Phylogenetically weighted regression: A method for modelling non-stationarity on evolutionary trees

T. Jonathan Davies1,2,3,4 | James Regetz5 | Elizabeth M. Wolkovich4,6,7 | Brian J. McGill8

1Department of Biology, McGill University, Montreal, Quebec, Canada
2African Centre for DNA Barcoding, University of Johannesburg, Johannesburg, South Africa
3Departments of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
4Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
5National Center for Ecological Analysis and Synthesis, Santa Barbara, California
6Arnold Arboretum, Boston, Massachusetts
7Organismic & Evolutionary Biology, Harvard University, Cambridge, Massachusetts
8School of Biology and Ecology & Sustainability, Solutions Initiative, University of Maine, Orono, Maine

Correspondence
T. Jonathan Davies, Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.
Email: j.davies@ubc.ca

Funding information
Division of Environmental Biology, Grant/Award Number: 0072909

Editor: Andrew Kerkhoff

Abstract

Aim: Closely related species tend to resemble each other in their morphology and ecology because of shared ancestry. When exploring correlations between species traits, therefore, species cannot be treated as statistically independent. Phylogenetic comparative methods (PCMs) attempt to correct statistically for this shared evolutionary history. Almost all such approaches, however, assume that correlations between traits are constant across the tips of the tree, which we refer to as phylogenetic stationarity. We suggest that this assumption of phylogenetic stationarity might be often violated and that relationships between species traits might evolve alongside clades, for example, owing to the effects of unmeasured traits or other latent variables. Specific examples range from shifts in allometric scaling relationships between clades (e.g., basal metabolic rate and body mass in endotherms, and tree diameter and biomass in trees) to the differing relationship between leaf mass per area and shade tolerance in deciduous versus evergreen trees and shrubs.

Innovation: Here, we introduce an exploratory modelling framework, phylogenetically weighted regression, which represents an extension of geographically weighted regression (GWR) used in spatial studies, to allow non-stationarity in model parameters across a phylogenetic tree. We demonstrate our approach using empirical data on flowering time and seed mass from a well-studied plant community in southeastern Sweden. Our model reveals strong, diverging trends across the phylogeny, including changes in the sign of the relationship between clades.

Main conclusions: By allowing for phylogenetic non-stationarity, we are able to detect shifting relationships among species traits that would be obscured in traditional PCMs; thus, we suggest that PWR might be an important exploratory tool in the search for key missing variables in comparative analyses.

KEYWORDS
flowering time, geographically weighted regression, latent variables, non-stationarity, phylogenetic comparative methods, phylogenetically weighted regression, trait correlations
1 | INTRODUCTION

Beginning in the 1980s, a growing recognition of statistical non-independence among species contributed to the rise of phylogenetic comparative biology and its related methods [phylogenetic comparative methods (PCMs), Harvey & Pagel, 1991]. Most widely used PCMs, including phylogenetically independent contrasts (Felsenstein, 1985) and Grafen’s standard regression (Grafen, 1989), assume a model of evolutionary change to correct for the non-independence of taxa by estimating the expected divergence in trait values as a function of time. For continuous traits, a Brownian motion model of evolution is frequently assumed, in which trait variance increases linearly with the sum of phylogenetic branch lengths between taxa on a dated phylogenetic tree (Felsenstein, 1985), resulting in a multivariate normal distribution of trait values at the tips of the tree. These important methods contain various assumptions of stationarity, including stationarity of rates of evolution, stationarity of covariances and stationarity of model coefficients (correlation structure between traits).

Real-world data are often not truly stationary. Recognition of this non-stationarity is playing a growing role in multiple fields, from hydrology (Jain & Lall, 2001; Milly et al., 2008) to ecology (da Silva Cassemiho, De Souza Barreto, Rangel, & Diniz-Filho, 2007; Foody, 2004; Wimberly, Yabsley, Baer, Dugan, & Davidson, 2008). As a result, an increasing number of techniques have been developed to deal with non-stationarity. Mixed and hierarchical models allow assumptions of constant variance to be relaxed. Recently, statistical approaches that relax assumptions of stationarity to accommodate local variation over space (Cho, Lambert, & Chen, 2010; Fotheringham, Brunsdon, & Charlton, 2002) or time (Maurer, Koch, Hammerl, Hammerl, & Pokorny, 2009; Schleip, Rutishauser, Luterbacher, & Menzel, 2008; Torrence & Compo, 1998; Torrence & Webster, 1999) have become available, permitting exploration of important shifts in trends.

Many forms of evolutionary non-stationarity are addressed by new phylogenetic models. Examples include the use of maximum likelihood approaches to find the best transformation of the underlying variance–covariance matrix to model the correlation structure of traits and tree (e.g., Freckleton, Harvey, & Pagel, 2002; Martins & Hansen, 1997; Mazel et al., 2016), and allowing deviation from strict Brownian assumptions and variation in evolutionary rates of character divergence across the branches of the phylogenetic tree (Beaulieu, Jhwueng, Boettiger, & O’Meara, 2012; Butler & King, 2004; Harmon et al., 2010; O’Meara, Ané, Sanderson, Wainwright, & Hansen, 2006; Revell, Mahler, Peres-Neto, & Redelings, 2012). Nonetheless, unless modelled explicitly, most PCMs still assume that correlations between traits are constant across the tips of the tree (and, indeed, among ancestral lineages), an assumption we refer to here as phylogenetic stationarity [but see the recent paper by Uyeda, Pennell, Miller, Maia, & McClain (2017) for an important exception].

An assumption of phylogenetic stationarity may be perfectly justifiable; for example, when shifts in the relationship between traits are either weak or simply represent noise. However, we suggest that important non-stationarity in the correlation between traits across a phylogeny might be common and biologically meaningful in some datasets [see Supporting Information Appendix S2, Box S1 and Table S1]. Specifically, phylogenetic non-stationarity might be common when there are interactions (in a statistical regression sense) between several variables, only some of which are being analysed. As a simple example, consider a scenario where life habit (annual versus perennial) evolves according to a random walk. Now consider two additional traits, seed size and first flowering date. If, in a perennial plant, an early first flowering date leads to an early release of seeds, requiring the seed to have a longer juvenile stage to survive, this could be counter-balanced by having larger seeds. This offset would result in a negative relationship between first flowering date and seed mass (early flower = large seed). However, in annual plants there is often greater reliance on seed banking of seeds that remain dormant underground for several years before germinating, allowing large-seeded plants to grow quickly and reach an early flowering date, once they have germinated. As a consequence, the correlation between first flowering date and seed mass would be positive in regions of the phylogeny that have evolved an annual life habit and negative in regions that have evolved a perennial life habit. This is not a hypothetical example. Although the evolutionary mechanisms we invoked are speculative, Bolmgren and Cowan (2008) found exactly this empirical pattern. In a full phylogenetic analysis, they found a negative correlation between first flowering date and seed mass because perennials dominated their data numerically, but when analysed separately as annual, perennial or woody, they found positive, negative and neutral relationships, respectively.

In this example of seed size and flowering time, the variable leading to non-stationarity was simple (life habit having only three states) and known, and methods exist to analyse the data using an interaction between life habit and seed mass in a phylogenetic extension of ANCOVA. However, it is easy to imagine several scenarios where the interaction variable is not obvious (Supporting Information Appendix S2, Box S1), but might nonetheless be meaningful biologically. Identification of missing or latent variables can thus not only improve model fit, but also provide new insights into trait relationships and species ecology. Given that we wish to avoid the pitfalls of data dredging (Anderson, 2008), it is inappropriate simply to throw in a large number of potential variables in the hope that an important variable causing non-stationarity will be included and turn the non-stationarity into an interaction term. In such scenarios, it is more inferentially conservative to adopt an assumption of non-stationarity and let phylogenetic position serve as both a surrogate for excluded variables and a guide to their probable identity.

Here, we propose a new phylogenetic comparative method, which we term phylogenetically weighted regression (PWR), that allows for phylogenetic non-stationarity in modelling phylogenetic correlation. Put simply, PWR is a way of looking for changes in modelled relationships over different parts of the evolutionary tree. Our method parallels the more widely used method of geographically weighted regression (GWR, Charlton, 2002; Fotheringham, Charlton, & Brunsdon, 1998), which has been applied throughout
economics (Cho et al., 2010), geography (Fotheringham et al., 1998) and ecology (Foody, 2004, 2005) to understand how relationships vary across space. Both GWR and PWR estimate local regression coefficients by weighting data from surrounding observations based on distance. Phylogenetically weighted regression, however, weights the local taxonomic neighbourhood using phylogenetic distance rather than geographical distance. Conceptually, PWR shows similarities to the method of Uyeda et al. (2017), but rather than looking for discrete shifts that may match to “adaptive zones”, we allow for continuous variation in parameters across the branches of the phylogenetic tree.

We suggest that our new method of PWR might be a useful exploratory step in phylogenetic comparative analysis. Given that the number of parameters in PWR can be large [larger than for phylogenetic generalized least squares (PGLS)], users should be mindful of overfitting. Overfitting, however, can be avoided by using a cross-validation method that optimizes predictive accuracy on held-out data (i.e., data not used in estimation of model parameters). In addition, we note that PWR and PGLS should not be considered as alternative models; the goals of PWR versus PGLS are fundamentally different and have different inferential contexts. When the goal is to assess the statistical significance of a relationship assumed (or demonstrated) constant across the phylogeny with all important variables included, PCMs, such as PGLS or phylogenetic independent contrasts, should be used. Phylogenetically weighted regression, however, provides a tool for evaluating potential non-stationarity and missing variables in comparative data before the use of more traditional PCMs.

2 METHODS AND RESULTS

2.1 Model details

Following the general approach implemented in GWR, PWR involves the repeated application of weighted regression, in which individual observations are allowed to vary in terms of their influence on each fit. In our implementation, given a phylogeny with \( n \) tips, PWR proceeds by fitting \( n \) weighted regressions, one for each species \( i \), with the weights in each case calculated such that the \( i \)th regression yields coefficient estimates that we can regard as “phylogenetically local” to that species. Specifically, in the local regression for focal species \( i \), this species itself is given a weight of one, and the weights for all other species \( (j \neq i) \) are calculated as some decreasing function of phylogenetic distance from species \( i \). Conceptually, this means that more distantly related taxa have less influence on a given local fit.

Phylogenetic generalized least squares is a well-characterized method for accommodating phylogenetic correlation between species by appropriately parameterizing the covariance (off-diagonal) terms of the residual variance–covariance matrix (Freckleton, 2002; Grafen, 1989). In particular, in a regression of \( y \) on \( x \) PGLS can be written as:

\[
y_i = b_0 + b_1 x_i + e_i, \quad \text{where } \text{cov}(e_i,e_j) = \sigma_{ij}
\]  

In PWR, the coefficients are calculated using a weighted least squares (WLS) method where the weights, \( w_i \), are high for phylogenetically close data points and low for phylogenetically distant data points. The WLS is run once for each different data point, \( i \), producing as many estimates of \( b_1 \) and potentially \( b_0 \) as there are data points, thus for each point \( i \) solve:

\[
y_i = b_0 + b_1 x_i + e_i, \quad \text{where } \text{cov}(e_i,e_j) = \begin{cases} 0 & \text{off-diagonal} \\ w_{ij} & j^{th} \text{row of diagonal} \end{cases}
\]

The exact weight functions, \( w_{ij} \), are described in more detail in the next section.

Thus, one can regard PGLS as a method that models correlated errors (given by non-zero off-diagonal terms), but which treats regression coefficients themselves as constant over the phylogeny (single values calculated for \( b_0 \) and \( b_1 \) across the whole phylogeny). One of the main reasons to model the error structure is to ensure proper assessment of the true degrees of freedom and resulting estimates of significance versus a null hypothesis. In contrast, PWR dispenses with correlated errors (off-diagonal terms are zero), instead allowing the regression relationship itself to vary over the phylogeny (multiple \( b_0 \) and \( b_1 \)). The evolutionary model is used to define the weightings of the other data points in each local regression. It is worth noting that PGLS and PWR are both extensions of the simple ordinary least squares (OLS) model, but they are extensions in different directions. Thus, the intersection of PGLS and PWR occurs only when the error covariance matrix in PGLS converges on a constant \( \sigma \) times the identity matrix and when the weightings of all points in PWR are equal. These two scenarios converge on OLS. Thus, although both PGLS and PWR address phylogenetic structure, their intersection is simple OLS, an approach that is generally inappropriate for phylogenetic regression. This relatively low overlap in the two models is reflective of the fact that despite sharing phylogenetic thinking, their goals are fundamentally different and have different inferential contexts (hypothesis testing versus exploratory statistics).

2.1.1 Distance weighting functions

In geographically weighted regression, weights are calculated as a decreasing function of geographical distance (e.g., using the Gaussian function). For PWR, we replace geographical distance with phylogenetic distance. Although in principle one could use the same weighting functions conventionally used in GWR, here we introduce two alternatives related to terms of evolutionary models commonly used in comparative phylogenetics (Brownian and Ornstein–Uhlenbeck/Hansen & Martins). Provided branch lengths correspond to evolutionary time, the phylogeny can be represented as a covariance matrix \( \Sigma \), in which off-diagonal terms give the expected covariance in trait values between pairs of species. Indeed, these are the terms used to impose a correlation...
structure in PGLS. In PWR, we instead use these terms as local regression weights, \( w \). In the corresponding GLS implementation, this equates to constraining the residual variances associated with the fit for species \( i \) to be proportional to the vector \( 1/w_i \), where the \( w_i \) values are obtained from column \( i \) of the evolutionary covariance matrix, \( \Sigma \).

Adapting the notation of Paradis, Claude, and Strimmer (2004), under a Brownian motion model we expect the covariances to take the following form:

\[
\sum = \sigma^2 T_a
\]  \hspace{1cm} (3)

where \( T_a \) is the distance between the root of the tree and the most recent common ancestor of species \( i \) and \( j \). Note that the covariance decreases as a simple linear function of evolutionary distance (Hansen & Martins, 1996). For the purposes of specifying local weights, we can ignore the variance term \( (\sigma^2) \), which is constant with respect to the distance between species, leaving us with the following simple specification for PWR weights:

\[
w_{ij} = T_a
\]  \hspace{1cm} (4)

2.1.2 | Optimal bandwidth calculation

The Brownian model yields a fixed set of weights depending only on the structure of the tree, but more commonly the distance weighting function used for local regression will contain a free parameter that determines how quickly weights decline with increasing distance (Figure 1). For example, in the O-U case, values of \( \alpha \) closer to zero will yield fits that are more inclusive of distant relatives, and in the limit it will yield estimates identical to those obtained by standard regression. In the GWR literature, this free parameter is typically referred to as the bandwidth, and we adopt the same convention here, thus distinguishing it from the constraint parameter, \( \alpha \), fitted in the O-U model. Unless there are a priori reasons to select a particular bandwidth (see Discussion), a simple optimization procedure can identify the bandwidth that minimizes the sum of squared predictive error, \( \sum (\hat{y}_i - y_i)^2 \), where \( \hat{y}_i \) is the predicted value of \( y_i \) for species \( i \) at the optimal bandwidth, \( \hat{w}_i \).
is the predicted value of the response variable for species i based on a locally weighted regression w leaving out the focal species i itself, and \( y_i \) is the corresponding observed value of the response variable. This jackknife-like assessment of optimal bandwidth, successively leaving out each target species, is designed to avoid overfitting. Choosing model complexity (essentially the number of parameters in the model, and hence fitted in PWR the rate at which the regression coefficient changes over the phylogeny) by testing prediction on “out-of_bag” (i.e., hold-out) data has proved to be a robust methodology to avoid overfitting in machine learning.

2.2 | Application of PWR

We describe the performance of PWR using computer simulations and on an empirical dataset of flowering times and seed size. We used R (R Core Team, 2011) to carry out all PWR computations. For a given tree, we obtain a matrix of all pairwise phylogenetic weights using the Brownian and O-U phylogenetic correlation structures implemented in the ape package (Paradis et al., 2004). We then fit the local regression model for each species using the weighted regression. In order to determine the optimal bandwidth, we apply the general purpose subplex optimization function available in the subplex package (King, 2008). All code is provided in Supporting Information Appendix S1.

2.2.1 | Illustration using a simple tree

We first illustrate PWR using a hypothetical phylogeny with eight terminal taxa, each associated with given predictor and response variables (Figure 2), which we model via a simple linear relationship:

\[
y = \alpha + \beta x + \epsilon
\]  

(7)

Here, \( x \) and \( y \) are the predictor and response variables, \( \alpha \) and \( \beta \) are the intercept and slope coefficients, and \( \epsilon \) is the residual error term. We focus on the slope term, \( \beta \), because it represents the effect of the predictor on the response. For the purposes of illustration, we deliberately chose values of the variables as might arise from a deep bifurcation in the tree with respect to the relationship between predictor and response. Note that this example is intentionally chosen to violate the stationarity assumptions of PGLS strongly and conform to the type of non-stationarity for which PWR is designed. It is used as an illustrative example and not an assessment of techniques.

First, we obtained coefficient estimates for the above model using conventional PGLS. This approach fits a global model in the sense that it assumes that the relationship between the predictor and response is constant throughout the tree (phylogenetic stationarity), although it does take into account the phylogenetic non-independence among taxa by allowing non-zero covariances in the residual error structure. This model returns a (single) slope coefficient estimate of \( \hat{\beta} = 0.31 \), with the 95% confidence interval covering (−0.33, 0.95). Fitting PWR yields a more nuanced picture and returns phylogenetically local estimates that differ in both sign and magnitude on either side of the basal split (Figure 3); a large positive effect is evident in one clade (\( \hat{\beta}_1 = 1.0 \) for tips A–D), and a modest negative effect is evident in the other (\( \hat{\beta}_2 = −0.3 \) for tips E–H). In addition, the uncertainty associated with the local PWR estimates is considerably smaller than that of the global PGLS. In our example here, PWR estimates are similar under all variants of the weighting function because the shifting relationship between the predictor and response variables is closely associated with tree structure.

2.2.2 | An empirical example using data on flowering times and seed size in plants

To illustrate the application of PWR to a real-world dataset, we explored the relationship between average seed size and flowering time in plant species from southeastern Sweden, using a fully resolved phylogeny containing 125 of the 126 species analysed by Bolmgren and Cowan (2008); one species (Hieracium umbellatum) has now been synonymized with other species within the genus and was thus excluded from our analysis. We used the mean date of flowering onset as our flowering metric, derived from phenological surveys in Uppland and Södermanland provinces between 1873 and 1917 (Arnell, 1923); seed mass data were compiled from the literature (see Bolmgren & Cowan, 2008). We refer to these data as the Arnell dataset.

The regression itself is a simple linear model relating seed mass to phenology. The response variable is mean first flowering date (ffd), measured as the integer day-of-year starting from 1 January, and the predictor variable is the natural logarithm of seed mass measured in milligrams:

\[
ffd = \alpha + \beta_i \text{(seed mass)} + \epsilon
\]  

(8)
As with our illustration above, we first fit this model using conventional PGLS. Here, we proceed with an O-U covariance matrix, because this evolutionary model, which includes an additional parameter that measures the strength of return towards a theoretical optimum, was favoured over Brownian motion when comparing PGLS model fits (difference between AIC values, \( \Delta \text{AIC} = 54.8 \)).

The PGLS model yields an estimated \( \hat{\beta} \) (representing number of days change per log unit increase in seed mass) of \(-1.77\%\) and 95% confidence interval covering \((-3.78, 0.25)\). Next, we fit PWR using O-U-based distance weights [Equation (6)]. The median value of all 125 local \( \hat{\beta} \) estimates is \(-1.85\), closely matching the global PGLS estimate. However, we see considerable variation in the local values themselves, with individual PWR \( \hat{\beta} \) estimates ranging from \(-25.0\) to \(20.4\). Although we may not necessarily take the specific local estimates at face value, when viewed together they provide a powerful picture of how the relationship between seed mass and phenology varies across the phylogeny (Figure 4). Several distinct clades exhibit a stronger negative relationship between seed mass and first flowering date than is suggested by the global model, whereas others are associated with a positive relationship.

The optimized O-U bandwidth was \( \alpha = 19.92 \) (per branch-length unit), which yields fits that are highly localized. To put this value in context, even if species \( i \) shared 99% of the evolutionary history with species \( j \) dating back to the root node, the relative weight for \( j \) would be only ca. 0.67. This weight drops to ca. 0.14 for species that diverged after 95% of the time since the root and is a negligible < 0.0001 with 75% shared history.

If instead we were to use Brownian-based distance weights [Equation (4)], the median value of the local PWR \( \hat{\beta} \) estimates is only \(-1.40\). In addition, the individual local estimates vary over a much smaller range (from \(-2.45\) to \(-0.68\)); this is not surprising given that the Brownian weights decline only linearly with distance and are considerably larger than in the O-U case, resulting in individual estimates that are much less localized.

### 2.2.3 Performance of PWR across varying tree topologies

We explored performance of the PWR procedure further by considering variation in phylogenetic tree topology, which underlies the
FIGURE 4  Phylogenetically weighted regression (PWR) estimates (dots) and 95% confidence intervals (horizontal lines) across the Arnell phylogeny using Ornstein–Uhlenbeck-based distance weighting, for a simple regression of first flowering date on seed size. Continuous and dashed vertical grey lines show the estimate and 95% confidence intervals from an analogous global model estimated using phylogenetic generalized least squares (PGLS) [Colour figure can be viewed at wileyonlinelibrary.com]
structure of the covariance matrix. To do this, we use the following heuristic and fivefold cross-validation:

1. Generate 1,000 random tree topologies of size \( n = 32 \), which we found to be a reasonable compromise between computational tractability and species richness.
2. For each tree, simulate tip values, \( x \), and a slope coefficient value, \( \beta \), assuming a Brownian model and a single ancestral root-state, zero.
3. For each species apply Equation (7) (in this case ignoring the error term) to predict \( y \) using the known values of \( x \) and the slope coefficient, \( \beta \), setting a single global intercept (\( \alpha \)) to zero.
4. Randomly partition tips into five approximately equal-sized groups (folds).
5. For each group:
   a. Subset the full tree to exclude the tips in that group.
   b. Separately fit PWR and PGLS to this tree using the simulated \( x \) and \( y \) values.
   c. Use the fitted models to generate predicted \( y \) values for each tip in the hold-out group and local (PWR) or global (PGLS) \( \beta \) slope coefficients.
   d. Calculate the root mean squared errors (RMSEs) between predicted and actual (simulated) \( y \) values in the hold-out group.
6. Calculate the mean of the RMSE for PWR and PGLS.

Tree topologies were simulated following Davies, Cooper, Diniz, Thomas, and Meiri (2012), in order broadly to span empirical estimates of topological imbalance (Shao & Sokal, 1990) and stemminess (Fiala & Sokal, 1985; Mooers & Heard, 1997). The contrast with estimates from PGLS allows us to evaluate sensitivity of PWR to varying tree topology and is not intended to represent a formal model comparison between methods.

This cross-validation framework yields two numbers for each tree, one being a cross-validation error estimate for PWR and the other being a cross-validation error estimate for PGLS. In \( > 71\% \) of cases, cross-validation errors for the slope coefficient were lower for PWR than for PGLS, and in \( > 86\% \) of cases PWR provided more accurate estimates of \( y \), indicating that PWR is able to capture reliably the modelled non-stationarity in the underlying data. Relative model performance (RMSE) was not obviously biased by either of our two axes of tree shape (see Supporting Information Appendix S2, Fig. S1). However, PWR errors tended to be larger for more tippy trees (see Supporting Information Appendix S2, Fig. S2), that is, trees with longer terminal branches, most probably because cophenetic distances between tips will be longer on average for such trees, and thus even sister taxa will inform local model estimates only weakly.

3 | DISCUSSION

There is growing appreciation that the tempo and mode of trait evolution can vary across a phylogeny, and a number of evolutionary models have been developed to account for such variation (Blomberg, Garland, & Ives, 2003; Butler & King, 2004; Freckleton & Harvey, 2006; Hansen, 1997). Here, we have illustrated a new method, PWR, that also allows trait correlation coefficients to vary across the branches of a phylogenetic tree. Our approach is a phylogenetic extension of GWR (Fotheringham et al., 1998), substituting spatial distance for evolutionary distance as measured on a phylogenetic tree, and we present it here as an exploratory modelling framework. Our method shares similarities with more traditional phylogenetic comparative methods (PCMs), such as phylogenetic regression (Grafen, 1989), in that differences between species are predicted to accrue in proportion to divergence times, but here we consider non-stationarity in correlations between species traits across phylogenetic trees.

3.1 | Comparison of PWR with traditional phylogenetic comparative methods

We believe that PWR provides a complement to traditional PCMs. As with most statistical tools, one approach should not be seen as superior to the other. Instead, different tools should be selected depending on the goals and how well the data match the assumptions underlying the tool. For example, although new methods allow the fitting of complex, but unknown, evolutionary trajectories by transforming the variance–covariance matrix to adjust for model heterogeneity within PGLS (e.g., Mazel et al., 2016), information on the underlying biology is lost. In contrast, PWR, like the method of Uyeda et al. (2017), is able to identify shifting correlations between traits of interest across the phylogeny and thus help in the search for key missing variables influencing trait evolutionary relationships. If PWR does not identify strong deviations from phylogenetic stationarity (i.e., strong violations of the PGLS assumptions), then standard PGLS methods can be applied, fitting appropriate transformations of the variance–covariance matrix, for example, using Pagel’s \( \lambda \).

A perceived concern about using PWR [or its related spatial and temporal forms, GWR and temporally weighted regression (TWR)] is that it is prone to overfitting. It is true that the number of parameters (i.e., model degrees of freedom) is much larger in PWR than PGLS. In a simple example studying the linear regression between two traits, PGLS may only have four parameters (two coefficients, \( \alpha \) and correlation decay), whereas PWR might have an order of magnitude more. For this reason, PWR should not be used on small datasets. However, it is a fallacy that PWR is an overfit model. In a machine learning context, overfitting has a formal definition of using too many parameters (or equivalently too flexible a functional form), so that noise in the data is fitted rather than signal. The general technique used to avoid overfitting is to “train” the model on a subset of the data and then assess the goodness-of-fit of the model on excluded or held-out “test” data. Given that the test data are not used in calibrating the model, it is impossible to fit the noise in the held-out data, overfitting on the train data will result in an assessment of poor fit (Hastie, Tibshirani, & Friedman, 2001). Various enhancements are common. One example is to fit a variety of model complexities and to choose the one that has the best predictive power on the held-out
data (maximizing the signal-to-noise ratio). Another is to use \( k \)-fold cross-validation, where different fractions of the data are held out sequentially. These are exactly the techniques we present here (and used in GWR). The bandwidth is optimized to give the best prediction on the target (held-out point) using a jackknife or \( n \)-fold cross-validation (where \( n \) is the number of data points). This jackknife validation determines in a mathematically precise way the optimal degree of localness versus globalness in the fitting of the data.

### 3.2 Performance and generality of PWR

Using an empirical example and computer simulations, we showed that PWR is able accurately to recover phylogenetically local patterns in the relationship between a species-level predictor and response variable under assumptions of non-stationarity. In such a situation, conventional PGLS is not appropriate because it will tend to produce a slope coefficient effectively averaged across the tree (as shown in our example). In addition, we showed that PWR performance was not biased by phylogenetic tree shape, as indexed by topological imbalance (the symmetry in species richness between sister clades), nor by "stemminess"—the relative distribution of internodal distances from the root to the tips of the tree). However, we note that local estimates for evolutionary distinct taxa should be treated with greater caution because they are informed by fewer data points, and in some cases the correlation between estimated and empirical model coefficients was poor.

### 3.3 Selection of evolutionary model for weighting function and bandwidth considerations

We demonstrated the use of PWR using weighting functions based on two widely used models of evolution, in contrast to the weighting functions based on geographical distances traditionally used for GWR. In GWR, the local neighbourhood is typically defined using a bi-square or Gaussian weighting function, such that nearby observations are given a greater weight than those located further away. In our simulations, we assumed a Brownian motion weighting function that has a natural evolutionary basis, describing the expected divergence in a continuous character through time under neutral or fluctuating selection (Felsenstein, 1985). In our empirical analysis of plant flowering times, we additionally explored a weighting function matching to an O-U process (Martins & Hansen, 1997), and it would be straightforward to transform the underlying phylogenetic branch lengths to consider alternative rate-variable models (see “Distance weighting function” in Methods and Results). In practice, however, results of GWR have been found to be relatively insensitive to the choice of weighting function (Charlton, 2002), and in the simple example depicted in Figure 3, PWR accurately recovers the correct local regression coefficients using a Brownian motion weighting function, even though the evolutionary shift was modelled as a single switch at the first split in the tree.

Within the GWR framework, specification of an appropriate bandwidth (the rate at which weights decline with distance) is perhaps more important than the identification of the weighting function (Charlton, 2002). Under Brownian motion, weights decrease simply as a monotonic function of evolutionary distance, and bandwidth is ignored. However, in our simulations assuming an O-U process, we determined bandwidth using an optimizing function that minimized the sum of squared errors. We suggest that this bandwidth might be broadly interpreted as indicative of the evolutionary rate of change in model coefficients and can be likened to the phylogenetic half-life concept of Hansen (1997). A narrow bandwidth may reflect faster rates of change, such that modelled evolutionary relationships are divergent even among closely related species. Conversely, a broad bandwidth would indicate slower rates of change. However, because too narrow a bandwidth may result in model instability, particularly for evolutionarily isolated species, a pragmatic approach might be to explore a range of bandwidths and compare shifts in model coefficients. An alternative approach would be to allow an adaptive bandwidth, such that the phylogenetic distance defining the local neighbourhood itself varies across the tree. Although the fitting of an adaptive bandwidth adds an additional layer of complexity, equivalent methods have been developed in the GWR literature (Charlton, 2002).

### 3.4 Extensions to PWR

We have illustrated the performance of PWR by examining variation in the slope of the relationship between species traits, but the approach can be adapted for other uses. Specifically, PWR is an iterated form (one iteration per tip) of WLS regression. Therefore, any extension to univariate regression that works with WLS is straightforward to incorporate in PWR. For example, although we focus on the slope, which is often of most interest in ecological studies, PWR also allows for non-stationarity in the modelled intercept. Likewise, extension to a WLS multivariate regression framework with multiple explanatory variables (and interactions) with each coefficient showing non-stationarity would not be complicated (see Wimberly et al., 2008, for an example of multivariate WGR in ecology). In theory, it should be possible to extend PWR to a generalized least squares (GLS) framework with a full covariance matrix, but this extension has not been developed within the GWR literature, and we are therefore hesitant to introduce it in the context PWR until it can be thoroughly tested and benchmarked.

Phylogenetically weighted regression may also be a useful method for imputation and possibly for handling intraspecific trait variation. Traditional PCMs that incorporate intraspecific variation (e.g., Felsenstein, 2008; Ives, Midford, & Garland, 2007) are still not widely used; however, heteroscedastic variance across species (non-constant variances on the covariance matrix diagonal) can be converted to a WLS regression and should also be addressable within the PWR framework. Adapting PWR to more recent Bayesian approaches for handling intraspecific variation, such as Markov chain Monte Carlo generalized linear mixed models (Hadfield, 2010), would be more difficult. In the area of phylogenetic methods for data imputation, PWR could be an especially useful new approach. By optimizing the bandwidth that informs the local regression, PWR might
overcome some of the shortfalls of existing phylogenetic imputation methods, which often have low prediction accuracy (Molina-Venegas et al., 2018), and imputation is already embedded within the cross-validation heuristic of the method.

3.5 Application and utility of PWR

Like the GWR method that inspired it, PWR provides a useful tool for exploratory analyses. As an exploratory tool, PWR offers an additional option for modellers to examine trait correlations, one that uniquely relaxes the assumption of a static relationship between traits across a phylogeny. Strong evidence for phylogenetic non-stationarity might indicate interactions with unmeasured traits or environments. Phylogenetically weighted regression reveals the phylogenetic structure of these missing variables. Researchers using PWR will need to consider carefully their tree’s shape and structure and to be cautious in interpretation of local estimates for species on isolated phylogenetic branches and how varying bandwidths influence the output of PWR; such considerations are particularly important when less is known about the underlying factors driving evolutionary relationships.

ACKNOWLEDGMENTS

Thanks to D. Orme, N. Salamin, N. Kraft and K. Bolmgren, whose comments on earlier drafts greatly improved the manuscript. This work was conducted as part of the Forecasting Phenology working group supported by the National Center for Ecological Analysis and Synthesis (NCEAS), a Center funded by NSF (DEB-0072909), the University of California at Santa Barbara and the state of California.

DATA ACCESSIBILITY

Associated R code is referenced as Supporting Information Appendix S1; this and all example data are available through https://github.com/lizzieinvancouver/pwr

ORCID

T. Jonathan Davies http://orcid.org/0000-0003-3318-5948

Brian J. McGill http://orcid.org/0000-0002-0850-1913

REFERENCES


BIOSKETCH

T. JONATHAN DAVIES’s research lies at the interface between ecology and evolution, making use of the information within phylogenetic trees to bridge between them. His research addresses questions on the distribution of biodiversity and the challenges posed to its conservation.

SUPPORTING INFORMATION

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