2.154)	0.25	2.168	999	1.478 (3.426)	-0.406	3.448	843
Tree 4	0.75 (0.812)	-1.135	5.305	979	0.533 (0.478)	-1.351	1.433	1000	0.431 (0.224)	-1.453	1.471	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	5.482 (7.764)	0.485	7.775	790	3.412 (3.715)	-1.585	4.037	997	2.406 (0.83)	-2.591	2.721	1000
Tree 2	3.346 (7.637)	-1.651	7.809	857	2.343 (1.264)	-2.654	2.939	996	1.837 (0.403)	-3.16	3.186	1000
Tree 3	8.472 (10.613)	3.475	11.158	534	9.458 (6.715)	4.461	8.058	959	4.566 (4.164)	-0.431	4.184	714
Tree 4	2.96 (3.097)	-2.037	7.709	807	2.112 (1.369)	-2.885	3.193	990	1.722 (0.622)	-3.275	3.334	1000

Table C.26: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from binary response models - Non-phylogenetic GLM (multivariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.101 (1.847)	0.06	1.847	996	-0.003 (0.436)	-0.044	0.438	1000	0.001 (0.304)	-0.041	0.306	1000
Tree 2	-0.024 (0.507)	-0.065	0.511	1000	0.007 (0.215)	-0.035	0.217	1000	0.004 (0.124)	-0.037	0.13	1000
Tree 3	0.194 (6.715)	0.153	6.712	780	0.028 (1.652)	-0.013	1.652	1000	-0.064 (4.118)	-0.105	4.117	794
Tree 4	0.016 (0.813)	-0.025	3.235	994	-0.003 (0.325)	-0.044	0.328	1000	-0.003 (0.208)	-0.044	0.213	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	1.287 (1.988)	-0.415	2.03	956	0.662 (0.526)	-1.041	1.166	1000	0.533 (0.345)	-1.17	1.22	1000
Tree 2	0.641 (0.897)	-1.061	1.389	998	0.457 (0.270)	-1.246	1.274	1000	0.38 (0.145)	-1.322	1.33	1000
Tree 3	1.8 (7.075)	0.098	7.071	772	1.87 (2.356)	0.167	2.361	995	1.764 (3.958)	0.062	3.956	765
Tree 4	0.719 (1.162)	-0.983	3.556	968	0.451 (0.393)	-1.251	1.311	999	0.365 (0.244)	-1.338	1.36	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	6.028 (7.490)	-0.064	7.482	455	6.24 (8.321)	0.148	8.317	831	3.953 (2.482)	-2.14	3.276	980
Tree 2	3.704 (4.736)	-2.388	5.300	430	4.153 (4.504)	-1.94	4.902	878	2.894 (1.694)	-3.199	3.619	996
Tree 3	5.087 (11.095)	-1.006	11.132	676	15.792 (16.276)	9.699	18.938	773	9.205 (9.595)	3.112	10.079	562
Tree 4	4.118 (12.289)	-1.975	12.430	453	3.832 (6.121)	-2.26	6.521	830	2.778 (2.575)	-3.315	4.197	986

Table C.27: **Mean effect size (M) with standard deviation in brackets, mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from binary response models - Non-phylogenetic GLM (multivariate model - binary predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.039 (3.886)	-0.107	3.886	996	0.024 (1.103)	-0.043	1.103	1000	0.029 (0.867)	-0.038	0.867	1000
Tree 2	-0.024 (1.761)	-0.092	1.763	1000	0.008 (0.724)	-0.06	0.726	1000	0.017 (0.483)	-0.05	0.485	1000
Tree 3	1.026 (30.054)	0.958	30.050	780	0.128 (1.685)	0.061	1.685	1000	-0.707 (11.041)	-0.774	11.062	794
Tree 4	0.038 (3.397)	-0.03	13.953	994	0.035 (1.204)	-0.033	1.204	1000	0.055 (0.851)	-0.013	0.851	1000

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	2.143 (6.338)	-0.168	6.337	956	0.802 (1.359)	-1.509	2.03	1000	0.678 (0.871)	-1.633	1.851	1000
Tree 2	1.251 (3.826)	-1.060	3.968	998	0.748 (0.746)	-1.563	1.732	1000	0.758 (0.510)	-1.553	1.635	1000
Tree 3	14.435 (32.033)	12.124	34.232	772	1.004 (2.458)	-1.307	2.783	995	1.346 (8.836)	-0.965	8.882	765
Tree 4	2.095 (6.631)	-0.216	16.464	968	1.013 (2.49)	-1.297	2.807	999	0.727 (0.884)	-1.584	1.814	1000

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	20.17 (19.071)	6.708	20.197	455	8.697 (11.014)	-4.765	11.995	831	6.247 (4.572)	-7.215	8.541	980
Tree 2	15.834 (17.796)	2.372	17.933	430	7.12 (7.200)	-6.342	9.592	878	5.574 (3.274)	-7.888	8.54	996
Tree 3	37.588 (22.607)	24.126	33.051	676	9.301 (10.166)	-4.161	10.979	773	10.909 (11.322)	-2.553	11.597	562
Tree 4	18.363 (21.374)	4.901	21.910	453	8.201 (10.197)	-5.261	11.468	830	5.918 (4.730)	-7.544	8.903	986

C.3.2 PGLMM

Table C.28: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from binary response models - PGLMM (univariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.263 (5.302)	0.184	5.301	596	0.067 (1.695)	-0.012	1.694	625	-0.016 (0.975)	-0.095	0.979	656
Tree 2	-0.175 (2.509)	-0.254	2.519	494	0.046 (1.114)	-0.033	1.113	625	0.017 (0.343)	-0.062	0.348	589
Tree 3	0.447 (11.439)	0.369	11.433	473	0.079 (4.786)	0.000	4.783	646	-0.163 (3.408)	-0.242	3.414	574
Tree 4	0.1 (2.965)	0.021	2.962	603	0.039 (1.065)	-0.04	1.065	651	0.006 (0.653)	-0.073	0.657	691

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	6.594 (6.707)	4.71	8.19	450	3.77 (2.329)	1.885	2.994	417	2.843 (1.156)	0.958	1.501	401
Tree 2	3.608 (2.847)	1.724	3.325	358	1.946 (1.296)	0.062	1.295	320	1.276 (0.871)	-0.609	1.061	350
Tree 3	6.719 (12.088)	4.835	13.007	448	8.651 (5.512)	6.767	8.724	479	4.194 (3.752)	2.309	4.402	460
Tree 4	3.457 (3.073)	1.573	3.450	466	2.319 (1.367)	0.435	1.433	427	1.814 (0.756)	-0.07	0.759	381

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	14.181 (8.089)	9.184	12.231	355	7.269 (2.982)	2.272	3.745	318	5.836 (1.613)	0.839	1.816	290
Tree 2	7.883 (4.285)	2.886	5.162	351	4.414 (1.970)	-0.583	2.053	415	2.904 (0.929)	-2.093	2.29	471
Tree 3	17.621 (13.969)	12.624	18.814	368	18.706 (7.466)	13.71	15.606	361	8.604 (4.976)	3.608	6.141	335
Tree 4	8.191 (4.166)	3.194	5.245	368	4.555 (1.777)	-0.442	1.829	336	3.926 (1.204)	-1.071	1.61	343

Table C.29: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from binary response models - PGLMM (multivariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.786 (8.128)	0.744	8.155	547	-0.028 (2.325)	-0.07	2.325	600	0.099 (1.231)	0.058	1.231	685
Tree 2	-0.17 (3.887)	-0.211	3.888	424	-0.014 (1.291)	-0.056	1.291	503	-0.012 (0.492)	-0.053	0.494	530
Tree 3	0.114 (16.326)	0.072	16.308	452	0.151 (6.058)	0.11	6.054	621	-0.16 (3.685)	-0.201	3.687	559
Tree 4	0.202 (4.376)	0.161	4.375	548	0.005 (1.374)	-0.036	1.373	644	0.029 (0.75)	-0.012	0.75	682

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	5.684 (7.611)	3.981	8.582	460	3.37 (2.371)	1.668	2.897	457	2.762 (1.359)	1.059	1.722	438
Tree 2	3.433 (3.77)	1.731	4.144	380	1.956 (1.686)	0.254	1.702	309	1.356 (0.887)	-0.347	0.951	299
Tree 3	6.089 (16.903)	4.387	17.445	444	7.720 (6.075)	6.017	8.546	484	3.641 (3.988)	1.939	4.43	479
Tree 4	3.427 (4.444)	1.724	4.762	500	2.234 (1.705)	0.532	1.784	468	1.751 (0.899)	0.049	0.899	433

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	13.528 (9.860)	7.435	12.338	358	6.965 (2.811)	0.872	2.939	316	5.18 (1.515)	-0.913	1.766	326
Tree 2	8.122 (4.589)	2.029	5.011	333	5.507 (2.093)	-0.585	2.17	303	3.746 (1.054)	-2.346	2.572	345
Tree 3	9.342 (18.099)	3.249	18.366	396	18.475 (7.934)	12.382	14.7	334	6.671 (6.758)	0.578	6.774	370
Tree 4	9.813 (6.163)	3.721	7.196	1000	4.815 (1.795)	-1.278	2.201	333	3.587 (1.078)	-2.506	2.728	322

Table C.30: **Mean effect size (M) with standard deviation in brackets, mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from binary response models - PGLMM (multivariate model - binary predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.623 (15.434)	-0.69	15.435	544	0.091 (6.012)	0.024	6.007	585	0.094 (3.52)	0.027	3.517	662
Tree 2	0.535 (12.348)	0.468	12.342	425	0.052 (4.864)	-0.016	4.859	490	-0.019 (2.061)	-0.087	2.061	514
Tree 3	1.315 (37.815)	1.248	37.794	455	0.329 (5.617)	0.262	5.619	638	0.143 (14.503)	0.076	14.489	531
Tree 4	-0.600 (13.807)	-0.667	13.810	503	0.227 (5.680)	0.16	5.677	623	0.033 (2.859)	-0.035	2.857	712

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	4.443 (13.839)	2.132	13.988	497	3.23 (5.816)	0.919	5.883	585	2.529 (3.438)	0.218	3.442	605
Tree 2	3.186 (11.31)	0.876	11.33	419	2.986 (4.483)	0.675	4.528	431	2.154 (2.527)	-0.156	2.529	431
Tree 3	3.909 (43.492)	1.598	43.473	446	2.518 (5.345)	0.207	5.344	619	2.17 (14.659)	-0.141	14.646	525
Tree 4	4.837 (13.736)	2.527	13.954	544	2.895 (5.355)	0.584	5.382	579	2.217 (2.811)	-0.094	2.81	652

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	18.273 (16.174)	4.811	16.856	416	9.496 (4.650)	-3.967	6.107	362	7.951 (2.895)	-5.511	6.223	323
Tree 2	16.234 (12.149)	2.771	12.444	347	9.600 (4.485)	-3.862	5.913	322	7.414 (2.436)	-6.048	6.519	313
Tree 3	28.466 (36.901)	15.004	39.794	416	9.555 (6.804)	-3.907	7.839	423	13.022 (13.067)	-0.44	13.06	478
Tree 4	16.302 (12.041)	2.840	12.360	438	9.84 (5.325)	-3.622	6.434	368	7.692 (2.600)	-5.77	6.328	367

C.3.3 PLR

Table C.31: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from binary response models - PLR (univariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.08 (0.539)	1.295	1.402	99	0.029 (0.425)	1.244	1.314	97	0.016 (0.234)	1.231	1.253	97
Tree 2	0.054 (0.301)	1.269	1.304	100	0.007 (0.135)	1.222	1.23	100	-0.006 (0.099)	1.209	1.213	100
Tree 3	0.022 (1.666)	1.237	2.068	97	-0.095 (1.097)	1.12	1.564	92	0.112 (1.682)	1.327	2.133	71
Tree 4	-0.024 (0.394)	1.191	1.254	99	0.039 (0.187)	1.254	1.268	100	0.034 (0.125)	1.249	1.255	100

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.912 (0.96)	-0.973	1.363	97	0.592 (0.359)	-1.293	1.341	99	0.656 (0.314)	-1.228	1.267	100
Tree 2	0.588 (0.476)	-1.297	1.38	100	0.51 (0.212)	-1.374	1.391	100	0.46 (0.129)	-1.425	1.43	100
Tree 3	1.684 (2.274)	-0.2	2.27	86	1.177 (1.206)	-0.707	1.393	89	1.224 (1.323)	-0.66	1.472	86
Tree 4	0.572 (0.482)	-1.312	1.397	96	0.493 (0.215)	-1.391	1.408	100	0.396 (0.158)	-1.488	1.497	100

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	2.791 (1.931)	-2.206	2.923	72	2.41 (1.071)	-2.587	2.798	92	2.497 (0.907)	-2.5	2.658	98
Tree 2	2.16 (1.334)	-2.837	3.131	84	1.893 (0.535)	-3.104	3.15	98	1.79 (0.325)	-3.207	3.224	100
Tree 3	5.753 (4.693)	0.756	4.721	71	3.712 (2.283)	-1.285	2.604	62	4.071 (2.512)	-0.925	2.662	79
Tree 4	1.718 (0.999)	-3.279	3.426	73	1.548 (0.563)	-3.449	3.494	94	1.491 (0.495)	-3.506	3.54	100

Table C.32: **Mean effect size (M) with standard deviation in brackets (SD), mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from binary response models - PLR (multivariate model - continuous predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.038 (0.706)	-0.003	0.702	99	0.035 (0.521)	-0.006	0.519	97	0.015 (0.26)	-0.027	0.26	95
Tree 2	0.051 (0.407)	0.01	0.405	100	-0.012 (0.138)	-0.053	0.147	100	-0.006 (0.126)	-0.047	0.134	100
Tree 3	0.021 (2.565)	-0.02	2.551	92	-0.223 (1.646)	-0.264	1.658	86	0.015 (1.526)	-0.026	1.515	69
Tree 4	-0.079 (0.622)	-0.12	0.631	98	0.044 (0.22)	0.002	0.219	100	0.031 (0.129)	-0.01	0.129	100

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	1.029 (1.7)	-0.673	1.82	93	0.532 (0.378)	-1.17	1.229	98	0.601 (0.319)	-1.101	1.146	100
Tree 2	0.469 (0.632)	-1.234	1.385	99	0.422 (0.231)	-1.281	1.301	100	0.378 (0.141)	-1.324	1.332	100
Tree 3	1.833 (3.293)	0.131	3.276	81	0.973 (1.738)	-0.729	1.875	76	1.189 (2.347)	-0.514	2.388	80
Tree 4	0.575 (0.8)	-1.127	1.380	95	0.42 (0.232)	-1.283	1.303	100	0.356 (0.169)	-1.346	1.356	100

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	4.839 (4.214)	-1.254	4.321	27	2.831 (1.239)	-3.262	3.485	48	3.191 (0.995)	-2.902	3.066	77
Tree 2	3.580 (2.195)	-2.513	3.315	34	2.264 (0.761)	-3.829	3.903	73	2.255 (0.504)	-3.837	3.87	91
Tree 3	6.930 (6.162)	0.837	6.131	35	4.124 (4.017)	-1.969	4.433	45	5.960 (2.543)	-0.132	2.52	49
Tree 4	1.643 (1.763)	-4.450	4.776	32	2.097 (0.782)	-3.996	4.07	67	2.124 (0.631)	-3.969	4.019	81

Table C.33: **Mean effect size (M) with standard deviation in brackets, mean error (ME), rooted mean squared error (RMSE) and number of simulations (N) from binary response models - PLR (multivariate model - binary predictor)**:
a) input correlation $r = 0.01$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	-0.058 (2.88)	-0.126	2.868	99	-0.154 (1.085)	-0.221	1.102	97	0.03 (0.531)	-0.038	0.53	95
Tree 2	0.037 (1.237)	-0.031	1.231	100	0.137 (0.638)	0.069	0.639	100	-0.003 (0.536)	-0.07	0.538	100
Tree 3	0.082 (5.632)	0.014	5.602	92	0.84 (6.119)	0.772	6.132	86	-0.726 (11.469)	-0.793	11.413	69
Tree 4	0.017 (2.506)	-0.05	2.494	98	-0.034 (0.866)	-0.102	0.868	100	0.041 (0.526)	-0.026	0.524	100

b) input correlation $r = 0.5$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	0.665 (2.605)	-1.646	3.069	93	0.607 (1.014)	-1.703	1.98	98	0.407 (0.585)	-1.904	1.991	100
Tree 2	0.994 (2.944)	-1.317	3.212	99	0.667 (0.662)	-1.644	1.771	100	0.64 (0.46)	-1.67	1.732	100
Tree 3	1.672 (6.225)	-0.639	6.219	81	2.212 (7.162)	-0.099	7.116	76	1.978 (11.161)	-0.332	11.096	80
Tree 4	1.028 (3.058)	-1.282	3.301	95	1.03 (2.662)	-1.281	2.942	100	0.441 (0.636)	-1.87	1.974	100

c) input correlation $r = 0.9$.

	Species 20				**Species 50**				**Species 100**			
	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N	M (SD)	ME	RMSE	N
Tree 1	9.049 (8.064)	-4.413	9.061	27	4.606 (3.254)	-8.857	9.424	48	4.395 (1.612)	-9.067	9.207	77
Tree 2	6.895 (6.022)	-6.567	8.850	34	4.1 (1.685)	-9.362	9.510	73	4.238 (0.931)	-9.225	9.271	91
Tree 3	10.991 (12.663)	-2.471	12.723	35	11.876 (15.078)	-1.586	14.994	45	8.044 (9.637)	-5.418	10.97	49
Tree 4	8.215 (9.948)	-5.247	11.090	28	5.332 (3.982)	-8.130	9.040	67	4.936 (2.090)	-8.526	8.775	81