

Background

Jason Pienaar, Bjørn Tore Kopperud, Kjetil Lysne Voje, Thomas F. Hansen (in no particular order)

2018-10-02

Contents

1	Comparative methods for adaptive hypotheses	2
1.1	Background	2
1.2	Phylogeny	2
2	Comparative methods built around an Ornstein-Uhlenbeck process	2
2.1	The optimum	3
2.2	Ornstein-Uhlenbeck process	3
2.3	Parameterization	3
2.4	Random continuous predictor	4
2.5	Categorical fixed predictors (niches, regimes)	4
2.6	Direct-effect predictors	4
2.7	Example: expectation with random covariates	4
2.8	Residual covariances	5
2.9	Intercept	5
3	Brownian-motion models	6
3.1	Regime-dependent trends	6
3.2	Continuous covariates	7
4	Parameter estimation	7
4.1	Model formulation	7
4.2	Least-squares	8
4.3	Maximum likelihood	8
5	Model selection	9
6	Estimation error	9
6.1	Measurement variance	9
7	Caveats and limitations	10
8	History & comparison with other packages	10
9	References	11

1 Comparative methods for adaptive hypotheses

1.1 Background

Comparing traits in groups of related species in different environments or time periods can provide powerful tests of adaptive hypotheses (e.g. Sober 2008; Hansen 2014). The distributions of trait values of extant species at the tips of a phylogeny (and fossil data or reconstructed ancestral states within the phylogeny if available) however, may reflect, adaptation, phylogenetic inertia, both of these, or none of these. We use the term phylogenetic inertia in the sense of resistance to adaptive change i.e. it comes about when related species inherit an inert trait from a common ancestor, which then lags adaptive change to new optima. Hence, if inertia is present, observed trait values may reflect both adaptive change as well as an influence of a maladapted ancestral trait. When testing an adaptive hypothesis with such data, the constraints provided by such resistance to change for current trait values as they evolve towards their optima need to be controlled for (and similarly, the effects of adaptation need to be controlled for when testing hypothesis of phylogenetic inertia).

Most contemporary comparative methodologies do not control for inertia in the sense defined above, but rather control for general effects of phylogeny. Their principal concern is to deal with statistical issues that arise from non-independent, correlated data that naturally arise from evolution along a branching phylogeny, and are usually used to remove these correlations before the actual analysis takes place. Phylogenetic effects, however, will also be present when traits adapt to variables that are themselves phylogenetically structured. Effects of adaptation should not be controlled for when studying adaptation and they need to be separated from inertia. Furthermore, most contemporary methods assume that traits evolve as a Brownian motion, a crucial assumption and one that provides the analytical techniques to remove the correlations (e.g. independent contrasts). Numerous studies however, have pointed out that comparative method that assumes a Brownian motion as the underlying evolutionary process are not suitable for studying adaptation towards optima (e.g. Hansen 2014). A Brownian motion is a stochastic process in which the expected mean value equals the common ancestral state at the base of the species phylogeny. There is no mechanism in the process that allows one to specify adaptation to or maintenance at specific optima. Even if the ancestral state is at the optimum and the optimum does not change through time, any deviations from the optimum relationship generated by the stochastic Brownian-motion process will be inherited and there is no mechanism that allows a trait in a species to return to its optimum value.

Hansen (1997), Butler & King (2004), Hansen & Orzack (2005) and Hansen *et al.* (2008) have argued that comparative methods based on an Ornstein-Uhlenbeck (OU) process are more suited to studying adaptive hypotheses using comparative data. The OU process, unlike a pure Brownian motion, consists of a deterministic component (which can be used to model adaptation of a trait towards an optimum) as well as a stochastic component (which models the random fluctuations of the trait as it evolves towards the optimum. If we take away the deterministic component of an OU process, we are left with a Brownian motion.

1.2 Phylogeny

These methods require a rooted phylogeny with branch lengths. Both ultrametric and non-ultrametric trees can be used, but this will have implications for reliable estimation of ancestral states. Polytomies and non-branching edges are not a problem but each branch must have a length. The units of the lengths are arbitrary and can be given in time units or number of nucleotide substitutions. A lot of the power of this method comes from having accurate branch lengths so some care should be taken in time-calibrating the phylogeny.

2 Comparative methods built around an Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck process describes stochastic evolution with a deterministic tendency to move towards a fixed state (Hansen & Martins 1996). To model adaptive evolution in a comparative data

set, we interpret this fixed state as a “primary optimum”, defined as the average fitness optimum that would be reached by a large number of independent species evolving for a sufficiently long time in a given niche to be free of any ancestral influence (Hansen 1997). The model assumes that primary optima exist at any point in time (on the phylogeny), and the idea is to test hypotheses about the effect of specified environmental variables on the primary optima. Currently, the software allows the primary optimum to be modeled as a multiple regression on several continuous “random” variables, continuous “direct-effect” covariates, or as a one-way ANOVA on fixed, categorical variables mapped as regimes on the phylogenetic tree.

2.1 The optimum

If we let $\theta(x_1, x_2 \dots)$, be the primary optimum as a function of environmental variables, $x_1, x_2 \dots$, then for the regression model, we have:

$$\theta(x_1, x_2 \dots) = b_0 + b_1x_1 + b_2x_2 \dots,$$

where b_0 is an intercept, and $b_1, b_2 \dots$ are regression slopes on the environmental variables. For the ANOVA model we let $x_1, x_2 \dots$ be indicator variables indicating different categorical states of the environment, and we write

$$\theta(x_1, x_2 \dots) = \theta_1x_1 + \theta_2x_2 \dots,$$

where $\theta_1x_1, \theta_2x_2 \dots$ are parameters describing the state of the primary optimum in the different environments or niches. The goal is to estimate the parameters b_i or θ_i to test whether particular aspects of the environment have an effect on the optimum or not. Our adaptive hypotheses thus takes the form: Given that there is a single optimum, the species are adapted to variable x if x has an (important) effect on the optimum. As we will see, the methods also allow for assessments of how strongly the species tend to evolve towards the optimum, and thus to reject the base assumption of adaptation altogether.

2.2 Ornstein-Uhlenbeck process

In both the fixed categorical and continuous random predictor cases, the evolution of the response trait, y , can be represented by the stochastic differential Ornstein-Uhlenbeck process equation:

$$dy = -\alpha(y - \theta)dt + \sigma_y dW_y,$$

which is interpreted as follows: dy is the change in y over a time step dt , α is a parameter measuring the rate of adaptation towards the optimum, dW_y is a white-noise process having independent, normally-distributed random changes with mean zero and unit variance. The standard deviation of the random changes is given by σ_y and θ are the fixed or random optima as defined above. Categorical predictor variables must be mapped on the phylogeny a priori, and the method treats these as fixed.

2.3 Parameterization

We note that we will work with the transformed parameters:

$$t_{1/2} = \frac{\log(2)}{\alpha}$$

$$v_y = \frac{\sigma_y^2}{2\alpha},$$

These are both easier to interpret than α and σ_y . The phylogenetic half life, $t_{1/2}$, is the time it takes for the expected trait value to move half the distance from the ancestral state to the primary optimum. The half life provides a useful metric to assess the strength of phylogenetic inertia as it is on the same linear scale as the phylogenetic branch lengths. If the half life is short relative to the total length of the

phylogeny, it means adaptation to the primary optimum is rapid in expectation, and if the half life is long, it suggests the presence of inertia and that ancestral trait values influence the species' observed trait values at the tips. Consequently, if the half life is large, we would expect the species traits to be poorly adapted to the primary niches. A half life of infinity corresponds to evolution governed by a Brownian-motion process in which there is no tendency to move towards the optimum, and a half life of zero corresponds to immediate adaptation. The parameter v_y is the variance expected when the adaptive and stochastic forces are in stochastic equilibrium.

2.4 Random continuous predictor

For the random predictors we have an additional stochastic differential Brownian-motion process to describe their evolution on a phylogeny:

$$dx = \sigma_x dW_y,$$

where σ_x gives the standard deviation of the random changes (hence, the rate at which the predictors change). These dynamical equations can be solved for the different assumptions about the optimum.

2.5 Categorical fixed predictors (niches, regimes)

With fixed, categorical predictor variables, the model predicts that the species data will be of the form:

$$y_i = c_{0i}y_a + c_{1i}\theta_1 + c_{2i}\theta_2 \cdots + r_i,$$

where y_a is a parameter describing the state of the trait at the root of the tree, the coefficients, c_{ji} , are determined by the time on the phylogeny the species i has spent in environment j , such that more recent environments are weighted more than more ancient environments. The weights are determined by the rate of adaptation, α . More specifically, c_{ji} is a sum over all time intervals in the past history of species i where this species experienced environment j , and each interval contributes a term $e^{-\alpha t} - e^{-\alpha s}$, where, starting at the root with $t_0 = 0$, s is the time when the species enters the environment, and t is the end-time of the environment (see Hansen (1997) for formal derivation). The coefficient for the ancestral state y_a is $c_0 = e^{-\alpha t}$, where t is the time from the root to species i . Note that if α is large, then only the most recent environment will contribute to the trait values of the species. Ancient environments only contribute to species' trait values if the rate of adaptation is low. SLOUCH will automatically compute these coefficients given a mapping of niches on the tree. The residual terms, r_i , are assumed to be normally distributed with mean zero and variances and covariances given later.

2.6 Direct-effect predictors

For some biological systems (e.g. allometry) it may be more realistic to model the influence of x on y as direct, non-lagged process. In this case, we can expand the Ornstein-Uhlenbeck differential equation:

$$dy = -\alpha(y - \theta)dt + bdx + \sigma_y dW_y,$$

where changes in the predictor (dx) has an immediate effect on changes in the response trait (dy). The direct-effect model makes no assumptions of any phylogenetic process that gives rise to x .

2.7 Example: expectation with random covariates

For random, continuous predictor variables, evolving as if by a Brownian motion, the model predicts:

$$y_i = k + \rho(\alpha t_i)b_1x_{1i} + \rho(\alpha t_i)b_2x_{2i} \cdots + r_i,$$

where x_{ij} is the current state of predictor variable j for species i , and r_i is the residual term. The intercept (k) discussed below. The phylogenetic correction factor (ρ) is another measure of the strength of inertia:

$$\rho(\alpha t_i) = 1 - \frac{1 - e^{-\alpha t_i}}{\alpha t_i},$$

where t_i is the time from root to species i . The correction factor varies between 0 and 1, and is identical for all species for ultrametric trees. While the optimal regression represents the relationship between predictor and the trait that we expect if adaptation is immediate, the evolutionary regression is the generalized least squares regression of a trait on the predictor, and is influenced by both adaptation and inertia. The optimal regression is accordingly an estimate of the optimal relationship between predictor and the trait totally free of constraints, while evolutionary regression is the ‘observed’ relationship. The magnitude of the difference between the optimal and the evolutionary regression (ρ) is therefore informative of the relative importance of inertia and adaptation between the predictor and the response trait. The evolutionary and optimal regression will be identical if there is no inertia ($\rho(\alpha t_i) = 1$). The optimal and evolutionary regressions are not alternative models, but rather two perspectives or aspects of the same model. In the random environment model, we also need to estimate the rate of evolution of the environmental variable, σ_x .

The intercept-only model is also useful for testing one of the assumptions of the OU-BM model; that the predictor variables evolve as a Brownian motion. The stationary variance ($v_y = \sigma_y^2/2\alpha$) of the joint OU-BM process is a measure of the relative influence of stochastic factors in the adaptive process relative to the primary adaptive force. Secondary stochasticity is generated by a combination of stochastic influences (such as changes in unmeasured selective forces, genetic drift, etc) on the response variable as it evolves towards the optimum or optima.

2.8 Residual covariances

The residual covariances of the fixed-effect model were derived in Hansen (1997):

$$\text{Cov}[r_i, r_j] = v_y e^{-\alpha t_{ij}} (1 - e^{-2\alpha t_a}),$$

where t_i is the time from root to species i , t_a is the time from root to the most recent common ancestor of species i and j , and t_{ij} is the time separating species i from species j (i.e. the sum of the distances from i and j to their most recent common ancestor). For the random-effect model, Hansen *et al.* (2008) derived the residual covariances as:

$$\text{Cov}[r_i, r_j] = \left(\frac{\sigma_\theta^2 + \sigma_y^2}{2\alpha} \right) (1 - e^{-2\alpha t_a}) e^{-\alpha t_{ij}} + \sigma_\theta^2 \left(t_a \frac{1 - e^{-\alpha t_i}}{\alpha t_i} \frac{1 - e^{-\alpha t_j}}{\alpha t_j} - \frac{1 - e^{-\alpha t_a}}{\alpha} \left(e^{-\alpha t_{ia}} \frac{1 - e^{-\alpha t_j}}{\alpha t_j} + e^{-\alpha t_{ja}} \frac{1 - e^{-\alpha t_i}}{\alpha t_i} \right) \right)$$

The parameter σ_θ^2 in the above equation equals $b_1^2 \sigma_{x_1}^2 + b_2^2 \sigma_{x_2}^2 + \dots$, assuming the predictor variables are not correlated. The instantaneous variances, $\sigma_{x_i}^2$, of the predictor variables are estimated in advance. Note that this is considerably more complex than with fixed effects, and that the residual covariances also depend on the regression parameters b_i .

2.9 Intercept

The intercept, k , in the random-effect model is given as:

$$k = e^{-\alpha t} y_a + (1 - e^{-\alpha t}) b_0 + (1 - e^{-\alpha t} - \rho(\alpha t)) (b_1 x_{a1} + b_2 x_{a2} + \dots),$$

where y_a and x_{aj} are the ancestral states of the trait, y , and the predictor variables, x_j , and b_0 is the true intercept of the optimal regression. As t is the time from root to the species, we also see that the intercept on a non-ultrametric tree will differ from species to species. Clearly, k is not a parameter of biological

interest. What we would like is to get estimates of its components, and in particular of the “optimal” intercept, b_0 . For species on a non-ultrametric phylogeny, it is technically possible to get independent estimates of b_0, y_a and the x_a ’s, but unless there is strong phylogenetic inertia and the phylogeny is strongly non-ultrametric, these would be poorly resolved, and the individual estimates will be inaccurate. We generally do not recommend separate estimates of these parameters, and for an ultrametric phylogeny, we can only obtain an estimate of the composite k . An estimate of the optimal intercept may, however, be obtained indirectly with an additional assumption. First, note that SLOUCH returns independent estimates of the ancestral states x_a based on the assumption that the predictors evolve as if by a Brownian motion. The problem is then only to get an estimate of the ancestral state of the trait, y_a . One reasonable way to do this is to assume that the common ancestor was optimal, and use $y_a = b_0 + b_1x_{a1} + b_2x_{a2} \dots$. From this we can derive an estimate of the optimal intercept as:

$$b_0 = k + (\rho(\alpha t) - 1)(b_1x_{a1} + b_2x_{a2} \dots),$$

where k, b ’s, x_a ’s and $\rho(\alpha t)$ are all output from the SLOUCH analysis. This may be useful for plotting of the optimal regression, etc. The error due to the ancestral-optimality assumption is likely to be small unless the phylogenetic half-life is large.

3 Brownian-motion models

The Ornstein-Uhlenbeck process is a generalization of Brownian motion. When $\alpha = 0$, the model collapses into a Brownian motion with no pull towards the optimum. Brownian motion is not distinguishable from an Ornstein-Uhlenbeck process where the phylogenetic half-life is much larger than the tree length ($t_{1/2} \gg t$). SLOUCH can also explicitly fit Brownian Motion models of trait evolution. The stochastic differential equation is

$$dy = \sigma_y dW_y,$$

where $dW_y \sim N(0, dt)$. For a zero-trend model, we have:

$$y_i = y_a + r_i,$$

where y_a is the ancestral state (or “phylogenetic mean”). The residual covariances are

$$\text{Cov}[r_i, r_j] = \sigma_y^2 t_a,$$

where t_a is the sum of shared branch lengths for species i and j . With this model, we only estimate one parameter with likelihood: σ_y^2 .

3.1 Regime-dependent trends

Keep in mind that, since there is no directional or stabilizing trend in a standard Brownian motion, the intercept is the same as the ancestral state y_a . Unlike the OU process, there are no optima (θ), and the rate of adaptation (α) is not a part of the model. We can, however, expand the model to include regime-dependent trends:

$$dy = \tau dt + \sigma_y dW_y$$

where τ is one or more trends. The trends (τ) in this model are equivalent to $\tau = \lim_{\alpha \rightarrow 0}(\theta\alpha)$ under an OU model. The model predicts:

$$y_i = y_a + \sum_s \tau_s t_s + r_i,$$

where t_s is the total time that species i has spent in regime s . With ultrametric trees, y_a can not be estimated independently of the trends τ , and as such y_a is not estimated with default settings. In other words, we assume that $y_a = 0$, and hence the trends τ_s can only be interpreted relative to each other. If

we do estimate y_a independently, however, the trends τ_s will represent the absolute expected direction of change. The residuals covariances are equal to the zero-trend model.

3.2 Continuous covariates

Next, we can expand the model to include continuous covariates. If we model the covariates to influence the response trait directly and immediately, the model implementation is similar to the previously described direct-effect model. Another option is to model the trend as a linear function of a random variable evolving as a Brownian motion ($\tau = a + bx$).

$$\begin{aligned} dy &= (a + bx)dt + \sigma_y dW_y \\ dx &= \sigma_x dW_x \end{aligned}$$

where a are the regime-dependent trends as previously denoted τ , and x is a random variable evolving as a Brownian motion. In the case of multiple independent predictors, we have

$$dy = (a + \sum_k b_j x_j)dt + \sigma_y dW_y,$$

where $\sum_j b_j x_j$ is also a Brownian motion with variances $\sigma_\theta^2 = \sum_j b_j^2 \sigma_j^2$. For each species, the model predicts:

$$y = y_a + \sum_s a_s t_s + \sum_j \rho(t) b_j x_j + r,$$

where y_a is the ancestral state, a_s is the trend in regime s , and t_s is the sum of time spent in regime s . x_j is the current (extant) state of predictor variable j , and $\rho(t)$ is the phylogenetic correction factor:

$$\rho(t) = \frac{t}{2},$$

where t is the time from the species to the root of the tree. Note that unlike previously, where ρ was unitless and in the interval $[0,1]$, the ρ for the Brownian motion continuous-trend model is positive and in units of time. The residual covariances for this model are

$$\text{Cov}[r_i, r_j] = \sigma_y^2 t_a + \sigma_\theta^2 t_a (t_a^2/12 + t_{ia} t_{ja}/4),$$

where the time from the most recent common ancestor (MRCA) of species i, j to species i is t_{ia} , and the time from the MRCA to species j is t_{ja} .

Comment: The brownian-motion section is a little thin. We can't just cite thomas' & krzysztof's notes for the interpretation and derivation, can we? Does this kind of material even fit in a manual, if not first introduced elsewhere?

4 Parameter estimation

4.1 Model formulation

First, we set up the model. As an example, we model log neocortex size (mm^2) as having three distinct optima (browsers, grazers, mixed feeders) and adapting in response to log brain size (g), where brain size evolves as if by a Brownian motion. For each species i , this is

$$y_i = c_{0i} y_a + c_{1i} \theta_{\text{browser}} + c_{2i} \theta_{\text{grazer}} + c_{3i} \theta_{\text{mixed feeder}} + b \rho(t_i) x_i + r_i.$$

For the moment, let's assume that y_a is equal to the optimum present at the root of the tree ($\theta_{\text{mixed feeder}}$). In matrix form, the equivalent is

$$\mathbf{Y} = \begin{bmatrix} c_{11} & c_{21} & c_{31} + c_{01} & \rho(t_1)x_1 \\ c_{12} & c_{22} & c_{32} + c_{02} & \rho(t_2)x_2 \\ \vdots & \vdots & \vdots & \vdots \\ c_{1n} & c_{2n} & c_{3n} + c_{0n} & \rho(t_n)x_n \end{bmatrix} \begin{bmatrix} \theta_{\text{browser}} \\ \theta_{\text{grazer}} \\ \theta_{\text{mixed feeder}} \\ b \end{bmatrix} + \mathbf{r},$$

where \mathbf{Y} is a vector of observed log neocortex size, $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$ is a vector of residuals, and n is the number of species. Note that the coefficients c are themselves functions of the rate of adaptation (α), the regime topology and the tree branch lengths. In the next section we will use the equivalent compact version:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{r},$$

where \mathbf{X} is the model matrix, and β is a vector of linear model parameters.

4.2 Least-squares

In SLOUCH, the regression parameters are estimated using the least-squares method. Conditional on the other model parameters (e.g. $t_{1/2}$, $\sigma_y^2/2\alpha$ and $\sigma_{x_j}^2$), we start with an ordinary least-squares estimate of the regression parameters:

$$\hat{\beta}_0 = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y}).$$

We require an initial estimate $\hat{\beta}_0$ since the regression coefficients enter the residual variance-covariance matrix \mathbf{V} when measurement error is included in the model. We use $\hat{\beta}_0$ and \mathbf{V} to obtain a generalized least-squares (GLS) estimate of the regression parameters as:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}).$$

We update \mathbf{V} and iterate the GLS estimation until convergence is reached. For models with random predictor variables, this estimate yields the optimal or primary regression, as the interpretation depends on the phylogenetic half-life. In calculating the evolutionary regression, $\hat{\beta}_{\text{evolutionary}}$, we reconstruct the model matrix \mathbf{X} by setting $\rho = 1$, and apply same above formula. The variances and covariances of the linear model parameter are estimated as:

$$\text{Var}[\hat{\beta}] = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1},$$

where the square root of the diagonal elements are the standard errors.

4.3 Maximum likelihood

We use the GLS estimates to we evaluate the log-likelihood (= support) function:

$$\log \mathcal{L}(t_{1/2}, \sigma_y^2/2\alpha) = -\frac{1}{2} (n \log(2\pi) + \log |\mathbf{V}| + (\mathbf{y} - \mathbf{X}\hat{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta})).$$

The regression parameters and likelihood are estimated in equal fashion for all types of models. SLOUCH includes two options for estimating $t_{1/2}$ and $\sigma_y^2/2\alpha$ using likelihood. A numerical optimization routine (L-BFGS-B, Byrd *et al.* 1995) is used as default, but a grid search is also implemented.

5 Model selection

SLOUCH output provides the user with various statistics in order to compare models with different combinations of predictor variables. These are the coefficient of determination (R^2), Akaike Information Criterion (AIC), the small-sample corrected AIC (AICc) and the stricter Schwarz Information Criterion (SIC) (Burnham & Anderson, 1998). For continuous predictor models, these criteria can be used to choose the number of informative predictors in a multiple regression, or to decide which is the best predictor of a given trait's evolution. These information criteria can also be used to decide whether a single or multiple optima Ornstein-Uhlenbeck process or a Brownian-Motion process best describes a trait's evolution. For models with fixed categorical predictors, these information criteria are useful for assessing which categories influence a trait's evolution (Butler & King 2004). In this way the user can either pool different categories or further subdivide a given category into biological meaningful categories in order to test various hypotheses concerning differential selective niches. We note however, that the information criteria will sometimes pick models with strange parameter estimates like optima or ancestral states that are way out of biological bounds, or stationary variances of zero. Such models should be regarded with skepticism, and one can use grid search to look at the support surfaces to see if there are other peaks in the likelihood surface with more reasonable estimates. Note that all of the model fit statistics, as well as the parameter search itself, is conditional on the *naive* GLS estimator, i.e. prior to any bias-correction routines.

6 Estimation error

The linear model coefficients are reported with standard errors, which can be used to construct confidence intervals. These are, however, conditional on the maximum-likelihood estimates of $t_{1/2}$ and v_y .

The estimation errors in $t_{1/2}$ and v_y are more difficult to estimate, however, as we use numerical techniques to optimize the likelihood. With numerical optimization, the default setting does not provide estimation error. An experimental option (`sloUCH.fit(...hessian=TRUE)`) uses numerical sampling to approximate the Hessian matrix at the likelihood peak, which can be used to report standard errors of $t_{1/2}$ and v_y . We note that this method can be numerically unstable, particularly with long half-lives, and may even crash the program for small half-lives.

A good way to inspect estimation error is to perform a grid search, which can be seen in the examples vignette. A grid search starts by manually providing vectors of potential values for each parameter to the program to find the combination that maximizes the likelihood. The best estimates of $t_{1/2}$ and v_y are a joint estimate and for this reason the support region (the three-dimensional log-likelihood plot) is given as a measure of uncertainty for $t_{1/2}$ and v_y jointly. The marginal lowest and highest values of $t_{1/2}$ within the support region, conditional on the best estimate of v_y , can also be obtained using the figure and the support-region tables in the output (and vice versa for v_y). The `sloUCH.fit` function can also take scalars rather than vectors, which may be useful for finding the marginal support region of $t_{1/2}$ conditional on v_y (and vice versa).

6.1 Measurement variance

Hansen & Bartozsek (2012) discuss how to incorporate measurement error into a phylogenetic comparative analysis. Most often, a comparative analysis involves using species means as input values. In this case the main component of measurement error is estimation error in the mean (or whatever other statistics are used) resulting from sampling a finite number of individuals. If each mean species value to be used in the regression was obtained from a large number of individuals (> 20 -30 or so) the square of the standard error (i.e. the estimation variance) can be entered into the measurement error column for that variable. If no measurement error is to be included, then simply add a column of zeros. We also refer the reader to the examples vignette for a more detailed example. SLOUCH deals with measurement error in the response variable by adding measurement variance to the diagonals of the residual variance matrix. Measurement variance in the random predictor variable(s) is incorporated by multiplying it by the square of the uncorrected slope parameter and then adding it to the diagonals of the residual variance matrix. If the observational errors in the predictor variables are non-zero, the generalized least squares estimator

may be biased. A bias correction for the regression coefficients is implemented according to Hansen & Bartoszek (2012).

7 Caveats and limitations

As with any statistical model, the quality of the parameter estimates is dependent on the quality of the data and certain assumptions concerning the residuals must be met. SLOUCH utilizes both measured trait values and the phylogenetic topology and branch lengths as data. With regards to the trait values, we assume that a linear relationship exists between expected optimal trait values and the predictor variables. Although SLOUCH can perform non-linear regression by entering quadratic terms in the model, this will most likely violate the assumption that the predictors evolve as a Brownian Motion. SLOUCH also assumes that the underlying evolutionary process is equal in all parts of the tree, as $t_{1/2}$ and σ_y^2 are constant. When studying a clade that is heterogeneous (e.g. Cetartiodactyla) or spans large timespans this assumption may be violated, and one may consider splitting the phylogeny or using methods that relax this assumption (e.g. Beaulieu *et al.* 2014). It should also be kept in mind that any inferences drawn about the rates of adaptation, phylogenetic inertia and its potential effects on trait values are conditional on the phylogeny used. SLOUCH does not include routines for evaluating the effects of phylogeny uncertainty in a systematic manner, but one can perform the analysis on plausible alternative phylogenetic hypotheses to examine the effects of phylogenetic uncertainty on parameter estimates. Simulation may also be useful here for examining potential interactions between true parameter values and alternative phylogenies.

8 History & comparison with other packages

This document describes an R-package for estimating parameters for a suite of comparative methods based on the Ornstein-Uhlenbeck process as a model of adaptive evolution. We refer to this program as SLOUCH 2.0 since it builds on the models in SLOUCH 1.0 (Hansen *et al.* 2008). SLOUCH 1.0 dealt with a continuous trait that evolves as an Ornstein-Uhlenbeck process around an optimal state that itself depend on an environmental variable that evolves as if by a Brownian motion. For the current version we have extended the model to deal with an optimal state that depends on multiple randomly evolving predictors and multiple direct effect predictors. We also describe in more detail how to include measurement error in the analyses. SLOUCH 2.0 also includes functionality designed to deal with traits that evolve in response to fixed categorical niches that can be mapped onto a phylogeny and then used to estimate niche optima for the response trait (Hansen, 1997; Butler & King, 2004). The multi-optima model is thus an alternative to Butler & King's (2004) OUCH program, but it differs in a few crucial aspects, as described below. SLOUCH 1.3 and earlier versions were built as an addition to OUCH, and data formats and some functions were borrowed from the original version of OUCH, either as is, or slightly modified. In SLOUCH 2.0 the phylogenetic tree format is revamped, and now uses the format used by the R-package APE (Paradis *et al.* 2004).

The `fixed.fit` function of SLOUCH 1.1 differs from the OUCH program as follows: a) it allows the user to estimate parameters using a non-ultrametric tree, such as those that include extinct species, b) it allows the user to incorporate measurement error in the response variable, c) it automatically calculates standard errors for estimated optima, goodness of fit as R^2 and can be used to compute support regions for rates of adaptation and trait variances, d) it includes functionality to use both grid search and the built-in R optimization routines to estimate parameters, and e) it includes various options to deal with intercept estimation, which can be problematic. [Do we need to compare with OUCH?]

SLOUCH 2.0 also includes Brownian-motion models where the optima are instead modeled as trends. This document is written to serve as an updated user manual for SLOUCH. It provides a summary of the rationale for the methods, and illustrates them with examples based on the ruminant neocortex data set. It extends the user manual for SLOUCH 1.0 by discussing multiple regression and ANOVA type of models for the optimum, and gives more details on modeling non-ultrametric trees and measurement error. The mathematical and conceptual basis of the adaptation-inertia methods are presented and discussed in Hansen (1997), Butler & King (2004), and Hansen *et al.* (2008). A user of SLOUCH should also consult these papers.

9 References

- Burnham, K. P. & Anderson, D. R. (1998). *Model selection and inference: A practical information-theoretic approach*. Springer.
- Butler, M. A., & King, A. A. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist*, 164(6), 683–695. <https://doi.org/10.1086/426002>
- Edwards, A. W. F. (1992). *Likelihood*. expanded edition Johns Hopkins University Press. Baltimore, MD.
- Escudero, M., Hipp, A. L., Hansen, T. F., Voje, K. L., & Luceño, M. (2012). Selection and inertia in the evolution of holocentric chromosomes in sedges (*Carex*, Cyperaceae). *New Phytologist*, 195(1), 237–247. <https://doi.org/10.1111/j.1469-8137.2012.04137.x>
- Grabowski, M., Voje, K. L., & Hansen, T. F. (2016). Evolutionary modeling and correcting for observation error support a 3/5 brain-body allometry for primates. *Journal of human evolution*, 94, 106–116.
- Hansen, T. F. (1997). Stabilizing Selection and the Comparative Analysis of Adaptation. *Evolution*, 51(5), 1341. <https://doi.org/10.2307/2411186>
- Hansen, T. F., & Martins, E. P. (1996). Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, 50(4), 1404–1417.
- Hansen, T. F., & Orzack, S. H. (2005). Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. *Evolution*, 59(10), 2063–2072.
- Hansen, T. F., Pienaar, J., & Orzack, S. H. (2008). A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, 62(8), 1965–1977. <https://doi.org/10.1111/j.1558-5646.2008.00412.x>
- Hansen, T. F., & Bartoszek, K. (2012). Interpreting the evolutionary regression: The interplay between observational and biological errors in phylogenetic comparative studies. *Systematic Biology*, 61(3), 413–425. <https://doi.org/10.1093/sysbio/syr122>
- Hansen, T. F. (2014). Use and misuse of comparative methods in the study of adaptation. In Garamszegi & L. Zsolt (Eds.), *Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology* (pp. 351–379). Springer. https://doi.org/10.1007/978-3-662-43550-2_14
- Harvey, P. H., & Pagel, M. D. (1991). *The comparative method in evolutionary biology* (Vol. 239). Oxford: Oxford university press.
- Martins, E. P., & Hansen, T. F. (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist*, 149(4), 646–667.
- O’Meara, B. C., & Beaulieu, J. M. (2014). Modelling stabilizing selection: The attraction of Ornstein-Uhlenbeck models. In Garamszegi & L. Zsolt (Eds.), *Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology* (pp. 381–393). Springer. https://doi.org/10.1007/978-3-662-43550-2_15
- Labra, A., Pienaar, J., & Hansen, T. F. (2009). Evolution of Thermal Physiology in *Liolaemus* Lizards: Adaptation, Phylogenetic Inertia, and Niche Tracking. *The American Naturalist*, 174(2), 204–220. <https://doi.org/10.1086/600088>
- Ridley, M. (1983). *The explanation of organic diversity: the comparative method and adaptations for mating*. Oxford University Press, USA.
- Toljagić, O., Voje, K. L., Matschiner, M., Liow, L. H., & Hansen, T. F. (2017). Millions of years behind: Slow adaptation of ruminants to grasslands. *Systematic Biology*, (318). <https://doi.org/10.1093/sysbio/syx059>
- Sober, E. (2008). *Evidence and evolution: The logic behind the science*. Cambridge University Press.
- 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(8), 2369–2383.