

# Trees for fishes: The neglected role for phylogenetic comparative methods in fisheries science

James T. Thorson 

Resource Ecology and Fisheries  
Management, Alaska Fisheries Science  
Center, National Marine Fisheries Service,  
NOAA, Seattle, Washington, USA

## Correspondence

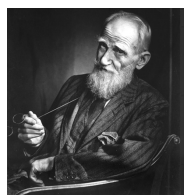
James T. Thorson, Resource Ecology and  
Fisheries Management, Alaska Fisheries  
Science Center, National Marine Fisheries  
Service, NOAA, Seattle, WA, USA.  
Email: [james.thorson@noaa.gov](mailto:james.thorson@noaa.gov)

## Abstract

Fisheries scientists compare processes among species to estimate species productivity, management reference points, and climate sensitivities. Ecologists have developed “phylogenetic comparative methods” (PCMs) to address these questions, but there is surprisingly little application of PCM within fisheries science. Here, I bridge this gap by introducing PCM (including Brownian motion, Ornstein–Uhlenbeck, and Pagel's kappa and lambda models for species covariance), thereby showing that PCM generalizes the nested taxonomic random effects that are commonly used in fisheries science. I next summarize phylogenetic structural equation models (PSEMs), which extend the linear models that are commonly used in fisheries. Finally, I re-analyse a high-quality database used to predict mortality rates from longevity and/or growth parameters. I specifically propose a PSEM that reverts to a longevity-based prediction when longevity information is available but uses phylogenetic corrected growth parameters otherwise. Using this single PSEM replaces the common practice of fitting and predicting using separate linear models depending upon what data are available for a given species. Cross-validation suggests that the relationship between log-mortality rate and longevity does not vary based on phylogeny, and therefore, linear models and PSEM both explain 82% of variance when longevity is available. When longevity is unavailable, by contrast, the linear model explains only 37% of variance while the PSEM explains 52% of variance, where this gain occurs from conditioning predictions on phylogenetic similarities. I therefore conclude that PCM and PSEM provide a general and user-friendly replacement for linear models and can improve performance for fisheries meta-analyses that are used for fisheries management applications.

## KEYWORDS

evolution, longevity, natural mortality, phylogenetic comparative methods, phylogenetic trait imputation, structural equation model



## Ghoti papers

Ghoti aims to serve as a forum for stimulating and pertinent ideas. Ghoti publishes succinct commentary and opinion that addresses important areas in fish and fisheries science. Ghoti contributions will be innovative and have a perspective that may lead to fresh and productive insight of concepts, issues and research agendas. All Ghoti contributions will be selected by the editors and peer reviewed.

## Etymology of Ghoti

George Bernard Shaw (1856–1950), polymath, playwright, Nobel prize winner, and the most prolific letter writer in history, was an advocate of English spelling reform. He was reportedly fond of pointing out its absurdities by proving that ‘fish’ could be spelt ‘ghoti’. That is: ‘gh’ as in ‘rough’, ‘o’ as in ‘women’ and ‘ti’ as in palatial.

## 1 | INTRODUCTION

Foundational questions in fisheries science and management are often addressed by comparing study results across species and stocks (termed “comparative methods”). For example, the concept of recruitment overfishing was popularized by analysing stock-recruit data and comparing results across stocks (Myers, 2001; Myers & Barrowman, 1996). This concept is then the foundation for the widespread use of biological reference point proxies (Clark, 2002; Mangel et al., 2013). Similarly, debates about the institutional and policy foundations of effective fisheries management are typically supported via comparing management outcomes across stocks (Melnichuk et al., 2012; Worm et al., 2009). Given that we have limited data to use in comparative analyses, there is a strong rationale for applying statistically efficient methods.

Many comparative analyses in fisheries science proceed by specifying a linear model for some response trait as a function of other easy-to-measure predictor traits (Gislason et al., 2010; Hoenig, 1983; Pauly, 1980). However, the residuals around this relationship are often correlated for groups of related species, and failing to account for this results in pseudo-replication (anti-conservative tests of significance and confidence intervals) and reduced precision when predicting traits for individual species. As a simple response, an analyst might respond by fitting the model repeatedly for different groups of taxa (Erickson & Nadon, 2021; Froese et al., 2014; Nadon & Ault, 2016). Alternatively, an analyst might instead include additional random effects representing variation among orders, families, and genera (Foss-Grant et al., 2016; Zhou et al., 2012), where this latter approach allows inference to taxa that have not previously been studied. I call this latter approach a “nested taxonomic model,” and it has been used to estimate life-history parameters within custom-built models (Thorson, 2020; Zhou et al., 2012).

Despite the ubiquity of comparative analyses (including these nested taxonomic models), fisheries science includes little cross-over with the related field of phylogenetic comparative methods (PCMs) that is ubiquitous within ecological theory and population genetics (Felsenstein, 1985). For example, searching the phrase “Phylogenetic comparative methods” using Google Scholar on 4 April 2023 returns three matches in the journal *Fish and Fisheries* and one in *CJFAS* (the latter only in the title of a cited article). By contrast, it returns 36 in *Methods in Ecology and Evolution* and 218 in *Systematic Biology*, and there are 96 R packages listed for PCM in the CRAN task view. PCM was developed to account for phylogenetic correlations when analysing patterns across taxa and therefore has many similarities to comparative methods in fisheries. This lack of exchange between PCM and fisheries comparative methods is unfortunate because PCM has developed a wide range of useful analytical methods, including (1) standardized software to produce, export, and visualize evolutionary relatedness among species; (2) analytical methods (and associated software) to learn about relationships while controlling for relatedness; (3) a nuanced vocabulary for discussing similarities among species, including tests for convergent evolution (Stayton, 2015) or evolutionary innovations that allow

adaptive radiation (Yoder et al., 2010); and (4) methods to distinguish between alternative evolutionary models, for example, changes in evolutionary rate among lineages (Drummond & Suchard, 2010).

I specifically highlight how PCM has developed procedures to test how multiple predictor traits and environmental treatments combine to cause changes in a given response trait (Martins, 2000). Given that predictor traits themselves evolve endogenously from changes in still other traits, these tests require specifying complex associations and multi-causal linkages among traits. Some (but not all) of these associations can be studied experimentally, so analysts must estimate the strength of remaining mechanisms using observation data and comparative methods. PCM accomplishes this using phylogenetic path analysis (van der Bijl, 2018; von Hardenberg & Gonzalez-Voyer, 2013) or, more recently, using phylogenetic structural equation models (PSEMs) (Thorson et al., 2023; Thorson & van der Bijl, 2023b). PSEM can exactly replicate results from a linear model, but also includes functionality to (1) account for missing data; (2) make predictions that condition upon evolutionary similarity; (3) represent trade-offs explicitly; (4) include dependencies on unmeasured (latent) variables and estimate composite variables as a weighted combination of measured traits; and (5) estimate exogenous covariance among predictor traits.

In the following, I seek to demonstrate two points:

1. Phylogenetic comparative methods (PCMs) provide a flexible and underutilized toolkit for conducting comparative analyses in fisheries science. In particular, I highlight how alternative evolutionary models result in substantial improvements in predictive performance for real-world analyses.
2. Phylogenetic structural equation models represent a natural solution to missing data in fisheries comparative analyses. In particular, I highlight that PSEM allows a single model to be fitted and then used to make predictions regardless of the quantity of data available for a given species.

To accomplish both points, I specifically highlight the predictive and inferential benefits when analysing the “gold standard” compilation of natural mortality and growth parameters for fishes (Then et al., 2015). This database has been re-analysed previously to demonstrate alternative estimators (Hamel & Cope, 2022), but our study is the first to use either PCM or PSEM for its analysis.

## 2 | A BRIEF INTRODUCTION TO PHYLOGENETIC COMPARATIVE METHODS

Phylogenetic comparative methods have been central to evolutionary and ecological studies since the 1980s (Felsenstein, 1985). To this day, PCM remains the subject of active and ongoing research, both in terms of improved analytical methods (von Hardenberg & Gonzalez-Voyer, 2013), user-friendly software (Tung Ho & Ané, 2014; van der Bijl, 2018), and concerns regarding its proper use and interpretation (e.g., Uyeda et al., 2018). At its core, PCM

involves specifying an additive “tree” that represents evolutionary relatedness among taxa under study. This tree includes a set of taxa (“tips”) and their evolutionary ancestors (“ancestral nodes”), where each node is connected by one “edge” to its immediate ancestor. This tree could be a case where each tip is equally distant from the “root” common ancestor (i.e., an ultrametric phylogeny) or simply a representation of taxonomy.

The analyst then specifies a stochastic process for how continuous or discrete characters are expected to evolve along the tree, and this combination of tree and evolutionary process (along with any parameters in this process) allows the analyst to calculate the expected correlation among taxa (Harmon, 2018). Including this correlation is then presumed to be necessary to get reliable inference about the association among traits, which are otherwise pseudoreplicated (Felsenstein, 1985).

Common evolutionary processes specified during phylogenetic PCM include the following:

1. *Brownian motion* (BM): This null model specifies that continuous trait  $Y$  evolves following a random walk, that is,  $Y_1 \sim \text{Normal}(Y_0, d\sigma^2)$ , where  $Y_1$  is the trait value for a taxon,  $Y_0$  is the value for its nearest ancestor,  $d$  is the branch length connecting the node and its ancestor, and  $\sigma^2$  is the evolutionary rate.
2. *Ornstein–Uhlenbeck* (OU): Alternatively, an analyst might specify that traits evolve around some evolutionarily optimal value  $\mu$  that presumably represents a peak in expected fitness (Lande, 1976). This alternative assumption then specifies a continuous first-order autoregressive process, which can be reparameterized as  $Y_1 \sim \text{Normal}(\mu + \rho^d(Y_0 - \mu), (1 - \rho^{2d})\sigma^2)$ , where  $\rho$  is the autocorrelation rate.
3. *Pagel's  $\lambda$* : An analyst might seek to determine whether the tree is useful in representing the correlation among taxa. To do so, they might specify that correlations arise from a mixture of a random-walk process and an independent value for each taxon, where the mixing rate  $\lambda$  is estimated as parameter (Pagel, 1999). This then results in  $Y_1 \sim \text{Normal}(Y_0, \lambda d\sigma^2)$  when  $Y_1$  is the trait value for an ancestral node, and  $Y_1 \sim \text{Normal}(Y_0, \lambda d\sigma^2 + (1 - \lambda)d_1\sigma^2)$  for the tips of the tree, where  $d_1$  is the distance from the root to taxon  $Y_1$ . A value of  $\lambda = 1$  then collapses to the Brownian motion model, while  $\lambda = 0$  indicates that the tree has no power to predict evolutionary relatedness.
4. *Pagel's  $\kappa$* : Finally, an analyst might seek to determine whether evolutionary distance is a useful predictor of relatedness. To do so, they replace branch length  $d$  with  $d^\kappa$ , where  $\kappa$  is an estimated non-linearity parameter (Pagel, 1999). In the limit that  $\kappa = 0$ , then all descendants are equally different regardless of evolutionary distance, while  $\kappa = 1$  reverts to the corresponding model but without Pagel's  $\kappa$  correction.

Additional models are reviewed extensively elsewhere (Harmon, 2018; Paradis, 2012), and these reviews typically advise practitioners to use model selection to select among these

options. However, previous research has shown that AIC can inappropriately favour complicated models, that is, phylogenetic breakpoint models that I do not review here (Ho & Ané, 2014). Future studies could instead use cross-validation to optimize predictive performance, although I do not do so here because I later use a jackknife to evaluate predictive performance. Therefore, I instead apply the Akaike information criterion (AIC) to select among the  $2 \times 2 \times 2$  factorial cross that arises from estimating Pagel's  $\lambda$ , Pagel's  $\kappa$ , or Ornstein–Uhlenbeck parameters relative to the null Brownian motion model. I also note that these models are also applicable to multi-generational evolutionary analysis using experimental and field data, sometimes called the “animal model” (Wilson et al., 2010), and this provides another rationale for learning these methods in detail.

### 3 | CONTRASTING PCM AND NESTED TAXONOMIC MODEL

By contrast, many fisheries analyses proceed instead by specifying a linear model for some response trait as a function of other easy-to-measure predictor traits (Gislason et al., 2010; Hoenig, 1983; Pauly, 1980). Using the standard “formula” notation in R (Wilkinson & Rogers, 1973), this might be specified as follows:

$$\text{response} \sim 1 + \text{trait1} + \text{trait2} + \dots \quad (1)$$

and this corresponds to fitting a linear model with estimated slopes:

$$\mu_i = \beta_0 + \sum_{j=1}^{n_j} \beta_j x_{ij} \quad (2)$$

where  $\mu_i$  is the predicted value for the response for sample  $i$ ,  $\beta_0$  is an estimated intercept,  $x_{ij}$  is predictor  $j$  (of  $n_j$  predictors) for sample  $i$ , and  $\beta_j$  is the estimated response for each predictor. However, the residuals around this relationship are often seen to covary by taxa. In response, an analyst might instead apply the “nested taxonomic model” by including additional random effects representing variation in the intercept  $\beta_0$  among orders, families, and genera (Foss-Grant et al., 2016; Zhou et al., 2012). Using the “nlme-formula” notation (Pinheiro & Bates, 2009), this might be specified as follows:

$$\text{response} \sim 1 + \text{trait1} + \text{trait2} + \dots + (1|\text{order}) + (1|\text{family}) + (1|\text{genus}) \quad (3)$$

where *order*, *family*, and *genus* are coded as factors and  $(1|\text{order})$  specifies a model with a random intercept for every level of factor *order*. This then corresponds to a linear mixed model:

$$\mu_i = \sum_{j=1}^{n_j} \beta_j x_{ij} + \sum_{k=1}^{n_k} \varepsilon_k Z_{i,k} \\ \varepsilon \sim \text{MVN}(0, \Sigma) \quad (4)$$

where  $\mathbf{Z}$  is an indicator matrix (containing only 1 and 0s) representing a nested effects structure and the random effects  $\varepsilon$  are estimated with blocked covariance  $\Sigma$ .

I call this latter approach “nested taxonomic levels.” It is closely related to the Brownian motion model when using a tree that is constructed from taxonomy, that is, where a Genus  $Y_0$  is treated as the ancestor of Species  $Y_1$  and  $Y_1 \sim \text{Normal}(Y_0, \sigma^2)$ , such that  $\sigma^2$  corresponds to the estimated variance for species within genus (the (1|genus) term on the right-hand side of Equation 3). In either case, the predicted distribution  $\hat{Y}_1$  for a species with no data will have mean equal to its nearest ancestor, that is,  $\mathbb{E}(\hat{Y}_1) = Y_0$ . For example, using the ultrametric phylogeny for actinopterygians (Rabosky et al., 2018) and retaining only six species, I visualize the covariance among taxa described using the following:

1. *Phylogenetic comparative methods*: The tree (and the resulting correlation among species) resulting from Pagel's  $\lambda$  model, when specifying  $\lambda = 0.8$  for illustration; or
2. *Nested taxonomic levels*: The tree (and resulting correlation) when treating order, family, genus, and species as random effects with the same variance for each hierarchical level.

Comparing these two confirms that the two approaches can result in very similar estimates of the correlation among species (Figure 1). In particular, both methods identify a greater covariance for species within a genus than among genera. In this case, the primary difference between methods is that species in genera *Merluccius* and *Lutjanus* are more closely related to one another than either is to species in genus *Engraulis*, and hence, species in the former genera have some block-wise correlation with one another when using PCM that is otherwise missing from the “nested taxonomic levels” approach. However, as outlined above, PCM also allows a broader set of alternative OU, lambda, and kappa models that can be tested for a given data set.

## 4 | LINEAR AND STRUCTURAL EQUATION MODELS

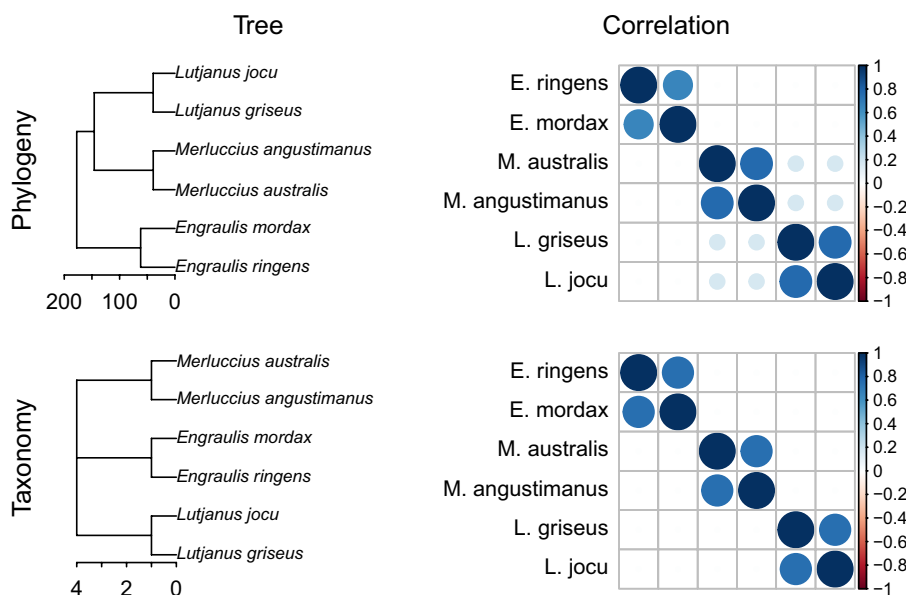
Before proceeding further, however, I also introduce the concept of a “structural equation model” (SEM). PCM is widely implemented using phylogenetic linear models (PLMs), that is, using R package *phylolm* (Tung Ho & Ané, 2014). However, phylogenetic SEMs (Thorson et al., 2023) were recently introduced and can fit any model that is possible using PLMs and additional models that are broadly useful in practice. I therefore introduce SEM and its phylogenetic extension and then list the practical advantages of phylogenetic SEM relative to PLMs.

SEM involves specifying a set of linear dependencies among variables  $\mathbf{y}$ . In particular, it specifies the following:

$$\mathbf{y} = \mathbf{\Gamma}\mathbf{y} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}) \quad (5)$$

where  $\mathbf{\Gamma}$  is a matrix of path coefficients (e.g., where  $\gamma_{2,1}$  in  $\mathbf{\Gamma}$  is the estimated linear effect of variable  $y_1$  on variable  $y_2$ ), and  $\text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$  is a zero-centred multivariate normal density function, where  $\boldsymbol{\Sigma}$  is the exogenous covariance among variables. SEM reduces to a conventional linear model when  $\mathbf{\Gamma}$  is composed of zeros for all rows except that which corresponds to the “dependent” (a.k.a. “response”) variable and when predictor variables are assumed to be measured without error. For example, consider an analysis of the von Bertalanffy growth rate  $K$ , asymptotic maximum length  $L_{\text{inf}}$ , and natural mortality rate  $M$ ,  $\mathbf{y} = \{\log(K), \log(L_{\text{inf}}), \log(M)\}$ . Specifying a linear model using R formula  $\log(M) \sim 1 + \log(K) + \log(L_{\text{inf}})$  could be obtained using a SEM by specifying the following:



**FIGURE 1** Illustration of the additive tree resulting from an ultrametric phylogeny and using a Pagel's  $\lambda$  model ( $\lambda = 0.8$ ) and the resulting correlation among six taxa (top row), compared with an additive tree resulting from a nested taxonomic covariance with four levels representing Class/Family/Genus/Species and assuming equal branch lengths at each level, and its resulting correlation matrix (bottom row). Trees are plotted using function *plot.phylo* in R package *ape* (Paradis & Schliep, 2019).

$$\mathbf{\Gamma} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \gamma_{K \rightarrow M} & \gamma_{L_{\text{inf}} \rightarrow M} & 0 \end{bmatrix} \quad (6)$$

However, the SEM also allows a wide range of alternative relationships among these variables (Mason et al., 2016; von Hardenberg & Gonzalez-Voyer, 2013). For example, SEM allows us to specify a recursive ("cyclical") dependency among variables (e.g.,  $\gamma_{L_{\text{inf}} \rightarrow K}$ ,  $\gamma_{K \rightarrow M}$ , and  $\gamma_{M \rightarrow L_{\text{inf}}}$ ). These recursive models cannot be specified using any sequence of linear models. Despite this greater flexibility, SEM still cannot identify two or more direct effects linking any two variables (e.g.,  $\gamma_{K \rightarrow M}$  and  $\gamma_{M \rightarrow K}$ ) without more information (e.g., a known value for  $\gamma_{K \rightarrow M}$  arising from a separate evolutionary experiment). These restrictions on identifiability are reviewed elsewhere (Kenny et al., 1998), but SEM software (including for phylogenetic SEM) will also generally indicate when a model specification is not identifiable.

Importantly, the SEM model can be used to compute the expected covariance  $\text{Var}(\mathbf{y}) = (\mathbf{I} - \mathbf{\Gamma}^t)^{-1} \mathbf{\Sigma} (\mathbf{I} - \mathbf{\Gamma})^{-1}$ . Conventional SEM then compares this estimated covariance with the sample covariance from a given data set. Assuming data are normally distributed, the sample covariance will follow a Wishart distribution, and this likelihood can be used to identify maximum-likelihood estimates for  $\mathbf{\Gamma}$  and  $\mathbf{\Sigma}$ . Alternatively, phylogenetic SEM is fitted to a matrix  $\mathbf{Y}$  of traits measurements  $y_{ij}$  for each taxon  $i$  and trait  $j$  and specifies the covariance among traits as a separable covariance, that is, the Kronecker product of the evolutionary covariance from PCM and the trait covariance from SEM (Thorson et al., 2023).

Usefully, PSEM is available as a high-level and fully documented R package *phylosem* that is available on CRAN (Thorson & van der Bijl, 2023a) and has been compared extensively with alternative PCM software (Thorson & van der Bijl, 2023b). This package requires three basic inputs:

1. *Structural dependencies using arrow notation*: The user specifies the structure of an SEM using the same format required by R package *sem* (Fox et al., 2020). In the following, I specify model structure using an "arrow notation" (Wright, 1921, 1934). Using this arrow notation, the user writes a text file that uses a one-headed arrow (e.g.,  $X \rightarrow Y$ ) to specify a path coefficient in  $\mathbf{\Gamma}$  and a two-headed arrow (e.g.,  $X \leftrightarrow X$ ) to specify an exogenous covariance in  $\mathbf{\Sigma}$ ;
2. *Evolutionary tree*: Similarly, the user specifies an evolutionary tree using a data structure available in R package *ape*;
3. *Trait data*: Finally, the user specifies a data frame of traits for each taxa, where missing data are indicated by an NA value.

Parameters are estimated using the R package TMB (Kristensen et al., 2016), and output can be converted to standard formats, including standard SEM plots using R package *sem* (Fox et al., 2020), model testing using R package *phylopath* (van der Bijl, 2018), a *phylo4d* object that merges the estimated traits and specified tree (Bolker et al., 2015), or a standard table of coefficients and standard

errors including the intercepts at the tree root (see Figure 2 for visual representation).

This PSEM has several advantages relative to the conventional approach of specifying a linear model to study associations among traits:

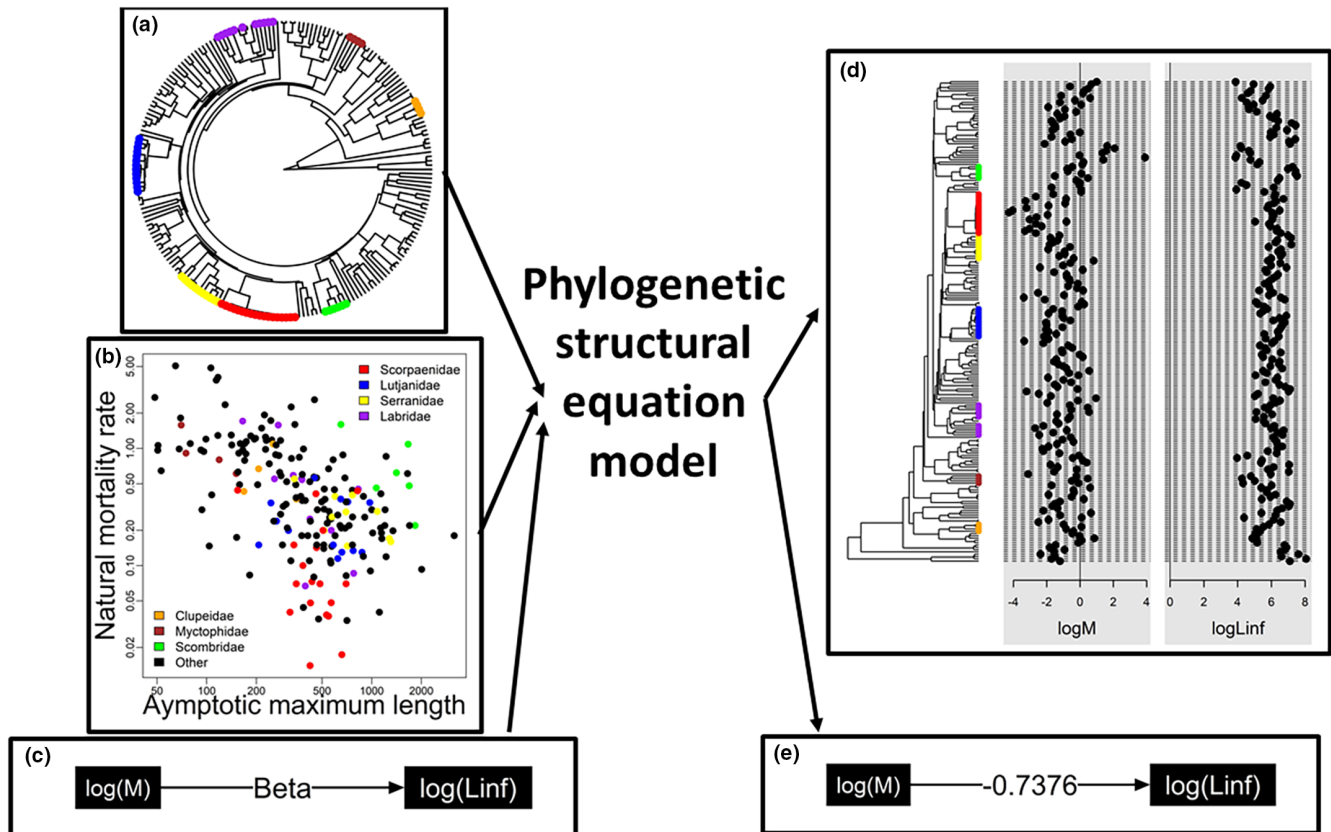
1. *Causal map*: The path diagram represented in  $\mathbf{\Gamma}$  allows an analyst to represent complicated hypotheses about evolutionary mechanisms. For example, if ecological theory and experiments support the claim that trait  $X \rightarrow Y$  and  $Y \rightarrow Z$ , and the analyst estimates these path coefficients using PSEM, then they can calculate the indirect effect of  $X \rightarrow Z$  via the product of these two estimated path coefficients. This allows them to predict how an exogenous change in one trait  $X$  can affect multiple subsequent traits  $Y$  and  $Z$ .
2. *Conditional independence*: Alternatively, an analyst might specify the path diagram for pragmatic reasons instead of using it to represent a causal map. For example, an analyst might want to predict  $Z$  using  $Y$ . However, in cases when  $Y$  is missing, they might have an alternative predictor  $X$ , which they only want to use in the absence of  $Y$ . This outcome can be accomplished by specifying the path diagram  $X \rightarrow Y$  and  $Y \rightarrow Z$ , such that  $Z$  is independent of  $X$  conditional upon a fixed value for  $Y$ .
3. *Missing data*: Finally, the PSEM provides a natural avenue to deal with missing data in comparative studies that specify a linear relationship among variables. In essence, the PSEM imputes missing values as random effects that follow a joint distribution defined by the PCM and SEM. This then solves two practical issues. First, it avoids the need to specify different models for species with a different combination of data, for example, using longevity when available and then reverting to growth parameters otherwise to estimate natural mortality rates (Hamel & Cope, 2022). Second, it avoids the common practice of imputing one value and then using the imputed value in a subsequent regression model. This "daisy chain" of predicted values for missing data is widely used, for example, when imputing values in FishBase (e.g., see Figure 1 of Thorson et al., 2014).

## 5 | CASE STUDY DEMONSTRATION: NATURAL MORTALITY RATES

Beyond these theoretical reasons to prefer PSEM to a conventional linear model for fisheries comparative analysis, I also study whether there is empirical evidence regarding the performance of PSEM vs. conventional linear models.

To assess empirical performance, I download the Then et al. (2015) database, which contains 230 records of natural mortality  $M$ , von Bertalanffy growth rate  $K$ , and asymptotic length  $L_{\text{inf}}$ , as well as longevity  $t_{\text{max}}$ . All records represent direct estimates (instead of predictions resulting from a previous comparative analysis), such that analysing these records will not be a circular use of life-history theory. I also download the most completed ultrametric phylogeny





**FIGURE 2** Conceptual figure showing the inputs (left side) and outputs (right side) for a phylogenetic structural equation model used for phylogenetic comparative analysis. Panel a shows an ultrametric phylogeny (plotted using package *ape* (Paradis & Schliep, 2019) and colouring species from selected families (colour legend in panel b). Panel b shows values for natural mortality (y-axis) and asymptotic maximum length (x-axis) from the Then et al. (2015) database. Panel c shows the specified structural model plotted using *semPlot* (Epskamp, 2022), including a single one-headed arrow “logLinf → logM” and two two-headed arrows representing independent evolutionary variance “logLinf ↔ logLinf” and “logM ↔ logM.” Panel d shows the estimated traits plotted using *phylosignal* (Keck et al., 2016) and again labelling selected families. Panel e shows the estimated path coefficients for the specified structural model (again plotted using *semPlot*).

for bony fishes currently available (Rabosky et al., 2018) and reduce the Then et al. database to the 201 taxa that are matched in both databases. Previous analyses (Hamel & Cope, 2022; Then et al., 2015) have suggested predicting log-mortality rate using log-longevity as the preferred method:

$$\log(M) \sim 1 + \log(t_{\max}) \quad (7)$$

However, the analyses have suggested that when longevity is unavailable a useful backup can be obtained by using growth parameters:

$$\log(M) \sim 1 + \log(K) + \log(L_{\text{inf}}) \quad (8)$$

As a third alternative, I propose using a PSEM. This model is designed to satisfy a few practical goals:

1. When information regarding longevity is available, the model does not use information about growth parameters  $K$  and  $L_{\text{inf}}$  to inform the prediction of mortality rate  $M$ . This implies that  $M$  is independent of  $K$  and  $L_{\text{inf}}$  conditional upon the value of  $t_{\max}$ , that is, there is no arrow directly connecting growth parameters to mortality rate.

2. Where possible, the model will estimate coefficients for  $\log(K)$  and  $\log(L_{\text{inf}})$  affecting other parameters. This then allows us to compare estimates of path coefficients with regression models that commonly include  $\log(K)$  and  $\log(L_{\text{inf}})$  as predictor variables.

These two goals then result in the PSEM that can be written using SEM “arrow notation” (Wright, 1921, 1934) as follows:

$$\log(K) \rightarrow \log(t_{\max})$$

$$\log(L_{\text{inf}}) \rightarrow \log(t_{\max})$$

$$\log(t_{\max}) \rightarrow \log(M) \quad (9)$$

In a loose sense, this specification treats  $\log(M)$  and  $\log(t_{\max})$  as more closely related than the other life-history parameters, and it therefore predicts that  $\log(M)$  will be more correlated with  $\log(t_{\max})$  than with the other parameters. This PSEM can be fitted either with an available longevity measurement or not, and I separately explore performance in these two circumstances. Future studies could explore alternative specifications involving a larger range of variables (e.g., Griffiths & Harrod, 2007; Thorson et al., 2023), although I restrict myself to these

variables to facilitate comparison with other analyses of this database (Hamel & Cope, 2022; Then et al., 2015).

## 5.1 | Evaluating case study performance

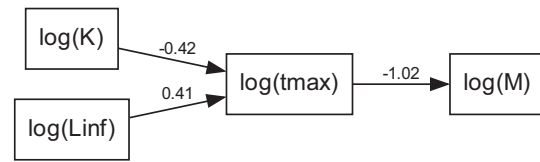
I first fit the PSEM (Equation 9) to all records and apply AIC to the  $2 \times 2 \times 2$  factorial cross of Pagel's  $\lambda$ , Pagel's  $\kappa$ , or Ornstein–Uhlenbeck transformations. I then proceed using this AIC-selected structure for the PSEM and compare its performance against the two linear models (Equations 7 and 8) in two scenarios (representing different types of missing data) using a jackknife experiment. This performance evaluation involves the following steps:

1. Replace the measurement of  $\log(M)$  with NA for the 1st record, fit the longevity method (Equation 7) and the PSEM (Equation 9) to all data, and record the predicted value for  $\log(M)$  for the 1st record for the linear model (termed “lmA”) or the PSEM (termed “psemA”);
2. Replace the measurements of both  $\log(M)$  and  $\log(t_{\max})$  with NAs for the 1st record, fit the growth-parameter method (Equation 8) and the PSEM (Equation 9) to all data, and again record the predicted value for  $\log(M)$  for the 1st record for the linear model (termed “lmB”) or the PSEM (termed “psemB”);
3. Repeat steps 1–2 for each record of the database;
4. Compare the true but withheld value of  $\log(M)$  from the original database with the predicted value.

I then compare the bias and mean-squared error of lmA, lmB, psemA, and psemB when averaging across all records. In particular, I compute the percent variance explained (PVE) for each method as follows:

$$\text{PVE} = \left( 1 - \frac{\sum_{i=1}^{n_i} (\log(\hat{M}_i) - \log(M_i))^2}{\sum_{i=1}^{n_i} (\log(M_i) - \mu)^2} \right) \times 100\% \quad (10)$$

where  $\log(M_i)$  is the withheld value and  $\log(\hat{M}_i)$  is the estimated value for record  $i$ ;  $\mu = n_i^{-1} \sum_{i=1}^{n_i} \log(M_i)$  is the true mean, and I have  $n_i = 201$  records for the jackknife experiment.  $\text{PVE} = 0\%$  indicates that a given method provides no improvement relative to using mean  $\mu$  as prediction, and  $\text{PVE} = 100\%$  indicates a perfect prediction with no residual errors. As sensitivity analysis, I also compare PSEM performance against a reduced PSEM that estimates two path coefficients,  $\log(L_{\text{inf}}) \rightarrow \log(M)$  and  $\log(K) \rightarrow \log(M)$ , i.e., bypasses any prediction of  $\log(t_{\max})$ . This sensitivity analysis then serves as bridge between



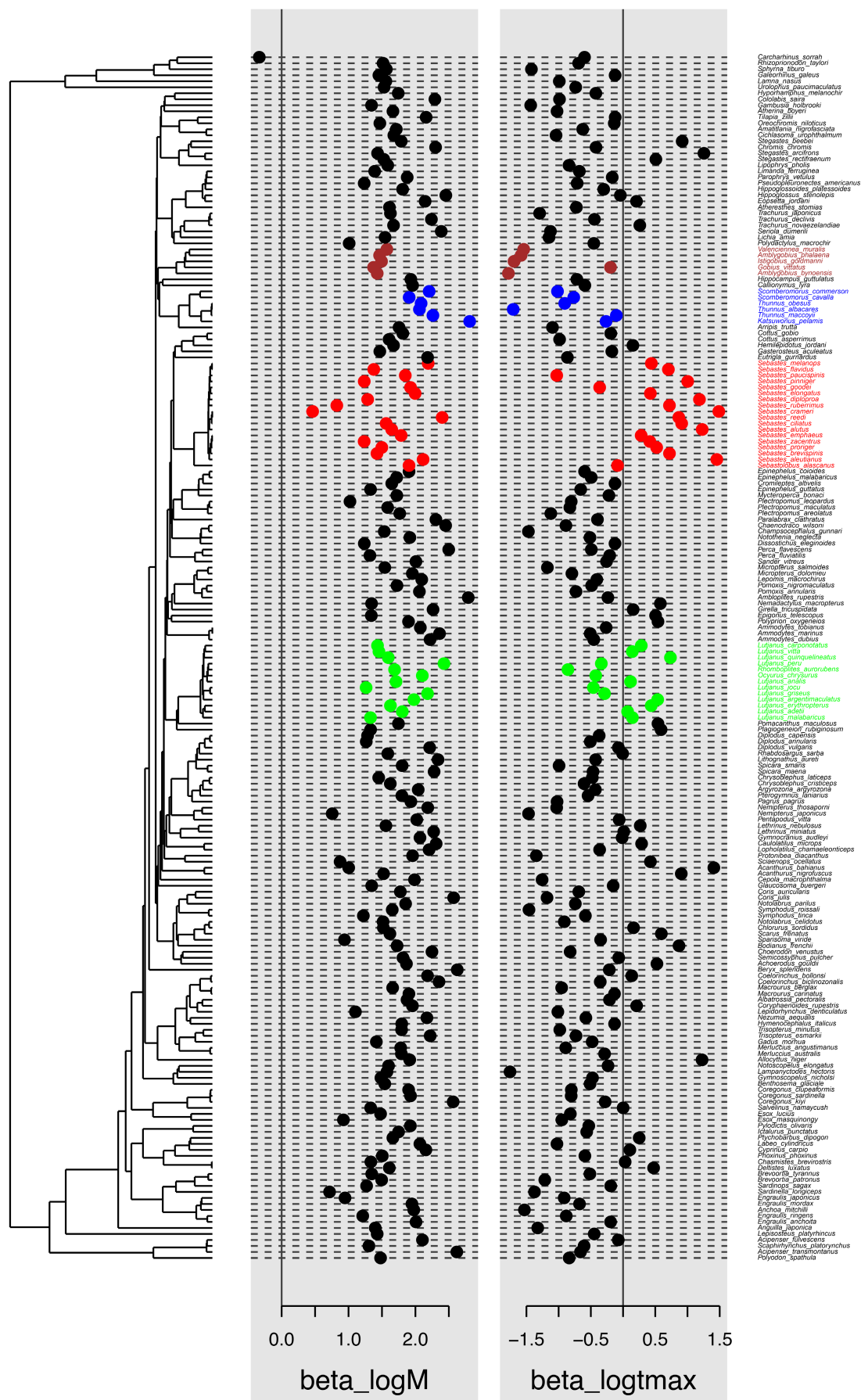
**FIGURE 3** Illustration of a structure equation life-history model fitted using *phylosem*, specified to ensure that  $\log(M)$  is independent of growth parameters  $\log(K)$  and  $\log(L_{\text{inf}})$  conditional upon longevity  $\log(t_{\max})$ . This model then has the behaviour of reverting to the longevity-based method when a longevity measurement is available, but using life-history information when not. Plot created using function *pathDiagram* in package *sem* (Fox et al., 2020).

the full PSEM (Equation 9) and the linear model using growth parameters (Equation 8).

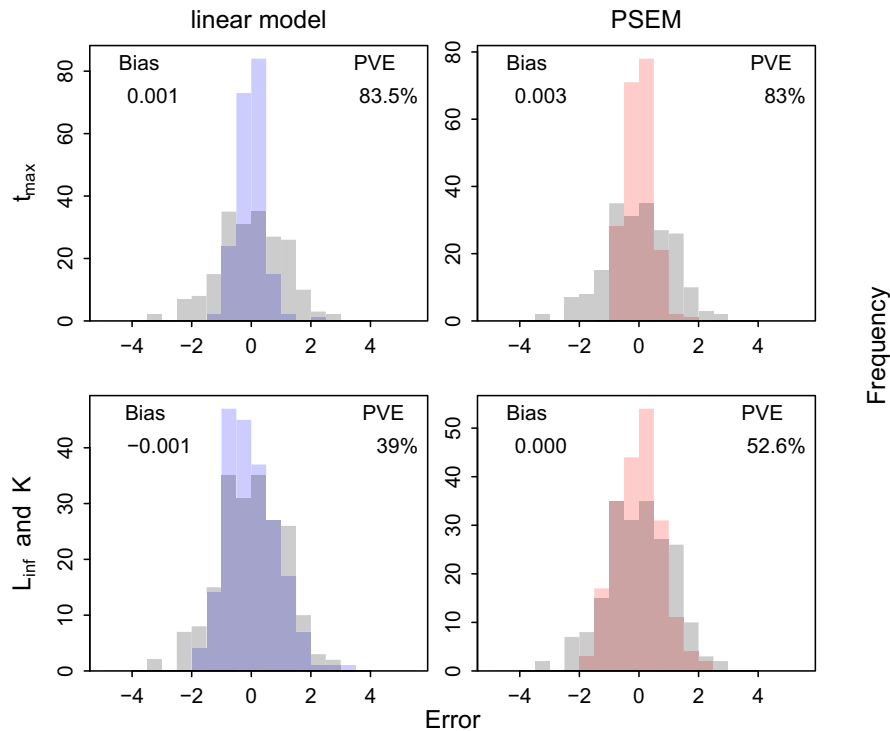
AIC identifies Pagel's  $\lambda$  as parsimonious ( $\lambda = 0.93$ ), suggesting that evolutionary relatedness predicts 93% of the covariance among species, and also identifies Pagel's  $\kappa$  as parsimonious ( $\kappa = 0.67$ ), suggesting that closely related species are more dissimilar than otherwise expected under a random-walk evolutionary process. As expected, mortality rate is essentially the reciprocal of longevity, that is, a  $-1.02$  slope for log-mortality and log-longevity (Figure 3). Increasing growth rate and decreasing asymptotic length are both associated with a decrease in longevity and a resulting increase in mortality rate. These estimated patterns are consistent with past analyses of this data set (Hamel & Cope, 2022; Then et al., 2015). However, the estimated intercept of the PSEM varies across the ultrametric phylogeny. Specifically, the estimated intercept used when predicting  $\log(M)$  as a function of  $\log(t_{\max})$  generally varies between 1 and 2.5 for all taxa, and visual inspection suggests that there is little phylogenetic signal (Figure 4, left column). By contrast, the estimated intercept used when predicting  $\log(t_{\max})$  as a function of  $\log(K)$  and  $\log(L_{\text{inf}})$  varies systematically among taxa (Figure 4, right column). For example, family Scorpaenidae (rockfishes) generally has a larger intercept and families Scombridae (tunas) and Gobiidae (gobies) have smaller intercepts than average. As a result, Scorpaenidae typically has a higher longevity (and lower mortality rate) than otherwise expected on the basis of growth parameters alone, and the opposite holds for Scombridae and Gobiidae.

Regarding model performance, the jackknife analysis suggests that a linear model using longevity (lmA in Equation 7) or the PSEM that includes longevity for a given species (psemA) both explain 83% of variance in the withheld value for log-mortality rate (Figure 5). However, when longevity is also withheld, the jackknife analysis suggests that the linear model using growth parameters (lmB in Equation 8) explains 39% of variance. By

**FIGURE 4** Ultrametric phylogeny (tree on left-hand side) and predicted variation in the intercept in the PSEM (Figure 3) when predicting  $\log(M)$  a function of  $\log(t_{\max})$  (left grey column) or  $\log(t_{\max})$  as a function of  $\log(K)$  and  $\log(L_{\text{inf}})$  (right grey column) for each of 201 taxa (right column) that are matched between the Then et al. (2015) longevity database and the Rabosky et al. (2018) tree, and highlighting individual families (Scorpaenidae: red; Scombridae: blue; Lutjanidae: green; Gobiidae: brown) to illustrate phylogenetic signal. Plot created using function *dotplot.phylo4d* in R package *phylosignal* (Keck et al., 2016).







**FIGURE 5** Histogram illustrating jackknife results for 201 taxa that can be matched against the phylogenetic tree (merging the Stein and modelled-Robosky trees), that is, showing prediction errors relative to the withheld value of log-natural mortality,  $Error_i = \log(\hat{M}_i) - \log(M_i)$ , either using longevity  $t_{max}$  for species  $i$  and instead using growth parameters  $K$  and  $L_{inf}$  (bottom row), evaluating estimates from a linear model (blue histogram, left column) or PSEM (red histogram, right column), and comparing each against the raw variation relative to the mean value  $\log(M_i) - n_i^{-1} \sum_{j=1}^{n_i} \log(M_j)$  (grey histogram), and listing the per cent variance explained ("PVE," top-right) and the bias (top-left) for each model.

contrast, the PSEM without longevity explains 53% of variance, that is, performance that is intermediate between the linear model with and without longevity information. Collectively, these results suggest that there is little useful phylogenetic information in the relationship between longevity and log-mortality (as also seen in the estimated intercepts, which show little phylogenetic pattern). However, the estimated intercept when using growth parameters is phylogenetically conserved, and conditioning predictions on this information allow for a substantial improvement relative to phylogenetic unconditional prediction. The sensitivity analysis PSEM (which bypasses longevity when using growth parameters to predict mortality) has almost identical performance to the full PSEM (Figure S1). This suggests that the full PSEM can be fitted and then used to predict mortality rates with or without longevity information, and these yield alternative "estimators" for mortality. Finally, the improved performance for PSEM relative to the conventional linear model remains when using a tree build from taxonomy (Figure S2), so I conclude that the residual correlations can be represented using either phylogeny or taxonomy in this case study.

## 6 | TREES FOR FISHES: REMOVING BARRIERS TO PASSAGE

In this paper, I have briefly introduced the large field of phylogenetic comparative methods (PCMs), which has received surprisingly little attention in fisheries science. I specifically contrast PCM with the nested taxonomic models that are commonly used in fisheries science, while also outlining the basic evolutionary models (OU

and Pagel's transformations) that can extend this approach. I then introduced phylogenetic structural equation models (PSEMs) and provided a real-world example where PSEM can improve predictive performance relative to current practices. I now discuss some of the outstanding obstacles to using PCM and PSEM in fisheries, highlighting the following: (1) lack of experience with PCM databases and analytical software; (2) lack of access to a comprehensive evolutionary tree; and (3) continued assembly of a comprehensive and high-quality database for life-history traits. I also discuss research to address these below.

Most obviously, PCM and PSEM require that the analyst specifies an evolutionary tree that represents relatedness, which is then used to estimate trait covariance. The Rabosky et al. (2018) tree for ray-finned fishes is available using R package "fishtree." It contains a phylogeny for 11,638 fishes based on direct (genetic) information or a modelled phylogeny for 31,516 fishes that is reconstructed using taxonomic information. The Stein et al. (2018) tree for Chondrichthyes includes a further 1192 fishes. Merging the Stein and modelled-Rabosky trees then represents approximately 96% of the >34,000 described fishes. However, there are logistical challenges in matching these trees against other fishery data sets, given inconsistent conventions for referencing species. For example, of the 386 species represented in the RAM Legacy Stock Assessment Database v4.495 (<https://zenodo.org/record/4824192>), accessing the Excel version and downloading the Taxonomy tab (Ricard et al., 2012), 301 are matched within the merged tree (i.e., 78%). As alternative, it is possible to use taxonomy to construct a tree and then proceed with PCM as per usual, and this ensures that the coverage is complete for any described species. However, this does not propagate information about

evolutionary relatedness when defining the correlation among taxa (see Figure 1). I therefore recommend adding a WoRMS identifier for all species listed in major fisheries data sets (e.g., the RAM Legacy Database) to facilitate merges with evolutionary trees. However, in some cases the additional information from a phylogeny might not improve performance relative to using taxonomy as a proxy to build a tree. For example, in our application the taxonomic and phylogenetic trees result in similar PCM performance (i.e., comparing Figure 5 and Figure S2).

Another impediment to increased use of PCM software is the lack of familiarity with major packages either for data access, visualization, or analysis. These tools have been extensively reviewed elsewhere (Harmon, 2018), so I focus here on outlining how a fisheries scientist could begin. I specifically recommend that fisheries scientists conducting research in comparative methods become familiar with the R package *ape* (Paradis & Schliep, 2019) or any later replacement, which includes functions to read, write, prune, merge, and plot trees, as well as converting taxonomy to a tree. I also recommend that fisheries scientists receive training in basic phylogenetic generalized linear models (PGLMs) using R package *phylolm* (Tung Ho & Ané, 2014) or any later replacement. Although I have emphasized PSEM in this study, I recommend further research to compare PSEM and simpler PGLMs for common types of fisheries meta-analysis.

Finally, I note that fisheries agencies, professional societies, and researchers must continue to invest in compiling comprehensive data for life-history, reproductive, metabolic, trophic, morphometric, and behavioural traits. FishBase was an early leader in compiling life-history information across all ecological taxa. However, researchers for other taxa have embarked on ambitious and ongoing efforts to compile ever-larger trait databases. Meanwhile, there are many databases of fish traits that are not available in FishBase or any other unified trait database as of April 2023, for example:

1. Species temperature index (STI) representing observed thermal preferences derived from habitat utilization (Burrows et al., 2019);
2. Experimental measurements of thermal tolerances for egg, larvae, juvenile, adult, and spawning stages of fishes (Dahlke et al., 2020);
3. Resting and active metabolic rates and resulting resting oxygen demand (Deutsch et al., 2020);
4. Morphometric information regarding body shape from museum records (Price et al., 2019);
5. Fisheries management reference points and proxies obtained from stock assessments (Ricard et al., 2012; Zhou et al., 2012).

I therefore recommend ongoing investment by the research community to ensure that these data are cross-linked (e.g., using taxonomic IDs from WoRMS) and easily available for researchers. I anticipate that consolidated research will continue to identify strong correlations among these traits (Thorson et al., 2023; Winemiller et al., 2015), allowing phylogenetic trait imputation to predict their values for use in fisheries research and management worldwide.

## ACKNOWLEDGEMENTS

I thank J. Sullivan and J. Cope for comments on an earlier manuscript draft and S. Barbeaux and M. Bryan for help in testing the analysis for real-world taxa. I also thank J. Hoenig and A. Then for making the Then et al. (2015) database publicly available and B. Babcock for previous discussions of the “nested taxonomy” model. Finally, I thank the many co-authors on previous papers developing FishLife and package *phylosem*, which has shaped my thinking regarding these topics.

## DATA AVAILABILITY STATEMENT

I have used the Then et al. (2015) database V1.0, available publicly online ([https://www.vims.edu/research/departments/fisheries/programs/mort\\_db/index.php](https://www.vims.edu/research/departments/fisheries/programs/mort_db/index.php)), and the Rabosky evolutionary tree (Rabosky et al., 2018), available via R package *fishTree* (Chang et al., 2019). I have also used information regarding length-at-maturity from FishBase (Froese, 1990) as compiled within FishLife (Thorson et al., 2023). Finally, I have conducted all analyses using R package *phylosem* release 1.1.0 (Thorson & van der Bijl, 2023b), publicly available via CRAN (Thorson & van der Bijl, 2023a) and Zenodo (<https://zenodo.org/record/8422292>). Simple code to replicate the analysis is available as vignette titled “PCM in fisheries.”

## ORCID

James T. Thorson  <https://orcid.org/0000-0001-7415-1010>

## REFERENCES

- Bolker, B., Butler, M., Cowan, P., de Vienne, D., Eddelbuettel, D., Holder, M., Jombart, T., Kembel, S., Michonneau, F., & Orme, B. (2015). *phylobase: Base package for phylogenetic structures and comparative data. R Package Version 0.8.0*.
- Burrows, M. T., Bates, A. E., Costello, M. J., Edwards, M., Edgar, G. J., Fox, C. J., Halpern, B. S., Hiddink, J. G., Pinsky, M. L., Batt, R. D., García Molinos, J., Payne, B. L., Schoeman, D. S., Stuart-Smith, R. D., & Poloczanska, E. S. (2019). Ocean community warming responses explained by thermal affinities and temperature gradients. *Nature Climate Change*, 9(12), 959–963. <https://doi.org/10.1038/s41558-019-0631-5>
- Chang, J., Rabosky, D. L., Smith, S. A., & Alfaro, M. E. (2019). An R package and online resource for macroevolutionary studies using the ray-finned fish tree of life. *Methods in Ecology and Evolution*, 10(7), 1118–1124. <https://doi.org/10.1111/2041-210X.13182>
- Clark, W. G. (2002). F 35% revisited ten years later. *North American Journal of Fisheries Management*, 22(1), 251–257.
- Dahlke, F. T., Wohlrab, S., Butzin, M., & Pörtner, H.-O. (2020). Thermal bottlenecks in the life cycle define climate vulnerability of fish. *Science*, 369(6499), 65–70. <https://doi.org/10.1126/science.aaz3658>
- Deutsch, C., Penn, J. L., & Seibel, B. (2020). Metabolic trait diversity shapes marine biogeography. *Nature*, 585(7826), 557–562. <https://doi.org/10.1038/s41586-020-2721-y>
- Drummond, A. J., & Suchard, M. A. (2010). Bayesian random local clocks, or one rate to rule them all. *BMC Biology*, 8(1), 114. <https://doi.org/10.1186/1741-7007-8-114>
- Epskamp, S. (2022). *semPlot: Path Diagrams and Visual Analysis of Various SEM Packages* (1.1.6) [Computer software]. <https://CRAN.R-project.org/package=semPlot>
- Erickson, K. A., & Nadon, M. O. (2021). An extension of the stepwise stochastic simulation approach for estimating distributions of missing life history parameter values for sharks, groupers, and other taxa. *Fishery Bulletin*, 119(1), 77–114.

- Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125(1), 1–15. <https://doi.org/10.1086/284325>
- Foss-Grant, A. P., Zipkin, E. F., Thorson, J. T., Jensen, O. P., & Fagan, W. F. (2016). Hierarchical analysis of taxonomic variation in intraspecific competition across fish species. *Ecology*, 97(7), 1724–1734. <https://doi.org/10.1890/15-0733.1>
- Fox, J., Nie, Z., & Byrnes, J. (2020). *Sem: Structural equation models. R package version 3.1-11* [Computer software]. <https://CRAN.R-project.org/package=sem>
- Froese, R. (1990). FishBase: An information system to support fisheries and aquaculture research. *ICLARM Fishbyte*, 8(3), 21–24.
- Froese, R., Thorson, J. T., & Reyes, R. B. (2014). A Bayesian approach for estimating length-weight relationships in fishes. *Journal of Applied Ichthyology*, 30(1), 78–85. <https://doi.org/10.1111/jai.12299>
- Gislason, H., Daan, N., Rice, J. C., & Pope, J. G. (2010). Size, growth, temperature and the natural mortality of marine fish. *Fish and Fisheries*, 11(2), 149–158.
- Griffiths, D., & Harrod, C. (2007). Natural mortality, growth parameters, and environmental temperature in fishes revisited. *Canadian Journal of Fisheries and Aquatic Sciences*, 64(2), 249–255. <https://doi.org/10.1139/f07-002>
- Hamel, O. S., & Cope, J. M. (2022). Development and considerations for application of a longevity-based prior for the natural mortality rate. *Fisheries Research*, 256, 106477. <https://doi.org/10.1016/j.fishres.2022.106477>
- Harmon, L. J. (2018). *Phylogenetic comparative methods: Learning from trees* (1.0 edition). CreateSpace Independent Publishing Platform.
- Ho, L. S. T., & Ané, C. (2014). Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*, 5(11), 1133–1146. <https://doi.org/10.1111/2041-210X.12285>
- Hoenig, J. M. (1983). Empirical use of longevity data to estimate mortality rates. *Fishery Bulletin*, 82(4), 898–903.
- Keck, F., Rimet, F., Bouchez, A., & Franc, A. (2016). phylosignal: An R package to measure, test, and explore the phylogenetic signal. *Ecology and Evolution*, 6(9), 2774–2780. <https://doi.org/10.1002/ece3.2051>
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 233–265). McGraw-Hill.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70(5), 1–21. <https://doi.org/10.18637/jss.v070.i05>
- Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 30, 314–334.
- Mangel, M., MacCall, A. D., Brodziak, J. K., Dick, E. J., Forrest, R. E., Pourzand, R., & Ralston, S. (2013). A perspective on steepness, reference points, and stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, 70(6), 930–940.
- Martins, E. P. (2000). Adaptation and the comparative method. *Trends in Ecology & Evolution*, 15(7), 296–299. [https://doi.org/10.1016/S0169-5347\(00\)01880-2](https://doi.org/10.1016/S0169-5347(00)01880-2)
- Mason, C. M., Goolsby, E. W., Humphreys, D. P., & Donovan, L. A. (2016). Phylogenetic structural equation modelling reveals no need for an 'origin' of the leaf economics spectrum. *Ecology Letters*, 19(1), 54–61. <https://doi.org/10.1111/ele.12542>
- Melnichuk, M. C., Essington, T. E., Branch, T. A., Heppell, S. S., Jensen, O. P., Link, J. S., Martell, S. J. D., Parma, A. M., Pope, J. G., & Smith, A. D. M. (2012). Can catch share fisheries better track management targets? *Fish and Fisheries*, 13(3), 267–290.
- Myers, R. A. (2001). Stock and recruitment: Generalizations about maximum reproductive rate, density dependence, and variability using meta-analytic approaches. *ICES Journal of Marine Science: Journal du Conseil*, 58(5), 937–951. <https://doi.org/10.1006/jmsc.2001.1109>
- Myers, R. A., & Barrowman, N. J. (1996). Is fish recruitment related to spawner abundance? *Fishery Bulletin*, 94(4), 707–724.
- Nadon, M. O., & Ault, J. S. (2016). A stepwise stochastic simulation approach to estimate life history parameters for data-poor fisheries. *Canadian Journal of Fisheries and Aquatic Sciences*, 73(999), 1–11.
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401(6756), 877–884. <https://doi.org/10.1038/44766>
- Paradis, E. (2012). *Analysis of phylogenetics and evolution with R* (Vol. 2). Springer.
- Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Pauly, D. (1980). On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. *Journal du Conseil International Pour l'Exploration de la Mer*, 39(2), 175–192. <https://doi.org/10.1093/icesjms/39.2.175>
- Pinheiro, J., & Bates, D. (2009). *Mixed-effects models in S and S-PLUS*. Springer.
- Price, S. A., Friedman, S. T., Corn, K. A., Martinez, C. M., Larouche, O., & Wainwright, P. C. (2019). Building a body shape morphospace of teleostean fishes. *Integrative and Comparative Biology*, 59(3), 716–730. <https://doi.org/10.1093/icb/icz115>
- Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M., & Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, 559(7714), 392–395. <https://doi.org/10.1038/s41586-018-0273-1>
- Ricard, D., Minto, C., Jensen, O. P., & Baum, J. K. (2012). Examining the knowledge base and status of commercially exploited marine species with the RAM legacy stock assessment database. *Fish and Fisheries*, 13(4), 380–398. <https://doi.org/10.1111/j.1467-2979.2011.00435.x>
- Stayton, C. T. (2015). The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution*, 69(8), 2140–2153. <https://doi.org/10.1111/evo.12729>
- Stein, R. W., Mull, C. G., Kuhn, T. S., Aschliman, N. C., Davidson, L. N. K., Joy, J. B., Smith, G. J., Dulvy, N. K., & Mooers, A. O. (2018). Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. *Nature Ecology & Evolution*, 2(2), 288–298. <https://doi.org/10.1038/s41559-017-0448-4>
- Then, A. Y., Hoenig, J. M., Hall, N. G., Hewitt, D. A., & Handling editor: Ernesto Jardim. (2015). Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. *ICES Journal of Marine Science*, 72(1), 82–92. <https://doi.org/10.1093/icesjms/fsu136>
- Thorson, J. T. (2020). Predicting recruitment density dependence and intrinsic growth rate for all fishes worldwide using a data-integrated life-history model. *Fish and Fisheries*, 21(2), 237–251. <https://doi.org/10.1111/faf.12427>
- Thorson, J. T., Cope, J. M., & Patrick, W. S. (2014). Assessing the quality of life history information in publicly available databases. *Ecological Applications*, 24(1), 217–226. <https://doi.org/10.1890/12-1855.1>
- Thorson, J. T., Maureaud, A. A., Frelat, R., Mérigot, B., Bigman, J. S., Friedman, S. T., Palomares, M. L. D., Pinsky, M. L., Price, S. A., & Wainwright, P. (2023). Identifying direct and indirect associations among traits by merging phylogenetic comparative methods and structural equation models. *Methods in Ecology and Evolution*, 14(5), 1259–1275. <https://doi.org/10.1111/2041-210X.14076>
- Thorson, J. T., & van der Bijl, W. (2023a). *phylosem: Phylogenetic Structural Equation Model* [Computer software]. <https://CRAN.R-project.org/package=phylosem>
- Thorson, J. T., & van der Bijl, W. (2023b). *phylosem: A fast and simple R package for phylogenetic inference and trait imputation using phylogenetic structural equation models. Journal of Evolutionary Biology*, 36(10), 1357–1364. <https://doi.org/10.1111/jeb.14234>

- Tung Ho, L.s., & Ané, C. (2014). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology*, 63(3), 397–408.
- Uyeda, J. C., Zenil-Ferguson, R., & Pennell, M. W. (2018). Rethinking phylogenetic comparative methods. *Systematic Biology*, 67(6), 1091–1109. <https://doi.org/10.1093/sysbio/syy031>
- van der Bijl, W. (2018). phylopath: Easy phylogenetic path analysis in R. *PeerJ*, 6, e4718. <https://doi.org/10.7717/peerj.4718>
- von Hardenberg, A., & Gonzalez-Voyer, A. (2013). Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. *Evolution; International Journal of Organic Evolution*, 67(2), 378–387. <https://doi.org/10.1111/j.1558-5646.2012.01790.x>
- Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 22(3), 392–399.
- Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. B., & Nussey, D. H. (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology*, 79(1), 13–26. <https://doi.org/10.1111/j.1365-2656.2009.01639.x>
- Winemiller, K. O., Fitzgerald, D. B., Bower, L. M., & Pianka, E. R. (2015). Functional traits, convergent evolution, and periodic tables of niches. *Ecology Letters*, 18(8), 737–751.
- Worm, B., Hilborn, R., Baum, J. K., Branch, T. A., Collie, J. S., Costello, C., Fogarty, M. J., Fulton, E. A., Hutchings, J. A., Jennings, S., Jensen, O. P., Lotze, H. K., Mace, P., McClanahan, T. R., Minto, C., Palumbi, S. R., Parma, A. M., Ricard, D., Rosenberg, A. A., ... Zeller, D. (2009). Rebuilding global fisheries. *Science*, 325(5940), 578–585.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20(7), 557–585.
- Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, 5(3), 161–215.
- Yoder, J. B., Clancey, E., Des Roches, S., Eastman, J. M., Gentry, L., Godsoe, W., Hagey, T. J., Jochimsen, D., Oswald, B. P., & Robertson, J. (2010). Ecological opportunity and the origin of adaptive radiations. *Journal of Evolutionary Biology*, 23(8), 1581–1596.
- Zhou, S., Yin, S., Thorson, J., Smith, T., & Fuller, M. (2012). Linking fishing mortality reference points to life history traits: An empirical study. *Canadian Journal of Fisheries and Aquatic Sciences*, 69(8), 1292–1301.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Thorson, J. T. (2024). Trees for fishes: The neglected role for phylogenetic comparative methods in fisheries science. *Fish and Fisheries*, 25, 168–179. <https://doi.org/10.1111/faf.12800>