Comparative Methods as a Statistical Fix: The Dangers of Ignoring an Evolutionary Model.
Author(s): Rob P. Freckleton, Natalie Cooper, and Walter Jetz
Source: *The American Naturalist*, Vol. 178, No. 1 (July 2011), pp. E10-E17
Published by: The University of Chicago Press for The American Society of Naturalists
Stable URL: http://www.jstor.org/stable/10.1086/660272
Accessed: 19/02/2014 13:44

E-Note

Comparative Methods as a Statistical Fix: The Dangers of Ignoring an Evolutionary Model

Rob P. Freckleton,[1,*] Natalie Cooper,[2] and Walter Jetz[2]

1. Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom; 2. Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520

Submitted June 14, 2010; Accepted December 16, 2010; Electronically published June 2, 2011

Online enhancement: appendix.

abstract: Comparative methods are widely used in ecology and evolution. The most frequently used comparative methods are based on an explicit evolutionary model. However, recent approaches have been popularized that are without an evolutionary basis or an underlying null model. Here we highlight the limitations of such techniques in comparative analyses by using simulations to compare two commonly used comparative methods with and without evolutionary basis, respectively: generalized least squares (GLS) and phylogenetic eigenvector regression (PVR). We find that GLS methods are more efficient at estimating model parameters and produce lower variance in parameter estimates, lower phylogenetic signal in residuals, and lower Type I error rates than PVR methods. These results can very likely be generalized to eigenvector methods that control for space and both space and phylogeny. We highlight that GLS methods can be adapted in numerous ways and that the variance structure used in these models can be flexibly optimized to each data set.

Keywords: generalized least squares (GLS), phylogenetic eigenvector regression (PVR), comparative methods.

Introduction

The comparative method is one of the most successful tools for testing theories in ecology and evolution (Maynard Smith 1978; Felsenstein 1985, 1988; Harvey and Pagel 1991; Bennett and Owens 2002; Garland et al. 2005; Freckleton 2009). In biology, the comparative method uses natural variation within and between groups of species as a cost-effective partial substitute for experimental manipulation. By using natural variation, comparative methods allow very broad hypotheses to be tested, in addition to hypotheses about factors that cannot easily be manipulated experimentally.

A drawback of the comparative method is that because the factors examined are usually uncontrolled, the results can be confounded in various ways. For example, species in a comparative analysis are related to each other and, as a consequence, may share similarities because they inherit them from their ancestors and not because of independent evolution (Felsenstein 1985; Harvey and Clutton-Brock 1985; Harvey and Pagel 1991). Thus, if data are analyzed by assuming that each species is independent in the statistical sense (implying that each species represents an independent evolutionary origin of each trait state), statistical tests will be compromised by phylogenetic relatedness among species (Grafen 1989; Martins and Garland 1991).

A suite of methods has been developed to deal with phylogenetic nonindependence in comparative data (e.g., Cheverud et al. 1985; Felsenstein 1985; Grafen 1989; Gittleman and Kot 1990; Lynch 1990; Pagel and Harvey 1992; Martins and Hansen 1997; Pagel 1997, 1999; Diniz-Filho et al. 1998; Housworth et al. 2004; Garland et al. 2005). These methods are closely related to statistical methods for addressing nonindependence used in other disciplines, including econometrics, spatial statistics, meta-analysis, and genetics (Gittleman and Kot 1990; Lynch 1990; Cressie 1993; Pagel 1997, 1999; Housworth et al. 2004; Dormann et al. 2007; Ives et al. 2007; references cited in Freckleton and Jetz 2009; Hadfield and Nakagawa 2010).

In broad terms, the current state of comparative methods can be summarized as follows: (1) Comparative methods use model-based analyses of trait evolution and emphasize the relationship between the evolutionary model and trait data (Felsenstein 1985; Hansen 1997; Pagel 1997, 1999; Harvey and Rambaut 2000; Hansen et al. 2008). (2) The approach uses comparisons of competing models to distinguish hypotheses (e.g., Pagel 1997, 1999; Harmon et al. 2003; Hansen et al. 2008). (3) The focus in comparative analysis has shifted more in favor of using phylogenetic information to improve our understanding of the evolutionary process and as a key tool in the study of macroevolution (Freckleton et al. 2003; Kelly and Price 2004; Paradis 2005; Freckleton and Harvey 2006; Thomas et al.

* Corresponding author; e-mail: r.freckleton@sheffield.ac.uk.

Am. Nat. 2011. Vol. 178, pp. E10–E17.
0003-0147/2011/17801-52241$15.00.
DOI: 10.1086/660272

2006). This is a shift in attitude, as the comparative method was largely regarded as a way of eliminating a "nuisance" in the data analysis: phylogeny was a problem that weakened analyses (Ricklefs and Starck 1996).

Although the methods cited above are related to methods for statistical analysis used in other areas of statistics, this does not mean that all methods are equally as useful or generally applicable. One of the fundamental tenets of the comparative method is that the relationships between species' traits result from common ancestry (Felsenstein 1985; Harvey and Pagel 1991). Consequently, in comparative analyses we can specify how we believe traits have evolved, model the process, and use the resultant model to specify how species' traits should relate to each other. Felsenstein (1973, 1985) was the first to do this in the context of comparative data, by modeling continuous traits using a Brownian process. In other disciplines this is not generally as straightforward: for example, in spatial analysis, a covariance structure has to be postulated, tested, and then refined or changed (Haining 1990). There is rarely an a priori model for such dependency, which may be very complex and difficult to characterize, and a fitted model may not be uniquely the best fit (Rohlf 2001).

Not all methods are equally capable of informing about the evolutionary process, and some approaches have been developed without an explicit evolutionary basis. Specifically, methods have been borrowed from other disciplines, in particular from spatial statistics. These include phylogenetic autocorrelation, based on spatial autocorrelation (Cheverud et al. 1985; Gittleman and Kot 1990), and phylogenetic eigenvector regression, based on spatial eigenvector regression (Diniz-Filho et al. 1998). The former is not widely used, presumably because it is very similar to generalized least squares (GLS) methods that have a more intuitive evolutionary basis (e.g., Martins and Hansen 1997; Pagel 1997, 1999). However, the latter method is argued to be an improvement over the autocorrelation approach (Diniz-Filho et al. 1998), and it has become increasingly frequently used in controlling for phylogenetic dependence (e.g., Küster et al. 2008), to look at the combined effects of spatial and phylogenetic effects in comparative data (e.g., Kühn et al. 2009), and in predicting conservation status (Safi and Pettorelli 2010). This is despite criticisms of the likely efficiency of the method in removing phylogenetic dependence from analyses (Rohlf 2001; Adams and Church 2011).

In this article, we use a comparison between the commonly used comparative methods of GLS and phylogenetic eigenvector regression (PVR) to highlight the problems of using techniques without an evolutionary basis as an ad hoc "fix" for phylogenetic structure in comparative data. These include reduced efficiency of estimation, enhanced variance in parameter estimates, inadequacy in dealing with phylogenetic signal, and increased Type I errors. These results can perhaps be generalized to apply to spatial eigenvector methods that have also been criticized in the literature (e.g., Beale et al. 2010). They have particular relevance for studies seeking to combine both phylogenetic and spatial eigenvectors (e.g., Kühn et al. 2009; Safi and Pettorelli 2010). Model-based approaches are more likely to succeed, as they offer better diagnostics and the ability to select between alternative formulations better tailored to the structure of the data.

Methods

Generalized Least Squares

The GLS model for data has been extensively described elsewhere in a comparative context (Grafen 1989; Martins and Hansen 1997; Pagel 1997, 1999; Garland and Ives 2000). In basic terms, this method fits a linear model to data in which a dependent variable, y, is modeled as a linear function of predictors x. The linear model relating X and Y, the observations of x and y, is

$$Y = X\beta + e. \quad (1)$$

The vector β describes the effects of the predictors, and e is an error term containing errors for the individual species. The errors are assumed to have a multivariate normal distribution with mean 0 and variance-covariance matrix $\sigma^2\mathbf{V}$.

The matrix $\mathbf{V}$ can be generated by making assumptions about the way traits evolve (Grafen 1989; Hansen 1997; Martins and Hansen 1997; Pagel 1997, 1999; Garland and Ives 2000). The simplest assumption is that traits accumulate variance as a linear function of time, that is, the Brownian model. In this model, trait variances and covariances are proportional to time. This basic assumption can be easily modified, for example, to allow for the speeding up or slowing down of evolution, speciational evolution, constraints on traits, or species-specific bursts of evolution, among others (e.g., Hansen 1997; Martins and Hansen 1997; Pagel 1997, 1999).

Once $\mathbf{V}$ has been specified, the theory for the linear model is very well elaborated (e.g., McCullagh and Nelder 1989). The parameters of the model can be estimated in a straightforward way and have desirable statistical properties, including being the best linear unbiased estimates (McCullagh and Nelder 1989). Essentially this means that the parameters estimated from this approach have the lowest variance among unbiased estimators; estimators with lower variance will be biased. Moreover, the estimated GLS model parameters (estimated means and variances) can be shown to be sufficient statistics for the true model, that is, making maximal use of information in the data (Casella

and Berger 2002). Put simply, if the structure of **V** is correct, then all other estimators of β will be worse than those estimated by GLS.

The GLS model is identical to the method of independent contrasts developed by Felsenstein (1985; see also Garland and Ives 2000; Freckleton and Harvey 2006; Freckleton and Jetz 2009). The mathematical relationship between the two is sketched by Felsenstein (1973). As a consequence, inasmuch as the method of contrasts is a special case of GLS, the GLS method is the most commonly used comparative method.

Phylogenetic Eigenvector Regression

PVR is a filtering method that uses covariates to remove dependencies from data. By including covariates with a strong phylogenetic signal, the effects of phylogenetic dependence can be statistically eliminated. The approach is straightforward. The eigenvector decomposition of an evolutionary distance matrix **D** (different from **V**, which is a matrix of shared path lengths, i.e., a similarity matrix) is used to generate a series of orthogonal covariates (eigenvectors) that summarize the phylogenetic structure. Some criterion (see below and Diniz-Filho et al. 1998) is then used to select the eigenvectors that optimize the description of the phylogenetic structure in the data. The selected eigenvectors are then included as predictors in a multiple regression of X on Y. The full recipe for the technique is given by Diniz-Filho et al. (1998).

The PVR method was criticized by Rohlf (2001), who pointed out that to fully capture the variation resulting from strong phylogenetic effects (e.g., the Brownian model), for a phylogeny of n species (or more generally, an $n \times n$ variance-covariance matrix), n independent eigenvectors would be required. The inclusion of all of these, using the method described by Diniz-Filho et al. (1998), would leave no extra degrees of freedom or variance for the inclusion of additional covariates or hypothesis testing. This approach therefore requires that some eigenvectors are not included. Because of this, the method will always fail to completely account for phylogenetic dependence. This procedure has never been justified, nor have the consequences of these problems for analyses using the method ever been explored.

Simulations

We performed 1,000 simulations of each parameter combination, varying the number of species included from 10 to 100, and we simulated data according to equation (1). We generated and analyzed phylogenies according to a birth-death model (using the packages TreeSim and APE in R; Paradis et al. 2004; R Development Core Team 2009; Stadler 2010). In generating the phylogeny, we set the ratio of extinction to speciation at 0.5. We found that changing this ratio between 0 and 0.9 did not affect the results we obtained. Very high values (or a coalescent model), however, substantially worsened the performance of the PVR method (see appendix in the online edition of the *American Naturalist*). Using the phylogeny, we generated a variance-covariance matrix, from which the function rmvnorm (from the R package mvtnorm; Genz et al. 2010) was used to generate random variates to describe the error term (e) in equation (1). The slope relating x to y was assumed to be 1.

We simulated two distributions of x. First, we assumed that there was no phylogenetic structure in x and drew x from a standard normal distribution. Second, we assumed that x was phylogenetically structured and generated x from a multivariate normal distribution, using rmvnorm. We assumed that the mean of x was 0 and the variance parameter was 1. In both cases, y was phylogenetically structured and generated from a multivariate normal distribution using rmvnorm. We note here that the GLS model makes no assumption about the phylogenetic distribution of x. Thus, in equation (1), the distribution of X is unspecified and may or may not show phylogenetic structure. The degree of phylogenetic structure in X is expected to affect the variance in estimates of model parameters. This is because the expected variance in the parameters is given by

$$\mathrm{Var}\,(\beta) = \sigma^2(XV^{-1}X')^{-1}. \qquad (2)$$

In the case of a single predictor, $XV^{-1}X'$ is a scalar and the greater the phylogenetic signal, the smaller this value will be. This is because phylogeny better predicts the value of X when the phylogenetic signal in the data is high. Consequently, when the phylogenetic signal is high, $(XV^{-1}X')^{-1}$ is large and the variance in β is greater; that is, increasing phylogenetic signal in the predictors will increase the variance in estimates of model parameters. However, with other aspects of the method, specifically Type I errors, efficacy of the method to account for phylogenetic nonindependence should be unaffected.

The first step in a PVR analysis is to select the optimum subset of eigenvectors for inclusion in a multiple regression of x against y. The literature on PVR methods is not clear on how to do this. Most suggest including all eigenvectors; however, for N species there will be $N - 1$ eigenvectors, so a regression of y on x and all eigenvectors will have insufficient degrees of freedom to allow for any statistical testing. We therefore included only the first 75% of eigenvectors, that is, those that explain the largest amount of variation in the distance matrix. We performed a multiple regression of y against x and the first 75% of eigenvectors, and then from these we selected the statistically

significant ones for inclusion in the final analysis (e.g., Diniz-Filho et al. 1998). The literature on PVR is also unclear as to whether the predictor variable, x, should be included in the initial eigenvector selection. Here, x was included in both analyses, as this should minimize sensitivity of output to possible collinearities (e.g., Freckleton et al. 2002). As a contrast, we also performed the simulation with another commonly used method: all eigenvectors were regressed on y alone to select the significant eigenvectors. These were retained and used to predict y from x. We do not report the results from this method, as it has the obvious problem of confounding phylogenetic signal in the data, the residuals, and the predictor (e.g., see Hansen et al. 2008 for an illustration).

For each set of simulated data, we used GLS and PVR to estimate the relationship between x and y. We recorded the mean and standard error of the estimated slope for the relationship, as well as the variance in the slope estimates across the 100 simulations, for each number of species. Although this is only a small range of all possible parameters, the values chosen are intended to be roughly representative of those estimated in comparative studies and are in no way unusual.

After the inclusion of eigenvectors, PVR model residuals should not contain any phylogenetic signal (Diniz-Filho et al. 1998). Therefore, to test the efficacy of the method in removing phylogenetic signal, we calculated Pagel's λ (Pagel 1999) for the PVR model residuals, using the approach described by Freckleton et al. (2002). This statistic varies between 0 (no phylogenetic signal) and 1 (Brownian motion). If the method removes all phylogenetic signal, as claimed, then the statistic should equal 0. If it is different from 0, then, in the case of Brownian error, the value of λ is the proportion of the phylogenetic signal that the method has failed to remove.

The final analysis of the simulations dealt with the issue of statistical errors: we calculated the probability of rejecting the null hypothesis when it is true (Type I error) by testing the hypothesis that the observed slope was different from the true value (in all of our simulations, this is assumed to be equal to 1). We did this with t-tests, using the estimated value of the slope and its standard error and the residual degrees of freedom. For the null model, the frequency of rejection should equal the nominal P value (we used the conventional threshold of $P = .05$). The code for running the analyses reported is available from R. P. Freckleton on request.

Results

GLS estimates are less variable than those obtained from the PVR method (fig. 1). The variability in estimates between simulations is always lower for GLS than for PVR (fig. 1*a*, 1*b*). When x is phylogenetically independent, the GLS estimates have a variance that is ~40% lower than the estimates of the PVR model (fig. 1*c*). Averaging across the simulations in figure 1*c*, the variance in estimates of the slopes from the GLS model was 55% lower than that estimated from the PVR model when x is phylogenetically structured (fig. 1*d*). As expected, the results from the GLS model are unaffected by the assumption about whether x is phylogenetically structured.

Estimating parameters for the PVR model has a statistical cost in terms of degrees of freedom. Each eigenvector included in the final model consumes 1 df, as shown in figure 1. This number increases as the size of the phylogeny increases, so that the proportion of total degrees of freedom required to characterize the effect of the phylogeny as sample size increases.

The low numbers of eigenvectors included in analyses of small phylogenies largely results from a lack of power to detect phylogenetic effects using P values in regression selection (fig. 2*a*). For small to moderately sized phylogenies, the PVR approach does not remove all phylogenetic signal, with the consequence that the residuals retain relatively high values of λ. One consequence is that, unless phylogenies are large, these residuals cannot be concluded to represent "species-specific" measures in any sense.

Type I error rates (i.e., probability of falsely rejecting the null hypothesis when it is true) of the PVR approach are high. Figure 2*b* and 2*c* shows Type I error rates for GLS and PVR methods. PVR methods have elevated Type I error rates relative to GLS methods largely because of the variability in slope estimates.

In the appendix, we show that these results are robust to the choice of model. For example, we show that the results are essentially the same if we use an Ornstein-Uhlenbeck model of trait evolution (Hansen 1997).

Discussion

Nonindependence is a common problem in statistical analysis, and techniques for dealing with nonindependence in spatial and timeseries data are especially well known (e.g., Haining 1990; Chatfield 1996). In both of these disciplines, a suite of techniques has been developed and a variety of methods are routinely used. With comparative analysis, we are in the unusual situation in that we can generate models for trait evolution and predict what the expected covariance in traits among species might be. The same is arguably not true elsewhere. For example, spatial covariance is difficult to predict and might arise from a range of direct and indirect drivers that have spatially complex distributions. Dealing with such covariation requires extensive data exploration and modeling (e.g., Haining 1990; Dormann 2007; Beale et al. 2010).

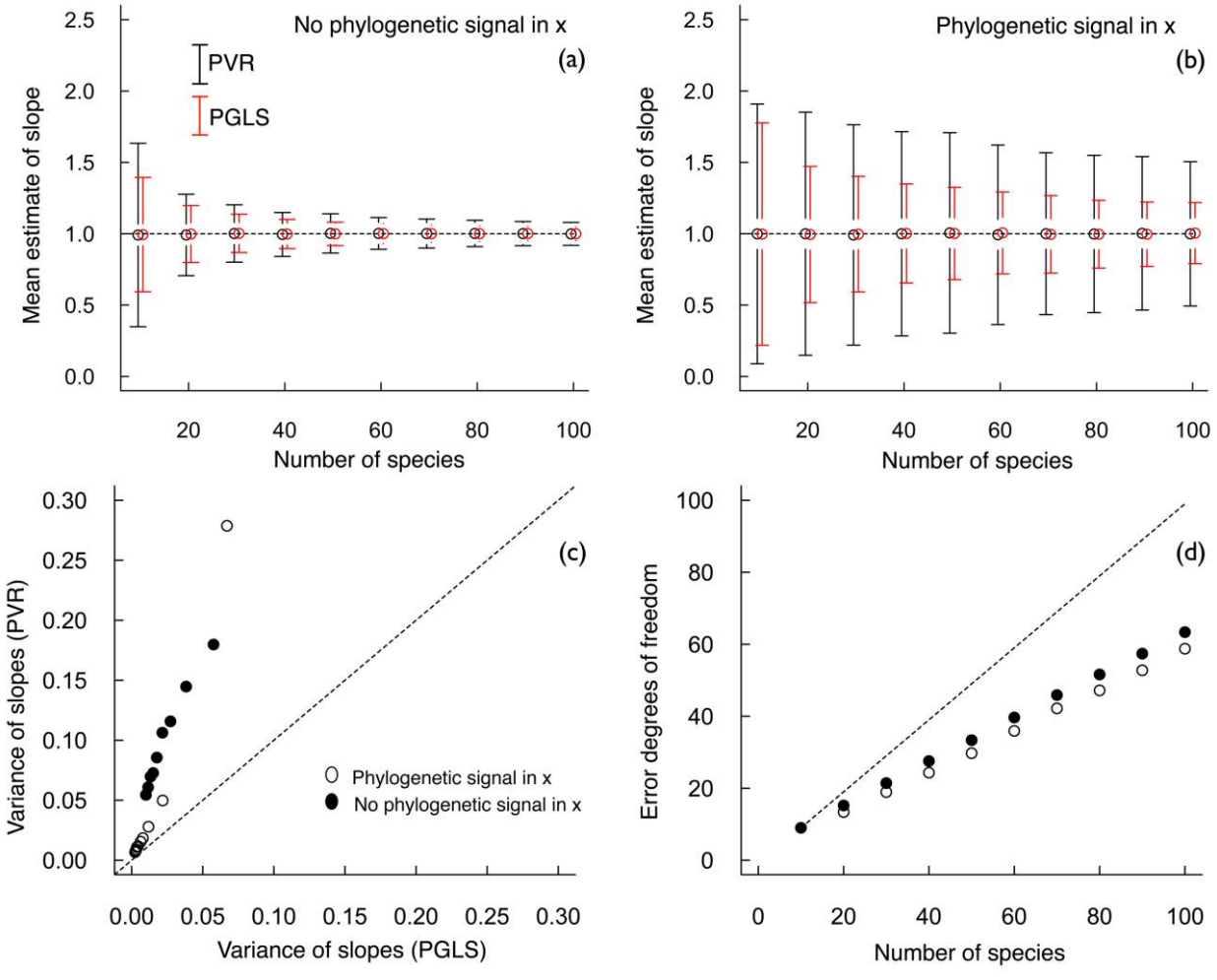

Figure 1: Comparisons of estimates of parameters, their variance, and residual degrees of freedom from generalized least squares (GLS) and phylogenetic eigenvector regression (PVR) models. *a*, Estimates of mean value of the slope for the regression of *y* on *x* when there is no phylogenetic signal in *x* for different-sized phylogenies. Error bars are 95% confidence intervals from 1,000 simulations. PVR estimates are in black; GLS estimates are in red. *b*, As in *a* but when *x* possesses phylogenetic signal. *c*, Variance in parameter estimates in *a* and *b* are compared. The variance in estimates from 1,000 simulations was calculated, and the values obtained from the PVR method were plotted against those from the GLS models. The dashed line represents the 1 : 1 relationship. *d*, Error degrees of freedom from the models. The dashed line is the GLS model; the points are means from the PVR method. In *c* and *d*, the filled circles are estimates from simulations in which there was phylogenetic signal in *x*, and the open circles are simulations in which there is no phylogenetic signal in *x*.

In comparative analysis, on the basis of the predictions of a Brownian model, the obvious starting point is to test whether species trait differences scale linearly with evolutionary distance or are independent of evolutionary distance (Garland et al. 1992; Freckleton et al. 2002; Freckleton and Harvey 2006). For one thing, the success of any comparative model depends on the accuracy of the phylogeny and the ability of the phylogeny to describe the data, so such tests are an important first step. From there, more complex models can be developed if required: the various transformations proposed in the literature (e.g., Grafen 1989; Hansen 1997; Pagel 1997, 1999; Hansen et al. 2008; Lavin et al. 2008) can change the assumed mode of trait evolution. However, in the case of the PVR, there is no clear way to accommodate or interpret such changes.

One claimed advantage of the PVR method is that it makes no assumptions about the evolutionary model that generated the data. Indeed, there is no element of model analysis or criticism in the PVR approach. However, this also generates a weakness: because there is no null model for the evolutionary process, there is no clear way to perform model criticism or diagnosis. GLS methods, includ-

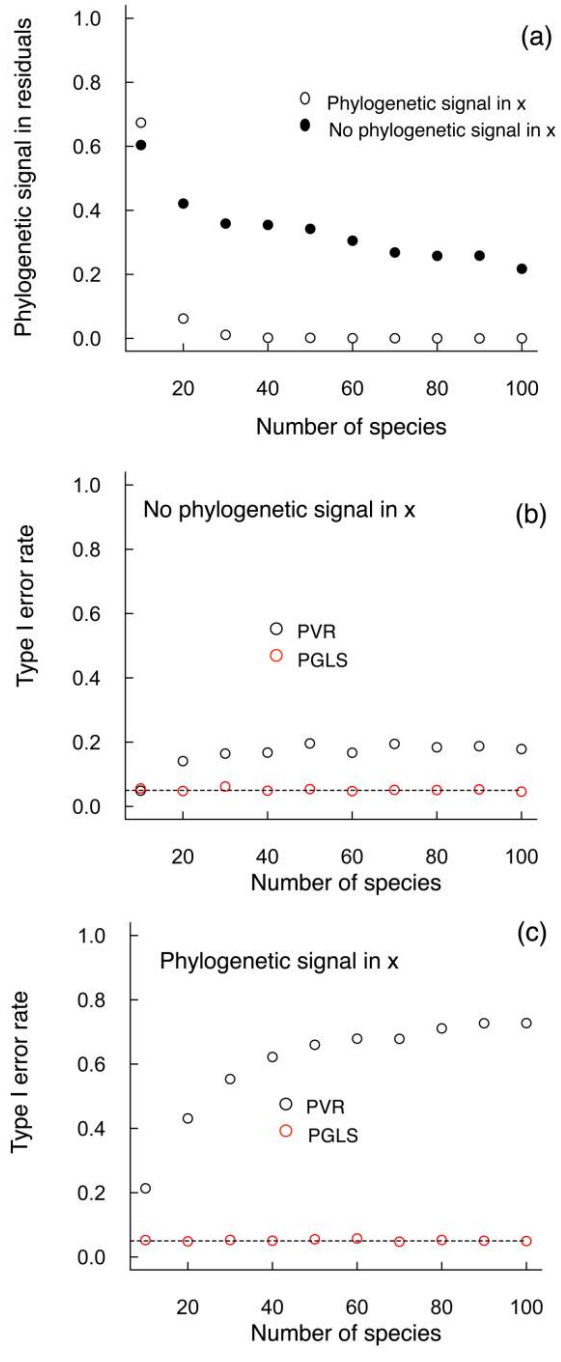

Figure 2: Phylogenetic signal in residuals and Type I error rates of phylogenetic eigenvector regression (PVR) methods. *a*, Estimates of Pagel's λ for residuals of PVR models from models based on phylogenies of different sizes and for data sets in which there is no phylogenetic signal in *x* (*open circles*) and in which there is a phylogenetic signal in *x* (*filled circles*). *b*, *c*, Type I error rates for phylogenetic generalized least squares (PGLS; *red symbols*) and PVR (*black symbols*) from data sets of varying sizes. In *b* there is no phylogenetic signal in *x*, whereas in *c* there is a strong phylogenetic signal in *x*.

ing the method of independent contrasts, have clear diagnostics and criteria for checking that the phylogenetic and other aspects of the model are behaving appropriately (e.g., Grafen 1989; Garland et al. 1992; Freckleton 2000, 2009). Because the PVR method does not state clearly the expected behavior of the phylogenetic component of the model, this is not possible to do. This conclusion is not specific to the Brownian model: we also simulated using Ornstein-Uhlenbeck models, and the PVR method performed equally as poorly (see appendix for examples).

Our analysis in figure 2*a* is the first to test the efficiency of the PVR method in removing known levels of phylogenetic signal. Rohlf (2001) pointed out that the PVR method may be inefficient in doing this, and our results indicate that appreciable phylogenetic signal may remain in the residuals. It has been suggested that the PVR method allows adaptive and nonadaptive variance in traits to be separated, with the residual term from the overall model representing the adaptive component of traits and the phylogenetic component representing nonadaptive components (Diniz-Filho et al. 1998). This interpretation is problematic, because such a variance decomposition implicitly assumes that most trait evolution is nonadaptive if traits show strong phylogenetic signal. Hansen et al. (2008) give an example, however, of how stabilizing selection on a trait evolving to an optimum that is itself subject to Brownian motion can give a distribution of traits that appears to be Brownian; that is, there is no nonphylogenetic component of trait variation. Our analysis, however, indicates that the PVR method produces an incorrect variance decomposition in any case (see also Adams and Church 2011).

The only conditions under which we have found the PVR method to produce statistically nearly acceptable results is when there is no phylogenetic signal in the data at all, including in the predictor, *x* (see appendix). Even if there is a phylogenetic signal in *x* but not in the residuals, the PVR approach fails to perform adequately. This, of course, is far from ideal, as the intention is for the PVR to allow a decomposition of the variance in the data into phylogenetic and nonphylogenetic components. This obviously cannot be achieved if the method is not reliable in the presence of phylogenetic signal. The phylogenetic generalized least squares approach works well in this situation if the λ parameter of Pagel (1997, 1999) is used to account for varying levels of phylogenetic signal in the residuals (see appendix; see also simulations in Freckleton et al. 2002).

The failure to specify a null model for trait evolution belies a philosophy that the phylogenetic signal in the data is purely a nuisance that is to be canceled out, with no further regard for the mechanism or nature of this dependence. As noted in the "Introduction," this runs

counter to current trends in ecology and evolutionary biology that emphasize the need for modeling data (e.g., Bolker 2008). Thus, if data show phylogenetic signal, or if the data deviate from a null model in a systematic way, then that is potentially of importance to the interpretation and underlying hypotheses.

The PVR method has been suggested as a technique that can deal simultaneously with both phylogenetic and spatial dependence in comparative data (e.g., Kühn et al. 2009). This is an important problem in comparative analysis, as many traits vary spatially as well as with phylogeny, and the question arises of how the spatial and phylogenetic contributions to trait variance may be decomposed (Freckleton and Jetz 2009). A recent article by Beale et al. (2010) compared the spatial analogue of the PVR method with GLS and other spatial statistical techniques in ecological analyses. They showed that the spatial eigenvector analysis was not able to reliably remove signal from the data and yielded more variable parameter estimates. Thus, the technique seems not to be best suited for either phylogenetic or spatial problems, and on this basis it would not seem to be a good option for dealing simultaneously with both. This is a particularly important point to consider if using these methods for predictions (e.g., Safi and Pettorelli 2010). Model-based methods would seem to offer more promise for achieving that, including parameters that measure the partition in variance between spatial and phylogenetic effects (Freckleton and Jetz 2009). Moreover, these alternative model-based techniques are computationally much more simple to implement.

In summary, the GLS method (in the widest sense) predates the PVR method in the comparative literature, has been used extensively, has better statistical properties, and is based on clear evolutionary models. The GLS method forms the basis for models in other areas of genetics and evolutionary biology, as well as in statistics. The PVR method was developed for spatial statistics, but it is not even widely used in that field, and recent evaluations in ecology have shown that it does not perform well in comparison with GLS methods. On that basis we would recommend that alternative methods to PVR be explored.

Acknowledgments

R.P.F. is funded by a Royal Society University Research Fellowship. N.C. is funded by a Seessel Postdoctoral Fellowship. We thank S. Blomberg, A. Diniz-Filho, A. Phillimore, and T. Webb for comments on an earlier version of this manuscript.

Literature Cited

Adams, D. C., and J. O. Church. 2011. The evolution of large-scale body size clines in *Plethodon* salamanders: evidence of heat-balance or species-specific artifact. Ecography 34, doi:10.1111/j.1600-0587.2011.06911.x.

Beale, C. M., J. J. Lennon, J. M. Yearsley, M. J. Brewer, and D. A. Elston. 2010. Regression analysis of spatial data. Ecology Letters 13:246–264.

Bennett, P. M., and I. P. F. Owens. 2002. Evolutionary ecology of birds: life histories, mating systems and extinction. Oxford University Press, Oxford.

Bolker, B. 2008. Ecological models and data in R. Princeton University Press, Princeton, NJ.

Casella, G., and R. Berger. 2002. Statistical inference. 2nd ed. Duxberry, Pacific Grove, CA.

Chatfield, C. 1996. The analysis of time series. Chapman & Hall, New York.

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

Cressie, N. A. C. 1993. Statistics for spatial data. Wiley, New York.

Diniz-Filho, J. A. F., C. E. R. de Sant'Ana, and L. M. Bini. 1998. An eigenvector method for estimating phylogenetic inertia. Evolution 52:1247–1262.

Dormann, C. F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography 16:129–138.

Dormann, C. F., J. M. McPherson, M. B. Araujo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30:609–628.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. American Journal of Human Genetics 25:471–492.

———. 1985. Phylogenies and the comparative method. American Naturalist 125:1–15.

———. 1988. Phylogenies and quantitative characters. Annual Review of Ecology and Systematics 19:445–471.

Freckleton, R. P. 2000. Phylogenetic tests of ecological and evolutionary hypotheses: checking for phylogenetic independence. Functional Ecology 14:129–134.

———. 2009. The seven deadly sins of comparative analysis. Journal of Evolutionary Biology 22:1367–1375.

Freckleton, R. P., and P. H. Harvey. 2006. Detecting non-Brownian trait evolution in adaptive radiations. PLoS Biology 4:2104–2111.

Freckleton, R. P., and W. Jetz. 2009. Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. Proceedings of the Royal Society B: Biological Sciences 276:21–30.

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

Freckleton, R. P., M. Pagel, and P. H. Harvey. 2003. Comparative methods for adaptive radiations. Pages 391–407 *in* T. M. Blackburn and K. J. Gaston, eds. Macroecology: concepts and consequences. Blackwell Scientific, Oxford.

Garland, T., Jr., and A. R. Ives. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. American Naturalist 155:346–364.

Garland, T., Jr., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Systematic Biology 41:18–32.

Garland, T., Jr., A. F. Bennett, and E. L. Rezende. 2005. Phylogenetic approaches in comparative physiology. Journal of Experimental Biology 208:3015–3035.

Genz, A., F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn. 2010. mvtnorm: multivariate normal and *t* distributions. R package, version 0.9-9. R Foundation for Statistical Computing, Vienna.

Gittleman, J. L., and M. Kot. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. Systematic Biology 39: 227–241.

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

Hadfield, J. D., and S. Nakagawa. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. Journal of Evolutionary Biology 23:494–508.

Haining, R. P. 1990. Spatial data analysis in the social and environmental sciences. Cambridge University Press, Cambridge.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

Hansen, T. F., J. Pienaar, and S. H. Orzack. 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62:1965–1977.

Harmon, L. J., J. A. Schulte II, A. Larson, and J. B. Losos. 2003. Tempo and mode of evolutionary radiation in iguanian lizards. Science 301:961–964.

Harvey, P. H., and T. H. Clutton-Brock. 1985. Life history variation in primates. Evolution 39:559–581.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.

Harvey, P. H., and A. Rambaut. 2000. Comparative analyses for adaptive radiations. Philosophical Transactions of the Royal Society B: Biological Sciences 355:1599–1605.

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

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

Kelly, C., and T. D. Price. 2004. Comparative methods based on species mean values. Mathematical Biosciences 187:135–154.

Kühn, I., M. P. Nobis, and W. Durka. 2009. Combining spatial and phylogenetic eigenvector filtering in trait analysis. Global Ecology and Biogeography 18:745–758.

Küster, E. C., I. Kühn, H. Bruelheide, and S. Klotz. 2008. Trait interactions help explain plant invasion success in the German flora. Journal of Ecology 96:860–868.

Lavin, S. R., W. H. Karasov, A. R. Ives, K. M. Middleton, and T. Garland Jr. 2008. Morphometrics of the avian small intestine, compared with nonflying mammals: a phylogenetic perspective. Physiological and Biochemical Zoology 81:526–550.

Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. American Naturalist 136:727–741.

Martins, E. P., and T. Garland Jr. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution 45:534–557.

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

Maynard Smith, J. 1978. Optimization theory in evolution. Annual Review of Ecology and Systematics 9:31–56.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Chapman & Hall, London.

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

———. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Pagel, M. D., and P. H. Harvey. 1992. On solving the correct problem: wishing does not make it so. Journal of Theoretical Biology 156: 425–430.

Paradis, E. 2005. Statistical analysis of diversification with species traits. Evolution 59:1–12.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

R Development Core Team. 2009. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Ricklefs, R. E., and J. M. Starck. 1996. Applications of phylogenetically independent contrasts: a mixed progress report. Oikos 77:167–172.

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

Safi, K., and N. Pettorelli. 2010. Phylogenetic, spatial and environmental components of extinction risk in carnivores. Global Ecology and Biogeography 19:352–362.

Stadler, T. 2010. Sampling through time in birth-death trees. Journal of Theoretical Biology 267:396–404.

Thomas, G. H., R. P. Freckleton, and T. Szekely. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proceedings of the Royal Society B: Biological Sciences 273:1619–1624.

Associate Editor: Gregory D. D. Hurst
Editor: Ruth G. Shaw