Multi-Response Phylogenetic Mixed Models: Concepts and Application

Ben Halliwell, Luke A. Yates, Barbara R. Holland

December 2022

Abstract

The scale and resolution of trait databases and molecular phylogenies is increasing rapidly. These resources permit many 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 analyses of trait evolution. While flexible and powerful, these models can be conceptually challenging. The literature describing PMM is also highly technical, creating additional barriers for many biologists. Here we present an accessible and practical guide to the PMM. We begin by outlining key concepts in mixed modelling, emphasising the use of covariance matrices to model correlative structure in the data. We present a simple notation for PMM, of which phylogenetic generalised least squares (PGLS) is a special case. We outline the limitations of single-response (SR) models such as PGLS for characterising patterns of trait variation, and emphasise that MR models will often be a preferable approach to analyses involving multiple species traits. Using simulated data and visual examples, we demonstrate the capacity of MR-PMM to partition trait covariance into phylogenetic and independent components. 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 offer a more nuanced understanding of trait correlations compared to SR approaches. Finally, we highlight common pitfalls and offer practical recommendations to analysts. To complement this material, we provide an online tutorial including additional examples, as well as side-by-side implementations in two popular R packages, `MCMCglmm` and `brms`.

1 Introduction

The motivation for developing phylogenetic comparative methods (PCM) is now broadly appreciated. For studies of trait evolution, PCM can address a range of fundamental questions (reviewed in Garamszegi 2014). For example, identify axes of phenotypic variation that are conserved over evolutionary time (Pagel, 1999; Blomberg et al., 2003); compare the fit of different models of evolution to better understand patterns of diversity (Harmon et al., 2008; Pennell et al., 2014); and evaluate the strength and direction of co-evolutionary relationships between traits (Housworth et al., 2004; Hadfield, 2010). These techniques can be applied to a diverse range of species traits, from morphology, to behaviour, to environmental tolerances and niche characteristics. Thus, as the tide of modern data collection continues to erode historical constraints on the scale and complexity of analyses, opportunities to apply these methods are increasing. Advances in climate modelling, remote sensing technology, and global collaborative data projects, in particular have led to a rapid increase in the resolution and availability of datasets suitable for comparative analyses of trait evolution. Progressive and pervasive genome sampling across the tree of life continues to provide more accurate and complete phylogenies. Finally, the computational resources required to pursue fully Bayesian approaches are now accessible to most researchers. Despite this coalescence of opportunities, practical and conceptual challenges in implementing modern methods create barriers to usage for many biologists.

A mature, albeit technical, literature on statistical models of trait evolution has existed for some time, including detailed mathematical treatments of phylogenetic models. In particular, development of the generalised linear mixed model framework has helped to bring phylogenetic comparative analyses of trait evolution into the mainstream (Lynch, 1991; Housworth et al., 2004; Hadfield, 2010). This approach has many benefits, not least the general familiarity with mixed models among researchers in ecology and evolution (Bolker et al., 2009). However, applications of this method are rarely pursued, or remain

relatively simplistic, even for recent large-scale analyses of species traits to which they would be well suited (Cássia-Silva et al., 2020; Grossnickle, 2020; Bruelheide et al., 2018; Díaz et al., 2016). Recent publications highlighting widespread misconceptions about common methods, poor statistical practices, and the potential for spurious inference (Freckleton, 2009; Cooper et al., 2016; Revell et al., 2008; Revell, 2010; Uyeda et al., 2018, Westoby et al. in prep.) confirm the need for more translational research.

This paper aims to help bridge the gap between theory and practice in phylogenetic analyses of trait evolution. We focus on Bayesian implementations of phylogenetic mixed models (PMM), particularly multi-response (MR) PMM, due to their flexibility and general utility in comparative biology. Though the benefits of multi- over single-response (SR) PMM in providing more nuanced analyses of trait evolution have long been recognised (Housworth et al., 2004; Hadfield, 2010), it is likely that practical barriers to implementation continue to prevent common usage of MR models. We begin with a conceptual introduction to mixed models, demonstrating how covariance matrices are used to model the correlated error structures present in phylogenetic effects. We define the basic SR-PMM and discuss its equivalence with PGLS, including the different parametrizations used to quantify phylogenetic signal from these models. After outlining the limitations of SR models for characterising patterns of co-evolution between traits, we shift focus to MR-PMM. In particular, we highlight the different components of covariance (phylogenetic versus independent, taxon-level versus trait-level) that arise in MR-PMM, how these error structures are specified using cross-covariance matrices, and how parametrizations of these matrices can be used to estimate phylogenetic and independent correlations between traits. In section 3, we explore data simulated from the MR-PMM to clarify the distinction between these correlations operating on different levels in the model hierarchy. In section 4 we discuss interpretation of the (co)variance components estimated by MR-PMM. 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 PMM, highlighting common pitfalls and providing recommendations to analysts.

To support our exposition of the material presented, we provide an online tutorial, including annotated code and additional examples. This tutorial demonstrates how to simulate multivariate trait data containing phylogenetic signal and specify, fit, and interpret corresponding MR-PMM in two popular and powerful R packages, `MCMCglmm` (Hadfield, 2010) and `brms` (Bürkner, 2017).

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:

$$y \sim 1 + X_1 + (1|\text{group}), \tag{1}$$

where y is the observed response, 1 specifies a population-level intercept, the measured covariate X_1 specifies a single fixed effect (although more effects could be added), and $(1|\text{group})$ specifies a random intercept at the group level. This syntax implies the provision of a data set comprising multiple observations of y and corresponding X_1 measurements and group assignments. In other words, we model y

as a linear combination of effects that account 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 make explicit the choice of statistical distributions. A model comparable to (1) can be expressed as,

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

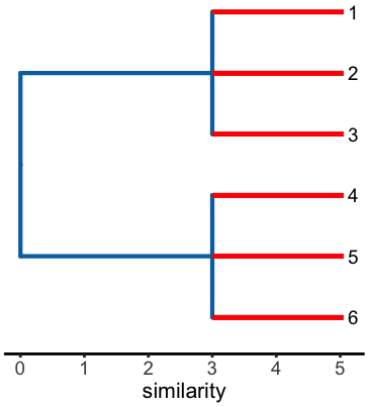
where, $\boldsymbol{\mu}$ is a vector of fixed effects including a population-level intercept, $\boldsymbol{b}$ a vector of group-level random effects, and $\boldsymbol{e}$ a vector of independent errors (unless otherwise stated, bold symbols are used to denote vectors with length equal to the number of observations). 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 n 1's on the diagonal and 0's in all off-diagonals) encodes independent error for each observation (see Figure 1a for an example). Finally, the parameters σ_b^2 and σ_e^2 denote the magnitude of the structured and independent variance components, respectively. In many instances, C is re-scaled to a correlation matrix (i.e., with 1's on the diagonal) which places it on the same scale as I and allows the relative proportion of total variance attributable to C and I to be calculated directly from σ_b^2 and σ_e^2.

We highlight the mapping of model terms from the mathematical formula in (2) to the R formula in (1): $\boldsymbol{\mu}$ maps to $1 + X_1$, $\boldsymbol{b}$ maps to (1|group), but the Gaussian error term $\boldsymbol{e}$ is not explicitly encoded in the R formula. When specifying these types of models in R, the error distribution and associated link function(s) are generally specified outside of the model formula to extend the use of the formula syntax in (1) to non-Gaussian distributions (e.g., for logistic regression, family = "binomial"; also see Box 1). We emphasise the mathematical format of (2), because it makes explicit the distributional assumptions of the model, and provides a broadly encompassing framework for the analysis of correlated data. Indeed, the model in (2) is equally applicable, whether covariance arises from sampling design or from phylogenetic history.

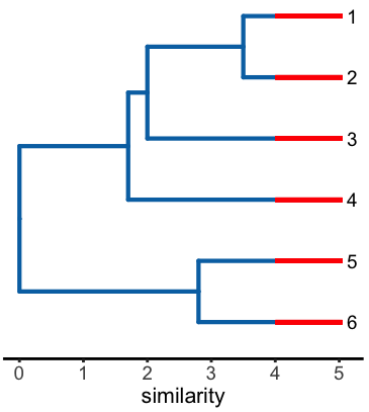
To illustrate, suppose we are analysing data collected from different transects using random-intercepts to model the correlation of observations within each transect. There are two different ways to view this model (i.e., two different models associated with the specification of the random-effect $\boldsymbol{b}$ for transect): the *conditional* model (2), where $\boldsymbol{e}$ is the residual given the predicted mean $\boldsymbol{\mu} + \boldsymbol{b}$, and the *marginal* model where $\boldsymbol{b} + \boldsymbol{e}$ is the residual given $\boldsymbol{\mu}$. The marginal model is obtained from the conditional model by averaging over the distribution of the random effects. For Gaussian distributed responses, marginalisation is analytically straightforward and the variance of the sum $\boldsymbol{b} + \boldsymbol{e}$ is simply the sum of variances; that is,

$$\begin{aligned} \boldsymbol{b} + \boldsymbol{e} &\sim \mathcal{N}(0, \Sigma) \\ \Sigma &= \sigma_b^2 C + \sigma_e^2 I. \end{aligned} \tag{3}$$

Thus, for the marginal model, the group structure and independent error are combined into a single covariance matrix, Σ, which governs the overall distribution of the marginalised residuals $\boldsymbol{b} + \boldsymbol{e}$. In this view, 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 graph (also see Ives 2018). This diagrammatic representation makes clear how covariance matrices can be used to model phylogenetic sources of structural dependence in the data.



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



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

Figure 1. 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.

2.2 Phylogenetic Mixed Models

In analyses of cross-species data, correlative structure 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). Within a mixed model, this phylogenetic structure can be handled in exactly the same way as a nested sampling design, by defining a covariance matrix that encodes an expectation of similarity between each observation. From this perspective, phylogenies can be viewed as diagrams of expected non-independence among species trait values (Figure 1b). Although important assumptions (e.g., model of evolution, accuracy of topology and branch lengths, etc.) are made when deriving a covariance matrix from a phylogenetic tree (see below), this analogy reveals how readily the mixed modelling framework is deployed to model phylogenetic structure in cross-species data.

An important assumption of PMM is that covariance among species trait values reflects phylogenetic structure. If true, we can derive the expected covariance matrix C directly, by assuming that some portion of the variance in $\boldsymbol{y}$ reflects Brownian Motion evolution (Felsenstein, 1985). Specifically, 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 (co)variances of trait values among species. Thus, Brownian motion gives a direct translation from phylogeny to

covariance matrix (Ives, 2018). The random effects, $\boldsymbol{b}$, in PMM are therefore expected to follow a multivariate normal distribution characterized by the phylogenetic correlation between taxa, C, and the phylogenetic variance (or, evolutionary rate parameter), σ_b^2 (2). In practice, the phylogeny used to derive C is supplied by the user and is considered known without error (although see Section 6), while 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. This is not the case for the off-diagonal elements. Instead, for a pair of taxa i and j, the expected phylogenetic covariance is given by C_{ij}, the sum of branch lengths from the root node to the most recent common ancestor of taxa i and j. In other words, the sum of shared branch lengths for taxa i and j.

To summarise, PMM is a special case of a mixed model (2) that incorporates phylogenetic relatedness via a covariance structure. Species trait values $\boldsymbol{y}$ are modelled as the sum of fixed effects (including an intercept) $\boldsymbol{\mu}$, a random effect based on phylogenetic position $\boldsymbol{b}$, and independent (observation-level) 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 Regression approaches

It is useful to compare PMM with related techniques to gain perspective on the relationships between methods. Perhaps the most well-known methods for phylogenetic regression are Felsenstein's (1985) phylogenetic independent contrasts (PIC) and phylogenetic generalised least squares (PGLS), first introduced by Grafen (1989). As with PMM, each of these methods seeks to estimate regression coefficients from cross-species data while accounting for dependence due to shared ancestry. Phylogenetic independent contrasts uses the reasoning that while trait values across species may not be independent, differences (contrasts) between species must be, because, given BM, speciation results in two independent evolutionary trajectories (Felsenstein, 1985). PGLS approaches the problem as a regression of correlated observations, with more closely related species expected to produce more similar residuals from the estimated regression line (Symonds and Blomberg, 2014). PGLS is equivalent to PIC when $\Sigma = \sigma_b^2 C$ (i.e., $\sigma_e^2 = 0$) (Blomberg et al., 2012) and is equivalent to ordinary least squares when $\Sigma = \sigma_e^2 I$ (i.e., $\sigma_b^2 = 0$). Typically however, both σ_b^2 and σ_e^2 are estimated in PGLS (see below) which scales the magnitude of phylogenetic effects based on the strength of signal in the data. In this general case, PGLS is equivalent to a marginal PMM for a single response trait.

2.2.2 Phylogenetic signal

A common objective of phylogenetic comparative methods is to quantify phylogenetic signal in species traits. Several metrics have been developed for this purpose, notably h^2 (Lynch, 1991), λ (Pagel, 1999) and K (Blomberg et al., 2003). Because h^2 (also known as phylogenetic heritability) 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 (e.g., Pagel 1999; Symonds and Blomberg 2014; Ives 2019), it is easily overlooked that these two statistics are in fact mathematically equivalent (Housworth et al. 2004). Indeed, each of these metrics expresses the relative proportion of variance attributable to phylogeny in cross-species trait data (also see Dingemanse and Dochtermann 2013 for a comparison with the quantity known as 'repeatability' calculated from variance decomposition of behavioural traits measured across individuals). Phylogenetic heritability, h^2, is naturally interpreted in the PMM framework and is expressed in terms of the estimated variances components as,

$$h^2 = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}. \tag{4}$$

This is precisely what $\lambda = h^2$ achieves in PGLS, which we now discuss.

2.2.3 Branch length transformations

In the original presentation of PGLS, the matrix C is taken as the total covariance matrix (Grafen, 1989), which assumes traits have evolved via BM on the phylogeny with no additional independent 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 $\boldsymbol{\theta}$. This is done to scale the tree according to the magnitude of phylogenetic signal detected in the residual structure of the model, or to test the fit of different evolutionary models 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

$$\Sigma = \sigma^2 C(\boldsymbol{\theta}). \tag{5}$$

The parameters σ^2 and $\boldsymbol{\theta}$ can be estimated using either maximum likelihood or Bayesian model-fitting approaches. A common instance is when $C(\theta)$ is a weighted sum of C and I characterised by a single parameter λ (Pagel, 1999)

$$C(\lambda) = \lambda C + (1 - \lambda) I \tag{6}$$

In this formulation, the additive terms express the relative proportions, scaled by λ, of the total variance attributable to phylogenetic and independent error. By substituting (6) into (5) and comparing to (4), it can be seen that $\sigma^2 = \sigma_b^2 + \sigma_e^2$ and $\lambda = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}$ (i.e., that λ is equivalent to h^2). As such, the property of λ optimisation in PGLS, where-by variance is attributed to phylogeny only to the extent that the observed variation reflects phylogenetic structure (Lynch, 1991; Housworth et al., 2004), is also a feature of PMM via the scaling of σ_b^2 and σ_e^2 during estimation.

2.3 Multi-Response Phylogenetic Mixed Models

2.3.1 Biological motivation

The motivation for PMM (2) so far has been to test the relationship between species traits while accounting for dependence due to shared ancestry (Symonds and Blomberg, 2014). This approach takes one focal trait as the response variable and models its relationship to other traits by including them as fixed effects within $\boldsymbol{\mu}$. This specification makes sense if we believe the predictor traits are a direct cause of the response trait. It is, however, misspecified if there is reciprocal causation between the traits, the traits are correlated due to a common causal mechanism, or there is phylogenetic signal in both traits. This poses an obvious problem because these features are precisely what we would expect in many real-world biological data sets. An alternative approach, which helps re-frame these challenges, is to treat both traits as response variables. That is, move all traits to the left-hand side of the model equation (7). This important shift in model assumptions simultaneously accounts for phylogenetic signal in all traits and permits the estimation and partitioning of correlations between traits, while accounting for any fixed effects. We argue this approach will be more meaningful for many ecological and evolutionary hypotheses.

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 an evolutionarily conserved relationship between these traits, with the remainder due to independent causes. In a single response model, 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 parameter. This means these separate (co)variance components are not recoverable from the model. Furthermore, phylogenetic

signal in the predictor traits is confounded with (phylogenetic components of) the residual covariance structure (Warton 2022; also see Wilson 2008 for a corollary from quantitative genetics). In other words, a single-response model cannot effectively separate covariance between traits that is phylogenetically structured (conserved correlation between x and y) from covariance between traits that is independent of the phylogenetic structure (independent correlation between x and y). A notable consequence of this confound is that inferences based on the slope parameters for predictor traits will be misleading when the 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., in prep). Thus, for cases when more than one trait contains phylogenetic signal, multi-response (MR) models are a preferable approach (Housworth et al. 2004; Hadfield 2010).

For MR models, the covariance of each trait pair (and its partitioning into phylogenetic and independent components) is the focus of analysis rather than the prediction of one trait as a function of the others. This is particularly advantageous when there are no clear *a priori* hypotheses about (unidirectional) causal relationships between traits (Warton et al., 2006). However, the joint-response perspective does not in fact limit our ability to make predictions from the model. On the contrary, by conditioning on $\boldsymbol{y}_2$, a predictive model for $\boldsymbol{y}_1$ given $\boldsymbol{y}_2$ can be generated from components of the full MR-PMM covariance matrix (Ives, 2022). Importantly, this conditional model derived from a MR-PMM differs from a SR-PMM with $\boldsymbol{y}_2$ included as a predictor, in that a matrix rather than a scalar operates on $\boldsymbol{y}_2$ in order to predict $\boldsymbol{y}_1$, leveraging all available information (see Section 5 for further discussion of predictive models). Thus, MR-PMM will often be a preferable approach for questions about trait co-evolution, even when prediction is an explicit objective of analyses.

In summary, many traits intended for use as predictors in SR models will themselves contain phylogenetic signal. If the true data-generating model derives from the 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 (co)variances into phylogenetic and independent contributions, MR-PMM allow us to decompose cross-species trait relationships into a more meaningful set of contributing factors. As partitioning trait variation and identifying functional trait relationships are principal objectives of the biological research program, the MR approach will often represent a better approximating model and thus offer more valuable inferences. Considering these benefits, MR-PMM appears to be underutilised in current research.

2.3.2 Specification and model parameters

As discussed earlier, all species traits in a MR-PMM are modelled jointly as response variables, which means they appear on the left-hand side of the model equation. With this design, both the phylogenetic random effects $\mathbf{b}$ and the independent errors $\mathbf{e}$ must also be modelled jointly, here using multivariate normal distributions (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

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 + \mathbf{b}_1 + \mathbf{e}_1 \\ \boldsymbol{\mu}_2 + \mathbf{b}_2 + \mathbf{e}_2 \end{pmatrix} \tag{7}$$

$$(\mathbf{b}_1, \mathbf{b}_2)^{\mathrm{T}} \sim \mathcal{N}(0, \Sigma^{\mathrm{phy}} \otimes C) \tag{8}$$

$$(\mathbf{e}_1, \mathbf{e}_2)^{\mathrm{T}} \sim \mathcal{N}(0, \Sigma^{\mathrm{ind}} \otimes I) \tag{9}$$

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

$$\Sigma^{\mathrm{phy}} = \begin{pmatrix} \Sigma^{\mathrm{phy}}_{11} & \Sigma^{\mathrm{phy}}_{12} \\ \Sigma^{\mathrm{phy}}_{21} & \Sigma^{\mathrm{phy}}_{22} \end{pmatrix} \qquad \text{and} \qquad \Sigma^{\mathrm{ind}} = \begin{pmatrix} \Sigma^{\mathrm{ind}}_{11} & \Sigma^{\mathrm{ind}}_{12} \\ \Sigma^{\mathrm{ind}}_{21} & \Sigma^{\mathrm{ind}}_{22} \end{pmatrix}. \tag{10}$$

and 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 and $n \times n$ 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

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

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 single-trait case (3), the total covariance matrix of the marginal model is the sum of the phylogenetic and independent components,

$$\Sigma \quad = \quad \Sigma^{\text{phy}} \otimes C + \Sigma^{\text{ind}} \otimes I. \tag{12}$$

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

$$\Sigma_{ij}^{\text{phy}} = \sigma_i^{\text{phy}} \sigma_j^{\text{phy}} \rho_{ij}^{\text{phy}} \qquad \text{and} \qquad \Sigma_{ij}^{\text{ind}} = \sigma_i^{\text{ind}} \sigma_j^{\text{ind}} \rho_{ij}^{\text{ind}}, \tag{13}$$

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

Our exposition emphasises a key distinction between SR and MR-PMM; the decomposition of (co)variance terms. As discussed earlier, in the SR case (2), C encodes phylogenetic covariance between observations—the tendency for closely related species to show similar values of a particular trait—and I encodes independent error associated with each observation. In the MR case, the Kronecker product (11) provides a means to include both phylogenetic and independent structures into 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 that is conserved over evolutionary time (i.e., a portion of trait co-variation that is associated with phylogeny), while $\widehat{\Sigma}_{ij}^{\text{ind}} \neq 0$ provides evidence for a component of correlation between traits that is independent of phylogeny (see Figure 2 for an illustration).

2.3.3 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). It would therefore be advisable to account for measurement methodology with a fixed effect covariate when attempting a comparative analysis with such data. The second reason is to explore hypotheses about the mechanistic nature of correlations identified from fitted models. For example, where known or hypothesised causal drivers of selection on particular response traits can be identified, these may be included as fixed effects to investigate whether the correlation between two traits provides evidence for independent responses to the same causative agent, or a more integrated association. A good example of this approach can be seen in Sanchez-Martinez et al. (2020), where the authors use intercept-only MR-PMM to partition the 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 to evaluate different hypotheses about trade-offs and integration between traits in the evolution of hydraulic systems (see also Kruuk and Hadfield 2007 for related examples from quantitative genetics models). We note that a similar procedure is possible within a multi-response framework. Specifically, in the event that proposed predictors display phylogenetic signal, these too can be fit as response variables and then conditioned upon to investigate the proposed hypothesis (see Section 5). If a predictor does not show signal, it should generally remain in the fixed effects to reduce model complexity and therefore the computational cost. In this case, the variance explained by the fixed effects must be included in the denominator of the h^2 (4) calculation (de Villemereuil et al., 2016).

3 Simulation

An instructive way to develop our understanding of the different covariance components estimated by MR-PMM is to experiment with simulating data under the assumptions of the model. Figure 2 shows simulated data sets from MR-PMM, for different values of phylogenetic and independent (co)variance, together with the phylogenenetic tree from which the structured random effects were generated. The ellipses within the scatter plot provide a visual representation of the simulated correlation components. The heat-map of trait data plotted against the phylogeny highlights the distinct signature each source of structural covariance leaves in the data.

In figure 2A, x and y were simulated with a strong positive correlation, operating entirely on the independent level with no phylogenetic signal in either trait. 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 2B shows the opposing situation, x and y are strongly positively correlated but now correlation manifests entirely with respect to the phylogenetic structure. The scatter plot in Figure 2B 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) will depend on the topology of the tree, with deep splits between major clades promoting separation along the major axis of correlation. We see this looking across clades in 2B, 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 2B shows clear phylogenetic structure corresponding to the different bivariate means of each subclade. Note that this structure weakens across clades from green, to orange to blue as the branching pattern becomes more deeply nested, i.e., as subclades become less clearly separated in evolutionary time.

In figure 2C, x and y were simulated with equal portions of phylogenetic and independent variance, and a strong positive correlation operating on both levels. This scenario, where partial correlations show the same sign, may be common for 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 relationship between traits. The data in 2C are more variable ($\sigma^b + \sigma^e = 2$), however we see attributes of both 2A and 2B present in the plots. For example, clades are distinguishable in the scatterplot but together conform more closely to the independent bivariate density imposed by ρ^e.

In Figure 2D, both traits contain phylogenetic and independent variance, but the correlation between traits only operates on the phylogenetic level. This shows how easily conserved correlations become obscured when random sources of variation contribute substantially to the total variance in each trait. Notably, figure 2D represents a set of conditions for which SR-PMM (i.e., PGLS) will typically not report a significant association between x and y. This is because, for the model $y|x$, part or all of the phylogenetically conserved component of variance in y is instead attributed to the (phylogenetic) residual structure, diminishing the influence of phylogenetic covariance between x and y on the estimate of the regression slope (Westoby et al., in prep). As conserved covariances may be the product of selection pressures that continue into the present day, this phylogenetic covariance may in fact be relevant for understanding ongoing associations between traits. In these situations, MR-PMM may represent a more meaningful approximating model.

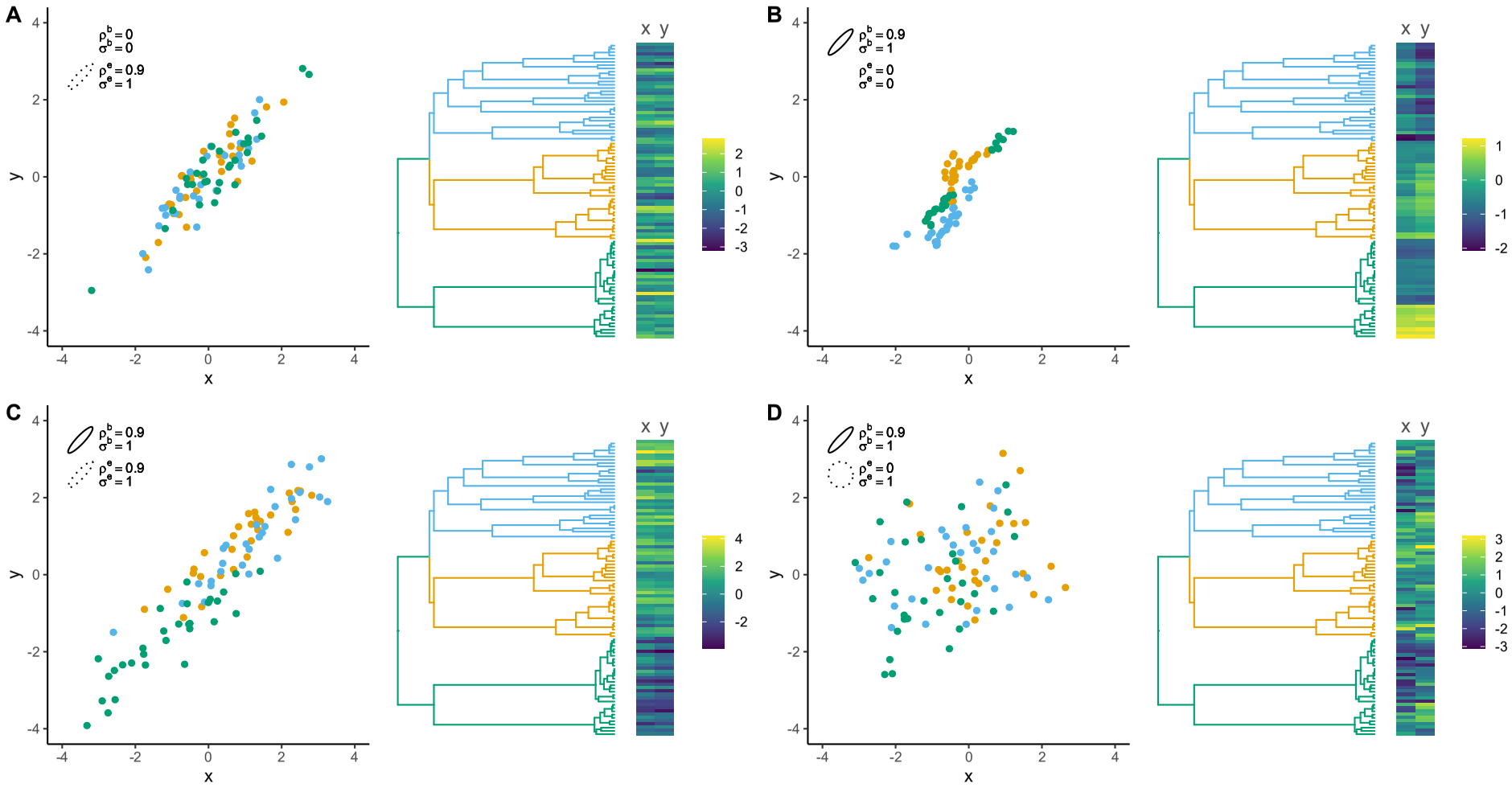


Figure 2: Bivariate trait data simulated from MR-PMM containing different levels of phylogenetic and independent (co)variance (A-D). Simulation conditions for each data set are inset in scatterplots. Trait data are plotted in scatterplots, as well as heatmaps arranged against the generating phylogeny.

Observations on simulated data are of course heuristics only; compared to real biological datasets, manifest from complex causal networks, the simulations presented in Figure 2 represent contrived evolutionary scenarios. Nonetheless, simulations provide concrete examples of the distinction between phylogenetic and independent covariance in cross species data and help clarify the different covariance structures we aim to partition with MR-PMM. More broadly, simulations help clarify what our models are really estimating. 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 methods 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 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 phylogenetic (co)variance components from PMM, and their utility for addressing specific ecological and evolutionary hypotheses, has been hotly debated (Freckleton et al., 2002; Westoby et al., 1995; Björklund, 1997; Harvey et al., 1995). While rigorous critique is necessary for methods development, the lack of resolution on this issue has hampered progress, as a balanced and unifying perspective is crucial to realise the full, operational potential of comparative methods such as MR-PMM.

We argue that phylogenetic (co)variances estimate the conserved component of species phenotypes given a model of evolution and a hypothesis of the phylogenetic relationship between taxa. However, the extent to which these conserved effects result from Brownian drift, or reflect constraints and adaptive responses to selection, will vary for different traits, between clades, across phylogenetic scales, and may even 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.". More specifically, the phylogenetic component of trait (co)variance in PMM will capture all variation attributable to the phylogenetic relationships among taxa. This may include contributions from both neutral and functional genetic differences between taxa accumulated over evolutionary history, but also 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 of independent trait variance 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 co-evolutionary change between traits i and j. Over subsequent divergences, such effects may come to reflect the underlying phylogenetic structure, and 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 co-variation resulting from strictly historical causes (Westoby et al., in prep).

Recognizing that phylogenetic and independent correlation can arise from multiple distinct processes also argues for caution when making specific 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 the 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. Furthermore, 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). Judicious use of fixed effects can help evaluate evidence for alternative causal hypotheses. For example, by testing whether phylogenetic correlation between traits remains after accounting for known environmental drivers of selection (e.g. Sanchez-Martinez et al. 2020, also see section 2.7). Ultimately however, it is rarely the pleasure of a comparative biologist to declare causation. Rather, to generate biologically meaningful hypotheses that can be tested with experimental research. In this regard, creative applications of MR-PMM have the potential to revolutionize our approach to understanding functional trait variation (Smith et al., 2020).

In summary, while MR-PMM offers a powerful tool for partitioning trait correlations in cross species data, the precise mechanisms driving observed covariances cannot be inferred. In many cases, mechanistic hypotheses for the association between traits have already been developed from theoretical or empirical work, which may guide interpretation (see section 6.2). Indeed, functional arguments for trait associations often motivate comparative work in the first place. Nonetheless, researchers should consider plausible alternative hypotheses both when selecting covariates and when offering an interpretation of the (co)variances observed in a given data set.

5 Prediction

Prediction is an important component of statistical modelling. In the context of PMM, the estimated correlation structure between taxa and traits can be used to predict new trait observations (e.g., for new species or missing data) by conditioning on the measured taxon-trait values. Even when the goals of the analysis are not to make predictions *per se*, the predictive distribution of a fitted model is useful for model validation, such as posterior-predictive checks, or model comparison using predictive assessment such as cross-validation scores or information criteria. We refer the reader to the accompanying tutorial for worked examples of how these predictive methods can be implemented in the R programming language.

5.1 Predicting to new or missing observations

The estimated phylogenetic and trait correlations for a fitted PMM can be used to predict to new taxa and traits given only partially observed data. For example, if only a subset of traits is measured for a new taxon then prediction of the unmeasured traits is improved by 1) the subset of trait measurements for the new taxon, provided at least one of corresponding trait correlations is non-zero, and 2) the existing measurements of the missing traits for the original taxa, provided at least one of the corresponding phylogenetic correlations is non-zero.

A predictive model based on correlation estimates is obtained from a fitted PMM by conditioning prediction on all available observations. To generate the conditional 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 (7)),

$$\begin{pmatrix} \boldsymbol{y}_{\text{obs}} \\ \boldsymbol{y}_{\text{new}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \widehat{\boldsymbol{\mu}}_{\text{obs}} \\ \widehat{\boldsymbol{\mu}}_{\text{new}} \end{pmatrix}, \begin{pmatrix} \widehat{\Sigma}_{\text{obs, obs}} & \widehat{\Sigma}_{\text{obs, new}} \\ \widehat{\Sigma}_{\text{new, obs}} & \widehat{\Sigma}_{\text{new, new}} \end{pmatrix} \right), \tag{14}$$

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.

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

$$\boldsymbol{y}_{\text{new}} | \boldsymbol{y}_{\text{obs}} \sim N(\widehat{\boldsymbol{\mu}}'_{\text{new}}, \Sigma'_{\text{new}}) \tag{15}$$

where

$$\widehat{\boldsymbol{\mu}}'_{\text{new}} = \widehat{\boldsymbol{\mu}}_{\text{new}} + \Sigma_{\text{new, obs}}\Sigma^{-1}_{\text{obs, obs}}(\boldsymbol{y}_{\text{obs}} - \hat{\boldsymbol{\mu}}_{\text{obs}}) \tag{16}$$

$$\widehat{\Sigma}'_{\text{new, new}} = \widehat{\Sigma}_{\text{new, new}} - \widehat{\Sigma}_{\text{new, obs}}\widehat{\Sigma}^{-1}_{\text{obs, obs}}\widehat{\Sigma}_{\text{obs, new}}. \tag{17}$$

The observed responses $\boldsymbol{y}_{\text{obs}}$ enter the conditional model (16) 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 (17). In a Bayesian setting, the set of posterior samples from the fitted PMM generates a corresponding distribution of $\widehat{\boldsymbol{\mu}}'_{\text{new}}$ and $\widehat{\Sigma}'_{\text{new, new}}$.

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 the Eucalyptus example (see Section 6 and Figure 3).

5.2 Model validation using posterior predictive checks

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 3 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 the trait predictions makes sense across the entire range of plausible temperature values for a 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 3).

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 (called the *training* set) and then assessing the models' predictive capacities on the remaining portion of the data (called the *test* set). 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 models such as PMM that are 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. These new approximate methods use the available set of posterior samples from a single model fit, combined with smoothing and importance sampling techniques, to estimate the predictive log likelihood of each data point as if it had been omitted from the training data (Vehtari et al., 2017). Remarkably, these methods have been shown to outperform information criteria while computing at comparable speeds. They can also be used with multivariate PMM-type data where the conditionally independent log likelihood of each (virtual) test datum is calculated using (15) (Bürkner et al., 2021). The approximate methods we describe here are implemented in the user-friendly R package loo (Vehtari et al., 2017), and we provide examples of their use in the tutorial.

6 Example - Leaf Economics in Eucalyptus

To demonstrate the 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. For brevity, we present only the results of the analyses here. The full details of the analyses 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 was available for three target leaf traits: specific leaf area (SLA); nitrogen content per dry mass of leaf tissue ($\mathbf{N}$); 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 trait values in many circumstances (Debastiani et al. 2021; Penone et al. 2014; new methods are even capable of simultaneously addressing missing data and phylogenetic uncertainty Nakagawa and De Villemereuil 2019). We calculated species mean values and associated standard errors for each trait by pooling all observations at the species level. Correctly propagating sampling error is an important consideration for PMM of species mean trait values, and should motivate the researcher, as it can improve both parameter estimates and hypothesis tests of phylogenetic (co)variance components (Ives et al., 2007). 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 3), 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 a reasonable proportion of the data, necessitating a significant number of additional model fits which we did not pursue here due to computational cost and to simplify the exposition (see tutorial for details).

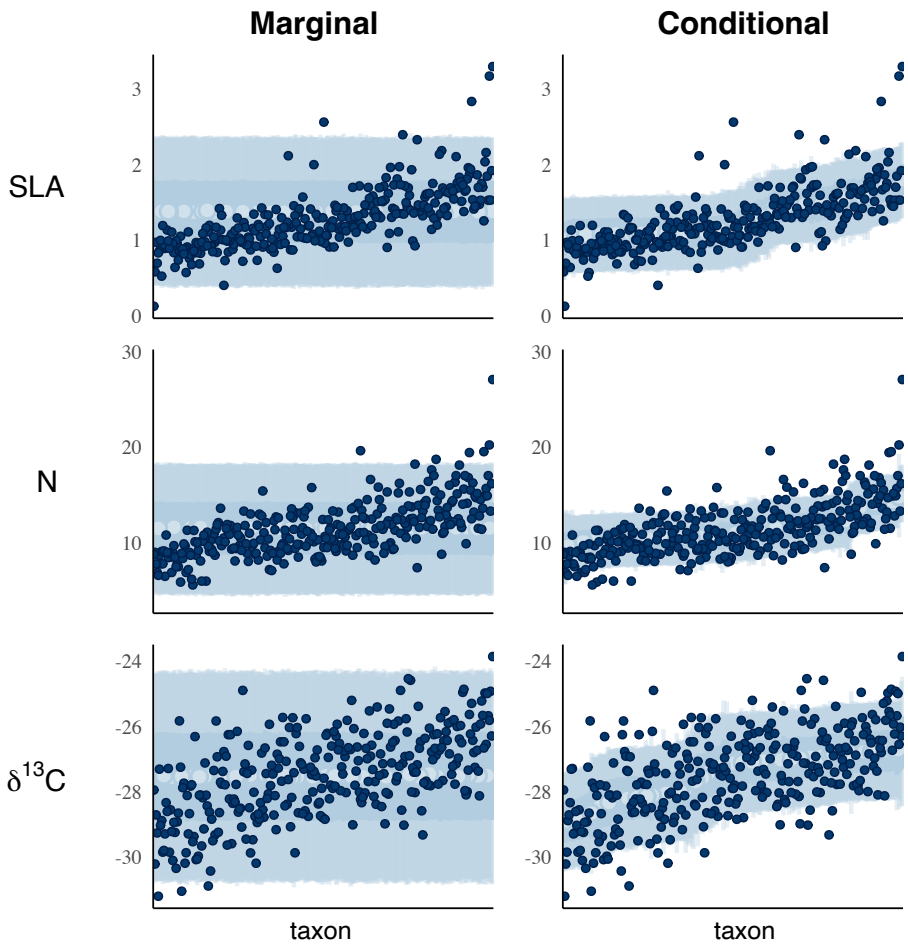


Figure 3. Posterior predictive checks from a fitted MR-PMM of leaf traits in Eucalyptus. Dark blue points represent observed species values. Five-point summaries (credible intervals: 0.025, 0.25, 0.5, 0.75, 0.975) in light blue show the posterior predictive distribution for each taxon. For the marginal 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.

6.2 Results and Discussion

Each trait showed strong phylogenetic signal (posterior mean and 95% credible interval for $\hat{h}^2$: SLA = 0.62 (0.54, 0.69), **N** = 0.67 (0.56, 0.78), δ^{13}**C** = 0.64 (0.54, 0.72)) and hence the tendency for similar values among closely related species (Figure 5). 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 **N**), were associated with lower water use efficiency (low δ^{13}**C**). This is consistent with predictions from leaf economic theory (Wright et al., 2004), that SLA, δ^{13}**C** and **N** represent components of a coordinated life history strategy which trade-off predictably as different regions of niche space are explored during species radiation and diversification. These co-evolutionary relationships are clearly supported by the fitted model, which reports a positive phylogenetic correlation between SLA and **N**, and negative phylogenetic correlations between these traits and δ^{13}**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.

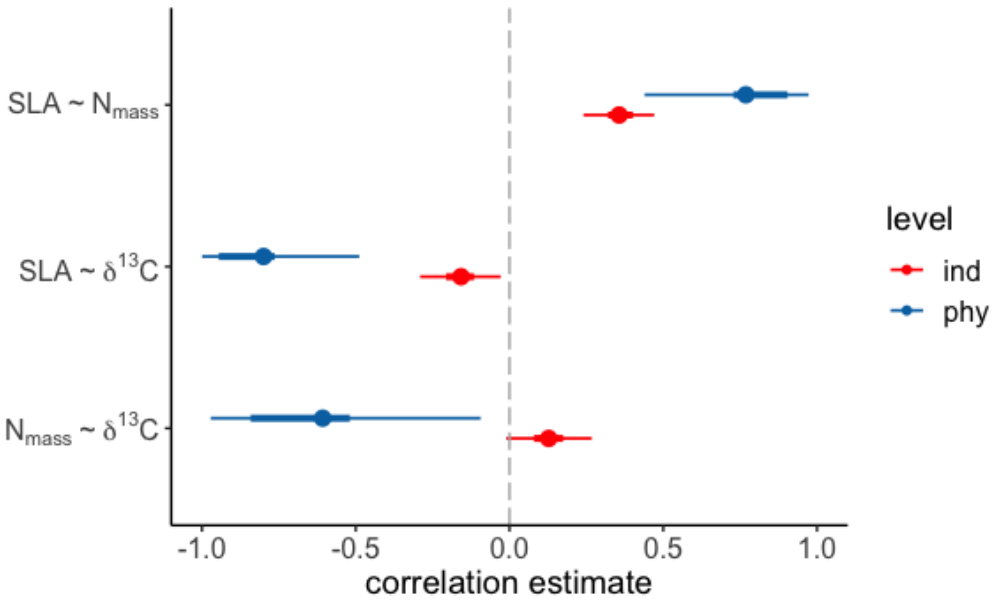


Figure 4. Partial correlation estimates from MR-PMM of leaf traits in Eucalyptus. Points represent the posterior median for each estimate, with 50% and 95% CIs represented by heavy and light wicks respectively.

A notable result is that for **N** and δ^{13}**C**, the independent correlation is positive while the phylogenetic correlation is negative (Figure 4). This situation highlights a core 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 these meaningful relationships in cross-species data (Figure 5). Given this observation, the subsequent objective is to posit plausible biological mechanisms to 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 δ^{13}**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 **N** and δ^{13}**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 **N** reflects variation in the carboxylating enzyme Rubisco (Evans and Clarke, 2019), greater **N** will be associated with increased demand for CO_2 in the mesophyll, relaxing isotope discrimination and increasing the δ^{13}**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 **N** tend to have greater δ^{13}**C** due to the influence of photosynthetic capacity on carbon discrimination.

One important consideration for this result, is that variation in leaf nitrogen has different eco-physiological 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; Figure SX). This suggests that part of the positive independent correlation observed between $\mathbf{N}_{\text{mass}}$ and δ^{13}**C** (Figure 4) represents variation in $\mathbf{N}_{\text{area}}$, corroborating our initial interpretation of results.

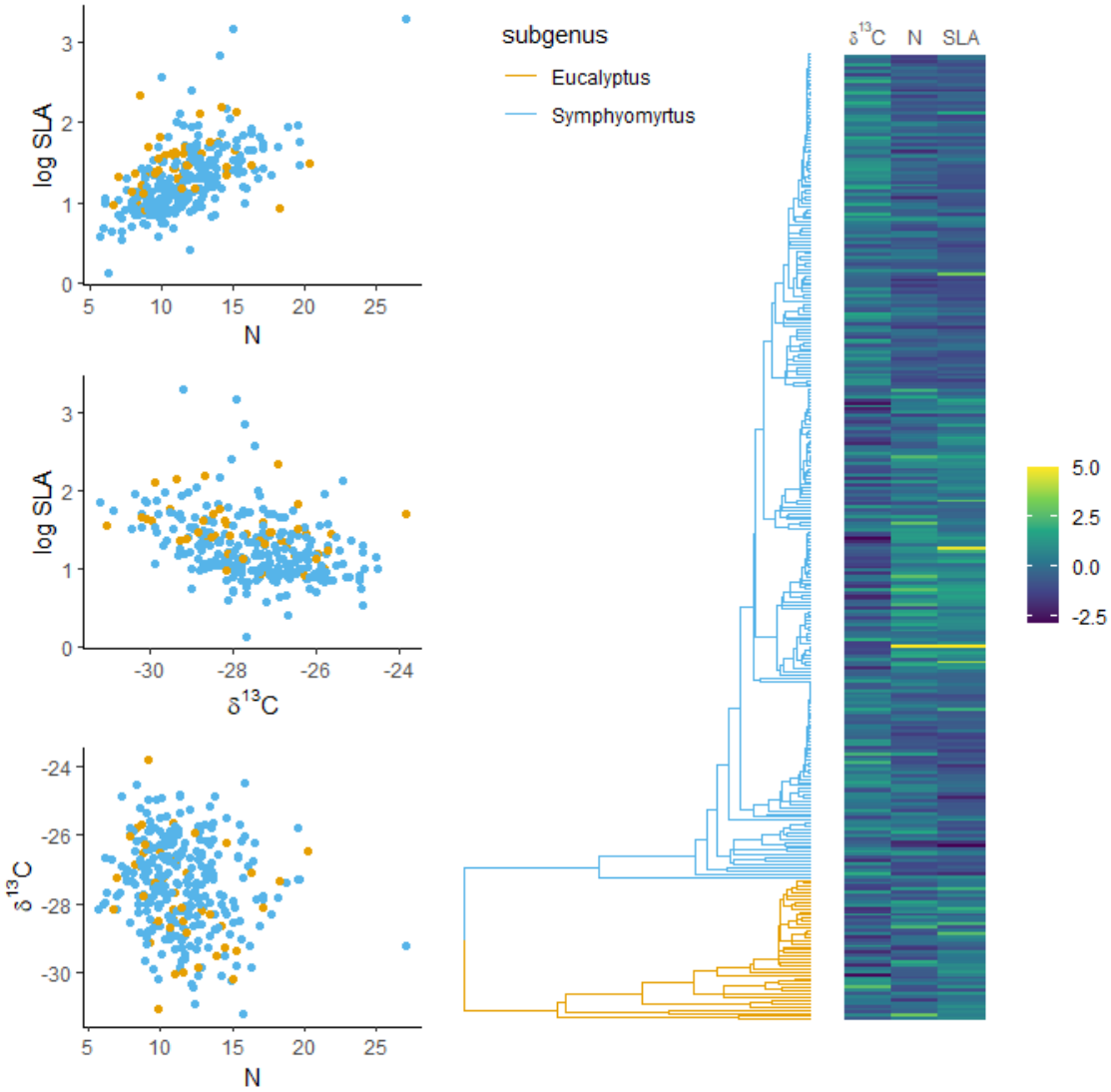


Figure 5. Scatter-plots (left) and heat-maps (right) of three leaf traits across 306 species of Eucalyptus. For heat-maps, values have been centred and scaled for each trait and are aligned with the corresponding species in the phylogeny (centre). For δ^{13}**C** and **N** (bottom left), opposing partial correlations (Figure 4) are obscured at the level of species phenotypes.

7 Discussion

7.1 Summary

Mixed models are a familiar, flexible and extendable framework for comparative analyses of species traits. Phylogenies are used to derive an expectation of dependence among species trait values, which enters the model as a covariance matrix. For a single response, modelling phylogenetic structure means the proportion of variance attributable to phylogeny can be estimated. However, single-response models with fixed effects may confound an ensemble of correlations if predictor and response variables both show phylogenetic signal. Multi-response 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, offering more nuanced inferences of trait evolution. We review these models, with a focus on concepts, implementation and interpretation in the context of phylogenetic comparative analyses. 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

A fundamental 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, others will clearly violate its assumptions (Losos, 2011; Blomberg et al., 2020). Furthermore, Brownian motion cannot account for many of the evolutionary dynamics of interest to biologists (e.g. directional macro-evolutionary trends, stabilising selection, heterogeneous rates of evolutionary change between clades or over time, etc.). The PMM framework does offer some flexibility to compare the fit of different models of evolution via transformations to the covariance matrix, as in (5). This makes PMM naturally extendable to a wide range of evolutionary models for continuous traits (Blomberg et al., 2003; Harmon et al., 2008). Nonetheless, modelling phylogenetic structure in this way entails assumptions about the causal network giving rise to interspecific trait variation that deserve careful consideration. For example, an important limitation of PMM is that it assumes the chosen evolutionary model is both constant through time and homogeneous across the tree, potentially making it vulnerable to singular evolutionary events (Uyeda et al., 2018). By contrast, techniques such as mixed Gaussian phylogenetic models allow for shifts, meaning parameters of the evolutionary model (e.g. the Brownian rate parameter σ_b^2), even the model itself (BM, Ornstein-Uhlenbeck, etc), can vary between clades and across time (Mitov et al., 2020, 2019). Approaches based on a more general class of stochastic diffusion models offer more detailed and dynamic models of trait evolution, including the potential to model evolution of the trait-level covariance matrix itself (i.e., $\Sigma = \Sigma(\theta, t)$) (Blomberg et al., 2020).

Broadly, however, the integration of different evolutionary models across traits remains poorly developed for the multi-response setting. For example, it is unclear how phylogenetic correlations between traits should be specified and interpreted when models contain many shifts or traits follow different evolutionary models. In such cases, a phylogenetic natural history approach may be preferable (e.g. Pagel et al. 2022; Ingram and Mahler 2013; Uyeda and Harmon 2014), which focuses on inferring the quality and position of major shifts in a multivariate generating process rather than quantifying partial correlations. Similarly, while the standard form of MR-PMM presented here is well suited to low dimensional trait analyses (e.g. <10 response traits), dimension-reduction techniques or alternative methods will be more appropriate for high dimensional data sets, such as those generated in studies of geometric morphometrics (Adams, 2014; Adams and Collyer, 2018; Collyer et al., 2015).

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 can be 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 (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 can require large sample sizes (i.e., observations on hundreds of taxa). Though to some extent this will depend on the data at hand, basic power analyses for phylogenetic covariance between two Gaussian traits indicate that performance may be unpredictable for smaller sample sizes (Housworth et al. 2004). However, a more comprehensive assessment of 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, and sample size. As in all cases, 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 mixed modelling is failure to include important fixed effects. This problem arises when omitted variables influence phenotypes at the observation level, adding error to estimates of (co)variance components. In practice, data on relevant covariates are rarely available for all observations, and it is common to use mean trait values pooled from multiple studies and observations per species. While species mean trait values based on very few replicates should be treated with caution (Ives et al., 2007), it is often possible to incorporate sampling error into analyses meta-analytically (Section 4; also see tutorial and `brms` vignettes). 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). 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. In particular, special care must be taken in specifying and validating models, and in the interpretation of results. Though we touch on each of these considerations (see tutorial), we recommend that readers delve further into these topics according to their specific research needs. For texts covering the broad application of mixed models in biology we recommend Bolker et al. (2009); Nakagawa and Schielzeth (2010), and for multi-response phylogenetic mixed models Hadfield (2010); Housworth et al. (2004); Dingemanse and Dochtermann (2013); Brommer et al. (2019). As a general recommendation for MR-PMM, variables that can reasonably be considered species traits belong on the left hand side of the model equation, where signal can be modelled when present. Predictors testing specific hypotheses and covariates relevant to sampling bias belong on the right hand side, in the fixed effects. This way, specific effects can be controlled for, with the remaining variation within and between traits partitioned among phylogenetic and independent covariance matrices.

7.4 Conclusion

Species phenotypes are the product of a complex network of causal processes. For some traits, variation is conserved over evolutionary time, creating patterns in cross-species data that reflect the phylogenetic relationships among taxa. 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. Single response PMM, such as PGLS, are unable to distinguish these components of trait correlation. In contrast, MR-PMM, in which the correlations between traits can be partitioned into phylogenetic and independent components, offer a more plausible and informative model of trait evolution under many circumstances. Combined with the flexibility of the mixed model framework to accommodate different error distributions and random effect structures, these properties make MR-PMM a unifying approach to many open questions in comparative biology. We expect more researchers will adopt these methods as barriers to implementation for non-specialists are broken down.

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. This work was funded by The Australian Research Council Centre of Excellence for Plant Success in Nature and Agriculture (CE200100015).

Box 1. Including diverse trait types

The MR-PMM formulation (7) 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

$$\begin{pmatrix} \boldsymbol{l}_1 \\ \boldsymbol{l}_2 \\ \vdots \\ \boldsymbol{l}_k \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 + \mathbf{b}_1 + \mathbf{e}_1 \\ \boldsymbol{\mu}_2 + \mathbf{b}_2 + \mathbf{e}_2 \\ \vdots \\ \boldsymbol{\mu}_k + \mathbf{b}_k + \mathbf{e}_k \end{pmatrix}, \tag{18}$$

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

$$\boldsymbol{y}_i \sim f_i(\eta_i^{-1}(\boldsymbol{l}_i), \boldsymbol{\phi}_i), \tag{19}$$

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 (19), the joint likelihood of the trait observations and the latent variables is

$$p(\boldsymbol{y}, \boldsymbol{l}) = \prod_{i=1}^{k} p(\boldsymbol{y}_i \mid \boldsymbol{l}_i, \phi_i)\, p(\boldsymbol{l} \mid \Sigma^{\text{phy}}, \Sigma^{\text{ind}}). \tag{20}$$

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 (4) 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

$$h^2_{\text{probit}} = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2_e + \sigma^2_\varepsilon} = \frac{\sigma^2_b}{\sigma^2_b + 2}. \tag{21}$$

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).

References

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

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

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

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

Björklund, M. 1997. Are 'comparative methods' always necessary? Oikos **80**: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**:717–745.

Blomberg, S. P., J. G. Lefevre, J. A. Wells, and M. Waterhouse. 2012. Independent Contrasts and PGLS Regression Estimators Are Equivalent. Systematic Biology **61**: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**: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**: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**: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. Efficient leave-one-out cross-validation for Bayesian non-factorized normal and Student-t models. Computational Statistics **36**:1243–1261.

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**:1–10.

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**: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**:34–56.

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**: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**:693–699.

de Villemereuil, P., H. Schielzeth, S. Nakagawa, and M. Morrissey. 2016. General Methods for Evolutionary Quantitative Genetic Inference from Generalized Mixed Models. Genetics **204**: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**: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**:5738–5743.

Dingemanse, N. J., and N. A. Dochtermann. 2013. Quantifying individual variation in behaviour: mixed-effect modelling approaches. Journal of Animal Ecology **82**: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**:82–94.

Evans, J. R., and V. C. Clarke. 2019. The nitrogen cost of photosynthesis. Journal of Experimental Botany **70**: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**:1–20.

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**:121–137.

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

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

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

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

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**:712–726.

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

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. Bayesian Data Analysis. Third edition. Chapman and Hall/CRC.

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

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

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. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: investigating evolutionary radiations. Bioinformatics **24**:129–131.

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

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

Ingram, T., and D. L. Mahler. 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. Methods in Ecology and Evolution **4**:416–425.

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**:234–251.

Ives, A. R. 2022. Random errors are neither: On the interpretation of correlated data. Methods in Ecology and Evolution **13**:2092–2105.

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

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

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**:19691–19695.

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

Losos, J. B. 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. Ecology letters **11**: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**:709–727.

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

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**:16921–16926.

Nakagawa, S., and P. De Villemereuil. 2019. A general method for simultaneously accounting for phylogenetic and species sampling uncertainty via Rubin's rules in comparative analysis. Systematic Biology **68**:632–641.

Nakagawa, S., P. C. D. Johnson, and H. Schielzeth. 2017. 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**:935–956.

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

Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature **401**: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**:1–12.

Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. 2014. geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics **30**:2216–2218.

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**:961–970.

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**:612–625.

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

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

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

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

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**:1599–1610.

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**:3931–3944.

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**:415–425.

Symonds, M. R., and S. P. Blomberg, 2014. A primer on phylogenetic generalised least squares. Pages 105–130 *in* Modern phylogenetic comparative methods and their application in evolutionary biology. 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**:29–48.

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

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**:902–918.

Uyeda, J. C., R. Zenil-Ferguson, and M. W. Pennell. 2018. Rethinking phylogenetic comparative methods. Systematic Biology **67**: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**:1413–1432.

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

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

Watanabe, S. 2010. 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**:1–14.

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

Westoby, M., L. Yates, B. Holland, and B. Halliwell. in prep. Phylogenetically conservative trait correlation: quantification and interpretation. Journal of Ecology **X**:pp–pp.

Wilson, A. 2008. Why h2 does not always equal VA/VP? Journal of evolutionary biology **21**:647–650.

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**: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**: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 page e1557.

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