Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2015, **6,** 340–346 doi: 10.1111/2041-210X.12337

Measurement errors should always be incorporated in phylogenetic comparative analysis

Daniele Silvestro[1,2,3]†, Anna Kostikova[2,3]†, Glenn Litsios[2,3], Peter B. Pearman[4,5] and Nicolas Salamin[2,3]*

[1]*Department of Plant and Environmental Sciences, University of Gothenburg, Carl Skottsbergsgata 22B, 413 19 Gothenburg, Sweden;* [2]*Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland;* [3]*Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland;* [4]*Department of Plant Biology and Ecology, University of the Basque Country, UPV/EHU, Apdo. 644, 48080 Bilbao, Spain; and* [5]*IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain*

Summary

1. The evolution of continuous traits is the central component of comparative analyses in phylogenetics, and the comparison of alternative models of trait evolution has greatly improved our understanding of the mechanisms driving phenotypic differentiation. Several factors influence the comparison of models, and we explore the effects of random errors in trait measurement on the accuracy of model selection.

2. We simulate trait data under a Brownian motion model (BM) and introduce different magnitudes of random measurement error. We then evaluate the resulting statistical support for this model against two alternative models: Ornstein–Uhlenbeck (OU) and accelerating/decelerating rates (ACDC).

3. Our analyses show that even small measurement errors (10%) consistently bias model selection towards erroneous rejection of BM in favour of more parameter-rich models (most frequently the OU model). Fortunately, methods that explicitly incorporate measurement errors in phylogenetic analyses considerably improve the accuracy of model selection.

4. Our results call for caution in interpreting the results of model selection in comparative analyses, especially when complex models garner only modest additional support.

5. Importantly, as measurement errors occur in most trait data sets, we suggest that estimation of measurement errors should always be performed during comparative analysis to reduce chances of misidentification of evolutionary processes.

Key-words: Brownian motion, comparative methods, macroevolution, measurement error, model selection, Ornstein–Uhlenbeck

Introduction

Comparative phylogenetic methods are the foundation of many ecological and evolutionary analyses. Recent advances have produced increasingly complex models that help to evaluate alternative macroevolutionary hypotheses on mechanisms of trait evolution (Beaulieu *et al.* 2012; O'Meara 2012). Such models of continuous trait evolution are used to address an increasing number of questions in evolutionary biology, such as key innovations in the ornamental traits in birds (Maia, Rubenstein & Shawkey 2013), rates of environmental niche evolution in fish (Litsios *et al.* 2012), evolution of niche preference in plants (Schnitzler *et al.* 2012; Kostikova *et al.* 2013), body mass evolution in mammals (Cooper & Purvis 2010; Lartillot & Delsuc 2012) and constraints on phenotypic change in lizards (Smith *et al.* 2011). Among the range of stochastic models proposed to describe the evolution of continuous traits, the Brownian motion model (BM; Edwards & Cavalli-Sforza 1985; Felsenstein 1985; Pagel 1999; Blomberg, Garland & Ives 2003), the Ornstein–Uhlenbeck process and accelerating vs. decelerating rates of character evolution model (ACDC; Blomberg, Garland & Ives 2003) are the most widely used. Although recent developments have opened the door to other options, such as jump processes (Eastman *et al.* 2013; Landis, Schraiber & Liang 2013) or time and lineage-dependent models (Eastman *et al.* 2011; Lartillot & Poujol 2011; Venditti *et al.* 2011; Beaulieu *et al.* 2012; Ingram & Mahler 2013), the simpler BM, OU and ACDC models remain remarkably popular.

The BM model describes the unconstrained random evolution of a trait along the branches of a phylogenetic tree, in which the variance of observed trait values of extant species is proportional to the elapsed time of evolution and a rate parameter σ^2. Alternatively, the OU model combines the stochastic component of BM with an attraction parameter (α) that leads trait values to and maintains them around an optimum (θ). Finally, the ACDC model allows temporal variation in the rate of evolution by varying σ^2 through time based on a

*Correspondence author. E-mail: nicolas.salamin@unil.ch
†Equal contributions.

parameter g. The resulting rate of evolution decreases with time if $g > 1$ and increases with time when $g < 1$ and the parameter g affects σ^2 exponentially. The evaluation of competing evolutionary hypotheses involves the comparison of models in a maximum-likelihood framework, using measures of model fit such as the Akaike information criterion (AIC) and likelihood ratio tests (LRT). The BM model is typically the null hypothesis that is tested against OU, ACDC and any other more complex models of trait evolution (O'Meara 2012). In comparative analyses, model inference is as important as the accurate estimation of the parameters because support for a particular model, for example OU over BM, fundamentally alters the interpretation of the evolutionary process studied.

All three models are widely used in comparative analysis to test alternative evolutionary hypotheses of trait evolution, although their biological interpretation remains to some extent ambivalent. For example, patterns of trait evolution resembling BM can result from both genetic drift and directional selection where the direction of selection fluctuates randomly through time (Felsenstein 1988). On the other hand, when natural selection towards both an optimum and random drift act on a phenotypic trait (Lande 1976; Hansen 1997; Butler & King 2004) or when neutral evolution occurs in a constrained part of morphospace (Wiens *et al*. 2010; Crisp & Cook 2012), then an OU model might provide a better fit to the data. The attraction parameter in the OU model is often interpreted as strength of selection acting on a trait, when OU is used as a model for stabilizing selection (Hansen & Martins 1996). The ACDC model represents a situation in which there is an increasing constraint on evolutionary change through the course of evolution as a result, for example, of niche filling (Blomberg, Garland & Ives 2003). Nevertheless, most frequently, the BM model represents a null (neutral) hypothesis of trait evolution and is tested against more complex OU and ACDC models.

Measurement error (ME) in trait values substantially affects parameter estimation in analyses of continuous trait evolution. It leads to underestimation of phylogenetic signal and overestimation of rates of evolution (Ives, Midford & Garland 2007; Felsenstein 2008; Revell & Reynolds 2012). Such ME, following the Ives, Midford & Garland (2007) definition, derives from the intraspecific variability of natural populations and from instrumental imprecision (Garamszegi & Møller 2010). In the following, we will refer to ME as the relative deviation of an estimated trait value from its true value (i.e. $(\mu_{sample} - \mu_{true})/\mu_{true}$). In particular, limited sample size combined with intraspecific variability has a strong impact on ME around the estimated mean of a species trait value. To provide an empirical example, the mean body mass of a crab-eating macaque (*Macaca fascicularis*) is 2·9 kg with intraspecific variability (standard deviation) of 1·2 kg (Fitch 2000). If we were to obtain the mean body mass of macaques by measuring only one individual and assuming that intraspecific variability is normally distributed ($N(\mu = 2{\cdot}9,\ \sigma = 1{\cdot}2)$), then we would expect an ME of 33% (here calculated as $(\mu_{sample} - 2{\cdot}9)/2{\cdot}9$). The expected ME would decrease to 15% and 11% with sample sizes of 5 and 10 individuals, respectively (Fig. 1). Only a sample larger than 45 individuals would result in an expected ME <5%. Notably, these expectations for ME do not incorporate instrumental imprecision or sampling biases. Thus, we can expect ME to be ubiquitously present in empirical data sets (Lynch 1991), potentially reaching substantial quantities.

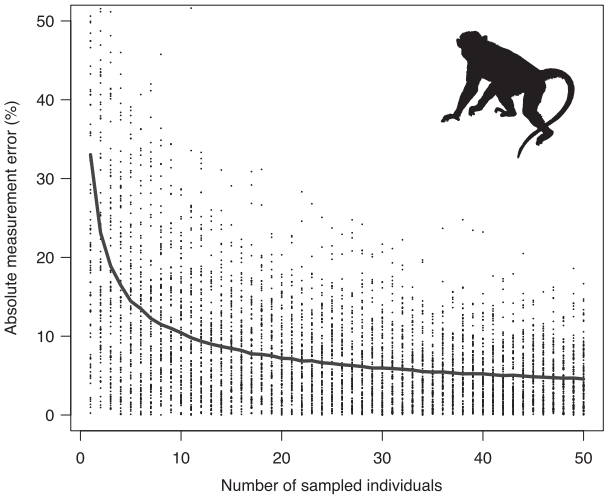

Fig. 1. Effect of sample size on the amount of measurement error (ME) when estimating a species mean trait value in the presence of intraspecific variability. The black line indicates the expected ME for a given sample size in *Macaca fascicularis* with true mean body mass equal to 2·9 kg, and known intraspecific variability is ±1·9 kg. Dots represent distribution of ME for the different sample sizes.

Recent advances in comparative methods make possible an explicit account of ME and significantly improve parameter estimation under different evolutionary models (Ives, Midford & Garland 2007; Felsenstein 2008; Revell & Reynolds 2012). While the biases on parameter estimates induced by ME are considerably reduced by these methods, the impact of ME (ignored or accounted for) on the accuracy of model selection is not well understood. In this study, we investigate ME-driven bias to determine whether the underlying BM model is reliably recovered, compared to alternative, parameter-rich OU or ACDC models. We perform maximum-likelihood optimizations of BM, OU and ACDC with different degrees of ME, which is simulated to reflect the expected quality of most phenotypic data sets. The analyses are carried out with special emphasize on model selection while (i) ignoring the presence of error, (ii) assuming error to be known and (iii) estimating error from the data. Based on our findings, we provide guidelines that evolutionary biologists should follow to reduce the risk of model misidentification when inferring evolutionary mechanisms during comparative analysis.

Methods

We simulated a continuous trait that evolved under Brownian motion along the branches of phylogenetic trees to examine how the presence of ME affects model selection. We simulated trees of different sizes using the TreeSim R package (R Development Core Team 2011; Stadler 2011) under a range of extinction and taxon sampling settings with 1000 replicates for each combination of parameter values (24 000 trees in total; see Table 1 for details). We rescaled the tree height of all trees to one to have comparable rates across trees (Revell 2010).

Table 1. Simulation settings

Type	Parameter values
Tree sizes	25, 50, 100, 250
Extinction rate	0, 0·3, 0·9
Sampling fraction	0·5, 1
Measurement error, %	0, 1, 5, 10, 15, 20, 25, 30, 40, 50

All pairwise combinations of parameters were used to simulate the trait data. Each simulation setting was replicated for 100 tree topologies.

On each tree of size n, we simulated a continuous trait with the R package phytools (Revell 2012) under a BM model with evolutionary rate parameter (σ^2) equal to 2·5 to generate a vector $a = [a_1, \ldots, a_n]$ of normally distributed trait values. We then exponentiated $\boldsymbol{a}$ to obtain a vector of log-normally distributed trait values, $\boldsymbol{x}$, as observed for most morphological traits (e.g. body size). We then altered the vector $\boldsymbol{x}$ to introduce different amounts of ME, and thus, ME operates at the scale of the original data rather than at the log-transformed scale. This was achieved by sampling for each species i a new trait value x'_i from a normal distribution centred on x_i with a standard deviation yielding on average ME is 1, 5, 10, 15, 20, 25, 30, 40 and 50% of the respective trait value (for the details see Appendix S1). We emphasize that while the average ME was fixed, this procedure resulted in different ME for each species, thus accounting for the fact that measurement errors are typically variable among extant species. Finally, we log-transformed the vector of altered trait values $\boldsymbol{x'}$ to obtain $\boldsymbol{a'}$ used in subsequent analyses.

We evaluated models using the Geiger package (Harmon *et al.* 2008) because it can readily incorporate predetermined ME or estimate it. In Geiger, ME is added to the respective variances in the phylogenetic variance–covariance matrix. All data sets were analysed by fitting a BM model (used to generate the original data), an OU model (with two additional parameters α and θ) and an ACDC model (also with two additional parameters g and z_0) using maximum likelihood. We compared the fit of the three models using two approaches: the likelihood ratio test (LRT) to assess whether OU or ACDC were significantly preferred over BM (P value <0·05) and the Akaike weights (Burnham & Anderson 2004) to quantify the relative fit of each model. We evaluated the models under three scenarios that assumed: (i) the simulated data held no ME, (ii) ME was estimated from the data, and (iii) the average ME across species is perfectly known, that is, using an average ME that was imposed upon the data during simulation. The code for simulating and analysing the data is available at http://www2.unil.ch/phylo/bioinformatics.

To demonstrate the importance of incorporating ME when reconstructing trait evolution in empirical data sets, we re-analysed a data set of primates (including the crab-eating macaque mentioned above) obtained from Harmon *et al.* (2010). The data set included a dated phylogeny and body mass values for 216 species. We used a cubic root and log transformation to normalize body mass (as in Harmon *et al.* 2010) and ran a maximum-likelihood optimization using Geiger under BM, ACDC and OU models. We initially ran the analyses assuming no error in the trait data and subsequently repeated them with ME estimated from the data. Finally, we used the AIC weights as a measure of

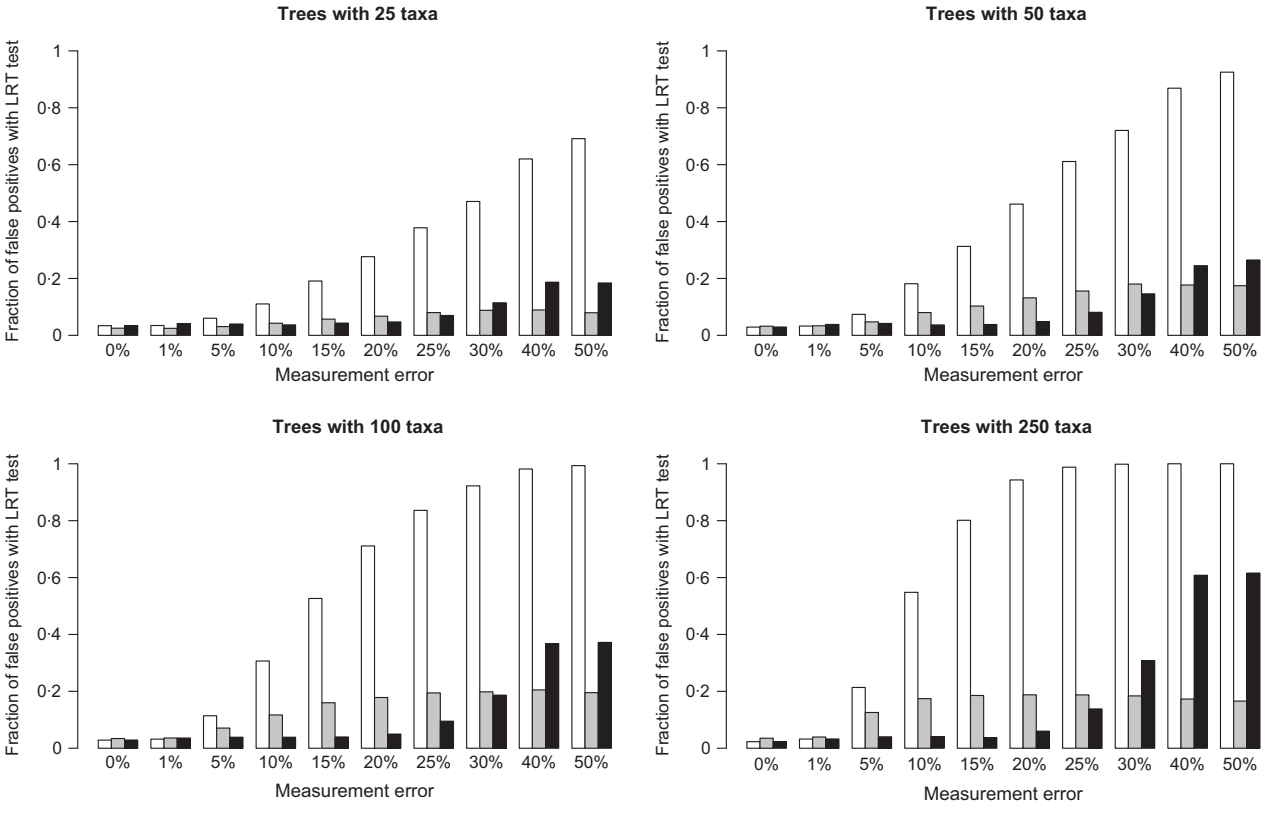

Fig. 2. Results of model selection for simulated data on phylogenetic trees containing 25, 50, 100 and 250 taxa. Data sets were simulated under the BM model, and both the BM and OU models were fitted to each simulated data set. False positives were identified as simulated data sets that rejected the true BM model using a likelihood ratio test with a significance threshold of 0·05. Bar plots indicate rates of false positives when measurement error is ignored (white), estimated (grey) or known (black).

model fit and compared the results obtained when ignoring and estimating ME.

Results

MODEL SELECTION

The accuracy of model selection is greatly affected by the presence of ME (Fig. 2 and Appendix S2). Indeed, the support for the alternative models compared to BM (i.e. the correct model) increases with increasing ME. We observe this bias across all simulations, regardless of the number of tips, taxon sampling and speciation process used to generate the trees (Appendix S2).

Among the three scenarios (i.e. ME ignored, estimated or known), the accuracy of model selection is poorest when ME is ignored (Fig. 2). For instance, the proportion of false positives in trees of 100 tips (regardless of the birth–death settings and taxon sampling) exceeds 0·1 with as little as 5% ME in the data. False positives increase with increasing ME and exceed 0·5 with ME = 15% and 0·8 with ME = 25%. The bias towards erroneously rejecting BM against more parameterized models also becomes stronger for larger trees. For example, with 250-species trees, the rate of false positives exceeds 0·2 even with ME as low as 5% (Fig. 2). The rate of false positives also appears to increase in the presence of high extinction rates. For instance, in trees of 100 tips, the rate of false positives raises from 0·08 to 0·2 as the extinction rate increases from 0 to 0·9 (Appendix S2). Incomplete (random) taxon sampling, on the other hand, does not appear to have a significant impact on the accuracy of model selection, and the rate of false positives is usually even lower than in trees with complete sampling (Appendix S2).

Much of the bias towards spurious support for the more parameter-rich models is, however, avoided by incorporating estimated or true ME in the model. Noticeably, rates of false positives remain low compared to those obtained when ignoring the ME (Fig. 2). When ME is known and accounted for in the analyses, the rate of false positives remains below 0·05 even with ME = 20%. However, rate of false positives rapidly increases as the ME exceeds 30%, and this is particularly evident on large trees (Fig. 2 and Appendix S2). Estimating error from the data also drastically reduces the rate of false positives (Appendix S2). While the rate of false positives generally exceeds 0·05 with ME around 10%, it stabilizes around 0·1–0·2 with large ME.

The results from the Akaike weights show that the relative likelihood of BM decreases in favour of OU as unaccounted ME increases (Fig. 3). The Akaike weight of the ACDC model is generally below 0·25 and is not affected by tree size or by

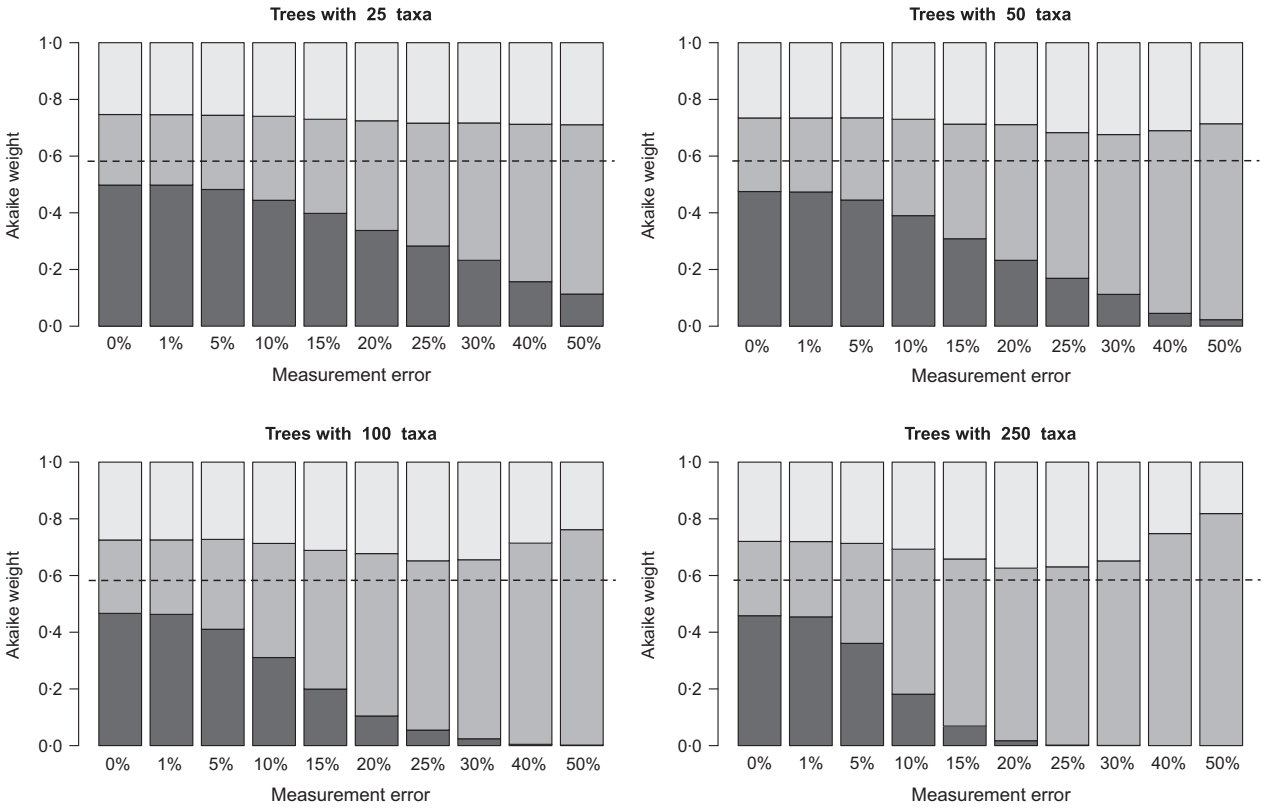

Fig. 3. Trends in Akaike weights for three evolutionary models as tree size increases. Black colour indicates AICc support for Brownian motion (BM) models, dark grey for Ornstein–Uhlenbeck models (OU) and light grey for accelerating/decelerating rates (ACDC) models. Dashed line represents the highest possible Akaike weight for BM model, which can be calculated since BM represents a special case of both OU and ACDC. This demonstrates that in the absence of ME, AIC can confidently distinguish between evolutionary models.

ME. The relative likelihood of OU can exceed 0·5 already at ME = 15% on a tree with as few as 100 tips (Fig. 3). When ME is accounted for (estimated or known), the Akaike weights consistently support the BM model over alternatives (Appendices S3 and S4).

PARAMETER ESTIMATION

Unaccounted ME on trait values biases the estimation of the model parameters. Under BM, the evolutionary rate parameter σ^2 is overestimated (Appendix S5). Furthermore, when an OU process is fitted to the data, the bias appears to affect both the estimated rate of trait evolution and the selection parameter α (Appendix S5). Indeed, we observe an upward bias of σ^2 coupled with an overestimation of α as ME exceeds 30% (Appendix S5). Correction for ME (known or estimated) substantially improves parameter estimation and yields estimates of the rate parameter of the BM model centred on the true value (Appendix S5).

Estimated errors generally provide an effective approximation to the true error in the data sets. However, we observe overestimation of errors as true ME exceeds 25% (Fig. 4). The precision of the estimation of ME is negatively correlated with the magnitude of the error. For example, with ME = 5%, the range of estimated errors is ~0–15%, while it ranges from ~0 to 42% with ME = 25% (Fig. 4).

CASE STUDY

The analysis of the empirical example of primate body size evolution ignoring ME showed strong evidence against the BM model (*P*-value <0·01, Akaike weight 0·028) and in favour of ACDC and OU. OU and ACDC obtained a similar support (Akaike weights 0·486 for both, Appendix S6). When correcting for ME, the statistical support in favour of OU drastically dropped (Akaike weight 0·058) and BM was not rejected against OU (*P*-value = 0·999). The ACDC model was strongly supported (Akaike weight 0·777) and outperformed BM albeit with a higher *P*-value (*P*-value < 0·023).

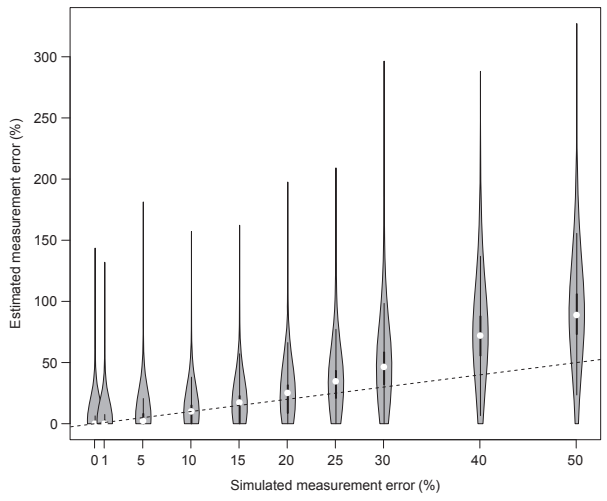

Fig. 4. Estimated measurement error (ME) across simulations. The mean estimated ME are shown for Brownian motion model.

Discussion and guidelines

Measurement error affects comparative analyses by biasing the estimation of model parameters (Ives, Midford & Garland 2007; Felsenstein 2008; Revell & Reynolds 2012). However, we demonstrate that it also biases model selection by leading to erroneous rejection of the simpler (but correct) BM model, especially in favour of the alternative OU model. The disrupting effect of ME on the covariances among species generated under a BM process results in overestimated evolutionary rates and underestimated phylogenetic signal (Ives, Midford & Garland 2007; Revell, Harmon & Collar 2008), as well as compromising the fit of the BM model. This increases rates of rejection of the neutral (BM) model and causes erroneous conclusions regarding the processes underlying trait evolution. This is problematic given that in most empirical studies, the support for the alternative OU model is often cited as an evidence of stabilizing selection.

Furthermore, the artifactual support for the OU model, paired with high values of the evolutionary rate parameter α, provides the false impression that the trait has evolved under strong selective pressure towards an optimal trait value, even though neither has played a role in the evolutionary history of the trait. The severity of the bias further increases in the presence of high relative extinction rates, whereas incomplete taxon sampling can, if anything, slightly reduce the bias. This is likely a consequence of the distribution of branching times, which tend to be closer to the tips under a birth–death process (Ricklefs 2007), while branching times are biased towards older ages in the case of random incomplete sampling (Yang & Rannala 2006). The disruption of the covariances, which will reduce any phylogenetic signal and lead to the acceptance of the OU model, as seen before, is therefore stronger for trees generated under a birth–death process. The relative likelihood of the ACDC model does not improve with increasing ME. This is probably because the underlying process behind the ACDC is still Brownian motion, albeit with temporal variation in the evolutionary rate, thus implying a level of phylogenetic signal similar to the BM model.

We show that incorporating ME (estimated or known) in the analysis significantly improves both model selection and parameter estimation. However, obtaining accurate ME estimates is difficult because ME is estimated from sparse data, that is a single observation per species. Better performance is observed when the ME is known and accounted for in the analysis, at least when ME <30%. For ME >30% estimated ME appears to reduce the number of false positives. This is possibly due to the overestimation of the ME (Fig. 4), making the phylogenetics signal stronger. Although we did not explore this, current implementations of comparative models allow the user to set species-specific ME, thus increasing the amount of information provided in the model. However, this is typically carried out by adding additional variance terms to the diagonal of the VCV matrix, but this is unlikely to be the appropriate

representation of the intraspecific variability. Future methodological advances should therefore consider incorporating explicitly multiple observations for each species to model changes of intraspecific variances along the phylogenetic trees (Salamin *et al*. 2010).

Intraspecific variation is present in all taxa, which we have illustrated with the case of the crab-eating macaque, and measuring one or very few individuals is far from optimal to estimate accurately a biological trait. Our simulations and the analysis of primate body mass evolution further show that comparative analyses are most reliable when there is no or very little ME in the data. With our empirical analysis of primate body size, we demonstrated that accounting for ME can significantly change the relative support for each model and decrease our confidence in rejecting BM. Therefore, the general recommendation is to seek large sample size (i.e. more individual measurements per species) to practically reduce the amount of ME present in the data set and account for the residual ME during comparative analysis. Our results suggest that ME should always be incorporated in comparative analyses. Information about ME can be retrieved from multiple observations, data bases, experiments and the literature. However, when this information cannot be retrieved from independent sources, ME should be estimated rather than ignored. In addition, awareness of the potentially high rate of false positives should prompt researchers to critically interpret the results of likelihood tests that reject BM in favour of more complex models.

Acknowledgements

DS received funding from the Wenner-Gren Foundation and Carl Tryggers stiftelse (Sweden). The project was supported by Swiss National Science Foundation Grant Nos 3100A0-122433 and 3100A0-138282 to P.B.P. and N.S., respectively. The Vital-IT facilities of the Swiss Institute of Bioinformatics were used for all computational aspect of this study.

Data accessibility

The data used in this paper are either already published (see text for the specific references) or simulated based on the scripts available at http://www.unil.ch/phylo/bioinformatics.

References

Beaulieu, J.M., Jhwueng, D.-C., Boettiger, C. & O'Meara, B.C. (2012) Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution*, **66**, 2369–2383.

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

Burnham, K.P. & Anderson, D.R. (2004) Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods and Research*, **33**, 261–304.

Butler, M.A. & King, A.A. (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist*, **164**, 683–695.

Cooper, N. & Purvis, A. (2010) Body size evolution in mammals: complexity in tempo and mode. *The American Naturalist*, **175**, 727–738.

Crisp, M.D. & Cook, L.G. (2012) Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes? *New Phytologist*, **196**, 681–694.

Eastman, J.M., Alfaro, M.E., Joyce, P., Hipp, A.L. & Harmon, L.J. (2011) A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution*, **65**, 3578–3589.

Eastman, J.M., Wegmann, D., Leuenberger, C. & Harmon, L.J. (2013) Simpsonian 'Evolution by Jumps' in an Adaptive Radiation of Anolis Lizards. arXiv preprint arXiv:1305.4216.

Edwards, A.W.F. & Cavalli-Sforza, L.L. (1985) Reconstruction of evolutionary trees. *Phenetic and Phylogenetic Classification*, Systematics Association Publication No. 6, London. Reprinted in Cladistic Theory and Methodology (ed. T.D.A.T.F. Stuessy), pp. 67–76. Van Nostrand Reinhold, New York, New York, USA.

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

Felsenstein, J. (1988) Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics*, **19**, 445–471.

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

Fitch, W.T. (2000) Skull dimensions in relation to body size in nonhuman mammals: the causal bases for acoustic allometry. *Zoology-Analysis of Complex Systems*, **103**, 40–58.

Garamszegi, L.Z. & Møller, A.P. (2010) Effects of sample size and intraspecific variation in phylogenetic comparative studies: a meta-analytic review. *Biological Reviews*, **85**, 797–805.

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

Hansen, T.F. & Martins, E.P. (1996) Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, **50**, 1404–1417.

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

Harmon, L.J., Losos, J.B., Jonathan Davies, T., Gillespie, R.G., Gittleman, J.L., Bryan Jennings, W. *et al*. (2010) Early bursts of body size and shape evolution are rare in comparative data. *Evolution*, **64**, 2385–2396.

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

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

Kostikova, A., Litsios, G., Salamin, N. & Pearman, P.B. (2013) Linking life-history traits, ecology, and niche breadth evolution in North American Eriogonoids (Polygonaceae). *The American Naturalist*, **182**, 760–774.

Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution*, **30**, 314–334.

Landis, M.J., Schraiber, J.G. & Liang, M. (2013) Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. *Systematic Biology*, **62**, 193–204.

Lartillot, N. & Delsuc, F. (2012) Joint reconstruction of divergence times and life-history evolution in placental mammals using a phylogenetic covariance model. *Evolution*, **66**, 1773–1787.

Lartillot, N. & Poujol, R. (2011) A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Molecular Biology and Evolution*, **28**, 729–744.

Litsios, G., Pellissier, L., Forest, F., Lexer, C., Pearman, P.B., Zimmermann, N.E. & Salamin, N. (2012) Trophic specialization influences the rate of environmental niche evolution in damselfishes (Pomacentridae). *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3662–3669.

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

Maia, R., Rubenstein, D.R. & Shawkey, M.D. (2013) Key ornamental innovations facilitate diversification in an avian radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 10687–10692.

O'Meara, B.C. (2012) Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 267–285.

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

R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. Foundation for Statistical Computing, Vienna, Austria.

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

Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

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

Revell, L. & Reynolds, G. (2012) A new Bayesian method for fitting evolutionary models to comparative data with intraspecific variation. *Evolution*, **66**, 2697–2707.
Ricklefs, R.E. (2007) Estimating diversification rates from phylogenetic information. *Trends in Ecology and Evolution*, **22**, 601–610.
Salamin, N., Wuest, R.O., Lavergne, S., Thuiller, W. & Pearman, P.B. (2010) Assessing rapid evolution in a changing environment. *Trends in Ecology and Evolution*, **25**, 692–698.
Schnitzler, J., Graham, C.H., Dormann, C.F., Schiffers, K. & Peter Linder, H. (2012) Climatic niche evolution and species diversification in the Cape flora, South Africa. *Journal of Biogeography*, **39**, 2201–2211.
Smith, K.L., Harmon, L.J., Shoo, L.P. & Melville, J. (2011) Evidence of constrained phenotypic evolution in a cryptic species complex of Agamid lizards. *Evolution*, **65**, 976–992.
Stadler, T. (2011) Simulating trees with a fixed number of extant species. *Systematic Biology*, **60**, 676–684.
Venditti, C., Meade, A. & Pagel, M. (2011) Multiple routes to mammalian diversity. *Nature*, **479**, 393–396.
Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B., Cornell, H.V. *et al.* (2010) Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, **13**, 1310–1324.
Yang, Z. & Rannala, B. (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, **23**, 212–226.

Received 20 November 2014; accepted 25 November 2014
Handling Editor: Tamara Münkemüller

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Deriving the standard deviation yielding predefined average ME.

Appendix S2. Proportion (%) of false positives in model selection with increasing measurement errors

Appendix S3. Akaike weights for models when ME are estimated.

Appendix S4. Akaike weights for models when ME are known.

Appendix S5. Estimated parameters σ and α parameters under BM and OU models.