Multi-Response Phylogenetic Mixed Models: Concepts and Application

Ben Halliwell[1,2], Luke A. Yates[1], and Barbara R. Holland [1]

[1]School of Natural Sciences, University of Tasmania, Hobart, Tasmania, Australia
[2]Corresponding author. Email: benjamin.halliwell@utas.edu.au

July 14, 2023

Abstract

The scale and resolution of trait databases and molecular phylogenies is increasing rapidly. These resources permit many open questions in macro-evolution and -ecology to be addressed with the right statistical tools. Phylogenetic mixed models (PMM), particularly multi-response (MR) implementations, offer great potential for multivariate analyses of trait evolution. While flexible and powerful, these methods are not often employed, reflecting a specialised and technical literature that creates barriers to usage for many biologists. Here we present a practical and accessible guide to MR-PMM. We begin by outlining key concepts in mixed modelling, emphasising the use of covariance matrices to model correlative structure in the data. We discuss the limitations of single-response (SR) models, such as phylogenetic generalised least squares (PGLS), for characterising patterns of trait covariation, and emphasise MR models as a preferable approach to analyses involving multiple species traits. Using simulated data and visual examples, we demonstrate the motivation for partitioning trait covariances into phylogenetic and independent components in MR-PMM. We discuss interpretation, prediction, and model validation, including methods based on leave-one-out cross validation. We then apply this approach to a real-world data set of leaf traits in Eucalyptus to show how MR-PMM can provide a more nuanced understanding of trait correlations compared to traditional SR approaches. Finally, we highlight common pitfalls and offer practical recommendations to analysts. To complement this material, we provide an online tutorial including side-by-side implementations of MR-PMM in two popular R packages, `MCMCglmm` and `brms`, as well as guidance for users looking to customize models further using the Stan programming language.

Contents

1 Introduction **5**

2 Models **6**

2.1 Mixed Models 6

2.2 Phylogenetic Mixed Models 8

2.2.1 Phylogenetic signal 9

2.2.2 Branch length transformations 9

2.2.3 Limitations of single response models 10

2.3 Multi-Response Phylogenetic Mixed Models 11

2.3.1 Considerations for fixed effects 12

2.3.2 Assessing conditional dependencies 13

2.4 Summary 14

3 Simulation **14**

4 Interpretation **15**

5 Prediction **17**

5.1 Predicting to new or missing observations 17

5.2 Model validation using posterior predictive checks 18

5.3 Predictive assessment: cross validation 19

6 Example - Leaf Economics in Eucalyptus **20**

6.1 Methods 20

6.2 Results and Discussion 21

7 Discussion **22**

7.1 Summary 22

7.2 Strengths and Weaknesses 22

7.3 Pitfalls and Recommendations 24

7.4 Conclusions 25

8 Tutorial **25**

9 Acknowledgements **26**

1 Introduction

The motivation for developing multivariate phylogenetic comparative methods is now broadly appreciated (Adams and Collyer 2018; Uyeda et al. 2015, 2018; Garamszegi 2014). Modern approaches aim to move beyond phylogenetic regression and variance partitioning of individual Gaussian variables (e.g., Lynch 1991; Pagel 1999; Blomberg et al. 2003), to methods capable of evaluating the strength, direction, and conservatism of co-evolutionary relationships within networks of continuous and discreet variables (Westoby et al., 2023; Haba and Kutsukake, 2019; Brommer et al., 2019; Hadfield, 2010). These multivariate techniques are applicable to a broad range of species traits, from morphology, physiology, and behavior, to environmental tolerance limits and niche characteristics. Furthermore, as the tide of 21st-century data collection erodes historical constraints on the scale and complexity of analyses, opportunities to apply these methods are increasing: Advances in climate modeling, remote sensing technology, and collaborative data projects are accelerating the availability and resolution of suitable datasets (Green et al., 2022; Herberstein et al., 2022; Falster et al., 2021; Kattge et al., 2020); the genomic revolution continues to provide more accurate and complete phylogenies (Young and Gillung, 2020; Laumer et al., 2019; Smith and Brown, 2018; Yeates et al., 2016); and the computational resources necessary to pursue fully Bayesian analyses are now accessible to most researchers. Despite this coalescence of opportunities, challenges in implementing modern methods create barriers to usage for many biologists.

A robust literature developing statistical models of multivariate trait evolution has existed for some time. However, new techniques are often slow to absorb into research practice, especially where they demand a technical or conceptual leap from users. Development of the generalised least-squares framework, helped to bring phylogenetic regression and correlative analyses of trait evolution into the mainstream (Rohlf, 2001; Martins and Hansen, 1997; Lynch, 1991; Grafen, 1989; Felsenstein, 1985). Subsequent extensions to phylogenetic generalised linear mixed models (PMM) provided additional benefits, including support for non-Gaussian traits, flexibility in the specification of hierarchical group effects, and a growing familiarity with mixed models among researchers in ecology and evolution (Ives and Helmus, 2011; Ives and Garland Jr, 2010; Bolker et al., 2009; Housworth et al., 2004). Currently, PMM also supports fully multivariate analyses, facilitating the estimation (and decomposition) of covariances between discrete and continuous traits (Hadfield, 2010). However, multivariate implementations of PMM remain relatively rare in the literature, even for datasets to which they would be well suited, such as recent large-scale comparative analyses of species traits (e.g., Cássia-Silva et al. 2020; Grossnickle 2020; Bruelheide et al. 2018; Díaz et al. 2016). A sobering list of publications highlighting widespread misconceptions about phylogenetic comparative methods, erroneous statistical practices, and the potential for spurious inference, confirm the need for ongoing translational research (Uyeda et al., 2018; Cooper et al., 2016; Uyeda et al., 2015; Revell, 2010; Freckleton, 2009; Revell et al., 2008). In particular, practical and accessible guidance for biologists wanting to apply multivariate methods that provide a more meaningful decomposition of trait relationships (Westoby et al., 2023).

This review aims to bridge the gap between theory and practice in phylogenetic mixed model

(PMM) analyses of trait evolution. We focus on Bayesian implementations of multi-response (MR) PMM, because this model class provides a highly flexible, multivariate framework with general utility for comparative biology. The capacity for MR-PMM to provide more nuanced analyses of trait (co)evolution has long been recognised (Housworth et al., 2004; Hadfield, 2010). However, we believe an incomplete conceptual understanding, along with practical barriers to implementation, continues to prevent common usage of MR-PMM among biologists. Thus, we begin with a broad conceptual introduction to mixed models, demonstrating how covariance matrices are used to model correlated error structures, such as those present in phylogenetic effects. We define the basic SR-PMM and discuss its equivalence with phylogenetic generalised least squares (PGLS), including the different parametrizations used to quantify phylogenetic signal. After outlining the limitations of SR models for characterising patterns of coevolution between traits, we shift focus to MR-PMM. In particular, we highlight the different components of trait covariance (phylogenetic versus independent) that arise in MR-PMM, how these error structures are specified using covariance matrices, and how parametrizations of these matrices are used to estimate phylogenetic and independent correlations between species traits. In section 3, we explore data simulated from MR-PMM to clarify the distinction between correlations operating on these different levels in the model hierarchy. In section 4 we discuss interpretation of these correlation coefficients. In section 5 we cover predictive assessment, including methods based on posterior predictive distributions and leave-one-out (LOO) cross-validation procedures. In section 6, we provide a worked example on a real-world data set of leaf traits in Eucalyptus. Finally, in section 7 we provide a summary of the strengths and weaknesses of MR-PMM, highlighting common pitfalls and providing recommendations to analysts.

To support our exposition, we provide an online tutorial, including annotated code and worked examples of matrix operations. This tutorial demonstrates how to simulate multivariate trait data containing phylogenetic signal and specify, diagnose, and interpret corresponding MR-PMM in two popular R packages, `MCMCglmm` (Hadfield, 2010) and `brms` (Bürkner, 2017). Finally, we provide an introduction to MR-PMM in Stan for more advanced users seeking greater flexibility in model parameterisation.

2 Models

2.1 Mixed Models

Biologists are familiar with structured sampling designs: these apply when each observation from a survey or experiment is a member of one or more nested groups. Acknowledging structure in our data is important, because observations within a group will often be more similar than can be explained by available predictors. For example, observations of species traits collected from the same transect may be more similar to each other than those collected from different transects, due to spatial heterogeneity in unobserved causal factors.

This covariance between observations—the tendency for members of a group to have similar trait values—may exist at various scales, producing a hierarchy of group-level effects. For example, observations (level 1) from transects (level 2) could be nested within sites (level 3). Hierarchical (or mixed) models aim to capture this structural hierarchy to more accurately model variation in the data.

In the simplest case, the hierarchical structure is limited to a single random (or group-level) intercept. For the popular `lme4` package in R, the syntax used to express this model would be:

$$y \sim 1 + X + (1|\text{group}), \tag{1}$$

where y is the observed response, 1 specifies a population-level intercept (often omitted from the model formula as fit by default), the measured covariate X specifies a fixed effect, and (1|group) specifies a random intercept at the group level, e.g., transect. In other words, we model y as a linear combination of effects, including a random effect that accounts for the fact that observations belong to distinct groups.

To express a more general form of this model and make precise the notion of a random effect, it is helpful to adopt formal mathematical notation. In particular, we need to be explicit about the choice of statistical distributions governing each random variable,

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{\mu} + \boldsymbol{b} + \boldsymbol{e} \\ \boldsymbol{\mu} &= \beta_0 1 + \beta_1 X_1 + ... + \beta_k X_k \\ \boldsymbol{b} &\sim \mathcal{N}(0, \sigma_b^2 C) \\ \boldsymbol{e} &\sim \mathcal{N}(0, \sigma_e^2 I), \end{aligned} \tag{2}$$

where, $\boldsymbol{\mu}$ is a vector of fixed effects including a population-level intercept and fixed effect coefficients associated with predictors X_i, $i = 1, .., k$, $\boldsymbol{b}$ a vector of group-level random effects, and $\boldsymbol{e}$ a vector of independent errors. The notation $\sim \mathcal{N}(\cdots)$ means "drawn from a (multivariate) normal distribution". For the random effects, an $n \times n$ covariance matrix C encodes the hierarchical structure among observations, and the identity matrix I (a matrix with 1's on the diagonal and 0's in all off-diagonals) encodes independent error for each observation (Figure 1A). Finally, the parameters σ_b^2 and σ_e^2 denote the magnitude of the structured (group-level) and independent (residual) variance components, respectively.

To illustrate, again suppose we are analysing data from species sampled in different transects, and we want to use random intercepts to model the covariance between observations within each transect. For Gaussian distributed errors, the two variance components can simply be added to from the total covariance:

$$\Sigma = \sigma_b^2 C + \sigma_e^2 I. \tag{3}$$

This *marginal* model is obtained by averaging over the distribution of the random effects,

which means treating $\boldsymbol{b} + \boldsymbol{e}$ as the (marginalised) residuals around a mean $\boldsymbol{\mu}$,

$$\boldsymbol{b} + \boldsymbol{e} \sim \mathcal{N}(0, \Sigma). \tag{4}$$

In this view, the hierarchical structure of our transect sampling design can be represented as a tree comprising a polytomous clade for each transect (Figure 1A). The corresponding covariance matrix Σ assumes a block-diagonal form for which the block sizes are determined by the number of observations within each transect, and the relative magnitudes of σ_b^2 and σ_e^2 correspond to relative lengths of the internal and external branches in the tree (also see Ives 2018). In many instances, C is re-scaled to a correlation matrix (i.e., a matrix with 1's on the diagonal) which places it on the same scale as I and allows the relative proportion of each variance component to be calculated from σ_b^2 and σ_e^2. Nonetheless, this diagrammatic representation of observation-level covariance makes clear how matrices are used to model sources of structural dependence in the data, including those arising from phylogenetic history (Figure 1B).

2.2 Phylogenetic Mixed Models

In analyses of cross-species data, dependence arises from phylogenetic effects: the tendency for closely related species to resemble each other phenotypically due to the hierarchical evolutionary history of life (Pagel, 1999; Blomberg et al., 2003). This phylogenetic structure is handled in exactly the same way as a nested sampling design, by defining a covariance matrix that encodes our expectation of dependence in the data (Figure 1B). This approach implies that conserved covariances among species trait values reflect the proposed phylogenetic structure, consistent with Brownian motion evolution (Felsenstein, 1985).

Brownian motion represents a stochastic random walk—trait values increasing or decreasing at a constant rate every time step—that branches at each node in the phylogeny. This process has the property that trait variance increases linearly with time, meaning that shared branch lengths are proportional to the expected (co)variances of trait values among species. Thus, for Brownian motion, there is a direct translation from phylogeny to covariance matrix (Ives, 2018). In PMM, phylogenetic structure enters the model as the covariance C of the random effects $\boldsymbol{b}$, where C is scaled by the phylogenetic variance (or, evolutionary rate) σ_b^2 (2). The phylogeny is supplied by the user and usually considered known without error (although in a Bayesian setting, phylogenetic uncertainty is easily incorporated by combining posterior samples across a set of candidate topologies). The rate parameter σ_b^2 is unknown and to be estimated from the data.

For n taxa in a phylogeny, C is an $n \times n$ matrix containing all pairwise similarity scores between taxa. Figure 1B shows a phylogeny, and associated C matrix, for 6 taxa. Note that C is symmetric, with values on the diagonal representing the sum of branch lengths for each path from root to tip. For an ultrametric tree (tips appear right-adjusted) like that presented in Figure 1B, all diagonal elements C_{ii} will be equal because all taxa were sampled in the

present day. In contrast, for a pair of taxa i and j, the expected phylogenetic covariance is given by the off-diagonal element C_{ij}, the sum of branch lengths from the root node to the most recent common ancestor of taxa i and j (see Rohlf 2001). In other words, the sum of shared branch lengths for taxa i and j.

To summarise, PMM is a special case of a linear mixed model (2) that incorporates phylogenetic relatedness via a covariance matrix, C. Species trait values $\boldsymbol{y}$ are modelled as the sum of fixed effects $\boldsymbol{\mu}$, a phylogenetic random effect $\boldsymbol{b}$, and independent (residual) error $\boldsymbol{e}$ (see Box 1 for non-Gaussian extensions). By default, the covariance structure of the random effects C assumes a Brownian motion model of evolution.

2.2.1 Phylogenetic signal

A common objective of phylogenetic comparative analyses is to quantify phylogenetic signal in species traits. Numerous metrics have been developed for this purpose, each with different assumptions and applications (Münkemüller et al., 2012). In PMM, phylogenetic heritability h^2 describes the proportion of variance in $\boldsymbol{y}$ attributable to phylogeny, and is expressed in terms of the estimated variance components as,

$$h^2 = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}. \tag{5}$$

Importantly, when fixed effects are specified in the model, the variance explained by these components must also be included in the denominator of the calculation (de Villemereuil et al., 2016).

2.2.2 Branch length transformations

Values of $h^2 < 1$ imply a residual error in $\boldsymbol{y}$ that cannot be explained by Brownian motion. In the original presentation of PGLS, the matrix C is taken as the total covariance matrix (Grafen, 1989), which assumes traits have evolved via Brownian motion with no additional residual error. This is equivalent to fixing $\sigma_e^2 = 0$, or ignoring red branch lengths in Figure 1B. With subsequent developments of the model, the covariance matrix C, or equivalently the branch lengths of the corresponding phylogenetic tree, is transformed as a function of one or more additional parameters $\boldsymbol{\theta}$, allowing different evolutionary models to be fit to the data (Symonds and Blomberg, 2014; Harmon et al., 2008). These parameters characterise branch-length transformations that produce a transformed phylogenetic covariance matrix $C(\boldsymbol{\theta})$, which is multiplied by a scalar parameter σ^2 to define the total covariance,

$$\Sigma = \sigma^2 C(\boldsymbol{\theta}). \tag{6}$$

A common instance is when $C(\theta)$ is a weighted sum of C and I characterised by a single parameter, λ (Pagel, 1999),

$$C(\lambda) = \lambda C + (1 - \lambda) I. \tag{7}$$

In this formulation, the additive terms express the relative proportions, scaled by λ, of the total variance attributable to phylogenetic and independent error. Because h^2 was originally framed in terms of a phylogenetic partial R^2 (Lynch 1991), and λ is frequently discussed in terms of a branch-length transformation in PGLS (e.g., Pagel 1999; Symonds and Blomberg 2014; Ives 2019), it is easy overlooked that these quantities are in fact mathematically equivalent (Housworth et al., 2004). Indeed, from the above it can be shown that $\sigma^2 = \sigma_b^2 + \sigma_e^2$ and $\lambda = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2} = h^2$. These equivalencies help clarify the relationship between PMM and other popular methods: PMM is equivalent to PIC (and PGLS without λ optimisation) when $\Sigma = \sigma_b^2 C$ (i.e., $\sigma_e^2 = 0$) (Blomberg et al., 2012) and to ordinary least squares when $\Sigma = \sigma_e^2 I$ (i.e., $\sigma_b^2 = 0$). With λ optimised, PGLS is equivalent to a marginal PMM for a single Gaussian response trait (3).

2.2.3 Limitations of single response models

PMM provides an elegant solution to the statistical problem of phylogenetic dependence in cross-species data. However, univariate implementations (2) have important limitations that arise from treating one species trait, often arbitrarily, as the response variable and other traits as predictor variables. This approach implies a causal relationship, where-by evolutionary changes in the response trait are driven by changes in predictor traits. More often, measured species traits are merely correlated due to a common causal mechanism. Even when two traits are functionally related, co-evolution may involve reciprocal causation, meaning strictly directed causal models are often not biologically defensible. Furthermore, the validity of inferences is compromised when predictor traits themselves contain phylogenetic signal due to confounding of the phylogenetic signal in the predictor with that in the residuals (Westoby et al., 2023).

Suppose we observe a phenotypic correlation between two traits, $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, across a number of species and that both traits display phylogenetic signal; a common occurrence in empirical data sets (Blomberg et al., 2003; Freckleton et al., 2002). Part of the covariance between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ across species is due to a phylogenetically conserved relationship between these traits (e.g., deep functional integration, or phylogenetic niche conservatism), with the remainder due to causes that are independent of phylogeny (e.g., local adaptation, or selection imposed by fluctuating environmental conditions). If we model these data using a single-response model (2), such as PGLS, treating $\boldsymbol{y}_2$ as a fixed predictor of $\boldsymbol{y}_1$ reduces this composite of correlations operating at the phylogenetic and independent levels to a single slope parameter. Not only are the separate (co)variance components not recoverable from the model, but phylogenetic signal in the predictor trait $\boldsymbol{y}_2$ is confounded with (phylogenetic components of) the residual covariance structure (Warton 2022; also see Wilson 2008 and Marques et al. 2022

for analogous issues in quantitative genetics and spatial statistics, respectively). In other words, SR models cannot effectively separate covariance between traits that is phylogenetically structured (conserved correlation between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$) from covariance between traits that is independent of the phylogenetic structure (independent correlation between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$). One consequence of this confound is that estimated slope parameters will be misleading when phylogenetic and independent correlations have different signs, or vary considerably in magnitude. In a worst-case scenario, these two correlation components may cancel each other out when reduced to a single (scalar) slope estimate, leading to the erroneous conclusion that the traits in question are not meaningfully related (Westoby et al., 2023).

An alternative approach, which re-frames this analysis as a multivariate statistical problem, is to treat both traits $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ as response variables. That is, move all traits to the left-hand side of the model equation (8). This shift to a multivariate framework has several distinct benefits: 1) it relaxes assumptions about biological causality between traits, 2) it allows phylogenetic signal to be modelled in all traits simultaneously, and 3) it permits a partitioning of trait correlations into phylogenetic and independent components (Housworth et al. 2004; Hadfield 2010). We argue this approach will often be more meaningful for comparative analyses of species traits addressing ecological or evolutionary hypotheses (Section 3, 4).

2.3 Multi-Response Phylogenetic Mixed Models

In MR-PMM, all species traits are modelled jointly as response variables, which means both the phylogenetic random effects $\mathbf{b}$ and the independent errors $\mathbf{e}$ must also be modelled jointly. For Gaussian response traits, this is achieved using multivariate normal distributions (but see Box 1 for non-Gaussian response types). In a natural extension of the single response model (2), each trait appears in a separate row, which for two traits can be written as

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 + \mathbf{b}_1 + \mathbf{e}_1 \\ \boldsymbol{\mu}_2 + \mathbf{b}_2 + \mathbf{e}_2 \end{pmatrix} \tag{8}$$

$$(\mathbf{b}_1, \mathbf{b}_2)^{\mathrm{T}} \sim \mathcal{N}(0, \Sigma^{\mathrm{phy}} \otimes C) \tag{9}$$

$$(\mathbf{e}_1, \mathbf{e}_2)^{\mathrm{T}} \sim \mathcal{N}(0, \Sigma^{\mathrm{ind}} \otimes I) \tag{10}$$

where between-trait covariances across both phylogenetic and independent components are modelled with the trait covariance matrices,

$$\Sigma^{\mathrm{phy}} = \begin{pmatrix} \Sigma^{\mathrm{phy}}_{11} & \Sigma^{\mathrm{phy}}_{12} \\ \Sigma^{\mathrm{phy}}_{21} & \Sigma^{\mathrm{phy}}_{22} \end{pmatrix} \qquad \text{and} \qquad \Sigma^{\mathrm{ind}} = \begin{pmatrix} \Sigma^{\mathrm{ind}}_{11} & \Sigma^{\mathrm{ind}}_{12} \\ \Sigma^{\mathrm{ind}}_{21} & \Sigma^{\mathrm{ind}}_{22} \end{pmatrix}. \tag{11}$$

where the matrix elements $\Sigma^{\mathrm{ind}}_{ij}$ and $\Sigma^{\mathrm{phy}}_{ij}$ are scalar covariance components between traits i and j. The Kronecker product $\otimes$ is a type of matrix multiplication used here to combine the $m \times m$ trait covariance matrix with the $n \times n$ species covariance matrix into an $nm \times nm$ covariance matrix. This larger matrix matches the dimensions of the multi-trait response

vector and correctly encodes the parameterised correlation structure across species and traits. For example, the Kronecker product for the phylogenetic component can be written as

$$\Sigma^{\text{phy}} \otimes C = \begin{pmatrix} \Sigma_{11}^{\text{phy}} C & \Sigma_{12}^{\text{phy}} C \\ \Sigma_{21}^{\text{phy}} C & \Sigma_{22}^{\text{phy}} C \end{pmatrix} \tag{12}$$

which has the effect of introducing phylogenetic relatedness, via C, into the covariance structure of each trait and each pairwise trait relationship (see tutorial for fully worked examples). Like the SR case (3), the total covariance matrix of the marginal model is the sum of the phylogenetic and independent components,

$$\Sigma \;=\; \Sigma^{\text{phy}} \otimes C + \Sigma^{\text{ind}} \otimes I. \tag{13}$$

Despite the large dimension of Σ, the number of estimated (co)variance parameters in MR-PMM is only $m(m+1)$ (e.g., 6 parameters for the two-trait case). These parameters are the standard deviations σ and the correlations ρ used to characterise the matrix elements of both Σ^{phy} and Σ^{ind}, expressed as

$$\Sigma_{ij}^{\text{phy}} = \sigma_i^{\text{phy}} \sigma_j^{\text{phy}} \rho_{ij}^{\text{phy}} \qquad \text{and} \qquad \Sigma_{ij}^{\text{ind}} = \sigma_i^{\text{ind}} \sigma_j^{\text{ind}} \rho_{ij}^{\text{ind}}, \tag{14}$$

for $i, j = 1, .., m$.

The key distinction between SR- and MR-PMM is the decomposition of trait (co)variances. In the MR case, the Kronecker product (12) provides a means to include both phylogenetic and independent structures in the modelled (co)variance term for each trait and trait pair. Specifically, the estimated quantity $\widehat{\Sigma}_{ij}^{\text{phy}} \neq 0$ provides evidence for a component of correlation between traits i and j that is conserved over evolutionary time (i.e., a portion of trait covariance that is associated with phylogeny), while $\widehat{\Sigma}_{ij}^{\text{ind}} \neq 0$ provides evidence for a component of correlation between these traits that is independent of phylogeny (see Figure 3 for an illustration).

2.3.1 Considerations for fixed effects

There are two main reasons for including fixed effects in a MR-PMM. The first is to account for sampling biases, such as differences in data-collection methods or laboratory protocols. For example, in the field of plant hydraulics, physiological traits relating to xylem vulnerability are often measured using different techniques that each have distinct assumptions and sources of error (Jansen et al., 2015). In such situations, it is advisable to account for measurement methodology with a fixed effect covariate. The second reason is to explore relationships between responses conditional on covariates, typically to evaluate hypotheses about the mechanistic basis of trait correlations. For example, Sanchez-Martinez et al. (2020) use this approach to partition correlations between a range of plant hydraulic traits into phylogenetic and independent contributions, then re-fit models with relevant climate covariates included as fixed effects. Whether or not trait correlations remained was used to evaluate

hypotheses about trade-offs and integration between these traits in the evolution of plant hydraulic systems (see also Kruuk and Hadfield 2007 for related examples from quantitative genetics models). Equivalent procedures are used to control for environmental variation in multivariate species co-occurrence models (Ovaskainen et al., 2017; Warton et al., 2015).

2.3.2 Assessing conditional dependencies

The practice of assessing conditional relationships between traits (i.e., after controlling for relevant covariates) is well-established for univariate models. For example, it is common for researchers to use multiple regression to estimate partial regression coefficients for each fixed effect predictor. However, in MR-PMM, it is often preferable to estimate conditional relationships from a joint model, where both traits and covariates are treated as responses variables and feature on the left-hand side of the model equation. This is needed because traits commonly treated as covariates, such as those characterising the climatic or environmental niche of species, often display strong phylogenetic signal (Liu et al., 2020; Kubota et al., 2017; Peixoto et al., 2017; Hof et al., 2010; Kozak and Wiens, 2010; Smith and Beaulieu, 2009; Hawkins et al., 2007). Treating covariates as response variables both avoids potential confounds between fixed and phylogenetic effects (see section 2.2.4) and allows conditional dependencies between traits to be estimated on the phylogenetic and independent levels, providing more nuanced interpretations. Conditional dependencies between response traits in MR-PMM are obtained by computing the matrix inverse of the estimated trait covariance matrices, known as precision matrices,

$$\Omega^{\text{phy}} = (\Sigma^{\text{phy}})^{-1} \qquad \text{and} \qquad \Omega^{\text{ind}} = (\Sigma^{\text{ind}})^{-1} \tag{15}$$

The use of precision matrices to characterise conditional dependencies casts MR-PMM in the framework of Gaussian graphical models (Popovic et al., 2019; Epskamp et al., 2018; Yuan and Lin, 2007), which have great potential for clarifying assumptions of phylogenetic comparative methods (Uyeda et al., 2018). Elements of the precision matrices relate directly to the existence of edges in a graphical network (Figure 2). For example, suppose we obtain data on three co-varying species traits x, y and z. The absence of an edge between x and y after accounting for z signifies conditional independence: that x and y are independent, given z. Precision matrices are sparser than their corresponding covariance matrices (i.e., contain more off-diagonal elements that are close to 0), which focuses our inference on candidate causal relationships. Outside of controlled experiments, it remains possible that edges in a graph could be explained by missing variables. Nonetheless, precision matrices are a powerful tool for generating candidate causal hypotheses from observational data (Popovic et al., 2019; Pearl, 1995). Furthermore, MR-PMM can be parameterised directly in terms of the precision elements, which naturally reduces model complexity and allows priors to be placed on these target parameters, e.g., for the purpose of regularization (see tutorial).

2.4 Summary

Many species traits intended for use as predictors or covariates in SR-PMM (2) will themselves contain phylogenetic signal. If the evolutionary processes generating data we observe are characterised by reciprocal co-evolution (or joint distribution) of two or more traits, then the prediction-focused SR model may be badly misspecified; it not only fails to adequately partition covariances but can also fail to identify the presence of significant correlative effects. By partitioning trait correlations into phylogenetic and independent contributions, MR-PMM (8) allows us to decompose cross-species trait correlations into a more meaningful set of contributing factors. This framework permits the estimation of conditional dependencies between response traits on both the phylogenetic and independent levels, which describe how pairs of traits are related after controlling for covariances with other response traits. As partitioning trait variation and identifying functional trait relationships are principal objectives of the biological research program, MR-PMM will often represent a better approximating model and thus offer more valuable inferences. Considering these benefits, MR-PMM appears to be underutilised in evolutionary and ecological research.

3 Simulation

To compare and contrast phylogenetic and independent trait correlations, and clarify their biological interpretation, we simulated data under MR-PMM for different values of phylogenetic and independent (co)variance between two Gaussian traits x and y. Figure 3 shows scatter plots of simulated data together with the tree from which phylogenetic random effects were generated. The ellipses within the scatter plot provide a visual representation of the simulation conditions. A heat map of trait data plotted against the phylogeny highlights the distinct signature each source of structural covariance leaves in the data.

In figure 3A, x and y were simulated with no phylogenetic signal ($\sigma^b = 0$), but a strong positive correlation operating on the independent level ($\rho^b = 0, \rho^e = 0.9$). Clades overlap completely in the scatter plot, together forming a distinguishable positive ellipse. The heatmap shows bands of colour across x and y that appear randomised with respect to the phylogeny (i.e., for any given species, low values of x tend to be observed together with low values of y, but the structure of the phylogeny is not reflected in the distribution of x and y values).

Figure 3B shows the opposing situation, x and y are strongly positively correlated but now this covariance manifests entirely with respect to the phylogenetic structure ($\rho^b = 0.9, \rho^e = 0$) and there is no residual variation in either trait ($\sigma^e = 0$). The scatter plot in Figure 3B shows clearly distinguishable clades, with a tendency for both within- and between-clade correlation. The extent of between-clade correlation (the tendency for clades to arrange along a positive slope) depends on the topology of the tree, with deep splits between major clades promoting separation along the major axis of co-variation. This phenomenon is apparent

across clades in 3B, but also within the green clade: a particularly early split between subclades is reflected in two groups of green points drawn out along a positive slope between x and y. The heatmap in figure 3B shows phylogenetic structure corresponding to the different bivariate means of each clade and subclade. This structure weakens across clades from green, to orange to blue as the topology becomes more deeply nested, i.e., as subclades become less clearly separated in evolutionary time.

In figure 3C, x and y were simulated with equal phylogenetic and independent variances, and a strong positive correlation operating on both levels. This scenario, where phylogenetic and independent correlations show the same sign and similar magnitude, is likely to be common for many pairs of species traits. For example, where selection pressures driving trait correlations over evolutionary time persist into the present day, and adaptation at the species level reflects the same functional relationships between traits. The data in 3C are more variable ($\sigma^b + \sigma^e = 2$) but show similarities to the patterns in both 3A and 3B. For example, clades are distinguishable in the scatterplot but together conform more closely to the independent bivariate density imposed by ρ^e.

In Figure 3D, both traits contain phylogenetic and independent variance, but the correlation between traits is only present on the phylogenetic level. This shows how easily conserved correlations may be obscured when random sources of variation contribute substantially to the total variance in species traits. Notably, figure 3D represents a set of conditions for which SR-PMM, such as PGLS, will typically fail to detect a significant association between x and y (Westoby et al. 2023). As conserved covariances may result from deep functional integration between traits, and can even be the product of selection pressures that continue into the present day, these phylogenetic effects will often be relevant for understanding ongoing ecological and evolutionary associations between traits (Westoby et al. 2023).

Visualizations of simulated data are powerful heuristics for biologists, as they provide tangible representations of the abstract covariance structures we aim to partition with MR-PMM. For more technical coverage of simulating multivariate data containing phylogenetic and independent covariance structure, we include detailed examples and R code in the tutorial.

4 Interpretation

Some researchers argue that we cannot make inferences on evolutionary processes from data collected in the present and that the interpretation of phylogenetic comparative analyses should be limited to patterns of variation, rather than the explicit processes that generated them (Losos, 2008, 2011; Ives, 2018; Revell et al., 2008). Others suggest that the integration of fossil evidence, path analyses, and simulation studies extend our epistemic reach to hypothesis tests of the processes generating observable variation (Uyeda et al., 2018; Quental and Marshall, 2010; Slater et al., 2012; Uyeda and Harmon, 2014). As a result, the biological interpretation of PMM (in particular PGLS), and its utility for addressing specific ecological

and evolutionary hypotheses, has been a subject of lively debate (Freckleton et al., 2002; Westoby et al., 1995; Björklund, 1997; Harvey et al., 1995). Reflecting on this constructive body of work, we believe MR-PMM emerges as a flexible, pluralistic framework for analyses of multiple species traits (Westoby et al., 2023). However, consensus on interpretation is critical to realise the full, operational potential of the method.

Phylogenetic (co)variances estimate the conserved component of species phenotypes given a model of evolution and a hypothesis of the phylogenetic relationship between taxa. The extent to which these conserved effects result from neutral Brownian motion, or reflect constraints and adaptive responses to selection, will vary for different traits, between clades, across phylogenetic scales, and is likely to fluctuate throughout the course of evolutionary history (Revell et al., 2008; Losos, 2008, 2011). As Housworth et al. (2004), point out "[PMM] envisions phenotypic evolution as being the result of a complex of forces including some that are retained over long periods of time, forming patterns in trait variation that reflect the underlying phylogenetic structure, and others that act more quickly, in bursts of change that are lost easily at new speciation events.". Expressed another way, the phylogenetic component of trait (co)variance in PMM will capture all variation attributable to the phylogenetic relationships among taxa. Examples include phenotypic variation arising from genetic differences between clades that have accumulated over evolutionary history, as well as non-genetic (e.g., cultural, environmental or developmental) effects that are phylogenetically structured for one reason or another. Conversely, the independent component will capture variation not attributable to phylogenetic structure, including effects from processes operating exclusively in the present day, as well as effects from historical processes that leave no phylogenetic signature in the data (e.g. because all species are equally affected, or because effects are random with respect to phylogeny). Examples include measurement error, phenotypic plasticity, and rapid evolutionary change following divergence from sister taxa (Housworth et al., 2004). The latter point is particularly interesting as, for the MR case, this means that Σ_{ij}^{ind} may partly reflect nascent phylogenetic covariance between traits i and j, i.e., over subsequent divergences through evolutionary time, such effects may come to reflect the underlying phylogenetic structure, and thus be attributed instead to Σ_{ij}^{phy}. This reinforces a key conceptual point that the distinction between phylogenetic and independent effects in PMM is as much about time as it is about mechanism (Housworth et al., 2004). Indeed, the term 'conservative trait correlation' may be preferable to 'phylogenetic trait correlation', as the former descriptor is mechanism-agnostic, accommodating the possibility that historical selective agents persist into the present day, while the latter implies trait correlation resulting from strictly historical causes (Westoby et al., 2023).

In MR-PMM, the covariance of each trait pair (and its decomposition into phylogenetic and independent components) is the focus of analysis rather than the prediction of one trait as a function of the others. This approach is clearly preferable when there are no *a priori* hypotheses about (unidirectional) causal relationships between traits (Warton et al., 2006). However, recognizing that phylogenetic and independent correlations can arise from multiple distinct causal processes also argues for caution when making mechanistic interpretations. For example, two traits may be phylogenetically correlated for a number of direct, or indirect, reasons: the association may be direct if the traits in question form part of a coordinated

life history strategy in which they functionally covary during adaptive niche diversification, or if conserved genes underlying a set of traits show linkage or pleiotropy, constraining the evolutionary potential for certain trait combinations. Alternatively, the association may be indirect, owing to similar selective pressures acting on both traits independently throughout evolutionary history, such as persistent environmental effects shared by members of a clade displaying phylogenetic niche conservatism. Once again, these mechanistic explanations are not mutually exclusive, as numerous processes may contribute to observed patterns of phylogenetic covariance between traits and across time (Losos, 2011; Revell et al., 2008). Deriving conditional dependencies can help evaluate evidence for alternative causal hypotheses in MR-PMM (see Section 2.3.2). Ultimately, however, it is rarely the pleasure of a comparative biologist to declare causation. Rather, we see the nuanced interpretive framework of MR-PMM as having three strong usage cases for biologists: 1) to provide a more meaningful decomposition of trait (co)variances, including phylogenetic and independent components; 2) to evaluate candidate hypotheses about functional relationships between species traits in a rigorous phylogenetic framework; 3) to generate new hypotheses about the ecological and evolutionary dynamics of trait coordination and niche evolution. With an emerging synthesis of evolutionary and ecological approaches, creative applications of MR-PMM have the potential to revolutionize our understanding of life history evolution, functional trait variation, and community assembly at numerous scales (Gallinat and Pearse, 2021; Smith et al., 2020).

5 Prediction

While the estimation of trait covariance, rather than trait prediction *per se*, is the focus of MR-PMM, the inclusion of traits as multiple responses rather than multiple predictors does not in fact limit our ability to make predictions from the model. On the contrary, the estimated covariance structure can be fully exploited for predictive goals through the use of conditional predictive distributions (described below). This approach actually improves on ordinary multiple regression for prediction as it makes use of the phylogenetic structure in the predictors. The predictive distribution of a fitted model is also useful for model validation, such as posterior-predictive checks, or model comparison using predictive assessment such as cross-validation. We now review conceptual aspects of predictive distributions in MR-PMM, and refer the reader to the accompanying tutorial for worked examples of how these predictive methods can be implemented in R.

5.1 Predicting to new or missing observations

The assumed phylogenetic relationships between taxa and estimated trait correlations from a MR-PMM can be used to predict to new taxa and traits given only partially observed data. To generate this predictive model, the new responses $\boldsymbol{y}_{\text{new}}$ (those to be predicted) are

appended to the observed responses $\boldsymbol{y}_{\text{obs}}$ to define a joint multivariate normal distribution (compare with (2) and (8)),

$$\begin{pmatrix} \boldsymbol{y}_{\text{obs}} \\ \boldsymbol{y}_{\text{new}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \widehat{\boldsymbol{\mu}}_{\text{obs}} \\ \widehat{\boldsymbol{\mu}}_{\text{new}} \end{pmatrix}, \begin{pmatrix} \widehat{\Sigma}_{\text{obs, obs}} & \widehat{\Sigma}_{\text{obs, new}} \\ \widehat{\Sigma}_{\text{new, obs}} & \widehat{\Sigma}_{\text{new, new}} \end{pmatrix} \right), \tag{16}$$

where the set of observed values includes both those used to fit the original PMM and any new observations that have been made. The components $\widehat{\boldsymbol{\mu}}$ and $\widehat{\Sigma}$ depend only on the estimated parameters of the fitted PMM and, if available, an extended phylogenetic covariance matrix that includes any new species. Here we restrict our attention to the Gaussian case, although the technique can be extended to the latent-variable formulation presented in Box 1; see also Ovaskainen and Abrego (2020, Chapter 7) for a detailed exposition in the context of joint species distributions.

It is a standard result from the theory of multivariate normal distributions (e.g., Tong, 2012) that the predictive distribution of the new responses $\boldsymbol{y}_{\text{new}}$ conditional on the observed data $\boldsymbol{y}_{\text{obs}}$ is

$$\boldsymbol{y}_{\text{new}} | \boldsymbol{y}_{\text{obs}} \sim N(\widehat{\boldsymbol{\mu}}'_{\text{new}}, \Sigma'_{\text{new}}) \tag{17}$$

where

$$\widehat{\boldsymbol{\mu}}'_{\text{new}} = \widehat{\boldsymbol{\mu}}_{\text{new}} + \Sigma_{\text{new, obs}} \Sigma^{-1}_{\text{obs, obs}} (\boldsymbol{y}_{\text{obs}} - \hat{\boldsymbol{\mu}}_{\text{obs}}) \tag{18}$$

$$\widehat{\Sigma}'_{\text{new}} = \widehat{\Sigma}_{\text{new, new}} - \widehat{\Sigma}_{\text{new, obs}} \widehat{\Sigma}^{-1}_{\text{obs, obs}} \widehat{\Sigma}_{\text{obs, new}}. \tag{19}$$

The observed responses $\boldsymbol{y}_{\text{obs}}$ enter the conditional model (18) as linear predictors of the new responses which adjusts the prediction from the unconditional mean $\widehat{\boldsymbol{\mu}}_{\text{new}}$ and reduces the predictive variance (19).

Conditional predictive distributions are useful not only in the context of prediction to new data, but also for the imputation of missing data and for cross-validation approaches to predictive assessment (see Section 5.3). We illustrate the use of conditional prediction for predictive assessment in an example analysis using real data in Section 6.

5.2 Model validation using posterior predictive checks

In Bayesian settings, posterior predictive checks use simulated data from the fitted model to test both the adequacy of the fit to the data and the plausibility of the model predictions (Gelman et al., 2013). These checks are typically visual plots that rely on qualitative assessments (Gabry et al., 2019). To test for adequacy of fit, one option is to superpose the observed data onto a plot of the distribution of the simulated data. For a PMM, this type of check could be performed using a separate plot for each trait type with the observed data point for each species plotted on top of a 5 point summary of the predicted distribution

(see Figure 4 for an example). For assessment of model plausibility, the current state of knowledge should be used to evaluate model predictions within ecologically plausible but perhaps unobserved ranges of the included covariates. For example, do predictions for $\boldsymbol{y}_1$ make sense across the entire range of plausible $\boldsymbol{y}_2$ values for a given set of taxa and/or region of interest?

The distinction between marginal and conditional models with respect to the phylogenetic random effects in PMM leads to two corresponding types of predictive distributions. The *marginal predictive distribution* concerns population-level prediction and the *conditional predictive distribution* concerns prediction to specific taxa conditional on the taxon-specific random effect estimate (Figure 4).

5.3 Predictive assessment: cross validation

Cross validation is the use of data splitting to estimate the predictive performance of one or more statistical models, usually for the purpose of model comparison and selection (Yates et al., 2022). Model selection is used when discrete decisions must be made about model structure. For example, whether or not to include various fixed effects (i.e., variable selection) or the choice of probability distribution (e.g., Poisson or negative binomial for count data). Predictive assessment tools such as cross validation are also useful to quantify or simply visualise how well a model can predict to new data (e.g., to new taxon-trait pairs in a PMM) which is distinct from the assessment of model adequacy which concerns prediction of data to which a model was fit.

Cross validation works by fitting each model to a subset of the available data and then assessing the models' predictive capacities on the remaining portion of the data. The splitting procedure is systematically iterated to select different test data and the overall predictive performance is summarised as a cross validation *score* (Arlot and Celisse, 2010). When the measure of predictive performance is the log likelihood of the test data then the predictive assessment is said to be *information theoretic*. Information criteria such as Akaike's Information Criteria (AIC, Akaike, 1973), or for Bayesian analyses the widely applicable information criteria (WAIC, Watanabe, 2010), approximate predictive log likelihood without data splitting by adding a bias correction to the log likelihood of the full data which requires only a single fit for each model. Information criteria are therefore faster to compute than cross-validation scores, however the latter are often preferred as they are less sensitive to violations of model assumptions and are readily combined with techniques to mitigate overfitting (Yates et al., 2021). For Bayesian MR-PMM estimated using Monte Carlo sampling, model fitting is often too slow to permit the use of ordinary cross validation, however recently developed approximate methods provide a rapidly computed and accurate alternative (Vehtari et al., 2017; Bürkner et al., 2021).

6 Example - Leaf Economics in Eucalyptus

To demonstrate potential applications of MR-PMM, we present an example analysis using data on leaf traits of Eucalyptus species from the AusTraits data set (Falster et al., 2021). Our intentions for this analysis were to evaluate evidence for: 1) phylogenetic signal in each trait; and 2) correlations between traits on the phylogenetic and independent level. Full details of the analysis including model specification, diagnostics, validation and inference, are available in the tutorial.

6.1 Methods

We derived the phylogenetic correlation matrix C from the maximum likelihood Eucalypt phylogeny presented in Thornhill (2019), pruning the tree to include only those species for which data were available for three target leaf traits: specific leaf area (SLA); nitrogen content per dry mass of leaf tissue ($\mathbf{N}_{\text{mass}}$); and the ratio of carbon isotopes 12 and 13 in leaf tissue ($\delta\,^{13}\mathbf{C}$). We chose to prune the tree to complete cases for simplicity, acknowledging work that suggests phylogenetic imputation may be a preferable approach to missing data in many circumstances (Debastiani et al. 2021; Penone et al. 2014). We calculated species mean values and associated standard errors for each trait by pooling all observations at the species level. Where transformations to trait data were necessary, these were applied to individual observations, before means and standard errors were calculated, to avoid subsequent approximations of the standard error (Garamszegi 2014, chap 7). Species for which the standard error of a trait could not be calculated because only a single observation was recorded were assigned a standard error equal to the 90th percentile of the standard error estimates for that trait across species. This represents a conservative approach to the assignment of unknown sampling variance in a meta-analytical context (Weir et al., 2018; Ives et al., 2007).

For model validation, we performed marginal and conditional posterior predictive checks. The proportion of data included in the nominal predictive intervals verified the capacity of the model to generate plausible data (Figure 4), although the proportion was lower for N and d13C than for SLA. The shortfall in predictive capacity for these traits may indicate the model is missing relevant covariates or that the chosen probability distributions are less well-suited to these two traits. We took additional steps toward predictive assessment using LOO-CV, but found the approximate method (Bürkner et al., 2021) failed for roughly 5% of the data, necessitating manual LOO re-fits of the model for these observations (see tutorial for details).

6.2 Results and Discussion

Each trait showed strong phylogenetic signal (posterior mean and 95% credible interval for $\hat{h}^2$: SLA = 0.73 (0.59, 0.84), $\mathbf{N}_{\text{mass}}$ = 0.81 (0.65, 0.93), $\delta^{13}\mathbf{C}$ = 0.76 (0.59, 0.88)) and hence the tendency for similar values among closely related species (Figure 6). The traits selected for this analysis are tightly linked to resource-use strategies and leaf economics (Reich, 2014; Wright et al., 2004). For example, Prieto et al. (2018) found that plant species in a Mediterranean woodland showing more resource acquisitive strategies (high SLA and $\mathbf{N}_{\text{mass}}$), were associated with lower water use efficiency (low $\delta^{13}\mathbf{C}$). This is consistent with predictions from leaf economic theory (Wright et al., 2004), that SLA, $\delta^{13}\mathbf{C}$ and $\mathbf{N}_{\text{mass}}$ represent components of a coordinated life history strategy which trade-off predictably as different regions of niche space are explored during species diversification. These co-evolutionary relationships are clearly supported by the fitted model, which reports a positive phylogenetic correlation between SLA and $\mathbf{N}_{\text{mass}}$, and negative phylogenetic correlations between these traits and $\delta^{13}\mathbf{C}$. Thus, in line with theory and empirical observations from other plant groups, these traits appear to co-vary predictably over evolutionary time in Eucalyptus.

A notable result is that for $\mathbf{N}_{\text{mass}}$ and $\delta^{13}\mathbf{C}$, the independent correlation is positive while the phylogenetic correlation is negative (Figure 5). This situation highlights an important strength of MR-PMM; the capacity to identify opposing correlations operating at the phylogenetic and independent level. Indeed, unlike the consistent correlations identified for other trait pairs, contrasting correlations may obscure meaningful relationships in cross-species data (Figure 6; Figure 3D). Given this observation, the subsequent objective is to identify plausible biological mechanisms that might explain these opposing relationships. Carbon fractionation within plant tissues is approximately proportional to the ratio of intercellular to ambient CO_2 mole fractions (ci:ca, see Farquhar et al. 1982). This inequality is strongly influenced by stomatal conductance, because stomatal aperture determines the rate of diffusion of CO_2 into leaves (Farquhar and Sharkey, 1982). As a result, leaf $\delta^{13}\mathbf{C}$ is generally considered a proxy for water use efficiency, with higher values found in species adapted to low soil moisture availability and showing greater stomatal regulation (Cernusak et al., 2013; Diefendorf et al., 2010; Kohn, 2010). This is consistent with the negative phylogenetic correlation observed between $\mathbf{N}_{\text{mass}}$ and $\delta^{13}\mathbf{C}$, i.e., as water use efficiency is necessarily traded-off against growth rate and resource acquisitiveness in the divergence of life history strategies. However, leaf nitrogen content can also influence ci:ca, and hence carbon discrimination, through effects on photosynthetic capacity. Specifically, as variation in $\mathbf{N}_{\text{mass}}$ reflects variation in the carboxylating enzyme Rubisco (Evans and Clarke, 2019), greater $\mathbf{N}_{\text{mass}}$ will be associated with increased demand for CO_2 in the mesophyll, relaxing isotope discrimination and increasing the $\delta^{13}\mathbf{C}$ of leaf tissue (Cernusak et al., 2013). This is consistent with the positive independent correlation observed, i.e. that after accounting for the apparent trade-off between these traits over evolutionary time, species with greater $\mathbf{N}_{\text{mass}}$ tend to have greater $\delta^{13}\mathbf{C}$ due to the influence of photosynthetic capacity on carbon isotope discrimination.

One important consideration for this result, is that variation in leaf nitrogen has different ecophysiological interpretations depending on whether it is expressed on a per mass or per area

basis (Wright et al., 2005; Dong et al., 2020). Specifically, carboxylation capacity is expected to be more strongly related to $\mathbf{N}_{\text{area}}$ than $\mathbf{N}_{\text{mass}}$ (Evans and Clarke, 2019). Interestingly, while $\mathbf{N}_{\text{mass}}$ and $\mathbf{N}_{\text{area}}$ were not correlated in this dataset (Pearson correlation coefficient: r = 0.09, P = 0.11), $\mathbf{N}_{\text{area}}$ was positively correlated with the residuals of $\mathbf{N}_{\text{mass}}$ from the fitted model (r = 0.29, P < 0.001). This suggests that part of the positive independent correlation observed between $\mathbf{N}_{\text{mass}}$ and $\delta^{13}\mathbf{C}$ (Figure 5) represents variation in carboxylation capacity, corroborating our initial interpretation of results.

7 Discussion

7.1 Summary

Mixed models are a familiar, flexible, and scalable framework for comparative analyses of species traits. Phylogenies are used to derive an expectation of similarity among species trait values, which enters the model as a covariance matrix. For a single response trait, modelling phylogenetic structure means the proportion of variance attributable to phylogeny can be estimated. However, SR models with fixed effects may confound an ensemble of correlations if predictor and response variables both show phylogenetic signal. MR-PMM provides a solution via the explicit decomposition of trait covariances. This allows correlations between response traits to be partitioned into phylogenetic and independent contributions, providing a more nuanced understanding of trait relationships. We review these models, with a focus on concepts, implementation and interpretation. Finally, we provide worked examples with supplementary code to help users get started and additional mathematical details for those seeking a more technical understanding.

7.2 Strengths and Weaknesses

An important feature of any phylogenetic comparative method is the assumed model of evolution. By default, PMM assumes the data respect an error structure given by Brownian motion on the supplied tree (and scaled by the residual variance). While Brownian motion will be an adequate model for some species traits (at some phylogenetic scales), others clearly violate its assumptions (Uyeda et al., 2018; Losos, 2011; Freckleton and Harvey, 2006). Many alternative evolutionary models have been proposed, including those with directional trends, stabilising selection, or changes in evolutionary rates over time (Harmon et al., 2010; Blomberg et al., 2003; Pagel, 1999). In SR-PMM, different evolutionary models may be implemented by transformations to the covariance matrix, as in (6), extending (2) to a range of evolutionary models for continuous traits besides Brownian motion (e.g., λ, δ, κ, early burst (EB), Ornstein-Uhlenbeck (OU), ACDC, etc.; Harmon et al. 2010, 2008; Blomberg et al. 2003; Pagel 1999). For the MR case, some transformations are directly

generalizable (e.g., ACDC, EB; Clavel et al. 2015) while others can be implemented only with constraints. For example, MR-PMM can handle a multivariate OU model if attraction to the optima is assumed to be equal and independent for all traits, i.e., the strength of selection matrix $\boldsymbol{A}$ is a multiple of the identity matrix (Goolsby et al., 2017; Tung Ho and Ané, 2014).

Looking beyond PMM, methods based on stochastic differential equations allow for multivariate evolutionary models too complex or dynamic to be expressed as branch-length transformations (Blomberg et al., 2020; Bartoszek et al., 2012; Butler and King, 2004). These include fully parameterised versions of the multivariate OU process, as well as multiple optima and shift models (see below), examples of which are implemented in mvMORPH (Clavel et al. 2015), with some models also estimable via a fast linear-time algorithm in Rphylopars (Goolsby et al., 2017). Like MR-PMM, these methods accommodate missing data and within-species variation, while offering more detailed and potentially more realistic evolutionary models. On the other hand, they do not naturally provide a decomposition of trait covariances and are restricted to continuous traits.

An important limitation of (8), is that it assumes the chosen evolutionary model is both constant through time and homogeneous across the tree, making it vulnerable to miss-specification under certain evolutionary scenarios (Uyeda et al., 2018; Clavel et al., 2015). Some alternative methods allow for shifts, meaning parameters of the evolutionary model (e.g. σ_b^2), even the model itself (Brownian motion, OU, etc), are allowed to vary between clades and across time (Pagel et al., 2022; Mitov et al., 2020, 2019; Clavel et al., 2015), including changes in the trait-level covariance matrix itself through time (i.e., $\Sigma = \Sigma(\theta, t)$) (Blomberg et al., 2020). Currently, it is not clear how these more dynamic models of trait evolution might be expressed in the PMM framework. Thus, while MR-PMM is well suited to ancestral state reconstruction (Section 5), other techniques will be more appropriate when trying to infer the exact timing or phylogenetic position of rare evolutionary events or shifts in the underlying processes generating trait variation (Pagel et al. 2022; Uyeda and Harmon 2014; Slater 2013).

A general strength of mixed models is the use of link functions to include (Gaussian) random effects in the modelled mean of traits that are themselves not Gaussian distributed (Nakagawa and Schielzeth, 2010; Hadfield, 2010). This capability is further exploited in MR-PMM to estimate both phylogenetic and independent correlations between Gaussian and non-Gaussian response variables where covariances are modelled on the link (or latent) scale (see Box 1). The challenge of modelling trait covariances on the link scale is their interpretation. Ideally, trait covariances would be modelled on the observed scale, however observation-scale covariances involving discrete variables are not easily characterised within a joint data distribution. It is possible to transform certain variance components from the link to the observation scale although the non-linearity of typical link functions prohibits an additive decomposition (de Villemereuil et al., 2016). Generalised joint attribute modelling (Popovic et al., 2018; Clark et al., 2017) offers a potential solution to some of these challenges using continuous data imputation and a censored observation model (generalising the binary threshold model) in lieu of classical link functions and discrete distributions.

One factor limiting the application of MR-PMM to some trait data sets is that reliable estimates of covariance components may require relatively large sample sizes, e.g., observations on hundreds of taxa. Though to some extent this will depend on the data and phylogeny at hand, basic power analyses for phylogenetic covariance between two Gaussian traits indicate that performance may be unpredictable for small sample sizes (Housworth et al. 2004). However, a more comprehensive assessment of statistical power and sensitivity in MR-PMM is overdue. In particular, it would be instructive to have clearer expectations around parameter uncertainty with respect to the number of response traits, the mixture of non-Gaussian distributions among responses traits, the relative magnitude of phylogenetic and independent (co)variances, the magnitude of sampling error, as well as sample size and phylogenetic uncertainty. MR-PMM is generally well suited to datasets with many species n relative to the number of traits p (i.e., high $n : p$ ratios). However, alternative methods, or rank-reduced versions, will be more appropriate for higher dimensional data sets with low $n : p$ ratios, such as those generated in studies of geometric morphometrics (Clavel and Morlon, 2020; Adams and Collyer, 2018; Collyer et al., 2015; Adams, 2014). As with any statistical model, the complexity of analyses should be guided by the data that are available (or reasonable to collect) in context of the hypotheses to be addressed.

7.3 Pitfalls and Recommendations

A common pitfall in phylogenetic comparative analyses is the failure to account for within-species variation. This problem arises when omitted variables influence phenotypes at the observation level, adding error to estimates of (co)variance components. In practice, data on all relevant covariates are rarely available and a meta-analytic approach to cross-species analyses is adopted instead. For example, it is common to use species mean trait values pooled from multiple observations or studies as the response vector, and model variation introduced by omitted variables as sampling error around these species mean values (Section 6; also see the tutorial). An alternative approach, assuming sufficient replication at the species level, is to model individual observations of species trait values directly, as opposed to pooled species means with associated standard errors (Felsenstein, 2008). This offers several benefits besides the inclusion of observation-specific covariates. For example, effective separation of phylogenetic and spatial effects (where observations on individual taxa are spread across spatially structured environmental gradients), and the estimation of within-species trait-level residual (co)variances, i.e. patterns of species-specific phenotypic plasticity and trait co-variation (Goolsby et al., 2017). With the current pace of comparative data collection, opportunities to fit and validate models with these detailed covariance structures are increasing.

Despite many advantages, MR-PMM remains a relatively complex and technical approach compared to common alternatives, such as PGLS, or multiple regression. In particular, special care must be taken during model specification to ensure models are structured as intended, as well as in the interpretation of results. Though we touch on each of these considerations here, the following texts covering mixed models in biology (Nakagawa and

Schielzeth, 2010; Wilson et al., 2010; Bolker et al., 2009), and specifically multi-response mixed models (Brommer et al., 2019; de Villemereuil, 2018; Dingemanse and Dochtermann, 2013; Hadfield, 2010; Housworth et al., 2004), are valuable reading.

7.4 Conclusions

1. Species phenotypes are the product of a complex network of causal processes that promote co-selection of traits, over both short and long time scales. For some traits, (co)variation is conserved over evolutionary time, creating patterns in cross-species data that reflect phylogenetic relationships (Section 2).

2. Distinguishing the influence of these conserved effects from those that are decoupled from phylogenetic history is a fundamental objective of evolutionary ecology, because it is the balance of these forces that defines whether constraints, trade-offs and coordinated strategies, should be understood in terms of deep evolutionary integration or labile responses to prevailing conditions (Section 3, 4).

3. Traditional approaches to cross-species analysis, such as PGLS, are not parameterised to distinguish between these components of trait correlation (Section 2), leading to potential confounds and inferential limitations. In contrast, MR-PMM, in which correlations between traits can be partitioned into phylogenetic and independent components, offer a more plausible and informative model of trait evolution under many circumstances (Section 3, 4, 6).

4. The ability to estimate conditional dependencies between traits, accommodate different error distributions (Box 1) and random effect structures, and employ conditional predictions for imputation and model validation (Section 5), makes MR-PMM a unifying approach to many open questions in comparative biology.

5. We expect more researchers will adopt these methods as barriers to implementation for non-specialists are broken down, underscoring the importance of accessible translational research.

8 Tutorial

Tutorial material relevant to this article can be found here.

9 Acknowledgements

We thank Ian Wright, Mark Westoby, Owen Powell, Chris Blackman, and Jonathan Mitchell for helpful comments during preparation of the manuscript. We thank Rachael Gallagher for consultation on the AusTraits database. This work was funded by The Australian Research Council Centre of Excellence for Plant Success in Nature and Agriculture (CE200100015).

References

Adams, D. C. (2014). A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. *Evolution 68*(9), 2675–2688.

Adams, D. C. and M. L. Collyer (2018). Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. *Systematic biology 67*(1), 14–31.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Ed.), *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pp. 267–281. Academiai Kiado, Budapest.

Arlot, S. and A. Celisse (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys 4*, 40–79.

Bartoszek, K., J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen (2012). A phylogenetic comparative method for studying multivariate adaptation. *Journal of Theoretical Biology 314*, 204–215.

Björklund, M. (1997). Are 'comparative methods' always necessary? *Oikos 80*(3), 607–612.

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

Blomberg, S. P., J. G. Lefevre, J. A. Wells, and M. Waterhouse (2012, January). Independent contrasts and PGLS regression estimators are equivalent. *Systematic Biology 61*(3), 382–391.

Blomberg, S. P., S. I. Rathnayake, and C. M. Moreau (2020). Beyond brownian motion and the ornstein-uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. *The American Naturalist 195*(2), 145–165.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution 24*(3), 127–135.

Brommer, J. E., B. Class, and G. Covarrubias-Pazaran (2019). Multivariate mixed models in ecology and evolutionary biology: inference and implementation in r. *EcoEvoRxiv Prepr 10*.

Bruelheide, H., J. Dengler, O. Purschke, J. Lenoir, B. Jiménez-Alfaro, S. M. Hennekens, Z. Botta-Dukát, M. Chytrỳ, R. Field, F. Jansen, et al. (2018). Global trait–environment relationships of plant communities. *Nature ecology & evolution 2*(12), 1906–1917.

Bürkner, P.-C. (2017). brms: An r package for bayesian multilevel models using stan. *Journal of statistical software 80*, 1–28.

Bürkner, P.-C., J. Gabry, and A. Vehtari (2021, 11). Efficient leave-one-out cross-validation for Bayesian non-factorized normal and Student-t models. *Computational Statistics 36*(2), 1243–1261.

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

Cássia-Silva, C., C. G. Freitas, L. P. Lemes, G. B. Paterno, P. A. Dias, C. D. Bacon, and R. G. Collevatti (2020). Higher evolutionary rates in life-history traits in insular than in mainland palms. *Scientific reports 10*(1), 1–10.

Cernusak, L. A., N. Ubierna, K. Winter, J. A. Holtum, J. D. Marshall, and G. D. Farquhar (2013). Environmental and physiological determinants of carbon isotope discrimination in terrestrial plants. *New Phytologist 200*(4), 950–965.

Clark, J. S., D. Nemergut, B. Seyednasrollah, P. J. Turner, and S. Zhang (2017). Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecological Monographs 87*(1), 34–56.

Clavel, J., G. Escarguel, and G. Merceron (2015). mvmorph: an r package for fitting multivariate evolutionary models to morphometric data. *Methods in Ecology and Evolution 6*(11), 1311–1319.

Clavel, J. and H. Morlon (2020). Reliable phylogenetic regressions for multivariate comparative data: illustration with the manova and application to the effect of diet on mandible morphology in phyllostomid bats. *Systematic biology 69*(5), 927–943.

Collyer, M. L., D. J. Sekora, and D. C. Adams (2015). A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity 115*(4), 357–365.

Cooper, N., G. H. Thomas, and R. G. FitzJohn (2016). Shedding light on the 'dark side'of phylogenetic comparative methods. *Methods in Ecology and Evolution 7*(6), 693–699.

de Villemereuil, P. (2018). Quantitative genetic methods depending on the nature of the phenotypic trait. *Annals of the New York Academy of Sciences 1422*(1), 29–47.

de Villemereuil, P., H. Schielzeth, S. Nakagawa, and M. Morrissey (2016, 11). General Methods for Evolutionary Quantitative Genetic Inference from Generalized Mixed Models. *Genetics 204*(3), 1281–1294.

Debastiani, V. J., V. A. Bastazini, and V. D. Pillar (2021). Using phylogenetic information to impute missing functional trait values in ecological databases. *Ecological Informatics 63*, 101315.

Díaz, S., J. Kattge, J. H. Cornelissen, I. J. Wright, S. Lavorel, S. Dray, B. Reu, M. Kleyer, C. Wirth, I. Colin Prentice, et al. (2016). The global spectrum of plant form and function. *Nature 529*(7585), 167–171.

Diefendorf, A. F., K. E. Mueller, S. L. Wing, P. L. Koch, and K. H. Freeman (2010). Global patterns in leaf 13c discrimination and implications for studies of past and future climate. *Proceedings of the National Academy of Sciences 107*(13), 5738–5743.

Dingemanse, N. J. and N. A. Dochtermann (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology 82*(1), 39–54.

Dong, N., I. C. Prentice, I. J. Wright, B. J. Evans, H. F. Togashi, S. Caddy-Retalic, F. A. McInerney, B. Sparrow, E. Leitch, and A. J. Lowe (2020). Components of leaf-trait variation along environmental gradients. *New Phytologist 228*(1), 82–94.

Epskamp, S., L. J. Waldorp, R. Mõttus, and D. Borsboom (2018). The gaussian graphical model in cross-sectional and time-series data. *Multivariate behavioral research 53*(4), 453–480.

Evans, J. R. and V. C. Clarke (2019). The nitrogen cost of photosynthesis. *Journal of Experimental Botany 70*(1), 7–15.

Falster, D., R. Gallagher, E. H. Wenk, I. J. Wright, D. Indiarto, S. C. Andrew, C. Baxter, J. Lawson, S. Allen, A. Fuchs, et al. (2021). Austraits, a curated plant trait database for the australian flora. *Scientific Data 8*(1), 1–20.

Farquhar, G. D., M. H. O'Leary, and J. A. Berry (1982). On the relationship between carbon isotope discrimination and the intercellular carbon dioxide concentration in leaves. *Functional Plant Biology 9*(2), 121–137.

Farquhar, G. D. and T. D. Sharkey (1982). Stomatal conductance and photosynthesis. *Annual review of plant physiology 33*(1), 317–345.

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

Felsenstein, J. (2008). Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *The American Naturalist 171*(6), 713–725.

Freckleton, R. (2009). The seven deadly sins of comparative analysis. *Journal of evolutionary biology 22*(7), 1367–1375.

Freckleton, R. P. and P. H. Harvey (2006). Detecting non-brownian trait evolution in adaptive radiations. *PLoS biology 4*(11), e373.

Freckleton, R. P., P. H. Harvey, and M. Pagel (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist 160*(6), 712–726.

Gabry, J., D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman (2019, January). Visualization in bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 182*(2), 389–402.

Gallinat, A. S. and W. D. Pearse (2021). Phylogenetic generalized linear mixed modeling presents novel opportunities for eco-evolutionary synthesis. *Oikos 130*(5), 669–679.

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

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013, 11). *Bayesian Data Analysis* (Third ed.). Chapman and Hall/CRC.

Goolsby, E. W., J. Bruggeman, and C. Ané (2017). Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods in Ecology and Evolution 8*(1), 22–27.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences 326*(1233), 119–157.

Green, S. J., C. B. Brookson, N. A. Hardy, and L. B. Crowder (2022). Trait-based approaches to global change ecology: moving from description to prediction. *Proceedings of the Royal Society B 289*(1971), 20220071.

Grossnickle, D. M. (2020). Feeding ecology has a stronger evolutionary influence on functional morphology than on body mass in mammals. *Evolution 74*(3), 610–628.

Haba, Y. and N. Kutsukake (2019). A multivariate phylogenetic comparative method incorporating a flexible function between discrete and continuous traits. *Evolutionary Ecology 33*, 751–768.

Hadfield, J. D. (2010). Mcmc methods for multi-response generalized linear mixed models: the mcmcglmm r package. *Journal of statistical software 33*, 1–22.

Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman, W. Bryan Jennings, K. H. Kozak, M. A. McPeek, F. Moreno-Roark, T. J. Near, et al. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution 64*(8), 2385–2396.

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger (2008). Geiger: investigating evolutionary radiations. *Bioinformatics 24*(1), 129–131.

Harvey, P. H., A. F. Read, and S. Nee (1995). Why ecologists need to be phylogenetically challenged. *Journal of Ecology 83*(3), 535–536.

Hawkins, B. A., J. A. F. Diniz-Filho, C. A. Jaramillo, and S. A. Soeller (2007). Climate, niche conservatism, and the global bird diversity gradient. *the american naturalist 170*(S2), S16–S27.

Herberstein, M. E., D. J. McLean, E. Lowe, J. O. Wolff, M. K. Khan, K. Smith, A. P. Allen, M. Bulbert, B. A. Buzatto, M. D. Eldridge, et al. (2022). Animaltraits-a curated animal trait database for body mass, metabolic rate and brain size. *Scientific Data 9*(1), 265.

Hof, C., C. Rahbek, and M. B. Araújo (2010). Phylogenetic signals in the climatic niches of the world's amphibians. *Ecography 33*(2), 242–250.

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

Ives, A. R. (2018). Mixed and phylogenetic models: a conceptual introduction to correlated data. *See https://leanpub.com/correlateddata*.

Ives, A. R. (2019). R s for correlated data: Phylogenetic models, lmms, and glmms. *Systematic Biology 68*(2), 234–251.

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

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

Ives, A. R., P. E. Midford, and T. Garland Jr (2007). Within-species variation and measurement error in phylogenetic comparative methods. *Systematic biology 56*(2), 252–270.

Jansen, S., B. Schuldt, and B. Choat (2015). Current controversies and challenges in applying plant hydraulic techniques. *New Phytologist 205*(3), 961–964.

Kattge, J., G. Bönisch, S. Díaz, S. Lavorel, I. C. Prentice, P. Leadley, S. Tautenhahn, G. D. Werner, T. Aakala, M. Abedi, et al. (2020). Try plant trait database–enhanced coverage and open access. *Global change biology 26*(1), 119–188.

Kohn, M. J. (2010). Carbon isotope compositions of terrestrial c3 plants as indicators of (paleo) ecology and (paleo) climate. *Proceedings of the National Academy of Sciences 107*(46), 19691–19695.

Kozak, K. H. and J. J. Wiens (2010). Accelerated rates of climatic-niche evolution underlie rapid species diversification. *Ecology letters 13*(11), 1378–1389.

Kruuk, L. E. and J. D. Hadfield (2007). How to separate genetic and environmental causes of similarity between relatives. *Journal of evolutionary biology 20*(5), 1890–1903.

Kubota, Y., B. Kusumoto, T. Shiono, and T. Tanaka (2017). Phylogenetic properties of tertiary relict flora in the east asian continental islands: imprint of climatic niche conservatism and in situ diversification. *Ecography 40*(3), 436–447.

Laumer, C. E., R. Fernández, S. Lemer, D. Combosch, K. M. Kocot, A. Riesgo, S. C. Andrade, W. Sterrer, M. V. Sørensen, and G. Giribet (2019). Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proceedings of the royal society B 286*(1906), 20190831.

Liu, H., Q. Ye, and J. J. Wiens (2020). Climatic-niche evolution follows similar rules in plants and animals. *Nature Ecology & Evolution 4*(5), 753–763.

Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology letters 11*(10), 995–1003.

Losos, J. B. (2011). Seeing the forest for the trees: The limitations of phylogenies in comparative biology: (american society of naturalists address). *The American Naturalist 177*(6), 709–727.

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

Marques, I., T. Kneib, and N. Klein (2022). Mitigating spatial confounding by explicitly correlating gaussian random fields. *Environmetrics 33*(5), e2727.

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

Mitov, V., K. Bartoszek, G. Asimomitis, and T. Stadler (2020). Fast likelihood calculation for multivariate gaussian phylogenetic models with shifts. *Theoretical Population Biology 131*, 66–78.

Mitov, V., K. Bartoszek, and T. Stadler (2019). Automatic generation of evolutionary hypotheses using mixed gaussian phylogenetic models. *Proceedings of the National Academy of Sciences 116*(34), 16921–16926.

Münkemüller, T., S. Lavergne, B. Bzeznik, S. Dray, T. Jombart, K. Schiffers, and W. Thuiller (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution 3*(4), 743–756.

Nakagawa, S., P. C. D. Johnson, and H. Schielzeth (2017, 9). The coefficient of determination r^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface 14*, 20170213.

Nakagawa, S. and H. Schielzeth (2010). Repeatability for gaussian and non-gaussian data: a practical guide for biologists. *Biological Reviews 85*(4), 935–956.

Ovaskainen, O. and N. Abrego (2020). *Joint Species Distribution Modelling: With Applications in R*. Ecology, Biodiversity and Conservation. Cambridge University Press.

Ovaskainen, O., G. Tikhonov, A. Norberg, F. Guillaume Blanchet, L. Duan, D. Dunson, T. Roslin, and N. Abrego (2017). How to make more out of community data? a conceptual framework and its implementation as models and software. *Ecology letters 20*(5), 561–576.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature 401*(6756), 877–884.

Pagel, M., C. O'Donovan, and A. Meade (2022). General statistical model shows that macroevolutionary patterns and processes are consistent with darwinian gradualism. *Nature communications 13*(1), 1–12.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika 82*(4), 669–688.

Peixoto, F. P., F. Villalobos, and M. V. Cianciaruso (2017). Phylogenetic conservatism of climatic niche in bats. *Global Ecology and Biogeography 26*(9), 1055–1065.

Penone, C., A. D. Davidson, K. T. Shoemaker, M. Di Marco, C. Rondinini, T. M. Brooks, B. E. Young, C. H. Graham, and G. C. Costa (2014). Imputation of missing data in life-history trait datasets: Which approach performs the best? *Methods in Ecology and Evolution 5*(9), 961–970.

Popovic, G. C., F. K. Hui, and D. I. Warton (2018). A general algorithm for covariance modeling of discrete data. *Journal of Multivariate Analysis 165*, 86–100.

Popovic, G. C., D. I. Warton, F. J. Thomson, F. K. Hui, and A. T. Moles (2019). Untangling direct species associations from indirect mediator species effects with graphical models. *Methods in Ecology and Evolution 10*(9), 1571–1583.

Prieto, I., J. I. Querejeta, J. Segrestin, F. Volaire, and C. Roumet (2018). Leaf carbon and oxygen isotopes are coordinated with the leaf economics spectrum in mediterranean rangeland species. *Functional Ecology 32*(3), 612–625.

Quental, T. B. and C. R. Marshall (2010). Diversity dynamics: molecular phylogenies need the fossil record. *Trends in ecology & evolution 25*(8), 434–441.

Reich, P. B. (2014). The world-wide 'fast–slow' plant economics spectrum: a traits manifesto. *Journal of ecology 102*(2), 275–301.

Revell, L. J. (2010). Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution 1*(4), 319–329.

Revell, L. J., L. J. Harmon, and D. C. Collar (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic biology 57*(4), 591–601.

Rohlf, F. J. (2001). Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution 55*(11), 2143–2160.

Sanchez-Martinez, P., J. Martínez-Vilalta, K. G. Dexter, R. A. Segovia, and M. Mencuccini (2020). Adaptation and coordinated evolution of plant hydraulic traits. *Ecology Letters 23*(11), 1599–1610.

Slater, G. J. (2013). Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the c retaceous-p alaeogene boundary. *Methods in Ecology and Evolution 4*(8), 734–744.

Slater, G. J., L. J. Harmon, and M. E. Alfaro (2012). Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution: International Journal of Organic Evolution 66*(12), 3931–3944.

Smith, S. A. and J. M. Beaulieu (2009). Life history influences rates of climatic niche evolution in flowering plants. *Proceedings of the Royal Society B: Biological Sciences 276*(1677), 4345–4352.

Smith, S. A. and J. W. Brown (2018). Constructing a broadly inclusive seed plant phylogeny. *American journal of botany 105*(3), 302–314.

Smith, S. D., M. W. Pennell, C. W. Dunn, and S. V. Edwards (2020). Phylogenetics is the new genetics (for most of biodiversity). *Trends in Ecology & Evolution 35*(5), 415–425.

Symonds, M. R. and S. P. Blomberg (2014). A primer on phylogenetic generalised least squares. In *Modern phylogenetic comparative methods and their application in evolutionary biology*, pp. 105–130. Springer.

Thornhill, A. H., M. D. Crisp, C. Külheim, K. E. Lam, L. A. Nelson, D. K. Yeates, and J. T. Miller (2019). A dated molecular perspective of eucalypt taxonomy, evolution and diversification. *Australian Systematic Botany 32*(1), 29–48.

Tong, Y. (2012). *The Multivariate Normal Distribution*. Springer Series in Statistics. Springer New York.

Tung Ho, L. s. and C. Ané (2014). A linear-time algorithm for gaussian and non-gaussian trait evolution models. *Systematic biology 63*(3), 397–408.

Uyeda, J. C., D. S. Caetano, and M. W. Pennell (2015). Comparative analysis of principal components can be misleading. *Systematic Biology 64*(4), 677–689.

Uyeda, J. C. and L. J. Harmon (2014). A novel bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic biology 63*(6), 902–918.

Uyeda, J. C., R. Zenil-Ferguson, and M. W. Pennell (2018). Rethinking phylogenetic comparative methods. *Systematic Biology 67*(6), 1091–1109.

Vehtari, A., A. Gelman, and J. Gabry (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing 27*(5), 1413–1432.

Warton, D. I. (2022). *Eco-Stats: Data Analysis in Ecology: From t-tests to Multivariate Abundances*. Springer Nature.

Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. Hui (2015). So many variables: joint modeling in community ecology. *Trends in ecology & evolution 30*(12), 766–779.

Warton, D. I., I. J. Wright, D. S. Falster, and M. Westoby (2006). Bivariate line-fitting methods for allometry. *Biological reviews 81*(2), 259–291.

Watanabe, S. (2010, dec). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res. 11*, 3571–3594.

Weir, C. J., I. Butcher, V. Assi, S. C. Lewis, G. D. Murray, P. Langhorne, and M. C. Brady (2018). Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. *BMC medical research methodology 18*(1), 1–14.

Westoby, M., M. R. Leishman, and J. M. Lord (1995). On misinterpreting the 'phylogenetic correction'. *Journal of Ecology 83*(3), 531–534.

Westoby, M., L. Yates, B. Holland, and B. Halliwell (2023). Phylogenetically conservative trait correlation: quantification and interpretation. *Journal of Ecology*.

Wilson, A. (2008). Why h2 does not always equal va/vp? *Journal of evolutionary biology 21*(3), 647–650.

Wilson, A. J., D. Reale, M. N. Clements, M. M. Morrissey, E. Postma, C. A. Walling, L. E. Kruuk, and D. H. Nussey (2010). An ecologist's guide to the animal model. *Journal of animal ecology 79*(1), 13–26.

Wright, I. J., P. B. Reich, J. H. Cornelissen, D. S. Falster, P. K. Groom, K. Hikosaka, W. Lee, C. H. Lusk, Ü. Niinemets, J. Oleksyn, et al. (2005). Modulation of leaf economic traits and trait relationships by climate. *Global Ecology and Biogeography 14*(5), 411–421.

Wright, I. J., P. B. Reich, M. Westoby, D. D. Ackerly, Z. Baruch, F. Bongers, J. Cavender-Bares, T. Chapin, J. H. Cornelissen, M. Diemer, et al. (2004). The worldwide leaf economics spectrum. *Nature 428*(6985), 821–827.

Yates, L. A., Z. Aandahl, S. A. Richards, and B. W. Brook (2022). Cross validation for model selection: a review with examples from ecology. *Ecological Monographs*, e1557. e1557 ECM22-0048.R1.

Yates, L. A., S. A. Richards, and B. W. Brook (2021). Parsimonious model selection using information theory: a modified selection rule. *Ecology 102*(10), e03475.

Yeates, D. K., K. Meusemann, M. Trautwein, B. Wiegmann, and A. Zwick (2016). Power, resolution and bias: recent advances in insect phylogeny driven by the genomic revolution. *Current Opinion in Insect Science 13*, 16–23.

Young, A. D. and J. P. Gillung (2020). Phylogenomics—principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology 45*(2), 225–247.

Yuan, M. and Y. Lin (2007). Model selection and estimation in the gaussian graphical model. *Biometrika 94*(1), 19–35.

Box 1. Including diverse trait types

The MR-PMM formulation (8) can be adapted to include trait types that are not appropriately modelled using a Gaussian distribution; for example, discrete counts or strictly positive values. To simultaneously include various trait types in a MR-PMM, a latent-variable formulation can be used to model between-trait covariance in the usual way via a multivariate Gaussian distribution while at the same time permitting the use of trait-specific probability distributions and link functions (Hadfield, 2010). For traits $i = 1, .., k$, the latent-variable formulation is written

$$\begin{pmatrix} \boldsymbol{l}_1 \\ \boldsymbol{l}_2 \\ \vdots \\ \boldsymbol{l}_k \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 + \mathbf{b}_1 + \mathbf{e}_1 \\ \boldsymbol{\mu}_2 + \mathbf{b}_2 + \mathbf{e}_2 \\ \vdots \\ \boldsymbol{\mu}_k + \mathbf{b}_k + \mathbf{e}_k \end{pmatrix}, \tag{20}$$

where $\boldsymbol{l}_i$ are the latent variables and $\mathbf{b}$ and $\mathbf{e}$ are drawn from multivariate normal distributions characterised by the phylogenetic (Σ^{phy}) and independent (Σ^{ind}) trait-correlation structures

$$(\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_k)^{\text{T}} \sim \mathcal{N}(0, \Sigma^{\text{phy}} \otimes C)$$
$$(\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k)^{\text{T}} \sim \mathcal{N}(0, \Sigma^{\text{ind}} \otimes I).$$

A model for the observations $\boldsymbol{y}_i$ is given in terms of a trait-specific probability distribution f_i and link function η_i

$$\boldsymbol{y}_i \sim f_i(\eta_i^{-1}(\boldsymbol{l}_i), \boldsymbol{\phi}_i), \tag{21}$$

where $\boldsymbol{\phi}_i$ includes any distribution-specific parameters (if required). For example, count data might be modelled using a negative binomial distribution with log link function and overdispersion ϕ. For the formulation (21), the joint likelihood of the trait observations and the latent variables is

$$p(\boldsymbol{y}, \boldsymbol{l}) = \prod_{i=1}^{k} p(\boldsymbol{y}_i \mid \boldsymbol{l}_i, \phi_i)\, p(\boldsymbol{l} \mid \Sigma^{\text{phy}}, \Sigma^{\text{ind}}). \tag{22}$$

Depending on the choice of distribution f_i, additional constraints must be imposed to ensure parameter identifiability. For example, using a probit threshold model for binary data, the mean μ_ε and variance σ^2_ε of the associated normal distribution must be fixed (typically to define a cumulative standard normal distribution) as well as the corresponding independent variance component Σ^{ind}_{ii} (typically fixed to one). It should be noted that these conventional, yet arbitrary, constraint choices re-scale the other parameters estimates accordingly. If f_i is Gaussian, then $\phi_i = \sigma_i$ must be fixed to make Σ^{ind}_{ii} identifiable.

Partitioning of the variance to define an analogue of h^2 (5) is possible for latent-variable models, but is more nuanced because there is a distinction between latent- and observation-scale partitioning where distribution-specific variances must be taken into account (Nakagawa et al., 2017). For example, using the above-stated constraints (i.e., $\sigma^2_e = \sigma^2_\varepsilon = 1$) in a probit model for a binary trait, the proportion of variance attributable to phylogeny on the scale of the threshold variable is

$$h^2_{\text{probit}} = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2_e + \sigma^2_\varepsilon} = \frac{\sigma^2_b}{\sigma^2_b + 2}. \tag{23}$$

Alternatives to the probit model include models based on the logistic and Gumbel distributions for which the variances are $\sigma^2_\varepsilon = \pi^2/3$ and $\sigma^2_\varepsilon = \pi^2/6$, respectively (see tutorial for a worked example that includes a logistic threshold model).

Variance terms are usually more easily interpreted when they are expressed on the scale of trait observations rather than that of the latent variables. For multi-response models involving non-Gaussian distributions there are generally no closed-form solutions for the observation-level variances and associated statistics, however numerical methods are readily applied given point estimates or a set of posterior samples of the model parameters (de Villemereuil et al., 2016).

Figure captions

Figure 1. A) Tree and matrix representations of the total covariance structure Σ for data collected using a nested sampling design ($\sigma_b^2 = 3$ and $\sigma_e^2 = 2$) B) Tree and matrix representations of the total covariance structure, Σ, among species trait values ($\sigma_b^2 = 4$ and $\sigma_e^2 = 1$). Blue and red text used to emphasise the translation of phylogenetic (co)variance and independent error variance respectively to the graphs on the left. Independent variance represents extensions to the terminal branches of the graph (diagonal elements in I), reducing the overall expectation of similarity between observations.

Figure 2. Correlations between species traits x, y and z (upper left), with the same information expressed qualitatively as a graph (lower left). To evaluate whether the relationship between x and y can be explained by a covariate z, we compute partial correlations from the precision matrix of trait covariances, which quantifies the relationship between x and y when controlling for z (upper right). In this example, despite strong covariance at the phenotypic level (upper left), the relationship between x and y is fully explained by a shared response to the covariate z.

Figure 3. Bivariate trait data simulated from MR-PMM containing different levels of phylogenetic and independent (co)variance (A-D). Simulation conditions for each data set are inset in scatterplots. Trait data are plotted in scatterplots, as well as heatmaps arranged against the generating phylogeny.

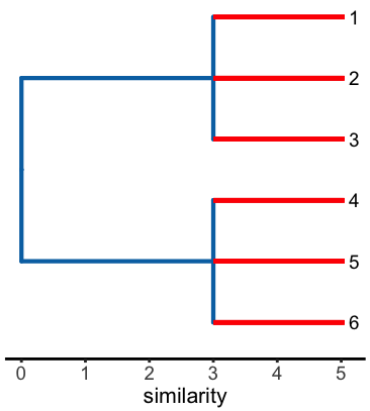
Figure 4. Posterior predictive checks from a fitted MR-PMM of leaf traits in Eucalyptus. Dark blue points represent observed species values. Five-point summaries (credible intervals: 0.025, 0.25, 0.5, 0.75, 0.975) in light blue show the posterior predictive distribution for each taxon. For the marginal model, predictions are made at the population level by marginalising over the distribution of phylogenetic random effects. For the conditional model, predictions are made at the taxon level by conditioning on the estimated phylogenetic random effects, which significantly reduces predictive uncertainty. For each trait and plot type, the taxon are ordered by the predictive mean of the conditional model.

Figure 5. Phylogenetic (blue) and independent (red) correlation coefficient estimates from MR-PMM of leaf traits in Eucalyptus. Points represent the posterior median for each estimate, with 50% and 95% CIs represented by heavy and light wicks respectively.

Figure 6. Scatter-plots (left) and heat-maps (right) of three leaf traits across 306 species of Eucalyptus. For heat-maps, values have been centred and scaled for each trait and are aligned with the corresponding species in the phylogeny (centre). For $\delta^{13}\mathbf{C}$ and $\mathbf{N}_{\text{mass}}$ (bottom left), opposing phylogenetic and independent correlations (Figure 4) are obscured at the level of species phenotypes.

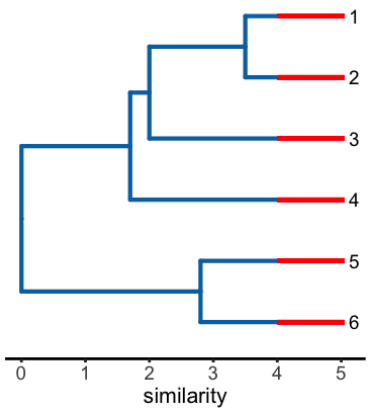
Figures

A



$$\rightarrow \Sigma = \sigma_{\mathbf{b}}^{\mathbf{2}} C + \sigma_{\mathbf{e}}^{\mathbf{2}} I = \begin{pmatrix} \mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 & 0 \\ \mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 & 0 \\ \mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{3} & \mathbf{3} & \mathbf{3} \\ 0 & 0 & 0 & \mathbf{3} & \mathbf{3} & \mathbf{3} \\ 0 & 0 & 0 & \mathbf{3} & \mathbf{3} & \mathbf{3} \end{pmatrix} + \begin{pmatrix} \mathbf{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{2} \end{pmatrix}$$

B

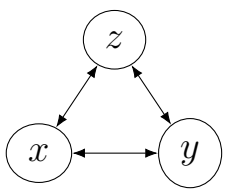


$$\rightarrow \Sigma = \sigma_{\mathbf{b}}^{\mathbf{2}} C + \sigma_{\mathbf{e}}^{\mathbf{2}} I = \begin{pmatrix} \mathbf{4} & \mathbf{3.5} & \mathbf{2} & \mathbf{1.7} & 0 & 0 \\ \mathbf{3.5} & \mathbf{4} & \mathbf{2} & \mathbf{1.7} & 0 & 0 \\ \mathbf{2} & \mathbf{2} & \mathbf{4} & \mathbf{1.7} & 0 & 0 \\ \mathbf{1.7} & \mathbf{1.7} & \mathbf{1.7} & \mathbf{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{4} & \mathbf{2.8} \\ 0 & 0 & 0 & 0 & \mathbf{2.8} & \mathbf{4} \end{pmatrix} + \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} \end{pmatrix}$$

Figure 1

correlation

$$\begin{array}{c c} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{pmatrix} 1 & 0.56 & 0.75 \\ 0.56 & 1 & 0.75 \\ 0.75 & 0.75 & 1 \end{pmatrix} \end{array}$$



partial

$$\begin{array}{c c} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{pmatrix} 1 & 0 & 0.6 \\ 0 & 1 & 0.6 \\ 0.6 & 0.6 & 1 \end{pmatrix} \end{array}$$

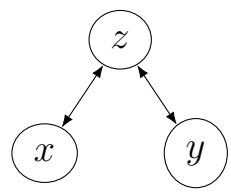


Figure 2

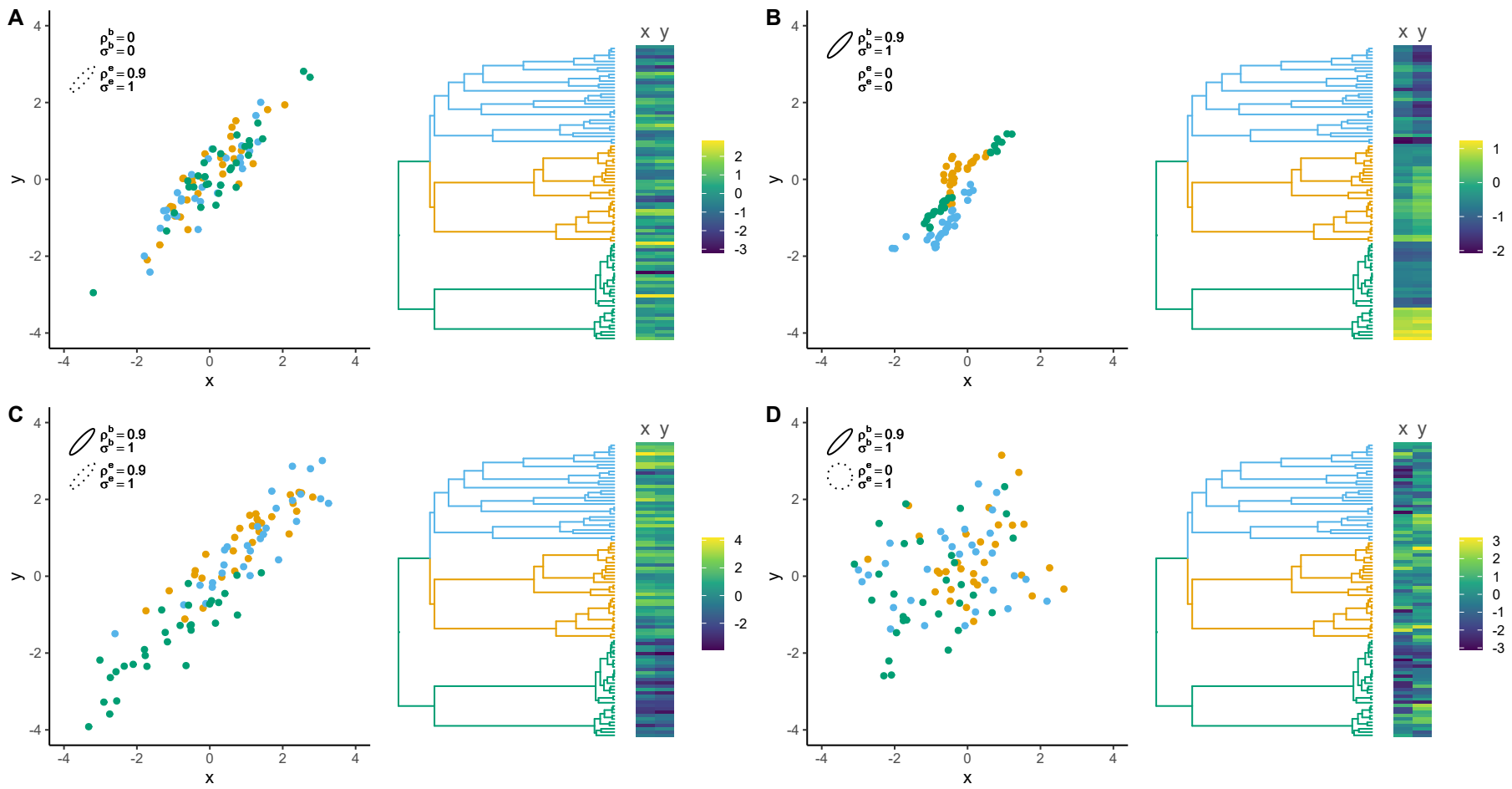


Figure 3

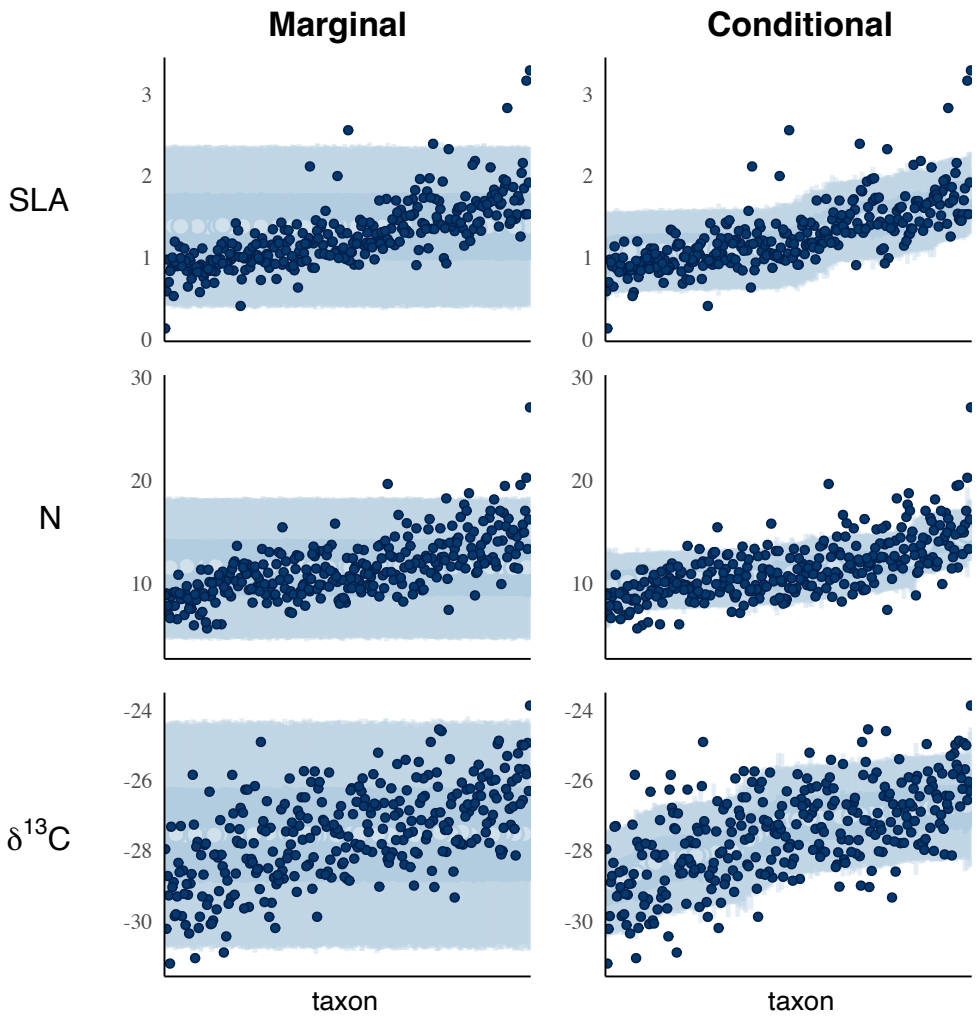

Marginal
Conditional
SLA
N
δ13C
3
2
1
0
30
20
10
-24
-26
-28
-30
taxon
taxon

Figure 4

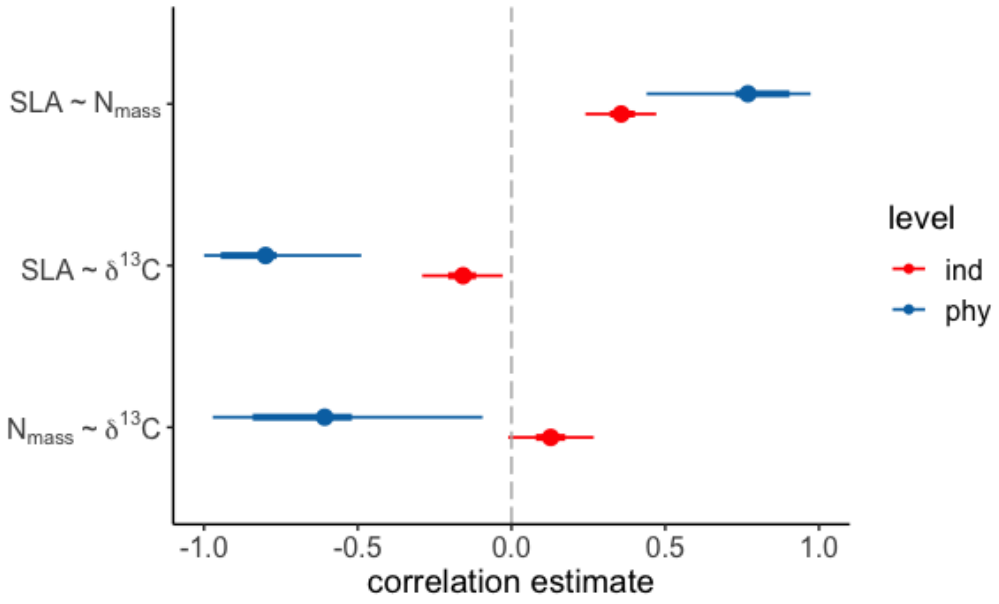

SLA ~ N_{mass}
SLA ~ $\delta^{13}C$
N_{mass} ~ $\delta^{13}C$
-1.0
-0.5
0.0
0.5
1.0
correlation estimate
level
ind
phy

Figure 5

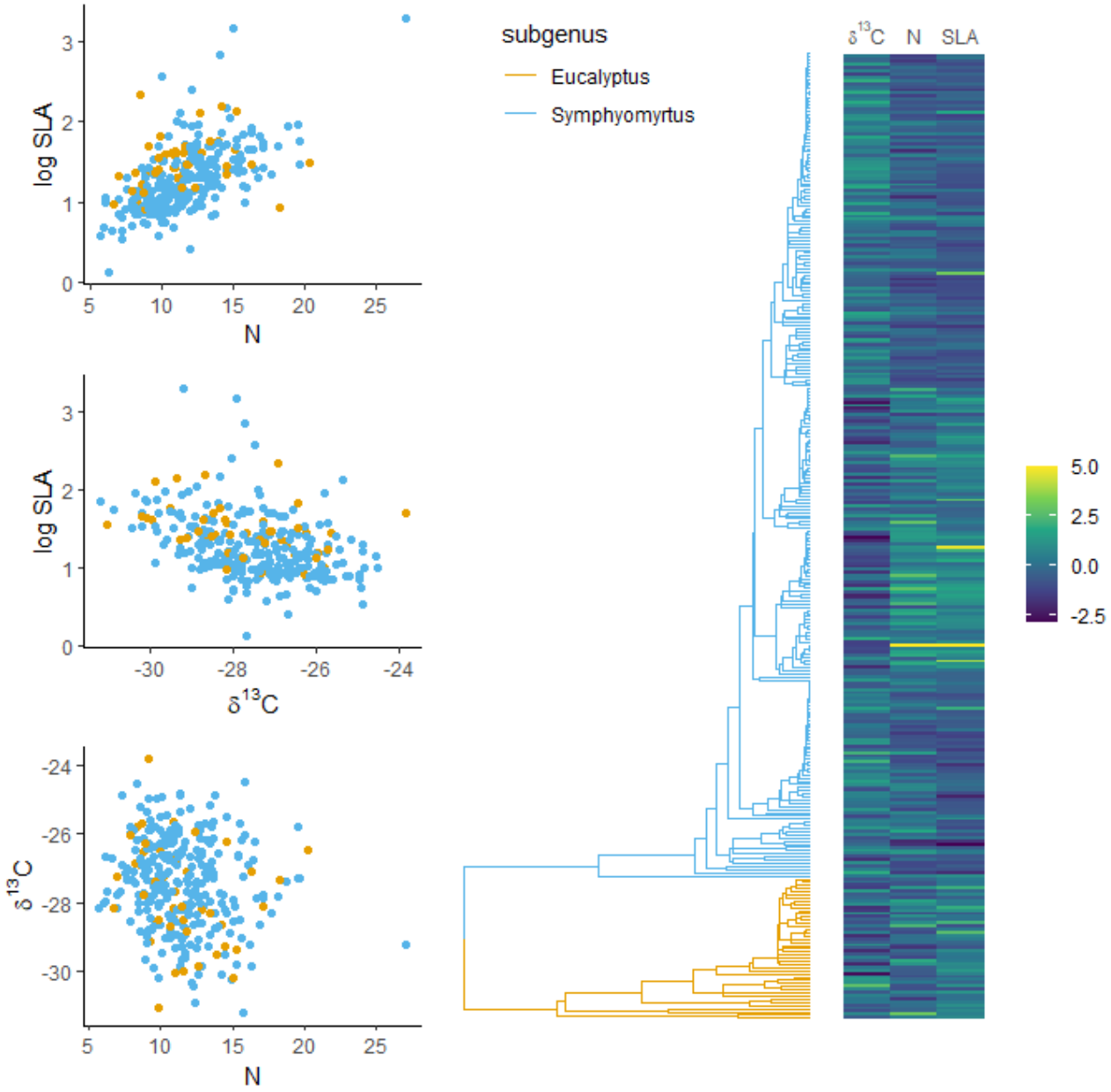

log SLA
N
log SLA
δ13C
δ13C
N
subgenus
Eucalyptus
Symphyomyrtus
δ13C
N
SLA
5.0
2.5
0.0
-2.5

Figure 6