Stochastic models in phylogenetic comparative methods: analytical properties and parameter estimation

KRZYSZTOF BARTOSZEK

CHALMERS | UNIVERSITY OF GOTHENBURG

Stochastic models in phylogenetic comparative methods: analytical properties and parameter estimation
Krzysztof Bartoszek

Author e-mail: `bartoszekkj@gmail.com`

Stochastic models in phylogenetic comparative methods: analytical properties and parameter estimation
Krzysztof Bartoszek

Abstract

Phylogenetic comparative methods are well established tools for using inter–species variation to analyse phenotypic evolution and adaptation. They are generally hampered, however, by predominantly univariate approaches and failure to include uncertainty and measurement error in the phylogeny as well as the measured traits. This thesis addresses all these three issues.

First, by investigating the effects of correlated measurement errors on a phylogenetic regression. Second, by developing a multivariate Ornstein–Uhlenbeck model combined with a maximum–likelihood estimation package in R. This model allows, uniquely, a direct way of testing adaptive coevolution.

Third, accounting for the often substantial phylogenetic uncertainty in comparative studies requires an explicit model for the tree. Based on recently developed conditioned branching processes, with Brownian and Ornstein–Uhlenbeck evolution on top, expected species similarities are derived, together with phylogenetic confidence intervals for the optimal trait value. Finally, inspired by these developments, the phylogenetic framework is illustrated by an exploration of questions concerning "time since hybridization", the distribution of which proves to be asymptotically exponential.

Stokastiska modeller för fylogenetiska komparativa metoder – analytiska egenskaper och parameteruppskattning
Krzysztof Bartoszek

Sammanfattning

Fylogenetiska jämförande metoder är väl etablerade verktyg för analys av fenotypisk evolution och adaptation baserat på variation mellan arter. Generellt är emellertid dessa metoder begränsade av övervägande univariata modeller och avsaknad av möjligheter att inkludera osäkerheter och mätfel i såväl fylogeni som i de analyserade egenskaperna. I den här avhandlingen behandlas alla dessa aspekter. Först genom att studera effekterna av korrelerade mätfel på en fylogenetisk regression–analys. Därefter genom att utveckla en multivariat Ornstein–Uhlenbeck–modell för adaptiv evolution, med parameter–skattningar i R. Denna modell tillåter, unikt och för första gången, direkt analys av adaptiv samevolution mellan egenskaper.

Att ta hänsyn till den ofta avsevärda fylogenetiska osäkerheten i jämförande studier kräver en explicit modell för det fylogenetiska trädet. Här används istället de nyutvecklade "betingade förgreningsprocesserna" i kombination med "Brownian motion" och "Ornstein–Uhlenbeck" processer s. a. s. "ovanpå". Ur detta härleds t. ex. "hur lika" arter kan förväntas vara, samt en sorts fylogenetiska konfidensintervall kring det optimala egenskapsvärdet.

Slutligen, inspirerad av ovanstående metodutveckling, ges en tillämpning på aktuella frågor kring "tid sedan hybridisering", vars fördelning visar sig vara asymptotiskt exponentiell.

Acknowledgments

I would like to thank my parents and Ewa Łączyńska for their constant support during the time of my PhD studies.

I would like to thank Staffan Andersson, Petter Mostad, Olle Nerman, Igor Rychlik and Serik Sagitov for their immense help in writing this thesis and Patrik Albin, and Thomas Ericsson for many helpful comments.

I would also like to thank Staffan Andersson, Mohammad Asadzadeh, Jakob Augustin, Wojciech Bartoszek, Joachim Domsta, Peter Gennemark, Thomas F. Hansen, Graham Jones, Michał Krzemiński, Petter Mostad, Bengt Oxelman, Pietro Lió, Jason Pienaar, Krzysztof Podgórski, Maria Prager, Małgorzata Pułka, Serik Sagitov, Jarosław Skokowski, Anil Sorathiya, Franciska Steinhoff, Anna Stokowska and Kjetil Voje for fruitful collaborations.

I finally thank my colleagues at the Department of Mathematical Sciences for a wonderful working environment.

<div align="right">

Krzysztof Bartoszek
Göteborg, August 5, 2013

</div>

List of Papers

The PhD thesis is based on the following papers,

I. Hansen T. F. and **Bartoszek K.** (2012). Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. Systematic Biology, **61**(3):413-425.

II. **Bartoszek K.**, Pienaar J., Mostad P., Andersson S. and Hansen T. F. (2012). A phylogenetic comparative method for studying multivariate adaptation. Journal of Theoretical Biology, **314**:204-215.

III. Sagitov S. and **Bartoszek K.** (2012). Interspecies correlation for neutrally evolving traits. Journal of Theoretical Biology, **309**:11-19.

IV. **Bartoszek K.** and Sagitov S. (*submitted*). Phylogenetic confidence intervals for the optimal trait value.

V. **Bartoszek K.** (*submitted*). Quantifying the effects of anagenetic and cladogenetic evolution.

VI. **Bartoszek K.**, Jones G., Oxelman B. and Sagitov S. (2012). Time to a single hybridization event in a group of species with unknown ancestral history. Journal of Theoretical Biology, **322**:1-6.

List of papers not included in this thesis

VIII. **Bartoszek K.** and Stokowska A. (2010). Performance of pseudo–likelihood estimator in modelling cells' proliferation with noisy measurements. Proceedings of the XII International Workshop for Young Mathematicians "Probability Theory and Statistics", Cracow: pp. 21-42.

IX. **Bartoszek K.**, Krzemiński M., and Skokowski J. (2010). Survival time prognosis under a Markov model of cancer development. Proceedings of the Sixteenth National Conference on Applications of Mathematics in Biology and Medicine, Krynica: pp. 6-11.

X. Asadzadeh M. and **Bartoszek K.** (2011). A combined discontinuous Galerkin and finite volume scheme for multi–dimensional VPFP system. AIP Conference Proceedings, **1333**:57-62.

XI. Steinhoff F. S., Graeve M., **Bartoszek K.**, Bischof K. and Wiencke C. (2012). Phlorotannin Production and Lipid Oxidation as a Potential Protective Function Against High Photosynthetically Active and UV Radiation in Gametophytes of *Alaria esculenta* (Alariales, Phaeophyceae). Photochemistry and Photobiology, **88**(1):46-57.

XII. **Bartoszek K.** (2012). The Laplace motion in phylogenetic comparative methods. Proceedings of the Eighteenth National Conference on Applications of Mathematics in Biology and Medicine, Krynica Morska: pp. 25-30.

XIII. **Bartoszek K.** and Krzemiński M. (*in press*). Critical case stochastic phylogenetic tree model via the Laplace transform. Demonstratio Mathematica.

Contents

1 Introduction

1.1 Phylogenetic comparative methods

In evolutionary biology one of the fundamental concerns is how different inherited traits evolve, depend on each other and react to the changing environment. The natural approach to answer questions related to these problems is to record trait measurements and environmental conditions, and subsequently analyze them.

There is however a problem that is not taken into account with just this approach. Namely, the sample is not independent as the different species are related by a common evolutionary history. The immediate consequence of this common evolutionary history is that closely related species will have similar trait values. As it is very probable that closely related species will be living in similar environments we could observe a false dependency between traits and environment. One needs to take into account this common phylogenetic history to be able to distinguish between effects stemming from similarity and evolutionary effects. This is of particular importance when studying the adaptation of traits towards environmental niches. Are species living in a similar environment similar because they have adapted to it or is it because they are closely related?

In statistical terms forgetting about the inter–species dependency means that we would be analyzing data under a wrong model. Such an analysis from the perspective of the correct model could lead to biased estimates and misleading parameter interpretations. It would certainly result in wrong confidence intervals and p–values.

The thesis presented here considers only a very specific part of the extremely wide field of phylogenetics — the mathematics and computations involved in modelling phenotype evolution and estimation of parameters of evolutionary models for continuous comparative data. A general overview of the field of phylogenetics has been presented by e.g. Felsenstein (2004); Lemey et al. (2009). Biological background for phylogenetic comparative methods has been discussed by e.g. Harvey and Pagel (1991) with methodologies for discrete data and simple continuous settings. Evolution of discrete characters is also discussed by Pagel (1994, 1999) and Paradis (2006) discusses a versatile R (R Core Team, 2013) package (ape) with many example analysis.

This thesis is based on six papers. The first two, concerning parameter estimation, each have R code attached to them.

Paper I Hansen and Bartoszek (2012) concerns the problem of measurement error (or observational variance, as most comparative studies take average values from a number of individuals) in phylogenetic regression studies. The problem of measurement error is widely addressed in the literature (see e.g. Fuller, 1987; Gleser, 1992; Buonaccorsi, 2010). However the general case of dependent errors in the predictor variables or

dependently evolving predictor variables seems to be lacking in literature. In Paper I we consider general covariance matrices $\mathbf{V}_d$ and $\mathbf{V}_u$ relating the predictors and measurement errors between observations. We show that if one assumes some structure on these matrices then substantial simplifications can be made in the bias formula. Correcting for the bias can increase the mean square error of the estimate so we propose a criterion, depending on the observed data, that indicates whether correction is beneficial or not.

Paper II (Bartoszek et al., 2012) presents and develops a multivariate model of trait evolution based on an Ornstein–Uhlenbeck (OU) (Felsenstein, 1988; Hansen, 1997; Butler and King, 2004; Hansen et al., 2008) type stochastic process, often used for studying trait adaptation, co–evolution, allometry or trade–offs. Included with the paper is an R software package to estimate the model's parameters. We are not limited to studying interactions between traits and environment. Within the presented multivariate framework it is possible to pose and rigorously test hypotheses about adaptation or trade–offs taking into account the phylogeny.

In the classical phylogenetic comparative methods setting it is assumed that the phylogeny is fully resolved (however this has not always been the case Edwards, 1970). This might seem reasonable as with the current wealth of molecular information we can obtain more and more accurate trees. However there is interest in tree–free methods (Bokma, 2008). In many cases the tree itself might not be of direct interest, in fact it could actually be in some situations a nuisance parameter, motivating methods that preserve distributional properties of the observed phenotypic value without a dependence on a particular tree (Bokma, 2010). We still have unresolved clades, for example in the Carnivora order (used for an example analysis by Crawford and Suchard, 2013). Another example for the usefulness of tree–free methods is in the analysis of fossil data (Slater et al., 2012, underline the importance of incorporating fossil data in comparative studies). There may be available rich fossilized phenotypic information but the molecular material might have degraded so much that it is impossible to infer evolutionary relationships.

Crawford and Suchard (2013) in their discussion suggest that with the advancement of our knowledge of the tree of life interest in tree–free methods may possibly diminish along with the number of unresolved clades. Firstly, estimation of parameters of complex (even as "simple" as the Ornstein–Uhlenbeck one) evolutionary models can take intolerable amounts of time for large phylogenies. Therefore unless there will be a major speed–up in estimation algorithms and computing power (but this will also probably come at a financial and energetic cost) tree–free methods should be considered as a viable alternative. Secondly, in situations where we can do a tree–based analysis "integrating" over the phylogeny can be

used as a sanity check, whether the conclusions based on the inferred phylogeny deviate much from those from a "typical" phylogeny. Tree–based methods very rarely use analytical estimators, they are currently nearly always either a numerical optimization of the likelihood function or MCMC/simulation based (Approximate Bayesian Computations). Such estimation algorithms, especially in a high–dimension parameter space are always at risk of falling into a local minimum. Therefore thirdly, contradictory tree–free results will give a reason for investigation. Additionally numerical and Bayesian methods nearly always require a seed/prior distribution to start the estimation. Calculating these using tree–free methods might be an attractive and fast option for the researcher.

Tree–free methods naturally require a modelling framework for the phylogeny. The one of conditioned (on the number of extant tips) branching processes is an appealing option. We know how many species we have observed. This statement of course raises another statistical question that still needs to be addressed — what if this number is only a random draw from the true number of contemporary species, this can be the case especially with lower orders i.e. below simple invertebrates or primitive plants. Conditioned branching processes, motivated amongst other by phylogenetic questions, have received significant attention in the past decade. In particular the research of Steel and McKenzie (2001); Popovic (2004); Aldous and Popovic (2005); Gernhard (2008a); Stadler (2008, 2009, 2011b); Hartmann et al. (2010); Lambert (2010); Mooers et al. (2012) has provided tools for us to develop tree–free phylogenetic comparative methods.

Despite the long time that phylogenetic comparative methods have been around not many analytical statistical properties of them have been found. Ané (2008); Ho and Ané (2013) could be the most notable exceptions where the consistency of phylogenetic maximum likelihood estimators is studied. However they differ from our work by a different model of tree growth and treatment of the ancestral state. This multitude of possible modelling assumptions all having their own biological justification makes phylogenetic comparative methods exciting to study and we hope that our conditioned branching process approach will aid in understanding them.

In line with the current comparative methods framework we consider three evolutionary models on top of our unobserved phylogeny. In Paper III (Sagitov and Bartoszek, 2012) we consider the Brownian motion model (Felsenstein, 1985). This was the first continuous comparative model introduced but Edwards (1970) had already considered the likelihood for a Brownian motion on top of an unobserved pure birth–tree. For it we calculate the interspecies correlation coefficient — how similar do we expect two randomly sampled species to be. Paper IV moves on to the Ornstein–Uhlenbeck model (Felsenstein, 1988; Hansen, 1997; Butler

and King, 2004; Hansen et al., 2008; Bartoszek et al., 2012) interpreted by us as an adaptive model. We assume a single–peak model, calculate for the pure–birth tree phylogenetically corrected confidence intervals for the optimal trait value and derive central limit theorems for the sample mean. Interestingly the form of the correction and central limit theorem depends on the value of α (the adaptation rate) with a phase transition occurring at α equalling half the speciation rate.

In Paper V we add to the Brownian motion and Ornstein–Uhlenbeck evolutionary models a jump component at speciation events (Mooers and Schluter, 1998; Mooers et al., 1999; Bokma, 2002, 2008, 2010). Our study indicates that speciation jump events decorrelate species and we also postulate that using quadratic variation is convenient to compare effects of gradual and punctuated evolution.

To derive analytics of tree–free models one has to study probabilistic properties of coalescent times of phylogenetic trees generated by birth–death models. These results, some of which, like the Laplace transform of the height of a pure birth tree (Paper IV) we believe are novel themselves, open the way to study the statistical properties of various tree indices (e.g. the total cophenetic index) and metrics on the space of trees. Such properties are needed so one can rigorously compare trees and ask if they fit a particular speciation–extinction model (Pybus and Harvey, 2000; Felsenstein, 2004). These have been of course studied previously in particular by Cardona et al. (2010); Mulder (2011); Cardona et al. (2012); Mir et al. (2013). Their approach however has been heavily influenced by number theory resulting in very general (in terms of models) but lengthy proofs. Our approach on the other hand usually results in derivations which are very short but dependent on knowing the properties of coalescent times under a particular branching process (see Papers III, V).

The final paper underlying this thesis, Paper VI (Bartoszek et al., 2013), is in a different spirit but has been motivated by our results concerning conditioned branching processes. We model a single hybridization event inside a pure–birth tree. We assume that for each fixed n there is a constant hybridization rate, β_n between each pair of species. Then if $n\beta_n/\lambda \to 0$, where λ is the speciation rate, we obtain that the distribution of the time till the hybridization event converges to the exponential distribution with rate 2λ.

1.2   Stochastic differential equations

There are two components of our tree–free framework. The first one is the model for the randomly evolving phenotype. The two currently standard models (see Section 2.2 for their definitions) are the Brownian motion model (Felsenstein, 1985) and the Ornstein–Uhlenbeck model (Hansen, 1997). These two stochastic processes appear as solutions of stochastic

differential equations (SDEs) given in the form,

$$\mathrm{d}X(t) = \mu(t, X(t))\mathrm{d}t + \sigma(t, X(t))\mathrm{d}W(t), \tag{1.1}$$

with initial condition X_0. $W(t)$ is the standard Wiener process — independent increments with $W(t) - W(s) \sim \mathcal{N}(0, t - s), 0 \le s \le t$. Traditionally Eq. (1.1) is interpreted as,

$$X(t) = X(0) + \int_0^t \mu(u, X(u))\mathrm{d}u + \int_0^t \sigma(u, X(u))\mathrm{d}W(u), \tag{1.2}$$

where we understand $\int_0^t Y(u)\mathrm{d}W(u)$ as the Itô stochastic integral. We assume that $t \in [0, T]$ for some (large) value of T. The sample paths of the process $Y(u)$ (notice as $\sigma(u, X(u))$ depends on X it is a stochastic process) have to satisfy (apart from the measurability and adaptedness of Y),

$$P\left(\int_0^T Y^2(u)\mathrm{d}u < \infty\right) = 1$$

and the Itô integral is the limit in probability,

$$\int_0^t Y_n(u)\mathrm{d}W(u) \xrightarrow{P} \int_0^t Y(u)\mathrm{d}W(u),$$

for a sequence of simple (piecewise constant, adapted and measurable) processes (on a predefined partition sequence

$$0 = t_0 < t_1^n < \ldots < t_{k-1}^n < t_k^n < t_{k+1}^n = T, \quad t \in (t_r^n, t_{r+1}^n],$$

such that $\max_i(t_{i+1}^n - t_i^n) \to 0$ as $n \to \infty$) $Y_n(t)$ converging to $Y(t)$ in probability. The Itô integral $\int_0^t Y_n(u)\mathrm{d}W(u)$ is defined as,

$$\int_0^t Y_n(u)\mathrm{d}W(u) := \sum_{i=1}^r Y_n(t_{i-1})\left(W(t_i) - W(t_{i-1})\right) + Y_n(t_r)\left(W(t) - W(t_{r-1})\right).$$

It can be shown that the sequence Y_n of processes exists (Klebaner, 2007; Øxendal, 2007).

Adopting the convention that matrices are written in bold–face, vectors as columns in normal font with an arrow above $(\vec{Z})$ and scalar values in normal font, systems of stochastic differential equations can be written similarly,

$$\mathrm{d}\vec{X}(t) = \vec{\mu}(t, \vec{X}(t))\mathrm{d}t + \mathbf{\Sigma}(t, \vec{X}(t))\mathrm{d}\vec{W}(t)), \tag{1.3}$$

with initial condition $\vec{X}_0$. The difference is that now $\vec{X}(t), \vec{\mu}(t, \vec{X}(t))$ are vector valued process, $\mathbf{\Sigma}(t, \vec{X}(t))$ is a matrix valued process and $\vec{W}(t)$ is

a multidimensional standard Wiener processes. As before Eq. (1.3) is interpreted as,

$$\vec{X}(t) = \vec{X}(0) + \int_0^t \vec{\mu}(u, \vec{X}(u))\mathrm{d}u + \int_0^t \boldsymbol{\Sigma}(u, \vec{X}(u))\mathrm{d}\vec{W}(u)), \qquad (1.4)$$

where we understand $\int_0^t \vec{Y}(u)\mathrm{d}\vec{W}(u)$ as a vector of entry–wise Itô stochastic integrals.

A natural question to ask is whether Eqs. (1.1, 1.3, 1.2, 1.4) uniquely define a process (solution to the SDE) and does this process (solution) exist. The sufficient condition (Klebaner, 2007) for this is that the functions $\vec{\mu}(\cdot, \cdot)$ and $\boldsymbol{\Sigma}(\cdot, \cdot)$ satisfy a local Lipshitz condition,

$$\forall_{m\in\mathbb{N}}\exists_{K_m>0}\forall_{t\in[0,T]}\forall_{\vec{x},\vec{y}\in\mathbb{R}^{d_X}} \left\{ \|\vec{x}\|, \|\vec{y}\| \leq m \right.$$
$$\left. \Longrightarrow \|\vec{\mu}(t,\vec{x}) - \vec{\mu}(t,\vec{y})\|^2 + \sum_{i=1}^{d_X}\sum_{j=1}^{d_W} \left(\boldsymbol{\Sigma}_{i,j}(t,\vec{x}) - \boldsymbol{\Sigma}_{i,j}(t,\vec{y})\right)^2 \leq K_m\|\vec{x}-\vec{y}\|^2 \right\}$$

and a global linear growth condition,

$$\exists_{C>0}\forall_{t\in[0,T]}\forall_{\vec{x}\in\mathbb{R}^{d_X}} \|\vec{\mu}(t,\vec{x})\|^2 + \sum_{i=1}^{d_X}\sum_{j=1}^{d_W} \boldsymbol{\Sigma}_{i,j}(t,\vec{x})^2 \leq C(1 + \|\vec{x}\|^2),$$

where $\|\cdot\|$ is the Euclidean norm, d_X the dimension of $\vec{X}$ and d_W the dimension of $\vec{W}$. In the univariate case the conditions are the same with $d_X = d_W = 1$.

The above has of course been a very compact summary of the bare essentials of how we understand stochastic differential equations. In our modelling frameworks we do not focus on the mathematical theory of stochastic differential equations but rather exploit the Markovian nature and finite dimensional distributions of the solution (for very specific ones — the Brownian motion and Ornstein–Uhlenbeck/Vašiček process). A mathematical introduction to stochastic differential equations can be read from e.g. Klebaner (2007); Medvegyev (2007); Øxendal (2007); Iacus (2008). Additionally Iacus (2008) provides rich R code concerning stochastic differential equations.

1.3 Conditioned branching processes

The second part of a tree–free model is the model of the phylogenetic tree. We choose the framework of *conditioned branching processes* (conditioned — on number of contemporary species). This topic has received significant attention in the recent years (e.g. Aldous and Popovic, 2005;

Gernhard, 2008a; Stadler, 2009, 2011b; Lambert, 2010; Stadler and Steel, 2012; Mooers et al., 2012). The basic mathematical object is a branching process with constant birth and death rates (respectively $\lambda \geq \mu \geq 0$). The intuitive construction of such a process is that we start with a single particle at time 0 and then the particle either lives for an exponential (with rate μ) time and dies or lives for an exponential (with rate λ) time and splits into two particles behaving identically as the parent. From our perspective, where the branching process is a modelling tool, this informal definition is sufficient. A formal treatment can be found in e.g. Harris (1963); Athreya and Ney (2000) while Kimmel and Axelrod (2002); Haccou et al. (2007) discuss combining branching processes with biological questions.

We are interested in conditioning on n, the number of current species represented by evolving particles of our branching processes. The motivation behind this in the phylogenetic comparative methods setting is obvious. We have our comparative sample and we assume that we have measured all species in the clade. We note that this is of course a simplification and one should rather consider a framework (and this is one of the possible future directions of developing this thesis) where the number of observed species is just a random sample from the true (possibly unknown) number of species. It has already been mentioned in the introduction this can be an issue especially when studying organisms lower than simple invertebrates.

Gernhard (2008a) presents the important properties of conditioned branching processes. A key property is that conditional on the tree height the times to coalescent of extant nodes are independent, identically distributed random variables. As we don't know the tree height, only the number of contemporary nodes, the strategy is to introduce a prior (before observing the number of extant nodes) distribution for it. Using this prior and the above mentioned independence property we obtain posterior distributions of characteristics of interest e.g. tree height, times of speciation events. Aldous and Popovic (2005); Gernhard (2008a) propose to use an (improper) uniform prior distribution on $(0, \infty)$ for the tree height. In the case of the pure birth tree ($\lambda > \mu = 0$) this is equivalent (Hartmann et al., 2010; Stadler and Steel, 2012) to stopping the tree just before the n–th speciation event (i.e. just before there are $n+1$ species). Additionally the interspeciation times in a Yule tree can be elegantly characterized. Namely the time between the $(k-1)$–th and k–th speciation events are exponentially distributed with rate $k\lambda$ as this is the minimum of k rate λ exponential random variables (Feller, 1971; Gernhard, 2008b, and Papers IV, V, VI).

7

1.4   Properties of the gamma function

The gamma function appears in formulae for important properties of our tree–free models. Below we summarize the well known (and exploited by us) properties of this function. The gamma function is defined as,

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} \mathrm{d}t$$

and the function has the fundamental recursive property $\Gamma(z+1) = z\Gamma(z)$ resulting for integer z, $\Gamma(z) = (z-1)!$. There is a close relation between the gamma function and the beta function $B(x, y)$,

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} \mathrm{d}t = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

and for large x and fixed y, $B(x, y) \sim \Gamma(y) x^{-y}$. In Papers IV and V we rely heavily on the below summation formula (verifiable by induction if $z \neq y$ and directly if $z = y$),

$$\sum_{k=1}^{n-1} \frac{\Gamma(k+y)}{\Gamma(k+z+1)} = \begin{cases} \frac{\Gamma(n+z)\Gamma(y+1) - \Gamma(z+1)\Gamma(n+y)}{\Gamma(z+1)\Gamma(n+z)(z-y)}, & z \neq y, \\ \Psi(n+y) - \Psi(1+y), & z = y, \end{cases} \tag{1.5}$$

where $\Psi(z)$ is the polygamma function defined as $\Psi(z) = \Gamma'(z)/\Gamma(z)$. This function has the property that for natural $z > 0$, $\Psi(z) = \gamma + H_{z-1,1}$, where $\gamma \approx 0.577$ is the Euler–Mascheroni constant and $H_{n,k}$ is the k–th generalized harmonic number,

$$H_{n,k} = \sum_{i=1}^n \frac{1}{i^k}$$

(instead of $H_{n,1}$ in Paper III a_n is used, in Paper IV h_n, in Paper V H_n).

2 Interacting traits

2.1 Paper I — Measurement error in Phylogenetic Comparative Methods

Paper I concerns the issue of correcting for measurement error in regression where the predictor variables and measurement errors can be correlated in an arbitrary way. The bias caused by measurement error in regression is a well studied topic in the case of independent observations (see e.g. Fuller, 1987; Buonaccorsi, 2010). However as we show in Paper I everything complicates when the predictor variables are dependent. If in addition we allow for measurement errors to be dependent between observations we get an even more complicated situation. This is typically the situation with comparative data. The predictor variables are usually species' traits evolving on the phylogeny. The phenotypic data is commonly species' means from a number of observations. Attached to these means is the intra–species variability, being statistically the same as (phylogenetically) dependent measurement errors.

We adopt the following notation: a variable with subscript $_o$ will denote an observed (with error) variable, while the subscript $_t$ will mean the true, unobserved value of the variable. Additionally by $\mathbf{I}$ we denote the identity matrix of appropriate size. For a matrix $\mathbf{M}$, $\mathbf{M}^T$ denotes the transpose of $\mathbf{M}$ and $\mathbf{M}^{-1}$ the inverse. The operation $\text{vec}(\mathbf{M})$ means the vectorization, *i.e.* stacking of columns onto each other, of the matrix $\mathbf{M}$ and $\text{vec}^{-1}(\vec{Z})$ the inverse of vec operation (assuming matrix sizes are known). For two random variables X and Y the notation $X \perp\!\!\!\perp Y$ means that they are independent.

We consider the general linear model,

$$\vec{Y}_t = \mathbf{D}_t \vec{\beta} + \vec{r}_t, \quad \vec{r}_t \sim \mathcal{N}(\vec{0}, \mathbf{V}_t), \tag{2.1}$$

where $\vec{Y}_t$ is a vector of n observations of the dependent variable, $\vec{\beta}$ is a vector of m parameters to be estimated, $\mathbf{D}_t$ is an $n \times m$ design matrix and $\vec{r}_t$ is a vector of n noise terms with $n \times n$ covariance matrix $\mathbf{V}_t$.

To the general linear model of Eq. (2.1) we want to introduce a measurement error model. We write the model with errors in the design matrix and the response variables as,

$$\mathbf{D}_o = \mathbf{D}_t + \mathbf{U},$$
$$\vec{Y}_o = \vec{Y}_t + \vec{e}_y,$$

where $\mathbf{U}$ is a $n \times m$ matrix of random observation errors in the elements of $\mathbf{D}_t$, and $\vec{e}_y$ is a vector of length n of observation errors in $\vec{Y}_t$. Each column of $\mathbf{U}$ is a vector of observation errors for a predictor variable. Furthermore we assume that $\text{vec}(\mathbf{D}_t)$ and $\text{vec}(\mathbf{U})$ are zero–mean normal

random vectors, with covariance matrices equalling $\mathbf{V}_d$ and $\mathbf{V}_u$ respectively. Putting all the elements together, the regression with observation error model is the following,

$$
\begin{aligned}
\vec{Y}_t &= \mathbf{D}_t\vec{\beta} + \vec{r}_t, & \text{vec}(\mathbf{D}_t) &\sim \mathcal{N}(\vec{0}, \mathbf{V}_d), \quad \vec{r}_t \sim \mathcal{N}(\vec{0}, \mathbf{V}_t), \\
\vec{Y}_o &= \vec{Y}_t + \vec{e}_y, & \vec{e}_y &\sim \mathcal{N}(\vec{0}, \mathbf{V}_e), \\
\mathbf{D}_o &= \mathbf{D}_t + \mathbf{U}, & \text{vec}(\mathbf{U}) &\sim \mathcal{N}(\vec{0}, \mathbf{V}_u), \\
\vec{Y}_o &= (\mathbf{D}_o - \mathbf{U})\vec{\beta} + \vec{r}_t + \vec{e}_y.
\end{aligned}
$$

If we further assume that all the errors are independent of each other and the other variables, i.e. $\vec{r}_t \perp\!\!\!\perp \vec{e}_y$, $\vec{r}_t \perp\!\!\!\perp \mathbf{U}$, $\mathbf{U} \perp\!\!\!\perp \vec{e}_y$, $\mathbf{U} \perp\!\!\!\perp \mathbf{D}_t$ and $\vec{Y}_t \perp\!\!\!\perp \vec{e}_y$ then we can write the last equation as, $\vec{Y}_o = (\mathbf{D}_o - \mathbf{U})\vec{\beta} + \vec{r}$, where $\vec{r} = \vec{r}_t + \vec{e}_y$, so $\vec{r} \sim \mathcal{N}(\vec{0}, \mathbf{V})$, where $\mathbf{V} = \mathbf{V}_t + \mathbf{V}_e$. In Paper I we represent the noise as $\vec{r} = \vec{r}_t + \vec{e}_y - \mathbf{U}\vec{\beta}$ and consider the regression model $\vec{Y}_o = \mathbf{D}_o\vec{\beta} + \vec{r}$ but here we do not do this so that we do not have to consider iterative procedures. We do not assume any structure (in particular diagonality) of the covariance matrices $\mathbf{V}$, $\mathbf{V}_t$, $\mathbf{V}_e$, $\mathbf{V}_d$ or $\mathbf{V}_u$. The generalized least squares estimator of $\vec{\beta}$ can be written, as

$$
\begin{aligned}
\hat{\vec{\beta}} &= (\mathbf{D}_o^T\mathbf{V}^{-1}\mathbf{D}_o)^{-1}\mathbf{D}_o^T\mathbf{V}^{-1}((\mathbf{D}_o - \mathbf{U})\vec{\beta} + \vec{r}) \\
&= (\mathbf{D}_o^T\mathbf{V}^{-1}\mathbf{D}_o)^{-1}\mathbf{D}_o^T\mathbf{V}^{-1}(\mathbf{D}_t\vec{\beta} + \vec{r}).
\end{aligned}
$$

We want to compute the expectation of this estimator conditional on the observed predictor variables $\mathbf{D}_o$. We assumed $\vec{r}$ had zero mean and is independent of $\mathbf{D}_o$ so we have,

$$
\begin{aligned}
\mathrm{E}\left[\hat{\vec{\beta}}|\mathbf{D}_o\right] &= (\mathbf{D}_o^T\mathbf{V}^{-1}\mathbf{D}_o)^{-1}\mathbf{D}_o^T\mathbf{V}^{-1}\,\mathrm{E}\left[\mathbf{D}_t|\mathbf{D}_o\right]\vec{\beta} \\
&= (\mathbf{D}_o^T\mathbf{V}^{-1}\mathbf{D}_o)^{-1}\mathbf{D}_o^T\mathbf{V}^{-1}\,\mathrm{E}\left[\mathbf{D}_o - \mathbf{U}|\mathbf{D}_o\right]\vec{\beta} \\
&= (\mathbf{I} - (\mathbf{D}_o^T\mathbf{V}^{-1}\mathbf{D}_o)^{-1}\mathbf{D}_o^T\mathbf{V}^{-1}\,\mathrm{E}\left[\mathbf{U}|\mathbf{D}_o\right])\vec{\beta}.
\end{aligned}
$$

It remains to calculate $\mathrm{E}\left[\mathbf{U}|\mathbf{D}_o\right]$, to do this we need to work with its vectorized form, $\mathrm{E}\left[\text{vec}(\mathbf{U})|\mathbf{D}_o\right]$. Using common facts about the multivariate normal distribution and that $\mathbf{D}_o = \mathbf{D}_t + \mathbf{U}$,

$\mathrm{E}\left[\text{vec}(\mathbf{U})|\mathbf{D}_o\right]$
$= \mathrm{E}\left[\text{vec}(\mathbf{U})\right] + \text{Cov}\left[\text{vec}(\mathbf{U}), \text{vec}(\mathbf{D}_o)\right]\text{Var}\left[\text{vec}(\mathbf{D}_o)\right]^{-1}(\text{vec}(\mathbf{D}_o) - \mathrm{E}\left[\text{vec}(\mathbf{D}_o)\right])$
$= \text{Var}\left[\text{vec}(\mathbf{U})\right]\text{Var}\left[\text{vec}(\mathbf{D}_o)\right]^{-1}\text{vec}(\mathbf{D}_o) = \mathbf{V}_u\mathbf{V}_o^{-1}\text{vec}(\mathbf{D}_o)$,

where $\mathbf{V}_o = \mathbf{V}_d + \mathbf{V}_u$. Therefore we have that,

$$
\mathrm{E}\left[\hat{\vec{\beta}}|\mathbf{D}_o\right] = \mathbf{K}\vec{\beta}, \tag{2.2}
$$

where the matrix $\mathbf{K}$ (reliability matrix) is given by

$$
\mathbf{K} = \mathbf{I} - (\mathbf{D}_o^T\mathbf{V}^{-1}\mathbf{D}_o)^{-1}\mathbf{D}_o^T\mathbf{V}^{-1}\text{vec}^{-1}(\mathbf{V}_u\mathbf{V}_o^{-1}\text{vec}(\mathbf{D}_o)). \tag{2.3}
$$

The above is a general formula for $\mathbf{V}_o$ and $\mathbf{V}_u$ of any form. Imposing some structure on them can simplify the formula for $\mathbf{K}$ and we can notice that no assumptions are needed on the covariance matrices $\mathbf{V}_t$ and $\mathbf{V}_e$. Below we consider some special cases of $\mathbf{V}_o$ and $\mathbf{V}_u$. Other generalizations and including fixed effects are discussed by Bartoszek (2011).

Independent observations of predictors where $\boldsymbol{\Sigma}_d$ is the covariance matrix of the true (unobserved) predictors in a given observation and $\boldsymbol{\Sigma}_u$ is the covariance matrix of the measurement errors in predictors in a given observation,

$$\mathbf{K} = (\boldsymbol{\Sigma}_d + \boldsymbol{\Sigma}_u)^{-1}\boldsymbol{\Sigma}_d. \tag{2.4}$$

Independent predictors (i.e. the predictors and their errors are independent between each other but not between observations), now $\boldsymbol{\Sigma}_{d_i}$ is the covariance matrix of the true (unobserved) ith predictor between observations and $\boldsymbol{\Sigma}_{u_i}$ the covariance matrix of the measurement error and $\mathbf{d}_{o_i}$ denotes the ith column of $\mathbf{D}_o$,

$$\mathbf{K} = \mathbf{I} - (\mathbf{D}_o^T\mathbf{V}^{-1}\mathbf{D}_o)^{-1}\mathbf{D}_o^T\mathbf{V}^{-1}\left[\boldsymbol{\Sigma}_{u_1}\boldsymbol{\Sigma}_{o_1}^{-1}\mathbf{d}_{o_1}; \ldots; \boldsymbol{\Sigma}_{u_m}\boldsymbol{\Sigma}_{o_m}^{-1}\mathbf{d}_{o_m}\right]. \tag{2.5}$$

Single predictor

$$\mathbf{K} = \mathbf{I} - (\mathbf{D}_o^T\mathbf{V}^{-1}\mathbf{D}_o)^{-1}\mathbf{D}_o^T\mathbf{V}^{-1}\mathbf{V}_u\mathbf{V}_o^{-1}\mathbf{D}_o \tag{2.6}$$

In the case of a single predictor with independent observations it is well known that $\kappa \in (-1, 1)$ (one–dimensional counterpart of Eq. 2.4). Therefore the bias corrected estimator has a larger variance than the uncorrected one. This immediately implies a trade–off if we use the mean square error, $\mathrm{E}\left[(\hat{\beta} - \beta)^2\right]$, to compare estimators.

In Paper I we study the case of dependent observations of a single predictor. We transform the formulae for the mean square errors of both estimators into a criterion in terms of estimable quantities,

$$\sigma_{\hat{\beta}}^2/\kappa^2 < \sigma_{\hat{\beta}}^2 + (\kappa - 1)^2\beta^2,$$

where $\sigma_{\hat{\beta}}^2$ is the variance of the estimator of β. Unlike in the independent observations case κ can take values outside the interval $(-1, 1)$.

2.2 Paper II — Multivariate Phylogenetic Comparative Methods

Due to the increase in computational power there has been a recent development of phylogenetic Ornstein–Uhlenbeck models (e.g. Butler and King, 2004; Hansen et al., 2008; Labra et al., 2009; Beaulieu et al., 2012;

Ingram and Mahler, 2013). Paper II is a novel addition to this direction and considers an adaptive multivariate stochastic differential equation model of phenotype evolution. It includes an R software package, mvSLOUCH (*multiv*ariate *S*tochastic *L*inear *O*rnstein–*U*hlenbeck models for phylogenetic *C*omparative *H*ypotheses) covering nearly all cases of Hansen and Martins (1996); Hansen (1997); Martins and Hansen (1997)'s framework (except some cases where the drift matrix is singular or does not have an eigendecomposition).

The first modelling approach via stochastic differential equations for comparative methods in a continuous trait setting is due to Felsenstein (1985). He proposed that the phenotype evolves as a Brownian motion along the phylogeny. A trait X is evolving as a Brownian motion if it can be described by the following stochastic differential equation along a single lineage,

$$\mathrm{d}X(t) = \sigma \mathrm{d}W(t), \tag{2.7}$$

where $\mathrm{d}W(t)$ is white noise. This means that the displacement of the trait after time t from its initial value is $X(t) \sim \mathcal{N}(X(0), \sigma^2 t)$. This model assumes that there is absolutely no selective pressure on trait X whatsoever, so that the trait value just randomly fluctuates from its initial starting point.

A multi–dimensional Brownian motion model is immediate,

$$\mathrm{d}\vec{X}(t) = \boldsymbol{\Sigma}\mathrm{d}\vec{W}(t), \tag{2.8}$$

where $\boldsymbol{\Sigma}$ will be the diffusion matrix and $\vec{W}$ will be the multidimensional Wiener process. All components of the random vector are normally distributed as $\vec{W}(t) - \vec{W}(s) \sim \mathcal{N}(\vec{0}, (t-s)\mathbf{I})$ and so $\vec{X}(t) \sim \mathcal{N}(\vec{X}(0), t\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T)$. The covariance matrix of all of the phylogenetically dependent data will be $\mathbf{V} = \mathbf{T} \otimes (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T)$, where $\otimes$ is the Kronecker product and $\mathbf{T}$ is the matrix of divergence times (time from the origin of the tree to the point of divergence) of species on the phylogeny. Such Brownian trait evolution has the property of time reversibility so both $\vec{X}(t) - \vec{X}(0)$ and $\vec{X}(0) - \vec{X}(t)$ will be identically distributed. Using this one can find an independent sample on the phylogenetic tree, i.e. the contrasts between nodes such that no pair of nodes shares a branch in the path connecting them. This observation was exploited by Felsenstein (1985) in his independent contrasts estimation algorithm (but see also Freckleton, 2012).

The lack of the possibility of adaptation in the Brownian Motion model was discussed by Hansen (1997); Hansen and Orzack (2005) and therefore a more complex Ornstein–Uhlenbeck type of framework was proposed,

$$\mathrm{d}\vec{Z}(t) = -\mathbf{F}(\vec{Z}(t) - \vec{\Psi}(t))\mathrm{d}t + \boldsymbol{\Sigma}\mathrm{d}\vec{W}(t), \tag{2.9}$$

termed here the Ornstein–Uhlenbeck model. The $\mathbf{F}$ matrix is called the drift matrix, $\vec{\Psi}(t)$ the drift vector and $\boldsymbol{\Sigma}$ the diffusion matrix.

In Paper II we consider in detail a very important decomposition of $\mathbf{F}$ — the multivariate Ornstein–Uhlenbeck Brownian motion (mvOUBM) model (a multivariate generalization of the OUBM model due to Hansen et al., 2008)

$$
\begin{aligned}
\mathrm{d}\left[\begin{array}{c} \vec{Y} \\ \vec{X} \end{array}\right](t) = & -\left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right]\left(\left[\begin{array}{c} \vec{Y} \\ \vec{X} \end{array}\right](t) - \left[\begin{array}{c} \vec{\psi} \\ 0 \end{array}\right](t)\right)\mathrm{d}t \\
& +\left[\begin{array}{c|c} \boldsymbol{\Sigma}^{yy} & \boldsymbol{\Sigma}^{yx} \\ \hline \boldsymbol{\Sigma}^{xy} & \boldsymbol{\Sigma}^{xx} \end{array}\right]\mathrm{d}\mathbf{W}(t).
\end{aligned}
\tag{2.10}
$$

Assuming that $\mathbf{F}$ and $\mathbf{A}$ have positive real part eigenvalues then respectively the $\vec{\Psi}(t)$ function represents a deterministic optimum value for $\vec{Z}(t)$, Eq. (2.9) and $\vec{\psi}(t)$ the deterministic part of the optimum for $\vec{Y}(t)$ in Eq. (2.10). If we write out the SDEs in vector form then f_{ij}/f_{ii} and a_{ij}/a_{ii}, where f_{ij} are the elements of $\mathbf{F}$ and a_{ij} of $\mathbf{A}$, can be understood as effects of Z_j and Y_j on the primary optimum of Z_i and Y_i respectively (Hansen, 1997; Hansen et al., 2008). The eigenvalues, if they have positive real part, can be understood to control the speed of the traits' approach to their optima, while the eigenvectors indicate the path towards the optimum. The diffusion matrix in general represents stochastic perturbations to the traits' approach to their optima. These perturbations can come from many sources internal or external, e.g. they can represent unknown/unmeasured components of the system under study or some random genetic changes linked to the traits.

Paper II includes an R package, mvSLOUCH, that implements a (heuristic) maximum likelihood estimation method to estimate the parameters of the stochastic differential equation (2.9). The package also allows for the estimation of parameters of the special submodel defined by Eq. (2.10). The implementation of the estimation procedure requires combining advanced linear algebra and computing techniques with probabilistic and statistical results. These have been discussed in detail by Bartoszek (2011).

Paper II also includes a reanalysis of a Cervidae data set (deer antler length, male and female body masses, compiled by Plard et al., 2011). Our results generally confirm those of Plard et al. (2011), that (i) there is positive linear relationship between the logarithms of antler length and male body mass and (ii) the mating tactic does not influence directly the antler length nor the male body mass. The allometry between antler length and male body mass is greater than 1/3 indicating that there is more than just a proportional increase of antler length (1–dimensional) when body mass increases (3–dimensional). We can also observe that the estimates of the effect of breeding group size on the antler length and male body mass are larger with the increase in breeding group size.

However our analysis in addition shows that antler length and male body mass are adapting very rapidly to changes in female body mass.

We can also see that they adapt independently of each other, there is no direct influence of one variable on the primary optimum of the other and all dependencies between antler length and male body mass are due to the common female body mass predictor variable, and correlations in noise pushing them away from their respective optima. This adaptation result was not observed by Plard et al. (2011) as they only considered a Brownian motion evolutionary model.

3 Tree–free models

3.1 Paper III — Interspecies correlation for neutrally evolving traits

Paper III is a first step to combine branching process with stochastic models of phenotype evolution. We asked how similar do we expect species phenotypes to be.

To measure the phenotypic similarity we concentrate on the interspecies correlation coefficient, defined classically as,

$$\rho_n = \frac{\text{Cov}\left[X_1, X_2\right]}{\sqrt{\text{Var}\left[X_1\right]\text{Var}\left[X_2\right]}}, \tag{3.1}$$

where X_1 and X_2 are a randomly sampled pair of tip species. We also assume that all n species are contemporary, so Eq. (3.1) will simplify,

$$\rho_n = \frac{\text{Cov}\left[X_1, X_2\right]}{\text{Var}\left[X\right]}, \tag{3.2}$$

where X is a randomly sampled tip species. We have two sources of variation the randomness in the tree and the phenotypic evolution. As the considered phenotypic processes are Markovian and the phenotype is assumed not to influence the branching then using the laws of total variance and covariance we have,

$$
\begin{aligned}
\text{Var}\left[X\right] &= \text{E}\left[\text{Var}\left[X|U_n\right]\right] + \text{Var}\left[\text{E}\left[X|U_n\right]\right], \\
\text{Cov}\left[X_1, X_2\right] &= \text{E}\left[\text{Cov}\left[X_1, X_2|U_n, \tau^{(n)}\right]\right] \\
&\quad + \text{Cov}\left[\text{E}\left[X_1|U_n, \tau^{(n)}\right], \text{E}\left[X_2|U_n, \tau^{(n)}\right]\right] \\
&= \text{E}\left[\text{Cov}\left[X_1, X_2|U_n\right]\right] + \text{Var}\left[\text{E}\left[X_{a_{12}}|U_n, \tau^{(n)}\right]\right],
\end{aligned}
\tag{3.3}
$$

where $X_{a_{12}}$ is the most recent common ancestor of the two randomly sampled species, U_n (T in Papers III, V) is the height of the tree and $\tau^{(n)}$ (τ in Papers III, V) is the time to coalescent of a random pair of tip species. To calculate these values we need to assume a stochastic model for the phenotype evolution and for the tree. In Paper III we assumed the Brownian motion model. The Brownian motion model is characterized by two parameters X_0 the ancestral state and σ the diffusion coefficient.

Under it the conditional (on tree) mean, variance and covariance are,

$$
\begin{aligned}
\text{E}\left[X|U_n\right] &= X_0, \\
\text{Var}\left[X|U_n\right] &= \sigma^2 U_n, \\
\text{Cov}\left[X_1, X_2|U_n, \tau^{(n)}\right] &= \sigma^2(U_n - \tau^{(n)}).
\end{aligned}
\tag{3.4}
$$

From Eq. (3.3) we see that we need to know the expectation of U_n and $(U_n - \tau^{(n)})$ — these are dependent on the assumed model of phylogeny. Our phylogeny modelling approach is heavily based on the findings of

T. Stadler (neé Gernhard) concerning conditioned branching processes. The phylogenetic tree is modelled as a constant rate birth–death process (λ, μ being the birth and death rates) conditioned on n contemporary tips. We consider three regimes, $\mu = 0$ (Yule, pure–birth case), $0 < \mu < \lambda$ (supercritical case), $\mu = \lambda$ (critical case) and in each case calculate the expectations, obtaining the following results in the pure–birth $(0 = \mu < \lambda)$ case,

$$
\begin{array}{rcl}
\mathrm{E}\left[U_n\right] & = & \frac{1}{\lambda} H_{n,1}, \\
\mathrm{E}\left[U_n - \tau^{(n)}\right] & = & \frac{2}{\lambda(n-1)}\left(n - H_{n,1}\right),
\end{array}
\tag{3.5}
$$

and supercritical $(0 < \mu < \lambda)$ case,

$$
\begin{array}{rcl}
\mathrm{E}\left[U_n\right] & = & \frac{1}{\mu(\lambda/\mu-1)}\left(H_{n,1} + e_{n,\lambda/\mu} - \ln\frac{\lambda/\mu}{\lambda/\mu-1}\right), \\
\mathrm{E}\left[U_n - \tau^{(n)}\right] & = & \frac{2\left(n + ne_{n,\lambda/\mu} - \frac{\lambda/\mu}{\lambda/\mu-1}\left(H_{n,1} + e_{n,\lambda/\mu} - \ln\frac{\lambda/\mu}{\lambda/\mu-1}\right)\right)}{\mu(\lambda/\mu-1)(n-1)},
\end{array}
\tag{3.6}
$$

where, $e_{n,y} = \int_0^1 x^n/(y-x)\,\mathrm{d}x$. Taking the quotient gives the correlation coefficient, decaying to 0 in both cases as $2/\ln n$. The third critical case needs more careful treatment as here $\mathrm{E}\left[U_n\right] = \infty$. Therefore the variance is undefined. However we define a proper prior distribution for U_n, in Paper III $U_n \sim \mathrm{Uniform}[0, N]$, calculate the necessary expectations $(m := (N+1)/N)$,

$$
\begin{array}{rcl}
\mathrm{E}\left[U_n\right] & = & ne_{n,m} \\
\mathrm{E}\left[U_n - \tau^{(n)}\right] & = & 2ne_{n,m} - \frac{2}{n-1}\left(\frac{ne_{n,m}+n}{m-1} - \frac{m}{(m-1)^2}\left(H_{n,1} + e_{n,m} - \ln\frac{m}{m-1}\right)\right),
\end{array}
\tag{3.7}
$$

and study the interspecies correlation coefficient for two asymptotic regimes,

- $\rho_n = 1 - \frac{1}{2(\ln N - H_{n,1}) + o(1)}$ as $N \to \infty$

- $\rho_n \to 2 - \frac{2}{\alpha}\left(1 + \frac{1}{e^\alpha Ei(\alpha)}\right) + \frac{2}{\alpha^2}\left(1 + \frac{\ln\alpha + \gamma}{e^\alpha Ei(\alpha)}\right)$ as $N, n \to \infty$, $n/N \to \alpha$, $Ei(z)$ is the exponential integral and γ is the Euler–Mascheroni constant.

In addition to the correlation coefficient we are able to calculate two formulae useful from an estimation point of view — the variance of the sample mean and expectation of the sample variance. Both can be written in terms of the interspecies correlation coefficient,

$$
\begin{array}{rcl}
\mathrm{Var}\left[\overline{X}_n\right] & = & \frac{1 + (n-1)\rho_n}{n}\,\mathrm{Var}\left[X\right], \\
\mathrm{E}\left[S_n^2\right] & = & (1 - \rho_n)\,\mathrm{Var}\left[X\right]
\end{array}
\tag{3.8}
$$

and in the Brownian motion case they equal for the different regimes, $(0 = \mu < \lambda)$, $(0 < \mu < \lambda)$, $(0 < \mu = \lambda)$ respectively,

$$
\mathrm{Var}\left[\overline{X}_n\right] = \sigma^2 \begin{cases} \frac{1}{\lambda}\left(2 - \frac{H_{n,1}}{n}\right), \\ \frac{2}{\mu}\left(\frac{1+e_{n,\lambda/\mu}}{\lambda/\mu-1} - \frac{1}{n}\frac{\lambda/\mu+1}{n(\lambda/\mu-1)^2}\left(H_{n,1} + e_{n,\lambda/\mu} - \ln\frac{\lambda/\mu}{\lambda/\mu-1}\right)\right), \\ \frac{1}{\lambda}\left(e_{n,m}\left(2n - 2N + \frac{2N(N+1)}{n} - 1\right) - \frac{A_{n,N}}{n-1}\right), \end{cases}
\tag{3.9}
$$

16

where $A_{n,N} = 2N(n - (N+1)(H_{n,1} - \ln(N+1)))\,\mu, \lambda$ and

$$\mathrm{E}\left[S_n^2\right] = \sigma^2 \begin{cases} \frac{1}{\lambda}\left(\frac{n+1}{n-1}H_{n,1} - 2\frac{n}{n-1}\right), \\ \frac{2}{\mu(n-1)(\lambda/\mu-1)}\left(\left(\frac{2\lambda/\mu}{\lambda/\mu-1} + n\right)\left(H_{n,1} - \ln\frac{\lambda/\mu}{\lambda/\mu-1}\right)\right. \\ \left. + \left(\frac{2\lambda/\mu}{\lambda/\mu-1} - n\right)e_{m,\lambda/\mu} + 1 - 2n\right), \\ \frac{1}{\lambda}\left(e_{n,m}\left(\frac{2Nn}{n-1} - \frac{2N(N+1)}{n-1} - n\right) + \frac{A_{n,N}}{n-1}\right), \end{cases}$$

(3.10)

with respective asymptotic behaviour, in the critical case $n/N \to \alpha$,

$$\mathrm{Var}\left[\overline{X}_n\right] \sim \sigma^2 \begin{cases} \frac{2}{\lambda}, \\ \frac{2}{\mu(\lambda/\mu-1)}, \\ \frac{n}{\lambda}\left(\frac{2}{\alpha^2}\left(\alpha^2 - \alpha + 1\right)e^\alpha Ei(\alpha) - \alpha + \ln\alpha\right), \end{cases}$$

(3.11)

$$\mathrm{E}\left[S_n^2\right] \sim \sigma^2 \begin{cases} \frac{\ln n}{\lambda}, \\ \frac{\ln n}{\mu(\lambda/\mu-1)}, \\ \frac{n}{\lambda}\left(\frac{1}{\alpha^2}\left(2\alpha - \alpha^2 + 2\right)e^\alpha Ei(\alpha) + \alpha - \ln\alpha\right). \end{cases}$$

(3.12)

We can see that as $\mathrm{Var}\left[\overline{X}_n\right] \to 2$ the sample average is not a consistent estimator of the ancestral state (similarly obtained by Ané, 2008). Knowledge of the variance of the sample mean can be developed into phylogenetic confidence intervals for the ancestral state and based on the expectation of the sample variance we can construct an unbiased estimator for σ^2. In Paper IV we develop these concepts for the Yule–Ornstein–Uhlenbeck process.

3.2   Paper IV — Tree–free phylogenetic confidence intervals for an adaptive model

A natural continuation of Paper III was to consider the Ornstein–Uhlenbeck process for phenotype evolution,

$$\mathrm{d}X(t) = -\alpha(X(t) - \theta)\mathrm{d}t + \sigma\mathrm{d}B(t).$$

(3.13)

In our framework we interpret that the trait $X(t)$ is adapting (we assume $\alpha > 0$) and trying to approach its adaptive peak (optimal value) — θ. We assume in Paper IV that the value of θ is constant over the whole phylogeny but notice that this is not what is desired in comparative methods. In a common comparative methods setting we should be able to identify groups of species (and subclades of the phylogeny — if available) that have their own unique adaptive peaks. Therefore our results on the single adaptive–peak could be useful for analyzing subclades and we leave a tree–free model with multiple adaptive peaks as a possible development of Paper IV.

In Paper IV we consider two classical estimators, the sample mean

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the sample variance

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2.$$

It turns out that the sample mean is an asymptotically unbiased estimator of θ and the sample variance is an asymptotically unbiased estimator of the stationary variance of the Ornstein–Uhlenbeck process, $\sigma^2/2\alpha$. We also show that these estimators are consistent, $\text{Var}\left[\overline{X}_n\right]$, $\text{Var}\left[S_n^2\right] \to 0$. By this we are able to derive phylogenetic confidence intervals for θ that differ from the usual classical ones by a phylogenetic correction factor. This factor is greater than 1 as we have a smaller effective sample size due to dependencies. It turns out however that the form of these intervals (and of the variance of the sample mean and sample variance) depends on the value of α, with three possible regimes.

i) $\alpha > 0.5\lambda$ in this case (fast adaptation) the approximate confidence interval at level $1 - x$ will be of the form

$$\overline{X}_n \pm z_{x/2} \frac{\sqrt{S_n^2}}{\sqrt{n}} \sqrt{\frac{2\alpha+1}{2\alpha-1}}, \tag{3.14}$$

where z_x is the x level quantile of the standard normal distribution and we also have a central limit theorem that the standardized sample mean $(\overline{X}_n - \theta)/\sqrt{\sigma^2/2\alpha}$ is asymptotically mean–zero normally distributed with variance $\frac{2\alpha+1}{2\alpha-1}$.

ii) $\alpha = 0.5\lambda$, there seems to be a phase transition for the adaptation rate equalling half the speciation rate. We can derive the following approximate phylogenetic confidence interval

$$\overline{X}_n \pm z_{x/2} \frac{\sqrt{S_n^2}}{\sqrt{n}} \sqrt{2\ln n}, \tag{3.15}$$

and the central limit theorem will hold for the following normalization of the sample average $\sqrt{n/\ln n}(\overline{X}_n - \theta)/\sigma$ which will be asymptotically normal mean–zero with variance 2.

iii) $0 < \alpha < 0.5\lambda$, slow adaptation — in this case the situation is more complicated, the limiting distribution of the normalized $\frac{\overline{X}_n-\theta}{\sqrt{\sigma^2/2\alpha}}$ sample mean will depend on the starting position and we cannot expect it to be normal (compare to a very similar model due to Adamczak and Miłoś, 2011a,b). We can propose the following approximate confidence interval

$$\left(\overline{X}_n - q_\alpha(x/2)\sqrt{S_n^2}n^\alpha, \overline{X}_n + q_\alpha(1-x/2)\sqrt{S_n^2}n^\alpha \right) \tag{3.16}$$

but the quantiles, $q_\alpha(x/2)$, $q_\alpha(1-x/2)$, have to be currently obtained by simulation methods.

The same sort of phase transition at $\alpha = 0.5\lambda$ was noticed by Adamczak and Miłoś (2011a,b) in a related modelling setting. One way of thinking about this (after Adamczak and Miłoś, 2011a,b) is that for a small α local correlations will prevail over the ergodic properties of the Ornstein–Uhlenbeck process. Of course by "small" we mean relative to the value of λ.

In the very recent work Ho and Ané (2013) also consider a phylogenetic Ornstein–Uhlenbeck model. They show that the maximum likelihood estimator of θ is not consistent. This does not contradict our work as they differ from us on modelling assumptions. Firstly they have a different model of tree growth, they assume a nested sequence of ultrametric trees with bounded internal node heights. This means that tree $n-1$ is a subtree of tree n and the tree height is bounded. The second crucial difference is that the ancestral state is not fixed but is a random variable drawn from the stationary distribution of the Ornstein–Uhlenbeck process. This combined with the boundedness of the tree height implies that the correlation between any two species, regardless of the number of tips, is bounded away from 0 for all n and hence the variance of the estimator of θ cannot decrease to 0.

In order to derive the above (using as in Paper III the laws of total variance and covariance) we were required to compute the Laplace transforms of U_n and $\tau^{(n)}$. Using them one can calculate all the moments of U_n and $(U_n - \tau^{(n)})$ and also relate our work to the splitted nodal distance metric between phylogenies (Felsenstein, 2004; Cardona et al., 2010).

3.3 Paper V — Quantifying the effect of jumps at speciation

Papers III and IV assumed Brownian motion and Ornstein–Uhlenbeck evolution respectively. This is in line with the idea that macroevolution is the consequence of microevolutionary changes over many generations. However another possibility, indicated by fossil records, has been pointed out, that phenotypes enjoy long periods of stasis with rapid change concentrated in short time periods (Eldredge and Gould, 1972; Gould and Eldredge, 1993). The first type of evolution is termed gradual, anagenetic or phyletic gradualism, the second cladogenetic or punctuated equilibrium. Naturally one can imagine that evolution could be a combination of the two mechanisms and so a unified framework for them should be developed. Such a framework has been introduced and developed by Bokma (2002, 2008, 2010). Bokma (2002) assumed that the logarithm of the phenotype evolves as a Brownian motion and at each speciation event receives a normally distributed jump. Earlier Mooers et al. (1999) used both anagenetic and cladogenetic models in their Gruinae analysis. Very recently Eastman et al. (2013) model *Anolis* lizards evolution by a Brownian motion with jumps inside branches.

Paper V combines the tree–free approach of Papers III and IV with speciation jumps. We consider both the Brownian motion and Ornstein–Uhlenbeck models. In fact we show, as expected, that the Brownian motion results are limits of the Ornstein–Uhlenbeck ones when $\alpha \to 0$. We calculate the interspecies correlation coefficient and in addition we derive the probability generating functions of the tree height, U_n and time to coalescent of a random pair of tip species, $\tau^{(n)}$. One can see in the Brownian motion case that increasing the variance of the jumps decorrelates species while a numerical treatment of the correlation function indicates the same in the Ornstein–Uhlenbeck case.

One of the biological motivations for considering such models is to determine which sort of evolution is dominating for a trait. As Mooers et al. (1999) point out speciational change should be more relevant for mate choice traits while anagenetic for traits under continuous selection pressures often changing direction. Bokma (2002, 2008, 2010) proposes different methods to estimate parameters of the models, however to quantify effects of each type of evolution and to compare them we need to also know for how long each type of evolution has occurred, e.g. if speciation was extremely rare in some very old lineage then jumps might have had little opportunity to contribute. Therefore in Paper V we suggest that the expectation (as we don't know the tree height nor precise number of jumps in our framework) of the quadratic variation would be a good quantity. In fact we show that it can be decomposed into the sum of the anagenetic and the cladogenetic component and so we can calculate the (expected) proportion of change due to each mode of evolution. In Paper V we illustrate this by interpreting the Hominoidea related estimates of Bokma (2002); Bokma et al. (2012).

One of the requirements of this is that we know the number (or in our case expectation) of jumps occurring along the (random) lineage. We are able to calculate this for the pure–birth tree (see also Steel and McKenzie, 2001, for the same formulae) but had to resort to simulation methods when a death component was included.

Our results hold for a number of different jump models. The jump can have any distribution provided it is mean zero and has a finite variance. It can occur after speciation on both lineages, on a randomly selected one or on each one with a given probability.

Lévy processes such as the Laplace motion (Bartoszek, 2012; Landis et al., 2013) are another approach to the inclusion of evolutionary jumps however they are not as easy to interpret as "a jump at/due to/causing speciation".

3.4   Example application

To illustrate and motivate the usefulness of the tree–free approach we will conduct an example analysis of the logarithm of *Ploceus* female tarsus

length (Fig. 1 — measurements provided by Staffan Andersson, Department of Biological and Environmental Sciences, University of Gothenburg, private communication). The phylogeny of this genus is unknown so a usual comparative approach is not possible. We assume that the logarithm of the female tarsus evolves according to a single optimum Ornstein–Uhlenbeck process,

$$\mathrm{d}X(t) = -\alpha(X(t) - \theta)\mathrm{d}t + \sigma\mathrm{d}W(t)$$

and would like to make inference on the value of θ from species average measurements of 52 members of the *Ploceus* genus. We showed in Paper IV that the sample average is a consistent and asymptotically unbiased estimator of θ and from our data we obtain $\hat{\theta} = \overline{X}_{52} = 19.993$. We would then like to calculate the phylogenetic confidence interval presented in Paper IV. There is a problem as since we do not have the phylogeny, we cannot obtain an estimate of α and λ. However a 33 species phylogeny of a sister clade, the *Euplectes* genus (both are members of the Ploceidae family), is available (Fig. 1, Prager et al., 2008; Prager and Andersson, 2009) and we posses measurements of female tarsus length in 32 species in this genus (measurements, including intra–species variation, provided by Staffan Andersson, $\overline{X}_{32} = 20.991$, $S^2_{32} = 9.5$). Using the R
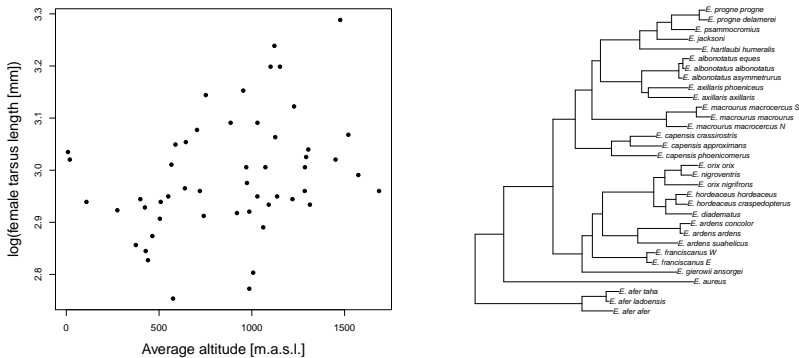

Figure 1: Left: logarithm of female tarsus length of the *Ploceus* genus plotted as a function of average altitude. Right: *Euplectes* phylogeny.

laser package (function pureBirth) we estimate $\hat{\lambda} = 6.146$ and using the mvSLOUCH package (Paper II) we obtain that in this genus (assuming the single optimum Ornstein–Uhlenbeck model) $\hat{\alpha} = 8.719$, $\widehat{\sigma^2} = 166.248$, $\hat{\theta} = 21.158 \pm 1.646$ (regression based confidence intervals conditional on α and σ). As a check the stationary variance equals $\widehat{\sigma^2}/(2\hat{\alpha}) \approx 9.534$

(very close to S_{32}^2 despite the small sample size, measurement error and that the phylogeny is not ultrametric). We assume that these estimates are good enough to plug into the *Ploceus* analysis (however the sample variance in the *Ploceus* genus is $S_{52}^2 = 5.328$ so the results have to be considered approximate). In Paper IV the phylogenetic confidence intervals were derived for $\lambda = 1$ so in our formula we have to take instead of α the value $\alpha^* = \alpha/\lambda$ estimated in our case as, $\hat{\alpha}/\hat{\lambda} \approx 1.419 > 0.5$. This means that we are in the fast adaptation regime and the 95% phylogenetic confidence interval will be given by the formula,

$$\overline{X}_{52} \pm \left(\sqrt{\frac{2\alpha^* + 1}{2\alpha^* - 1}} \right) \left(1.96 \frac{\sqrt{S_{52}^2}}{\sqrt{n}} \right) \approx 19.993 \pm 0.91.$$

The phylogenetic correction factor is $\sqrt{\frac{2\alpha^*+1}{2\alpha^*-1}} \approx 1.445$ so we can see that the phylogeny plays a role. The confidence interval is rather tight so we can say that 19.993 is a fair estimate of θ. We can also see that the confidence intervals in the two genera overlap so the female tarsus length might be under similar pressures in both.

Assuming that S_{52}^2 estimates the stationary variance well enough and taking $\hat{\alpha} = 8.719$ we obtain that $\widehat{\sigma^2} = 2\hat{\alpha}S_{52}^2 \approx 92.91$ in the *Ploceus* genus. This means that the main difference between the two genera (with respect to female tarsus length) is in the magnitude of the stochastic perturbations to the adaption towards the primary optimum. They are greater in the *Euplectes* and so we should expect more diversity in this clade.

The above analysis illustrates our tree–free method: in what situation it can be useful and where more work needs to be done. We can suspect that in the *Ploceus* the optimal value for the female tarsus length should be dependent on some environmental variables, e.g. Fig. 1 suggests the altitude of the species' habitat. The current method cannot take this into account so this indicates a direction of future development. The phylogenetic confidence interval depends on knowing α — we cannot estimate it as the phylogeny of our clade of interest is unknown. Therefore we have to obtain it by other means — here we estimated it from a sister clade with the hope that it would not differ substantially from the one in our clade of interest. It would therefore be desirable to incorporate uncertainty in the estimation of α into the method.

4 Phylogenetic networks

4.1 Introducing networks

Evolutionary relationships between species are traditionally portrayed in a tree like structure. We are however more and more aware that this is not the whole story. There are cases, especially with many plant species e.g. most breeds of wheat, peppermint (hybrid between spearmint and watermint), grapefruit (hybrid between pomelo and Jamaican sweet orange), where a more realistic description is a phylogenetic network due to hybridization events. In terms of phylogenetic comparative methods considering networks instead of trees can be useful in a situation where we are studying a phenotype evolving in separated populations (allowed however to mix from time to time) that could be under different environmental pressures.

The area of phylogenetic networks is a new topic. There is a large number of different types of networks. As in the case of trees one can consider rooted or unrooted networks. The fundamental concept related to unrooted networks is that of splits. A split is a bipartition of the set of taxa such that the two sets are disjoint. Combined they are the set of taxa. In a phylogenetic tree every edge uniquely corresponds to a split. In the network case every split is represented by a collection of edges. Rooted phylogenetic networks can be used to represent hybridization, recombination and reassortment events. They can be also used to reconcile species and gene trees if duplication, loss or transfer events took place. More generally they can be used to represent a set of competing trees on the same set or overlapping sets of taxa. Huson et al. (2010) in their very recent work discuss the different network types and the algorithms presently being used for their reconstruction. They concentrate on non–probabilistic network models and reconstruction methods. Another approach is of course a probabilistic one. Jones et al. (2013) develop a Bayesian methodology for hybridization network reconstruction. In such a case a prior distribution for the time to hybridization would be desirable and Paper VI is an attempt at this.

4.2 Paper VI — Time to hybridization is approximately exponential

In Paper VI we assume a very simple network setup. Firstly we assume that there are n contemporary diploid species linked by a pure–birth tree with speciation rate λ. Then we model hybridization by a Poisson process. If during a time period t there are $k \geq 2$ diploid species then the number of hybridization events, $N_k(t)$, in this time period is Poisson

distributed (parametrized by $\beta > 0$),

$$\mathrm{P}\left(N_k(t) = j\right) = \frac{\left(\binom{k}{2}\beta t\right)^j}{j!}e^{-\binom{k}{2}\beta t}, \quad j = 0, 1, 2, \dots \,. \qquad (4.1)$$

We condition on there being exactly one hybridization event inside the tree induced by the n diploid species. We make no statements about what occurs on the lineage(s) started from this hybridization event. There are two parameters in this model the speciation rate λ and hybridization rate β, however the distribution of the random variable τ_n — the time to the singled–out hybridization event can be characterized in terms of λ and the compound parameter (a relative hybridization rate),

$$\gamma := \frac{\beta}{2\lambda}. \qquad (4.2)$$

We define the random variable $\kappa_n \in \{2, \dots, n\}$ the number of species present when the hybridization event occurred and then decompose,

$$\tau_n = X + \sum_{j=\kappa_n+1}^{n} T_j, \qquad (4.3)$$

where T_i is the random time during which there were exactly i species present, distributed exponentially with rate $i\lambda$ as the minimum of i independent exponential random variables with rate λ, see Fig. 1 in Paper VI and Y is uniformly distributed on the interval $[0, T_{\kappa_n}]$. Using this decomposition we obtain all the moments of τ_n,

$$\mathrm{E}\left[\tau_n^r|\tau_n < \infty\right] = \frac{1}{\sum_{k=1}^{n-1}\frac{k\gamma}{1+k\gamma}}\frac{r!}{\lambda^r}\sum_{k=1}^{n-1}\frac{k\gamma}{1+k\gamma}$$
$$\times \sum_{i_1=k}^{n-1}\sum_{i_2=i_1}^{n-1}\cdots\sum_{i_r=i_{r-1}}^{n-1}\frac{1}{(1+i_1)(1+\gamma i_1)}\cdot\cdots\cdot\frac{1}{(1+i_r)(1+\gamma i_r)}. \tag{4.4}$$

In addition assuming that λ is constant, $n\gamma \to 0$ and proving the identity,

$$\sum_{k=1}^{n-1}k\sum_{i_1=k}^{n-1}\cdots\sum_{i_r=i_{r-1}}^{n-1}\left(\frac{1}{1+i_1}\cdot\cdots\cdot\frac{1}{1+i_r}\right) = 2^{-r}\binom{n}{2}, \qquad (4.5)$$

we obtain that conditional on $\tau_n < \infty$,

$$\tau_n \xrightarrow{\mathcal{D}} \mathrm{Exp}(2\lambda). \qquad (4.6)$$

This weak convergence is due to for each r, $\mathrm{E}\left[\tau_n^r|\tau_n < \infty\right] \to \frac{r!}{(2\lambda)^r}$, the r–th moment of the exponential with 2λ rate distribution. In Paper VI

we also show via simulations (histograms in Fig. 4 therein) that the exponential approximation is good for n even as small as 5 or 10.

In practice we cannot expect to know the parameter γ and it should be estimated from the observed data — n in our case. The method of moments estimator is $\tilde{\gamma}_n = 1/\binom{n}{2}$ and for $n \geq 4$ we have the following bounds on the maximum likelihood estimator, $\hat{\gamma}_n$,

$$\tilde{\gamma}_n = \frac{2}{n(n-1)} \leq \hat{\gamma}_n \leq \frac{2}{n(n-3)} \tag{4.7}$$

with the likelihood function given by,

$$\mathrm{P}_\gamma(\tau_n < \infty) = \prod_{i=1}^{n-1} \frac{1}{1+i\gamma} \sum_{k=1}^{n-1} \frac{k\gamma}{1+k\gamma}. \tag{4.8}$$

5   Future developments

Bartoszek (2011) discussed that one of the ways of building on the presented there work was to get around the need to condition on a pre–estimated phylogenetic tree, i.e. consider a tree–free model. Papers III, IV and V do exactly that. They can be of course further expanded by moving into a multivariate setting (with e.g. predictor/response traits) or allowing for (in the Yule–Ornstein–Uhlenbeck setting) multiple adaptive peaks developing as a Markov chain. Allowing for hypothesis testing concerning the number of adaptive peaks in this framework would be very desirable from a biological point of view, as current comparative analysis methods essentially condition on this and their layout on the tree. To work with a multitrait tree–free comparative model we would need to understand how the different traits influence each other along the unobserved tree and what information on these interactions is contained in the contemporary sample. A starting point is the study of the correlation coefficient in a bivariate neutrally (Brownian motion) evolving trait. Other analysis specific extensions have been discussed in the example application, Section 3.4.

One of the main modelling assumptions of the underlying framework of conditioned branching processes is that we know the number of contemporary species. While this seems reasonable at first, we are still in the process of discovering new (or reclassifying) species (primarily in the lower orders (e.g. simple invertebrates, primitive plants) but there is a very recent example of a new antelope Colyn et al., 2010). Therefore it might be desirable to consider that we have only observed a certain fraction of the species and so can only have an induced sub–tree. Stadler (2009, 2011a); Höhna et al. (2011); Lambert et al. (2013a,b) have already considered questions related to this in different settings. The first step in combining incomplete species sampling with phylogenetic comparative methods could be a study of the behaviour of the interspecies correlation coefficient in this situation. The study of inter–coalescent times of a subtree begun in Paper IV could be a starting point for this.

In Paper V we studied the effects of jumps on the sample mean and variance estimators. An interesting question is how do the jumps change the limit theorems of Paper IV and hence the form of the phylogenetic confidence intervals. The jump component in Paper V can have any distribution (provided it is mean 0 and finite variance). One question is what further properties would be needed for the central limit theorem to still hold in the case of fast adaptation with jumps.

Both Papers IV and V assume that there are no death events, this being naturally unrealistic. The compact form of the interspecies correlation coefficient of Paper III gives hope that further properties of a tree–free model with a death component can be derived analytically and so this is a further research path.

Paper VI can be immediately expanded by including death events, allowing for a specific number of hybridization events/contemporary hybrid species. Combining its results with the tree–free models would allow us to study phenotypes evolving on a network structure instead of a tree.

One can find in Papers III and V links between tree properties and particular phylogenetic indices/metrics. A natural continuation of this is a systematic approach to go through these and other common summary functions associated with trees and study how our results can be applied to them.

The other directions indicated by Bartoszek (2011) still remain open. The results of Paper I can be developed in the direction of studying the effects of measurement error on the estimation of covariance structures.

We can develop Paper II in a couple of directions. The first one concerns modelling trait evolution and interactions among the traits and the environment. An initial step can be to relax the regimes allowing them to change at random times thus forming a Markov chain. Secondly one can try to address the various statistical questions related to phylogenetically structured observations. Paper IV partially addresses this by deriving a phylogenetic confidence interval for a constant on the whole tree optimal trait value. Including the measurement error correction procedures of Paper I into comparative model estimation programs is also an interesting software development direction.

Both the traditional phylogenetic comparative approach and our discussed here tree–free approach assume that there is a one way flow of effects from speciation (phylogeny) to the phenotype. This is of course biologically unrealistic as one would expect some phenotypic explanation for speciation. Therefore the incorporation of a feedback mechanism from the trait process to the branching process in our tree–free framework is another future goal.

6 Bibliography

R. Adamczak and P. Miłoś. CLT for Ornstein–Uhlenbeck branching particle system. *ArXiv e-prints*, 2011a.

R. Adamczak and P. Miłoś. U–statistics of Ornstein–Uhlenbeck branching particle system. *ArXiv e-prints*, 2011b.

D. Aldous and L. Popovic. A critical branching process model for biodiversity. *Adv. Appl. Probab.*, 37(4):1094–1115, 2005.

C. Ané. Analysis of comparative data with hierarchical autocorrelation. *Ann. Appl. Stat.*, 2(3):1078–1102, 2008.

K. B. Athreya and P. E. Ney. *Branching Processes.* Dover Publications, New York, 2000.

K. Bartoszek. The Laplace motion in phylogenetic comparative methods. In *Proceedings of the Eighteenth National Conference on Applications of Mathematics in Biology and Medicine, Krynica Morska*, pages 25–30, 2012.

K. Bartoszek. *Multivariate Aspects of Phylogenetic Comparative Methods*, 2011. Licentiate thesis, University of Gothenburg, Göteborg, Sweden.

K. Bartoszek, J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen. A phylogenetic comparative method for studying multivariate adaptation. *J. Theor. Biol.*, 314:204–215, 2012.

K. Bartoszek, G. Jones, B. Oxelman, and S. Sagitov. Time to a single hybridization event in a group of species with unknown ancestral history. *J. Theor. Biol.*, 322:11–19, 2013.

J. M. Beaulieu, D. C. Jhwueng, C. Boettiger, and B. C. O'Meara. Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. *Evolution*, 66(8):2369–2383, 2012.

F. Bokma. Detection of punctuated equilibrium from molecular phylogenies. *J. Evol. Biol.*, 15:1048–1056, 2002.

F. Bokma. Detection of "punctuated equilibrium" by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. *Evolution*, 62(11):2718–2726, 2008.

F. Bokma. Time, species and seperating their effects on trait variance in clades. *Syst. Biol.*, 59(5):602–607, 2010.

F. Bokma, V. van den Brink, and T. Stadler. Unexpectedly many extinct Hominins. *Evolution*, 66(9):2969–2974, 2012.

J. P. Buonaccorsi. *Measurement Error Models, Methods and Applications*. CRC Press, 2010.

M. A. Butler and A. A. King. Phylogenetic comparative analysis: a modelling approach for adaptive evolution. *Am. Nat.*, 164(6):683–695, 2004.

G. Cardona, M. Llabrés, F. Rosselló, and G. Valiente. Nodal distances for rooted phylogenetic trees. *J. Math. Biol.*, 61(2):253– 276, 2010.

G. Cardona, A. Mir, and F. Rosselló. Exact formulas for the variance of several balance indices under the Yule model. *J. Math. Biol.*, 2012.

M. Colyn, J. Hulselmans, G. Sonet, P. Oudé, J. de Winter, A. Natta, Z. T. Nagy, and E. Verheyen. Discovery of a new duiker species (Bovidae: Cephalophinae) from the Dahomey Gap, West Africa. *Zootaxa*, 2637: 1–30, 2010.

F. W. Crawford and M. A. Suchard. Diversity, disparity, and evolutionary rate estimation for unresolved Yule trees. *Syst. Biol.*, 62(3):439–455, 2013.

J. M. Eastman, D. Wegmann, C. Leuenberger, and L. J. Harmon. Simpsonian "Evolution by Jumps" in an adaptive radiation of *Anolis* lizards. *ArXiv e-prints*, 2013.

A. W. F. Edwards. Estimation of the branch points of a branching diffusion process. *J. Roy. Stat. Soc. B*, 32(2):155–174, 1970.

N. Eldredge and S. J. Gould. Punctuated equilibria: an alternative to phyletic gradualism. In T. J. M. Schopf and J. M. Thomas, editors, *Models in Paleobiology*, pages 82–115. Freeman Cooper, San Francisco, 1972.

W. Feller. *An Introduction to Probability Theory and Its Applications Vol. II*. John Wiley & Sons, New York, 1971.

J. Felsenstein. Phylogenies and the comparative method. *Am. Nat.*, 125 (1):1–15, 1985.

J. Felsenstein. Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst*, 19:445–471, 1988.

J. Felsenstein. *Inferring Phylogenies*. Sinauer, 2004.

R. P. Freckleton. Fast likelihood calculations for comparative analyses. *Meth. Ecol. Evol.*, 3:940–947, 2012.

W. A. Fuller. *Measurement Error Models*. Wiley, 1987.

T. Gernhard. The conditioned reconstructed process. *J. Theor. Biol.*, 253:769–778, 2008a.

T. Gernhard. New analytic results for speciation times in neutral models. *B. Math. Biol.*, 70:1082–1097, 2008b.

L. J. Gleser. The importance of assessing measurement reliability in multivariate regression. *J. Am. Stat. Assoc.*, 87(419):696–707, 1992.

S. J. Gould and N. Eldredge. Punctuated equilibrium comes of age. *Nature*, 366:223–227, 1993.

P. Haccou, P. Jagers, and V. A. Vatutin. *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press, Cambridge, 2007.

T. F. Hansen. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, 51(5):1341–1351, 1997.

T. F. Hansen and K. Bartoszek. Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Syst. Biol.*, 61(3):413–425, 2012.

T. F. Hansen and E. P. Martins. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, 50(4):1404–1417, 1996.

T. F. Hansen and S. H. Orzack. Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. *Evolution*, 59(10):2063–2072, 2005.

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

T. E. Harris. *The Theory of Branching Processes*. Springer, 1963.

K. Hartmann, D. Wong, and T. Stadler. Sampling trees from evolutionary models. *Syst. Biol.*, 59(4):465–476, 2010.

P. H. Harvey and M. Pagel. *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford, 1991.

L. S. T. Ho and C. Ané. Asymptotic theory with hierarchical autocorrelation: Ornstein–Uhlenbeck tree models. *Ann. Stat.*, 41(2):957–981, 2013.

S. Höhna, T. Stadler, F. Ronquist, and T. Britton. Inferring speciation and extinction rates under different sampling schemes. *Mol. Biol. Evol.*, 28(9):2577–2589, 2011.

D. H. Huson, R. Rupp, and C. Scornavacca. *Phylogenetic Networks: Concepts Algorithms and Applications*. Cambridge University Press, Cambridge, 2010.

S. M. Iacus. *Simulation and Inference for Stochastic Differential Equations*. Springer, 2008.

T. Ingram and D. L. Mahler. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein–Uhlenbeck models with stepwise Akaike Information Criterion. *Meth. Ecol. Evol.*, 4(5): 416–425, 2013.

G. Jones, S. Sagitov, and B. Oxelman. Statistical inference of allopolyploid species networks in the presence of incomplete lineage sorting. *Syst. Biol.*, 62(3):467–478, 2013.

M. Kimmel and D. E. Axelrod. *Branching Processes in Biology*. Springer, Cambridge, 2002.

F. C. Klebaner. *Introduction to Stochastic Calculus with Applications*. Imperial College Press, London, 2007.

A. Labra, J. Pienaar, and T. F. Hansen. Evolution of thermal physiology in *Liolaemus* lizards: Adaptation, phylogenetic inertia, and niche tracking. *Am. Nat.*, 174(2):204–220, 2009.

A. Lambert. The contour of splitting trees is a Lévy process. *Ann. Appl. Probab.*, 38(1):348–395, 2010.

A. Lambert, H. K. Alexander, and T. Stadler. Phylogenetic analysis accounting for age-dependent death and sampling with applications to epidemics. *ArXiv e-prints*, 2013a.

A. Lambert, H. Morlon, and R. S. Etienne. The reconstructed tree in the lineage-based model of protracted speciation. *ArXiv e-prints*, 2013b.

M. J. Landis, J. G. Schraiber, and M. Liang. Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. *Syst. Biol.*, 62(2):193–204, 2013.

P. Lemey, M. Salemi, and A. Vandamme, editors. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press, Cambridge, 2009.

E. P. Martins and T. F. Hansen. Phylogenis and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.*, 149(4):646–667, 1997.

P. Medvegyev. *Stochastic Integration Theory*. Oxford University Press, Oxford, 2007.

A. Mir, F. Rosselló, and L. Rotger. A new balance index for phylogenetic trees. *Math. Biosci.*, 241(1):125–136, 2013.

A. Mooers, O. Gascuel, T. Stadler, H. Li, and M. Steel. Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. *Syst. Biol.*, 61(2):195–203, 2012.

A. Ø. Mooers and D. Schluter. Fitting macroevolutionary models to phylogenies: an example using vertebrate body sizes. *Contrib. Zool.*, 68:3–18, 1998.

A. Ø. Mooers, S. M. Vamosi, and D. Schluter. Using phylogenies to test macroevolutionary hypotheses of trait evolution in Cranes (Gruinae). *Am. Nat.*, 154:249–259, 1999.

W. Mulder. Probability distributions of ancestries and genealogical distances on stochastically generated rooted binary trees. *J. Theor. Biol.*, 280:139–145, 2011.

M. Pagel. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proc. Roy. Soc. B*, 255:37–45, 1994.

M. Pagel. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.*, 48 (3):612–622, 1999.

E. Paradis. *Analysis of Phylogenetics and Evolution with R.* Springer, 2006.

F. Plard, C. Bonenfant, and J. M. Gaillard. Revisiting the allometry of antlers among deer species: male–male sexual competition as a driver. *Oikos*, 120:601–606, 2011.

L. Popovic. Asymptotic genealogy of a critical branching process. *Ann. Appl. Probab.*, 14(4):2120–2148, 2004.

M. Prager and S. Andersson. Phylogeny and evolution of sexually selected tail ornamentation in widowbirds and bishops (*Euplectes* spp.). *J. Evol. Biol.*, 22:2068–2076, 2009.

M. Prager, E. I. A. Johansson, and S. Andersson. A molecular phylogeny of the African widowbirds and bishops, *Euplectes* spp. (Aves: Passeridae: Ploceinae). *Mol. Phylogenet. Evol.*, 46:290–302, 2008.

O. G. Pybus and P. H. Harvey. Testing macro–evolutionary models using incomplete phylogenies. *Proc. R. Soc. Lond. B*, 267:2267–2272, 2000.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org`.

S. Sagitov and K. Bartoszek. Interspecies correlation for neutrally evolving traits. *J. Theor. Biol.*, 309:11–19, 2012.

G. J. Slater, L. J. Harmon, and M. E. Alfaro. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution*, 66(12):3931–3944, 2012.

T. Stadler. Lineages–through–time plots of neutral models for speciation. *Math. Biosci.*, 216:163–171, 2008.

T. Stadler. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *J. Theor. Biol.*, 261(1): 58–68, 2009.

T. Stadler. Mammalian phylogeny reveals recent diversification rate shifts. *PNAS*, 108(15):6187–6192, 2011a.

T. Stadler. Simulating trees with a fixed number of extant species. *Syst. Biol.*, 60(5):676–684, 2011b.

T. Stadler and M. Steel. Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. *J. Theor. Biol.*, 297: 33–40, 2012.

M. Steel and A. McKenzie. Properties of phylogenetic trees generated by Yule–type speciation models. *Math. Biosci.*, 170:91–112, 2001.

B. Øxendal. *Stochastic Differential Equations*. Springer, 2007.