**Jointly Modeling the Evolution of Discrete and Continuous Traits**

James D. Boyko1\*, Brian O’Meara2, Jeremy M. Beaulieu1

*1Department of Biological Sciences, University of Arkansas, Fayetteville, AR, 72701, USA*

*2Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, USA*

*\*To whom correspondence should be addressed (jboyko@uark.edu).*

**Running head**: **Jointly Modeling Discrete and Continuous Traits**

**Abstract**

Identifiability.

Questions of how plant life habit influence genome size evolution (Beaulieu et al. 2012), how substrate use alters limb length evolution (Mahler et al. 2013), or how tooth morphology slowly changes in response to habitat and diet (Toljagić et al. 2018) are evidence of a recurring interest in testing whether evolutionary changes in one variable is causal in the change of another. To study questions such as these, one very common phylogenetic comparative approach is to employ an Ornstein-Uhlenbeck (OU) model, which assumes distinct regimes, described by the evolution of a discrete character, are known completely *a priori* (e.g., Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012), or assumes that “shifts” in regimes can be inferred directly from the distribution of the continuous trait (e.g., Ingram and Mahler 2013; Uyeda and Harmon 2014; Khabbazian et al. 2016). While these approaches are practical, they assume that the combination of independently estimated discrete and continuous models will produce a joint model, but with the discrete trait being the driving force underlying the evolution of the continuous character. However, evolution is rarely as simple as cause and effect, and a more realistic view of evolutionary change is one which takes a dialectical approach (Levins and Lewontin 1985) – namely, we suspect that as often as a discrete character causes change in the continuous character, continuous characters also influence discrete character evolution, or at the very minimum, help provide information about this evolution.

Progress along these lines has mostly involved acknowledging uncertainty in the evolution of the discrete character by way of stochastic maps. That is, a large set of alternative reconstructions of the discrete character are obtained completely uninformed by the continuous trait’s evolution, then the likelihood of the continuous character becomes the average of the likelihoods across these maps (e.g., Revell 2012). The advantage of this approach is that there is an explicit model for how regimes change through time, but the evolution of these regimes remains entirely independent of the continuous trait, and the probability of these regimes is not explicitly considered. For example, it may be possible that the model that best fits the discrete data generates stochastic maps which do not fit the continuous data. A promising approach was recently described for detecting adaptive codon evolution (Jones et al 2020), where a set of maps obtained for a discrete phenotype under a standard Markov process is optimized along with parameters associated with genotype properties, thus forcing an emergent dependency between the two. Similarly, May and Moore (2020) developed a joint model for discrete and continuous characters under a state-dependent Brownian motion model. Their approach takes advantage of a Bayesian framework through the use of priors in order to accommodate variation in the “background” rate of evolution in the continuous trait (i.e., rate variation across lineages that is independent of the discrete character under consideration). The novel Bayesian pipeline recently developed by Tribble et al. (2021) is the first attempt that we are aware of for jointly modeling discrete and continuous traits under an OU framework. Their approach samples discrete stochastic mappings informed solely by the discrete trait along with regime mappings which were informed solely by the continuous trait while accounting for the potential of hidden variation. This method allows for a more effective test of correlation between the discrete and continuous characters. One drawback, however, is that they do not explicitly account for the joint probability of the discrete and continuous parameter estimates together. They assume that the combination of independently estimated discrete and continuous models produces a joint estimate – the theoretical and practical implications of which have not been thoroughly explored.

Here we propose an explicitly joint estimate of the likelihood for a discrete and continuous character by combining the probability of the continuous character given a particular regime and the probability of that discrete regime painting, integrated over many regime paintings. Specifically, we combine hidden Markov models of discrete character evolution (Beaulieu et al. 2013) with generalized Ornstein-Uhlenbeck models (Hansen 1997; Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012; Ho and Ané 2014*b*). We demonstrate how our framework, which we call “hOUwie”, can be used to test hypotheses of correlated evolution between discrete and continuous characters while also accounting for hidden character states and unobserved variation. We also consider and explore the potential for Type-I error that may arise within these joint models. When the evolution of a continuous trait is heterogeneous and unlinked to a discrete focal character, are they more likely to support a character-dependent model than simple, single rate, character independent models? Finally, we apply several hOUwie models to test the correlated dynamics of the mode of seed dispersal and climatic niche evolution and compare our results to those that did not account for the potential joint evolution of discrete and continuous variables.

**Materials and Methods**

*The hOUwie model*

Our model is composed of two processes: one that describes the evolution of a discrete character and the other describes the evolution of a continuous character. To model the evolution of a single continuous character we use an Ornstein-Uhlenbeck (OU) model (Hansen 1997; Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012; Ho and Ané 2014*b*). Formally, the OU process is an Itô diffusion satisfying:

.

This model combines the stochastic evolution of a trait through time with a deterministic component which models the tendency for a trait to evolve towards an optimum. In this model, the value of a trait, , is pulled towards an optimum, , at a rate scaled by the parameter . The optimum, , is a piecewise constant on intervals and takes values in a finite set. This can represent the set of “selective regimes”, “regimes”, or Simpson’s “adaptive zones” (Cressler et al. 2015), though it is consistent with a variety of true underlying microevolutionary models (Hansen 2014). Additionally, random deviations are introduced by Gaussian white noise , which is distributed as a normal random variable with mean zero and variance equal to . Thus, is a constant describing the rate of stochastic evolution away from the optimum. We use the set of extensions introduced by Beaulieu et al. (2012) and implemented in the R package OUwie. This allows for multiple primary optima in which both the pull strength () and the rate of stochastic evolution () can vary across the phylogeny. However, the algorithm used to calculate the likelihood described in Beaulieu et al. (2012) involves matrix inversion -- a computationally costly procedure. Therefore, we implement a linear-time computation of the likelihood of Gaussian trait models following Ho and Ané (2014*b*). To do this, we first transform the phylogeny such that its variance covariance matrix,, is 3-point structured. We can write the variance covariance matrix of the untransformed phylogeny as , where following Beaulieu et al. (2012) and Ho and Ané (2014),

and,

where, is the distance from the root to the beginning of the selective regime () for the number of selective regimes along the path from the root to the last common ancestor of and , or from the root to the terminal tip. Our transformed phylogeny now has a variance covariance matrix and diagonal matrix . We can then apply this function to calculate the quadratic quantities and determinant of . The probability of our continuous trait is given by

where is the number of tips in the phylogeny (, is the continuous trait value of each species, and is the expected value of each species given the continuous trait model – calculated following equation (11) of Beaulieu et al. (2012), *D* is the discrete character data, is a particular regime mapping, and are the parameters of the hOUwie model.

Next, we describe the calculation of the probability of the underlying regime structure, , which is the joint probability of discret­e characters () and stochastic mapping (). This calculation is analogous to the pathway likelihood of Steel and Penny (2000). Recently, May and Moore (2020) suggested that the joint probability of a regime structure and the discrete character is the product of the probabilities of exponentially distributed waiting times. By this definition, branch lengths are the sum of waiting times. However, under this definition, when we calculate the probability of starting and ending a branch in state *i*, the likelihood of a regime structure was unaffected by the number of transitions (Supplemental Materials), even though the maximum likelihood estimate should be zero transitions (O’Meara 2008). We suspect this is an issue of treating the probability density functions independently, when their convolution is required, because the time of the second transition depends on the timing of the first as well as the total branch length available. Ultimately, the effect of this inaccuracy is that the number of transitions has no influence on the probability of a branch’s state reconstruction and that the sum of the joint probabilities will be greater than the marginal probability. For further details, and R code necessary to demonstrate these issues, we refer readers to the Supplemental Materials.

To calculate the probability of discret­e characters () and stochastic mapping () we use an approximation. Our approximation relies on a finite number of degree-2 internodes and uses the standard Chapman-Kolmgorov equation to calculate the probabilities of beginning in a particular state and ending in state (Pagel 1994). As the number of internodes increase, the amount of time between nodes decreases and the approximation improves. Formally, we calculate the joint probability of a regime structure and the discrete character as

where the instantaneous rate matrix (, is the phylogeny, is the root state probability (Pagel 1994; Yang 2006; Maddison et al. 2007), is the number of external nodes (tips), internal nodes, and internodes (degree-2 nodes) summed, indicates a particular branch, , where is an indicator function which ensures that we only use the probability of states indicated by the specific the regime mapping instead of summing over all possible state combinations. The continuous character probability requires the discrete state(s) to be defined along the entire branch, thus we place transitions halfway between any two nodes (which can be internal, external, or internodes).

Our likelihood function for the joint probability of a continuous and a discrete character is therefore,

where summing over all possible maps (), is the probability of the continuous character () given the discrete character data (), mapping (), hOUwie parameters (), and phylogeny (). is the joint probability of the discrete character data () and stochastic mapping () given the hOUwie parameters () and phylogeny ().

*The hOUwie model space*

Our simulation studies examined 22 possible hOUwie model structures for a binary discrete character, although the possible number of models is significantly higher. For the discrete component of the model, we assumed that transitions between the observed characters were equal, although when hidden states were included, different hidden states were allowed to have different rates. Furthermore, due to concerns with parameter estimation, we constrained transitions between hidden states to be the same for observed states, but this constraint can be relaxed if desired. The continuous model structures allowable in hOUwie are a generalized form of those allowed in OUwie and now include models in which only alpha varies (OU1A), only sigma varies (OU1S), and combinations of an OU and BM process (OUBM). The model structures range from completely character-dependent to character-independent. Character-dependent (CD) models are models in which any continuous OU parameter differs between observed discrete state, whereas character-independent models (CID) test whether observed discrete states can be described by the same OU parameters. There are two types of character-independent model (Fig. 1). First, character-independent models include structures where there are no differences between any OU parameters. Under this model the entire evolutionary history of the clade can be described by a single alpha, sigma, and optimum value (Fig. 1a). To combat this potentially unrealistic assumption we introduce a character-independent model which allows for differences in the OU parameters to depend upon an unobserved hidden state (CID+). This addition allows for heterogeneity within the evolutionary process without the necessity of it being linked to a focal trait (Fig. 1c). In total we examine 22 unique model structures (2 CID, 10 CD, and 10CID+). HYB models are not examined because of their increased parameter complexity. However, we find that they consistently fail to estimate well when applied to an empirical dataset, suggesting that further work needs to be done before they are viable options to include in the model set.

A screenshot of a cell phone

Description automatically generated with low confidence

**Figure 1**: A state-transition diagram describing the examined hOUwie model structures. It is possible to add more complexity, such as allowing transitions between hidden states to depend on the observed character (e.g., 0A to 0B differing from 1A to 1B). Colored circles indicate different states. Different shades of the same color are used to describe different observed states (light orange and light purple are state 0, while dark orange and dark purple are state 1). Different colors are used to represent different hidden states (orange indicates hidden state A and purple indicates hidden state B). Parameters which are shared between states are shown above the states, whereas parameters specifically associated with a particular state are shown to their left or right. Transitions between states are shown with colored arrows.

*Simulation study*

We evaluated the performance of the hOUwie model using 880 simulated phylogenetic datasets. For each dataset, we simulated a 100-tip pure birth phylogenetic tree with , and rescaled tree height to 1. We simulated a two discrete state dataset under a variety of BM and OU models. The root state was sampled in equal proportions of state 1 and 2. The parameters used to generate a phenotypic dataset depend on the structure of the generating model. For example, an OUM model and OU1 model can have identical , but they must differ in or else OUM will collapse into OU1. Thus, we set parameter “baselines” at , which were doubled if the model structure allowed the parameter to vary (consistent with the approach taken in Beaulieu et al. 2012). For example, a two-state OU1 model would be simulated with parameters , but an OUM model would be simulated with parameters . Once a phylogeny and phenotypic dataset were simulated, we fit our models to assess parameter estimation accuracy and model selection power. Although this represents a small subset of the potentially vast parameter space available to OU models, the behavior of these models has been thoroughly characterized and thus we chose parameters within the range of typical identifiability (Beaulieu et al. 2012; Ho and Ané 2014*a*; Cressler et al. 2015). Additionally, because hOUwie uses a variable number of mappings and internodes, we evaluate changing the number of stochastic maps and number of internodes included. Internodes are placed at a time interval specified by the user. For each branch, an internode will be placed a specified time interval apart and if the time slice is longer than the branch, no internodes are added. For our simulations, internodes were placed every 0.1, 0.5, and 1 MY. We then fit each model using 50, 100, and 200 stochastic mappings per likelihood evaluation. Each combination of the internode time interval and number of stochastic mappings per likelihood evaluation was evaluated for all 22 model structures.

To evaluate hOUwie’s performance and assess parameter estimation we fit models under two scenarios. First, when assessing internode placement and number of stochastic maps per likelihood, we fit the generating model to a dataset it created and examined differences between the simulating values and the maximum likelihood estimate using reduced mean squared error (RMSE). However, it is often the case that biological hypotheses do not depend on the exact value of a parameter, or it is difficult to make a statement about what value we expect or to take on. Instead, most hypotheses are put in relative terms. For example, we may expect that the rate of evolution will be greater for species with smaller range sizes than species with larger range sizes. Furthermore, if we do find a difference in the relative values of the parameters there is no guarantee that this difference is biologically meaningful. Thus, in addition to RMSE, we include assessments of Type-S (sign) errors in our evaluation of parameter estimates (Gelman and Carlin 2014). We use a simple calculation of Type-S error, evaluating the proportion of parameter estimated differences were inferred to be in the correct direction based on the true simulating values. Comparing ratios of parameters is also consistent with recommendations from Ho and Ané (2014*a*) to deal with identifiability issues.

Next, to evaluate model distinguishability, we simulate data under each of the 22 examined models and then fit each of those 22 models to all datasets generated. The simulation protocol for parameter values is identical to what was discussed above, but the number of stochastic maps was fixed at 100 and internodes were placed along branches every 0.1 units. We used the Akaike Information Criterion (AIC) to determine whether the best fitting model is the same as the generating model (Burnham and Anderson 2002). One flaw with this is that the best fitting model should not always be the generating model even under ideal circumstances: a trivial example is that for a tree with a single tip, the best continuous model is Brownian motion, even if the datum at that tip was generated using a more complex OU model. The simulations used here were for trees large enough that we would expect the best model to be the generating model if all is working as intended. Additionally, we can examine whether we find evidence of character dependence or character independence when each is the true scenario. Particularly of interest is when we simulate data under our CID+ model, whether we find false support for character dependence. We may expect this outcome because the CID+ model include heterogeneity in the evolutionary process that is unlinked to our focal trait and may produce patterns more consistent with character dependence than simple character independent models. This situation is analogous to the problem described in state dependent speciation-extinction (SSE) models by Rabosky and Goldberg (2015) and Beaulieu and O’Meara (2016) which demonstrated false support for character dependent models when speciation extinction dynamics are not linked to the focal character, but instead are heterogeneous.

*Joint probability for all possible mappings*

To evaluate the likelihood of a set of parameters, it would be preferable to marginalize the probability of the discrete and continuous characters over all possible stochastic mappings. However, this is computationally infeasible for most empirical datasets and thus our approach samples stochastic mappings proportional to the discrete character probabilities. Nonetheless, for a small phylogenetic tree and a modest number of internodes, it is possible to decompose the discrete and continuous probabilities for all potential stochastic mappings and explore the theoretical properties of jointly modeling the two character classes. Given the inconsistency of OU parameter estimates when regimes were associated exclusively with hidden states (Table 1, CID+), we briefly discuss the behavior of CD and CID+ models on a four-taxa phylogenetic tree to demonstrate that estimating the CID+ models is possible but may require a different approach to be reliable. We simulated a continuous character on the regime shown in Fig. 2 with parameters on a tree with height 1. These extreme parameter values are chosen to ensure a strong link between the underlying regime and the simulated continuous data. Regimes were then treated as either an observed (Fig. 2a) or hidden state (Fig. 2b). When treated as a hidden state, one of the two binary observed states was assigned to either tip within the hidden regime. We then fit the hOUwie model by integrating over all possible stochastic mappings, rather than a subset. For a CID model there are 8 possible mappings with no internodes, and for a CID+ model there are 1024 possible mappings with no internodes.

*A picture containing qr code

Description automatically generated*

**Figure 2**:Two scenarios for continuous and discrete character evolution. a) Continuous evolution and observed discrete character evolution are linked with the same underlying regime painting. Under this scenario, the observed discrete characters (circle and square) provide information about the regime structure. b) Continuous evolution and hidden discrete state evolution (green and purple) are linked. The observed state (circle and square) is not linked to the underlying regime or the continuous trait. Information to infer the underlying continuous regime comes directly from the continuous character distribution.

*The impact of climatic variables on seed dispersal*

An ideal way to address the complexity of a trait syndrome would be to enumerate all the phenotypes which comprise the seed dispersal syndrome and study the dynamics in-depth. However, on macroevolutionary scales, this approach is intractable due to the vast amount of species-specific data required and the increasing number of parameters introduced into models as the number of traits increases. Thus, we propose an alternative which will use a single feature of fruit morphology as a proxy for seed dispersal, but also acknowledges that not all lineages with the same fruit morphology are going to evolve in the same way. It is for this reason that we use hidden Markov models (HMM) as the basis for describing discrete character evolution. With HMMs we allow for the fact that dynamics within broadly defined categories are likely inadequate descriptions when treated homogeneously.

Here we examine three predictions outlined in Vasconcelos et al. (2021), but specifically measuring aridity. First, we expect that the climatic optima for abiotically dispersed seeds will be drier compared to biotically dispersed seeds (). Second, we expect that abiotically dispersed seeds will have faster rates of climatic niche evolution (). Finally, we expect that the climatic niches of biotically dispersed seeds will be more conserved through time (). We apply several hOUwie models to test these hypotheses and compare our results to those discussed in Vasconcelos et al. (2021). We expect that any differences found between this study and Vasconcelos et al. (2021) are because we can explicitly account for the joint probability of the discrete and continuous characters. We choose Ericaceae specifically because Vasconcelos et al. (2021) found two counter-intuitive results. First, they found that the phenotypic optima of abiotically dispersed seeds was more humid than biotically dispersed seeds. Second, they found that the rate of climatic evolution was greater in biotically dispersed seeds than abiotically dispersed seeds. Additionally, this previous work assumed either character dependence or simple character independence. As our simulations will show (see below), this assumption is not always valid, as heterogeneous character independence can give a false signal of character dependence. To account for this, we include the CID+ model within our model set.

We ran 27 hOUwie models: 2 CID, 10 CD, 10 CID+, and 5 HYB. *Gaultheria* is technically a dry-fruited genus within Ericaceae but has a persistent fleshy calyx that attracts frugivores (Stevens et al. 2004). For that reason, we code this as fleshy fruited within our dataset. We run each of the models with several, hOUwie specific, nuisance parameter sets. Models are evaluated using the sample size corrected Akaike Information Criterion (AICc) and model averaging is conducted when discussing how our results relate to our hypotheses (Burnham and Anderson 2002). Finally, we evaluate the expected values, the optima, and compare our results to Vasconcelos et al. (2021).

**Results**

*Parameter estimation given the generating model*

Parameter estimation in hOUwie did not vary greatly depending on the number of stochastic maps or number of internodes included (Fig. S2). Most variation in parameter estimation was instead related to the model complexity and the specific parameters which were allowed to vary (Table 1). Datasets which were generated with a variable alpha rate were never chosen as the best fitting model. This parameter estimation problem was amplified when alpha was allowed to vary alongside other parameters (as was the case in Beaulieu et al. 2012). Increasing the number of taxa has been shown to improve inference of OU models, but even with more taxa, if alpha is not large enough relative to sigma, estimation may remain poor. (Beaulieu et al. 2012; Ho and Ané 2014*a*; Cressler et al. 2015). CID+ parameter estimations were highly sensitive to model complexity. Although there are examples of low RMSE for each of the OU parameters, when these parameters were allowed to vary, error increased rapidly. This suggests that alternative approaches may be necessary to accurately estimate a hidden state OU model and possibly shift detection methods can supplement this gap in the interim (Uyeda and Harmon 2014; Khabbazian et al. 2016).

The sign error for variable theta was 0.05 or less, regardless of additional model complexity. Variable sigma models also performed well, with only models that also included a variable alpha (M7 and M12) having a sign error rate above 0.15 (Table 1). Finally, identifying the sign difference of alpha proved difficult for all variable alpha models. This is consistent with previous results which found the parameters of variable alpha models are generally difficult to estimate (Beaulieu et al. 2012). Sign errors cannot be evaluated for CID models because the parameter values associated with hidden states can freely switch between rate classes and have identical likelihoods. For example, if we were to simulate a dataset where states A and B are hidden states unlinked to the observed discrete character, , and , there would be no difference in likelihood if , and or , and . However, this could be distinguishable if both the observed discrete character and continuous character were linked to the hidden state.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| model | type\_1 | type\_2 | k | alpha\_free | sigma\_free | theta\_free | alpha\_sign\_error | sigma\_sign\_error | theta\_sign\_error | alpha\_rmse | sigma\_rmse | theta\_rmse | prop\_best | avg\_AICwt | prop\_CD | mean\_obs\_exp | sd\_obs\_exp |
| M1 | BM | CID | 3 | FALSE | FALSE | FALSE | NA | NA | NA | NA | 0.086 | 0.321 | 0.8 | 0.313 | 0.2 | 0.032 | 0.153 |
| M2 | BM | CD | 4 | FALSE | TRUE | FALSE | NA | 0.05 | NA | NA | 0.419 | 0.391 | 0.367 | 0.213 | 0.433 | -0.001 | 0.141 |
| M3 | BM | CID | 6 | FALSE | TRUE | FALSE | NA | NA | NA | NA | 97.669 | 0.283 | 0 | 0.007 | 0.3 | 0.026 | 0.188 |
| M4 | OU | CID | 4 | FALSE | FALSE | FALSE | NA | NA | NA | 1.88 | 0.344 | 0.096 | 0.333 | 0.179 | 0.433 | -0.011 | 0.098 |
| M5 | OU | CD | 5 | TRUE | FALSE | FALSE | 0.556 | NA | NA | 9.979 | 0.179 | 0.104 | 0.036 | 0.033 | 0.536 | 0.004 | 0.064 |
| M6 | OU | CD | 5 | FALSE | TRUE | FALSE | NA | 0.05 | NA | 1.44 | 0.644 | 0.102 | 0.167 | 0.139 | 0.533 | -0.018 | 0.105 |
| M7 | OU | CD | 6 | TRUE | TRUE | FALSE | 0.4 | 0.4 | NA | 2.49 | 4.18 | 0.101 | 0 | 0.039 | 0.5 | 0.013 | 0.064 |
| M8 | OU | CD | 5 | FALSE | FALSE | TRUE | NA | NA | 0.05 | 0.804 | 0.618 | 11.52 | 0.167 | 0.126 | 1 | -0.025 | 0.196 |
| M9 | OU | CD | 6 | TRUE | FALSE | TRUE | 0.75 | NA | 0 | 2.803 | 0.336 | 19.421 | 0.067 | 0.067 | 1 | 0.024 | 0.126 |
| M10 | OU | CD | 6 | FALSE | TRUE | TRUE | NA | 0.15 | 0.05 | 0.982 | 13.478 | 86.098 | 0.233 | 0.226 | 1 | 0.129 | 0.274 |
| M11 | OU | CD | 7 | TRUE | TRUE | TRUE | 0.5 | 0.2 | 0 | 2.667 | 20.419 | 53.158 | 0.3 | 0.271 | 1 | 0.183 | 0.267 |
| M12 | OU | CID+ | 7 | TRUE | FALSE | FALSE | NA | NA | NA | 23.696 | 1.733 | 0.75 | 0 | 0.001 | 0.345 | -0.18 | 0.067 |
| M13 | OU | CID+ | 7 | FALSE | TRUE | FALSE | NA | NA | NA | 3.942 | 89.838 | 0.122 | 0 | 0.004 | 0.433 | 0.012 | 0.102 |
| M14 | OU | CID+ | 8 | TRUE | TRUE | FALSE | NA | NA | NA | 35.524 | 59.036 | 1.277 | 0.038 | 0.039 | 0.308 | 0.488 | 0.102 |
| M15 | OU | CID+ | 7 | FALSE | FALSE | TRUE | NA | NA | NA | 1.397 | 2.091 | 31.886 | 0 | 0.001 | 0.667 | 0.001 | 0.422 |
| M16 | OU | CID+ | 8 | TRUE | FALSE | TRUE | NA | NA | NA | 2.878 | 4.048 | 57.683 | 0 | 0 | 0.667 | 0.276 | 0.483 |
| M17 | OU | CID+ | 8 | FALSE | TRUE | TRUE | NA | NA | NA | 1.472 | 80.775 | 49.334 | 0 | 0.002 | 0.467 | -0.081 | 0.395 |
| M18 | OU | CID+ | 9 | TRUE | TRUE | TRUE | NA | NA | NA | 3.227 | 150.971 | 51.689 | 0 | 0 | 0.533 | 0.349 | 0.537 |
| M19 | mixed | CD | 4 | TRUE | FALSE | FALSE | NA | NA | NA | 9.085 | 2.38 | 0.218 | 0.036 | 0.071 | 0.393 | 0.027 | 0.103 |
| M20 | mixed | CD | 5 | TRUE | TRUE | FALSE | NA | 0.05 | NA | 1.024 | 6.754 | 0.314 | 0.138 | 0.174 | 0.724 | 0.043 | 0.169 |
| M21 | mixed | CID+ | 6 | TRUE | FALSE | FALSE | NA | NA | NA | 22.298 | 2.631 | 0.199 | 0 | 0.002 | 0.267 | 0.015 | 0.11 |
| M22 | mixed | CID+ | 7 | TRUE | TRUE | FALSE | NA | NA | NA | 16.703 | 82.599 | 2.181 | 0 | 0.003 | 0.429 | -0.13 | 0.164 |

**Table 1**: Model is model name. type\_1 is the type of continuious process. type\_2 is whether the model is character dependent (CD) or independent (CID). k is the number of parameters. Alpha, sigma, and theta free are whether or not that parameter is free to vary in the model. sigma, theta, alpha sign error are the proportion of times the incorrect sign difference was inferred compared to simulated.theta, sigma, alpha rmse are the reduced mean squared errors of the best fitting model compared to the simulated values. Prop\_best is the proportion of fitted models that matched the generating model. Avg aic weight is the average weight of the generating model to its simulated data when fit and compared to all other models. prop\_CD is the proportion of models that had any CD model as the best fitting. Mean\_obs\_exp is the average difference between the expected and observed values for the best fitting model. SD\_obs\_exp is the standard deviation of the differences between observed and expected values (taken per species per simulation).

Despite variable theta models having the lowest sign errors and highest successful detection of character dependent models, theta estimates still had a high RMSE. One advantageous property of theta is that, unlike sigma and alpha, it describes a directly observable feature of the organism with easily interpretable units (the same as the continuous phenotype). Thus, we can examine the distribution of the simulated character and compare these distributions with the inferred optimal values and expected values to understand if the estimates are meaningful given the dataset. This post-hoc analysis revealed an interesting feature of our simulated dataset and hOUwie’s potential to find variation outside of the observed data (Fig. 3). We find that, although there is significant error around the optima estimate, the simulated data rarely included any species that had a phenotypic optimum near the simulated optima (Fig. 3b). This suggests that hOUwie can estimate phenotypic optimum outside of the observed phenotypic distributions. We are also able to examine the expected continuous trait values according to the fitted model and determine how well they match the observed phenotypic values (Fig. b,c; Table 1). The general match between observed and expected values suggests a reasonable approximation of the data by the model. Furthermore, although the RMSE in theta is large, it is likely a consequence of a lack of direct phenotypic observations around the optimum. This may be indicative of the potential for hOUwie to detect optima well outside the range of observed phenotypes. While the exact estimate of the phenotypic optima will not be highly accurate, its direction compared to the observed phenotypic distribution is still interesting and may be useful for hypothesis testing.

![Chart, bar chart

Description automatically generated]()

**Figure 3**: The two columns are data simulated with an initial discrete root state of 1 or 2 which will correspond to an initial continuous value of 12 or 24. a) Model averaged parameter estimates for theta. The dashed blue and red lines represent the simulating values of theta for states 1 and 2 respectively. b) the observed continuous character frequency. When data is simulated with root state 1 initially, most continuous data tends to be clustered around the initial value of 12. The same is true of root state 2, but most continuous data are clustered around the initial value of 24. c) The model expectation of the phenotypic value of the tips. The exact equations for calculating the expected value are described in Butler and King (2004) and Beaulieu et al. (2012) but are essentially the weighted average of time spent in a particular regime state given OU parameters and proportionally weighted by the nearness of the regime state to the present given an initial value. Summarized here are all non-hidden Markov models for data simulated under a variable theta CD model. HMMs are not included because model averaging hidden state associated parameter values is difficult outside the context of tip averages.

*The power to detect the generating model*

Simple character independent (CID) and character dependent (CD) models performed the best in our power analyses (Fig. 4). Of the character dependent models, variable alpha, and variable sigma models (M5, M7, M9) had low AIC weights across simulations even when they were the generating modes (Table 1). This reinforces our previous findings that the more complex variable alpha or variable sigma models are the most difficult to estimate and should be applied and interpreted with caution. Conversely, CD models which varied theta always found evidence for character dependence when the generating model included a link between discrete and continuous characters. Power analysis also suggests that hOUwie can detect character independence when that is the true model. However, when character independent models are heterogeneous (i.e., there is variation in how a continuous trait evolves unlinked to the focal character) hOUwie either favored simpler versions of character independent models or, more concerningly, character dependent models (Fig. 4c). This suggests that even in cases where our focal discrete and continuous characters are independent, if the evolutionary process is heterogeneous, we may still find false support for character dependence.

![Chart, box and whisker chart

Description automatically generated]()

**Figure 4**: Boxplots of summed AICwt for each model class for three simulating scenarios. Models are summarized into three classes (CD, CID, or CID+) although within each class models there are different model structures (e.g., variable theta vs variable alpha – see table 1). a) The simulating model is character dependent. The percent best fit (proportion from each class with dAIC = 0) for each model class was CD, CID, and CID +. b) The simulating model is character independent. The percent best fit for each model class was CD, CID, and CID +. c) The simulating model is character independent. The percent best fit for each model class was CD, CID, and CID +.

*Joint probability for all possible mappings*

Using all possible regime mappings, we were able to successfully distinguish between character dependent and character independent models (Table 2). When data was generated following a variable theta character dependent scenario, a character dependent model was preferred with a dAIC of 5.7, suggesting strong evidence for character dependence (Burnham and Anderson 2002). Interestingly, the maximum likelihood parameter estimates for the CD and CID+ models were identical in this scenario. The reason that a CD model will fit better despite having the same parameter estimates is the additional uncertainty when estimating hidden states in the CID+ model. Conversely, when continuous traits were unlinked to the underlying regime, a hidden state character independent model was preferred with strong evidence (dAIC=15.4). Taken together, this points to the theoretical possibility of estimating OU processes even when unlinked to an observed character. However, a different approach may be necessary as integrating over all possible mappings is not feasible for most empirical datasets.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| generating model | fit model | lnlik | k | aic | dAIC | discrete rate | alpha | sigma2 | theta\_1 | theta\_2 |
| CD | CD | -4.8094 | 5 | 19.61883 | 0 | 1.75 | 10 | 5 | 5 | 10.9375 |
| CD | CID+ | -7.6937 | 5 | 25.38755 | 5.76872 | 1.75 | 10 | 5 | 5 | 10.9375 |
| CID+ | CD | -11.924 | 5 | 33.84911 | 15.4376 | 2.03125 | 0.625 | 11.322632 | 6.234131 | 7.1875 |
| CID+ | CID+ | -4.2057 | 5 | 18.41151 | 0 | 48.99516 | 50 | 2.895839 | 4.991589 | 10 |

*Seed dispersal and climatic evolution*

We found evidence of a character dependent model over either a simple or hidden state character independent model, suggesting a link between the climatic niche of Ericaceae lineages and their fruit type (Table 3). The results presented in Table 3 are summarized into three broad categories (CD, CID, CID+, and HYB), but we tested 27 different scenarios and the overall best model was M11, in which alpha, sigma2, and theta were all allowed to vary (AICwt=0.87; Supplemental Materials). Both fleshy and dry fruits had phenotypic optima which correspond to non-arid environments. However, the 0.59 AI of fleshy fruits suggests dryland as their optimal habitat, and the 1.044 AI of dry fruits corresponds to a non-dryland environment (Middleton and Thomas 1997). Both and interact to create tip variance, so in addition to , we measured stationary variance . We found that Ericaceae lineages with dry fruits were less variable in their climatic niche evolution (, ) compared to fleshy fruits (, ). Additionally, the phylogenetic signal of fleshy fruited lineages was greater than dry fruited lineages (). This corresponds to phylogenetic half-lives of and which are 15% and 25% of the total tree height respectively. Transitions to fleshy fruit occurred at 0.0017 transitions per million years which is more than 6 times faster than transitions to dry fruits (0.0003 transitions per million years). But in both cases, the average waiting time to the next transition is much greater than the height of the tree (117.3 MY) suggesting few transitions between fruit types throughout the Ericaceae’s history. Taken together with the relatively short half-lives of each fruit type, our modeling suggests that the climatic optima of Ericaceae lineages is relatively stable. Finally, we note that the difficulty in estimating complex models is evident in this empirical application. Models with higher number of parameters should provide a more likely explanation of the data, but that was not always the case (Supplemental Materials, Table 3). This points to the failure of more complex models finding a maximum likelihood estimate. However, this is consistent with our earlier simulation study which suggests that without integrating over all possible mappings, it may not be possible to reliably find the MLE when the hidden states are associated with the OU process.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| model | np | lnLik | DiscLik | ContLik | AIC | dAIC | AICwt |
| M1 | 4 | -265.91 | -32.54 | -233.36 | 539.81 | 69.63 | 0 |
| M2 | 5 | -263.85 | -32.54 | -224.52 | 537.69 | 67.51 | 0 |
| M3 | 6 | -267.67 | -34.31 | -229.3 | 547.35 | 77.17 | 0 |
| M4 | 5 | -233.63 | -32.55 | -201.08 | 477.27 | 7.09 | 0.03 |
| M5 | 6 | -257.11 | -32.55 | -222.62 | 526.22 | 56.04 | 0 |
| M6 | 6 | -233.15 | -32.55 | -196.19 | 478.3 | 8.12 | 0.01 |
| M7 | 7 | -241.35 | -32.57 | -205.82 | 496.7 | 26.52 | 0 |
| M8 | 6 | -231.8 | -32.53 | -196.32 | 475.6 | 5.42 | 0.06 |
| M10 | 7 | -231.32 | -32.55 | -196.01 | 476.65 | 6.47 | 0.03 |
| **M11** | **8** | **-227.09** | **-32.56** | **-192.87** | **470.18** | **0** | **0.87** |
| M12 | 7 | -235.42 | -34.34 | -197.12 | 484.84 | 14.66 | 0 |
| M13 | 7 | -309.15 | -34.57 | -271.72 | 632.31 | 162.13 | 0 |
| M15 | 7 | -330.78 | -50.77 | -264.39 | 675.56 | 205.37 | 0 |
| M16 | 8 | -3931.97 | -35.42 | -2430.46 | 7879.94 | 7409.76 | 0 |
| M17 | 8 | -259.76 | -34.36 | -222.04 | 535.52 | 65.33 | 0 |
| M18 | 9 | -391.77 | -35.98 | -354.05 | 801.54 | 331.36 | 0 |
| M19 | 5 | -257.09 | -32.54 | -222.73 | 524.18 | 54 | 0 |
| M20 | 6 | -241.15 | -32.55 | -208.48 | 494.3 | 24.12 | 0 |
| M21 | 6 | -235.4 | -34.32 | -197.68 | 482.81 | 12.62 | 0 |
| M22 | 7 | -336.79 | -35.17 | -299 | 687.59 | 217.41 | 0 |
| M23 | 8 | -596.26 | -51.8 | -544.45 | 1208.51 | 738.33 | 0 |
| M24 | 9 | -307.66 | -42.7 | -260.65 | 633.32 | 163.14 | 0 |
| M25 | 12 | -296.3 | -43.01 | -247.55 | 616.6 | 146.42 | 0 |
| M26 | 12 | -362.65 | -61.05 | -281.89 | 749.3 | 279.12 | 0 |
| M27 | 15 | -354.15 | -66.97 | -268.04 | 738.3 | 268.12 | 0 |

Supplemental Table – Modeling results from Ericaceace seed dispersal and aridity.

**Discussion**

Phylogenetic comparative methods have been widely applied to study discrete and continuous characters separately, but there are surprisingly few options which simultaneously use both classes of character (Felsenstein 2012). Here we describe the hOUwie model and demonstrate how it can be used to test hypotheses of correlated evolution between discrete and continuous characters while accounting for hidden character states and unobserved variation. Our model jointly models discrete and continuous characters by linking both via a common regime painting. However, unlike other similar methods, our likelihood formula explicitly calculates the probability of the underlying regimes. This has the advantage of describing the discrete character evolution probabilistically and allows information from the discrete and continuous characters to jointly contribute to the overall likelihood.

*The significance of jointly modeling characters*

Most implementations of an OU process within PCMs assume that the regimes are known or estimate them without considering discrete character evolution (Uyeda and Harmon 2014). These approaches leave little room for inference about the regimes themselves and how they change through time, nor the possibility that the evolution of the continuous character could influence change in the regimes or vice versa. The primary difference between hOUwie and other applications of the OU process is the explicit joint modeling of the underlying regime structure and continuous trait. There are three main advantages to jointly modeling discrete and continuous traits. First, it represents a more realistic view of trait evolution. The evolution of any discrete or continuous trait is not independent of all other traits a lineage possesses. Second, the mutual information of multiple variables makes our inferences more robust (Cover and Thomas 1991). If we are certain of one variable (seed size), but uncertain about another variable (climate), we can use the knowledge of the relationship between our variables to draw more accurate inferences about the uncertain variable. Finally, using hidden Markov models allows for the discovery of hidden variation. There is little reason to believe that any relationship between discrete and continuous variables will be homogenous throughout the history of a lineage. With hOUwie, it is theoretically possible to discover regions of the phylogeny where different types of relationships emerge.

A joint model also makes the parameter estimates related to continuous trait evolution independent of pre-defined regime paintings. The use of pre-defined regime mappings can be often be useful for hypothesis testing (Butler and King 2004; Beaulieu et al. 2012). For example, testing for evidence for character displacement by mapping either a sympatric or allopatric evolutionary history (Butler and King 2004). However, when using this approach, the mapping is taken as an absolute certainty. There is no room for inference of other potential patterns nor is there the acknowledgement that the *apriori* mapping is one of many possibilities. Instead of contrasting mappings, hOUwie’s methodology integrates over the uncertainty of alternative regimes and relies on the interpretation of parameter estimates from contrasting model structures to find evidence for hypotheses. To that end, modeling the joint probabilities puts our models into a likelihood framework. This allows for model comparison, model averaging, and other tools to be utilized when comparing macroevolutionary hypotheses (Burnham and Anderson 2002). This framework may be preferable to testing whether a particular hypothesis is better than a trivial null model (Beaulieu and O’Meara 2016; Caetano et al. 2018). Detecting a significant correlation between two traits is an important first step in establishing an evolutionary relationship, however a richer understanding may be achieved through examining the models’ parameterization in relation to specific hypotheses. Model averaged parameters and tip rates are a way to both test preexisting hypotheses and begin developing new hypotheses based on interpretation (misse). Additionally, because the relative explanatory power of each model is related to parameter estimates, by using model averaging we account for model structure uncertainty in addition to parameter estimation uncertainty (Burnham and Anderson 2002).

*Seed dispersal and climatic niche evolution in Ericaceae*

For sedentary organisms, such as plants, dispersal is mainly limited to a brief stage of their life cycle and mediated mainly through the movement of seeds (Levin et al. 2003). Generally, the expectation is that seeds dispersed by frugivores are going to be dispersed to environments more like their parents’ environment, whereas abiotically dispersed seeds are likely to be more erratic in their dispersal patterns (Schupp 1993; Westoby et al. 1996). Furthermore, it has been proposed that adaptations for frugivorous dispersal is linked to tropical and subtropical biomes. This is because in these warmer and wetter habitats, large trees create shady environments where competition for light is more important. A shadier habitat then imposes a selective pressure for larger seeds because more nutrients are needed for germination (Foster and Janson 1985). However, the evolution of larger seeds comes with a tradeoff as they have a significantly lower dispersal potential (Howe and Smallwood 1982). Thus, we might expect that the climatic variables of a habitat influence the probability of transitioning between abiotic and biotic modes of dispersal, with transition rates from abiotic to biotic being greater in less arid environments. Here we revaluated three hypotheses related to climatic niche evolution and seed dispersal: (1) the climatic optima for abiotically dispersed seeds will be drier compared to biotically dispersed seeds (), (2) abiotically dispersed seeds will have faster rates of climatic niche evolution (), and (3) climatic niches of biotically dispersed seeds will be more conserved through time (). Consistent with previous findings we did not find evidence of these hypotheses in Ericaceae.

*Caveats*

There are three important caveats of our model. First, our discrete mapping probability ( is merely an approximation. What we calculate here is the probability of starting in a particular state and ending a particular state , summed over all possible paths. However, the continuous model probability is based off a particular pathway history which is defined throughout the entire branch (Hansen 1997). Our justification for this approximation is that as the length between internodes decreases this approximation improves because the probability of more than a single transition (or no transitions) decrease (see appendix). The second caveat is that we do not force hOUwie to sum over all possible mappings . This is because the number of mappings will grow exponentially as the number of nodes and internodes increases and the computation will quickly become infeasible. Instead, we simulate node, internodes, and tip states (tip states only in the case of hidden Markov models) using the stochastic mapping procedure described in Bollback (2006) and modified to include the possibility of internodes. We explored this caveat in series of simulation studies and demonstrate that this sampling procedure helps identify high probability mappings which contribute the most to the overall joint probability. Our final caveat is that because we generate node mappings using the stochastic mappings procedure, they would typically be based solely on the discrete character process. This can be particularly problematic for character independent models where the discrete character and continuous character are decoupled. An imperfect solution to this problem is to weight tip probabilities when first calculating the conditional likelihoods (Bollback 2006). Typically, tip values for hidden Markov models are given a weight of 1 for each state (Beaulieu et al. 2013). However, we can modify this value to be proportional to the probability of sampling a particular tip’s continuous trait value given the models theta and sigma value for each discrete state. This is the approach we take; we assume that the conditional probability of a tip being in a particular hidden state is not equal, but rather proportional to ; the probability of sampling the continuous trait value under a normal distribution for each possible state . In cases where there are no differences between parameters, conditional probabilities have equal weights.

*Future extensions*

We demonstrate a potentially serious problem when fitting OU models without allowing for heterogeneity in the evolution process. In cases where discrete and continuous characters are unlinked, but there is heterogeneity in the dynamics continuous character, we find high false-positive support for character dependent models. Although hOUwie is designed to mitigate these concerns, due to the caveats previously discussed it is far from a complete solution. The joint estimation procedure implemented in hOUwie has high parameter estimation error when focal discrete characters are unlinked to the continuous phenotype. There are several avenues to improve the hOUwie model that are actively being explored. Statistical work on Markov-Modulated Ornstein-Uhlenbeck models represent an intriguing possibility but have yet to be applied in phylogenetic comparative biology (Huang et al. 2016). However, if implemented in a phylogenetic context, it would completely remove the need for a simulation approach. Finally, it is possible to extend hOUwie to include state-dependent speciation and extinction dynamics. This extension would require a different calculation of the underlying regime mapping probability but would be relatively straightforward. A challenging aspect of this extension would be generating high joint probability mappings, although approaches for stochastically mapping SSE models exist (Freyman and Höhna 2019).

Conclusion (in progress)

The use of pre-defined discrete character mappings can be useful for testing hypotheses which rely on distinct, well-defined differences in the evolutionary histories of lineages. However, this approach assumes that the underlying mapping is known with complete accuracy. This approach inflates confidence in the mappings because it ignores the probabilistic nature of them. hOUwie’s methodology integrates over the uncertainty of high probability regimes and relies on the interpretation of parameter estimates from contrasting model structures to find evidence for hypotheses.

Joint modeling is realistic and theoretically important. The strong correlation between discrete and continuous suggest that practically, using existing methods should be okay. This was reinforced by our findings in seed dispersal – nearly identical parameter estimates. Hidden trait evolution has shown how heterogeneity in the evolutionary process can lead to false signal character dependent signals. hOUwie is meant to solve this but sadly comes short. Future work should focus on creating a model sets robust to this problem, which seems to be pervasive throughout phylogenetic comparative methods. Additionally, it is becoming important to evaluate what exactly hidden states are. The theoretical understanding is there, but what do they mean biologically? In some cases, we have a clear expectation of what they could be (refs.), but often we are left wondering if this is a real biological thing or simply noise in measurement.

Works Cited

Beaulieu, J. M., D.-C. Jhwueng, C. Boettiger, and B. C. O’Meara. 2012. Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. Evolution: International Journal of Organic Evolution 66:2369–2383.

Beaulieu, J. M., and B. C. O’Meara. 2016. Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. Systematic Biology 65:583–601.

Beaulieu, J. M., B. C. O’Meara, and M. J. Donoghue. 2013. Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character: The Evolution of Plant Habit in Campanulid Angiosperms. Systematic Biology 62:725–737.

Bollback, J. P. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC bioinformatics 7:88.

Boyko, J. D., and J. M. Beaulieu. 2021. Generalized hidden Markov models for phylogenetic comparative datasets. (N. Cooper, ed.)Methods in Ecology and Evolution 12:468–478.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach (2nd ed.). Springer, New York.

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

Caetano, D. S., B. C. O’Meara, and J. M. Beaulieu. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models: HMM AND THE ADEQUACY OF SSE MODELS. Evolution 72:2308–2324.

Cover, T. M., and J. A. Thomas. 1991. Elements of information theory. Wiley series in telecommunications. Wiley, New York.

Cressler, C. E., M. A. Butler, and A. A. King. 2015. Detecting Adaptive Evolution in Phylogenetic Comparative Analysis Using the Ornstein–Uhlenbeck Model. Systematic Biology 64:953–968.

Felsenstein, J. 2012. A Comparative Method for Both Discrete and Continuous Characters Using the Threshold Model. The American Naturalist 179:145–156.

Foster, S., and C. H. Janson. 1985. The relationship between seed size and establishment conditions in tropical woody plants. Ecology 66:773–780.

Freyman, W. A., and S. Höhna. 2019. Stochastic Character Mapping of State-Dependent Diversification Reveals the Tempo of Evolutionary Decline in Self-Compatible Onagraceae Lineages. Systematic Biology 68:505–519.

Gelman, A., and J. Carlin. 2014. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. Perspectives on Psychological Science 9:641–651.

Hansen, T. F. 1997. Stabilizing Selection and the Comparative Analysis of Adaptation. Evolution 51:1341–1351.

———. 2014. Use and Misuse of Comparative Methods in the Study of Adaptation. Pages 351–379 *in* L. Z. Garamszegi, ed. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Springer Berlin Heidelberg, Berlin, Heidelberg.

Hansen, T. F., J. Pienaar, and S. H. Orzack. 2008. A Comparative Method for Studying Adaptation to a Randomly Evolving Environment. Evolution 62:1965–1977.

Ho, L. S. T., and C. Ané. 2014*a*. Intrinsic inference difficulties for trait evolution with Ornstein‐Uhlenbeck models. Methods in Ecology and Evolution 5:1133–1146.

Ho, L. si, and C. Ané. 2014*b*. A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models. Systematic Biology 63:397–408.

Howe, H. F., and J. Smallwood. 1982. Ecology of Seed Dispersal. Annual Review of Ecology and Systematics 13:201–228.

Huang, G., H. M. Jansen, M. Mandjes, P. Spreij, and K. D. Turck. 2016. Markov-modulated Ornstein-Uhlenbeck processes. Advances in Applied Probability 48:235–254.

Ingram, T., and D. L. Mahler. 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein‐Uhlenbeck models with stepwise Akaike Information Criterion. Methods in ecology and evolution 4:416–425.

Khabbazian, M., R. Kriebel, K. Rohe, and C. Ané. 2016. Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. Methods in Ecology and Evolution 7:811–824.

Levin, S. A., \*Helene C. Muller-Landau, \*Ran Nathan, and \*Jérôme Chave. 2003. The Ecology and Evolution of Seed Dispersal: A Theoretical Perspective. Annual Review of Ecology, Evolution, and Systematics 34:575–604.

Levins, R., and R. Lewontin. 1985. The dialectical biologist. Harvard University Press.

Maddison, W. P., P. E. Midford, S. P. Otto, and T. Oakley. 2007. Estimating a Binary Character’s Effect on Speciation and Extinction. Systematic Biology 56:701–710.

Mahler, D. L., T. Ingram, L. J. Revell, and J. B. Losos. 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. Science 341:292–295.

May, M. R., and B. R. Moore. 2020. A Bayesian Approach for Inferring the Impact of a Discrete Character on Rates of Continuous-Character Evolution in the Presence of Background-Rate Variation. Systematic Biology 69:530–544.

Middleton, N., and D. Thomas. 1997. *World atlas of desertification.. ed. 2*. Arnold, Hodder Headline, PLC.

O’Meara, B. 2008. Using Trees: Myrmecocystus Phylogeny and Character Evolution and New Methods for Investigating Trait Evolution and Species Delimitation (PhD Dissertation). Nature Precedings 1–1.

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

Rabosky, D. L., and E. E. Goldberg. 2015. Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation. Systematic Biology 64:340–355.

Schupp, E. W. 1993. Quantity, quality and the effectiveness of seed dispersal by animals. Vegetatio 107:15–29.

Steel, M., and D. Penny. 2000. Parsimony, Likelihood, and the Role of Models in Molecular Phylogenetics. Molecular Biology and Evolution 17:839–850.

Stevens, P. F., J. Luteyn, E. G. H. Oliver, T. L. Bell, E. A. Brown, R. K. Crowden, A. S. George, et al. 2004. Ericaceae. Pages 145–194 *in* K. Kubitzki, ed. Flowering Plants · Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales, The Families and Genera of Vascular Plants. Springer, Berlin, Heidelberg.

Toljagić, O., K. L. Voje, M. Matschiner, L. H. Liow, and T. F. Hansen. 2018. Millions of Years Behind: Slow Adaptation of Ruminants to Grasslands. Systematic Biology 67:145–157.

Uyeda, J. C., and L. J. Harmon. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Systematic biology 63:902–918.

Westoby, M., M. Leishman, and J. Lord. 1996. Comparative ecology of seed size and dispersal. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 351:1309–1318.

Yang, Z. 2006. Computational molecular evolution (Vol. 284). Oxford University Press Oxford.

Zucchini, W., I. L. MacDonald, and R. Langrock. 2017. Hidden Markov models for time series: an introduction using R. Chapman and Hall/CRC.

NOTES AND OTHER

Ornstein-Uhlenbeck models have been conceptualized and interpreted in many ways. Quantitative genetic interpretations are reliant on adaptation towards an optimum and the parameters of the model can have very clear evolutionary interpretations. They have been used in stock prices as models of deterministic stochasticity where a stock price is expected to reach a certain value, but with some uncertainty along the way. However, in macroevolution, perhaps because of our ties to microevolutionary theory, interpretations of the OU process have been focused on stabilizing selection and adaptation towards an optimum. Strictly speaking, this is not a correct interpretation. The work of Thomas Hansen and others has shown that the OU model when parameterized to macroevolutionary data does not match what is expected at the microevolutionary level. Thus, the OU model has begun to take on interpretations of simply phylogenetic signal. But here we may be throwing the baby out with the bath water. Phylogenetic signal exists because of descent with modification. Brownian motion at a macro-level can be seen as microevolutionary adaptation so long as we expect that all lineages phenotypic optima drift randomly on large timescales. OU models do not describe the phenotypic optima of a population evolving at the microevolutionary level. However, they can describe and give insights into the long-term trends of those microevolutionary dynamics. The question then becomes how do we want to describe changes in the dynamics? Typically, OU models are codified with hypotheses informed by intuition of scientists or ancestral state reconstruction. However, we do not know the ancestral state. We can only make a probabilistic statement about what it may be. Most methods until now have not allowed for uncertainty in this reconstruction. Here we describe these dynamics via a hidden Markov model. The microevolutionary interpretation of this model notwithstanding, this model is tractable and widely used. In truth, it may be more appropriate to study macroevolution in ways that are interpretable only in a macroevolutionary setting.

Imagine microevolutionary dynamics are in equilibrium on a macroevolutionary landscape (like a period of stasis as described by Gould and others). The dynamics are loosely described by the OU models in which drift around the phenotypic optima occurs so long as movement is not too far away. From this point how does change occur? There is abundant evidence that a change should occur given the theory of punctuated equilibrium. Most OU modeling is done such that the positions of change are *apriori* decided. This is not a probabilistic statement about where we think change is to occur, it is a statement of fact that a change did occur. However, we do not know where changes occur or what the ancestral state was for certain, and thus a probability statement is far more appropriate. A theory in-line with punctuated equilibrium would be one which allows for a sudden shift to a new microevolutionary landscape. Our model describes such sudden shifts using hidden Markov models.

Several phylogenetic comparative methods (PCMs) have been developed to address the potential of correlated evolution. As early as Pagel (1994), models of jointly evolving discrete characters were used to test for evidence of correlation between characters. Since then, these simple Markov models have been extended to allow for hidden-state variation, state-specific diversification dynamics, and generalized for any number of discrete states and characters (refs.). A similar set of extensions has been developed for continuous characters. There are now continuous trait model s which allow for multivariate character evolution, state-dependent diversification dynamics, and more complex modeling of the phenotypic dynamics (refs.). However, despite the widespread application of PCMs to study discrete and continuous characters, there are surprisingly few options which simultaneously use both classes of character (Felsenstein 2012).

One of the most important advances made by Tribble et al. (2021) was the development of an explicit null model. Their pipeline explicitly acknowledges the potential for spurious correlations between datasets even when the focal traits are unlinked. However, their methodology assumes that the parameters describing the non-focal discrete characters are described by the same maximum likelihood estimate of the observed discrete characters. There is little if any reason to suspect this. Instead it is theoretically possible to estimate hidden variation based solely on an continuously varying process. This is because the information for detecting hidden states comes from the differences in how the observed states change. If the transitions between observed states of different lineages are more adequately described by several Markov processes rather than a single process, there will be information to detect hidden states (Boyko and Beaulieu 2021). Since the state-dependent process of hOUwie is a joint model of discrete and continuous character we can parse whether differences in continuous or discrete variation are due to the influence of our observed characters ( and ) or hidden states.

Hidden Markov models have a hierarchical structure that can be broken down into two components: a “state-dependent process” and an unobserved “parameter process” (Zucchini et al. 2017). Under an HMM, observations are generated by a given state-dependent process, which in turn depends on the state of the parameter process. In other words, the observed data are the product of several processes occurring in different parts of a phylogeny and the parameter process is way of linking them. It is initially unknown what the parameter process corresponds to biologically, hence the moniker “hidden” state. Nevertheless, the information for detecting hidden states comes from the differences in how the observed states change. If the transitions between observed states of different lineages are more adequately described by several Markov processes rather than a single process, there will be information to detect hidden states (Boyko and Beaulieu 2021). Since the state-dependent process of hOUwie is a joint model of discrete and continuous character we can parse whether differences in continuous or discrete variation are due to the influence of our observed characters ( and ) or hidden states.

First, we expect that the climatic optima for abiotically dispersed seeds will be drier compared to biotically dispersed seeds (). Second, we expect that abiotically dispersed seeds will have faster rates of climatic niche evolution (). Finally, we expect that the climatic niches of biotically dispersed seeds will be more conserved through time ().

If we do not model joint probabilities explicitly, we are still able to test against null models of correlated evolution. The issue with this is that it does not leave room for discovery and null hypothesis testing has been criticized elsewhere extensively. Whether traits are correlated is an important initial step in understanding the evolution of syndromes but could also lead to false positives as we have seen in other comparative methods. This is accounted for in hOUwie because it includes entirely hidden state nulls as well as the more generic trivial null hypotheses present in traditional studies. But more exciting than testing null hypotheses is interpreting the parameters from resulting models. Hybrid models which mix variation due to focal and hidden traits open ways for identification of interesting regions of the phylogeny for future studies to assess.

The problem of heterogeneous character independent models leading to focal dependence is certainly troubling. Shift detection approaches may be well suited for this task. although they do not explicitly calculate the joint probability of the discrete mapping and continuous character and are thus biased in similar ways towards the discrete character. A future extension could combine the discrete character methods developed here with the discrete independent models for the best model.

These sorts of complex models are only appropriate for large phylogenetic datasets (>500 species) since there needs to be substantial heterogeneity in the processes. It is also possible to fit complex models if the signal is strong enough, however.

Screw nulls, who cares if traits are correlated? They may be or may not be, but we should be interested in the model parameters instead.

When we see variation outside of what’s expected, how should we interpret it? Our simulations unintentionally demonstrated the ability to detect optima outside of the range of observed phenotypes. Did we find this in seed dispersal? We did find that the sigma parameters are different from what we saw previously. How can we relate the discrete process to the continuous process? By looking at the waiting time and comparing that to the half-life perhaps we can gain new insights. It is also important to discuss the OU parameters and their interpretation. This has been widely discussed other places, so we will be brief. The dialectical nature of joint modeling.

It is possible and likely that not all the adaptive features of a hypothesis are going to be measured. This is where the utility of hidden Markov models is most apparent – not all things are going to evolve in the same way at all times. The evolution of a male coloration may be dependent on the existence of an underlying female preference (ref.), the evolution of seed dispersal mode may depend on the climatic conditions of the lineage (ref.), and the evolution of extra-floral nectaries may depend on hidden precursors (ref.). However, evolutionary dynamics are rarely as simple as X causes Y, and there is likely an interplay between two or more variables in how each evolves. It is possible that female preference evolved independently of male coloration and later drove the evolution of male coloration. This is the realm of causative modeling and is an interesting aspect and growing research agenda of many talented scientists (ref.). But finding causative hypotheses may be best achieved by identifying regions where the association between variables is not perfect. Once we can generate testable predictions, we can readily supplement comparative analyses with experimental and observation evidence.

Our simulations demonstrated the potential to discover phenotypic optima outside observed phenotypic values. Even when data is character dependent the character independent model with multiple hidden states can find a high level of support. This makes sense as hidden states could be assigned at tips to match the observed states (albeit with less certainty) and thus should approach the probability of the character dependent model. This serves as a reminder that although we are focusing on a single focal trait, any trait that matches the distribution at the tips would produce identical results and we should be cautious with over interpretation.

Rates of evolution have changed throughout the history of life and produced the rich diversity of morphology, behaviour, and ecology that characterizes the biodiversity we see today. As such, models favoring variation in rates and states has fast become the rule, as opposed to the rare exception in empirical applications of comparative methods. The challenge then becomes how to model this heterogeneous process with the limitation that, for most comparative biologists, our data comes exclusively from extant species. Our model, which we call hOUwie, uses hidden Markov models as a statistically rigorous way to learn about variation in the evolutionary process. hOUwie is composed of two processes: one describes the evolution of a discrete character (such as whether a mammal is a carnivore, herbivore, or omnivore) and the other the evolution of a continuous character (such as mammalian body size). To model the evolution of our continuous character we have chosen to use an Ornstein-Uhlenbeck (OU) model. In evolutionary biology, we use it to model the change in traits through time, which is the product of two antagonistic forces continuously pulling trait values towards some optimal value and randomly moving away from the optimum. The dynamics of these forces can also vary by so-called “regimes”. Under a typical OU model, these underlying regimes are taken as a given and are defined a priori leaving no room for inference about the evolution of the regimes themselves. To resolve this problem, we will model the evolution of these regimes as a discrete character via the aforementioned hidden Markov model. One of the most important properties we will utilize is the inclusion of hidden states. Hidden states give us a way to learn about the variation in the evolutionary process without the necessity of it being linked to one of our explanatory variables.

Generally, organisms will track their preferred environmental conditions and climatic variables are thought to be some of the most important components of the abiotic environment (Donog 2008; Donog and Edw 2013, ref.). However, whether seed dispersal occurs via abiotic or biotic means could influence how species track their environmental conditions and their potential to adapt to changing environmental conditions (ref.). This suggests that abiotically dispersed seeds may have higher rates of climatic niche evolution, while biotically dispersed seeds are more likely to have conserved climatic niches.

As datasets have begun to expand, modelers have been given the luxury of increasing the complexity and biological realism of phylogenetic comparative methods available for use. There are a diverse set of phylogenetic comparative methods which can model the long-term evolution of phenotypes, whether they are discrete or continuous. These methods have seen several extensions and improvements including the addition of hidden-state variation, state-specific diversification dynamics, and multivariate evolutionary dynamics (refs.). However, despite widespread interest in modeling discrete and continuous characters, there are surprisingly few options available which simultaneously use information from both classes of character (Felsenstein 2012).

One underexplored option is the threshold model originally proposed by Wright (1934) and developed by Felsenstein (2012). This model proposes that an unobserved continuous variable underlies an observed discrete variable, but this method has not seen much use (Maddison and FitzJohn 2014). The more common approach to linking discrete and continuous characters is described by Revell (2013). With this approach, a class of Markov model is used to maximize the likelihood of the parameters associated with a discrete character (ref.). Then stochastic mapping (refs.) are used to generate several regime paintings which are simulated under the maximum likelihood estimate of the discrete parameters. Finally, a model of continuous evolution is fit in which the continuous model parameters are allowed to vary based on the underlying discrete regime. The number of stochastic mappings and how they should be summarized has received little attention.

For example, seed dispersal is often broadly classified into an abiotic or biotic mode (ref.). Our expectation is that biotically dispersed seeds will tend to move in a more predictable fashion, dispersing to environments more like their parents while abiotically dispersed seeds will be less predictable and lineages will experience a greater variety of environmental conditions through time (ref.). We might therefore expect that the evolution of the climatic niche of a lineage depends on its mode of seed dispersal. However, the evolution of seed dispersal model also changes through time depending on environmental conditions. It has been proposed that adaptations for frugivorous dispersal is linked to tropical and subtropical biomes (ref.). This is because in these warmer and wetter habitats, large trees create shady environments where competition for light is more important. A shadier habitat then imposes a selective pressure for larger seeds because more nutrients are needed for germination (ref.). However, the evolution of larger seeds comes with a tradeoff as they have a significantly lower dispersal potential (ref.). Thus, we might expect that the climatic variables of a habitat influence the probability of transitioning between abiotic and biotic modes of dispersal, with transition rates from abiotic to biotic being greater in less arid environments.

Finally, it is interesting to note that our study has found that the probability of a discrete character mapping is roughly correlated with the probability of the continuous trait given model parameters. This is important for the hOUwie model because of the simulation-based approach taken, but also has implications for other models currently being used and developed. Previous attempts to jointly model discrete and continuous characters have relied on stochastic maps generated by the MLE of discrete characters. Had the probability of the discrete characters as defined by the MLE been decoupled from the total joint probability of discrete and continuous characters, estimates related to the continuous character would have been biased towards an explanation of the discrete character. As we have shown here, albeit with a limited set of simulations, the probabilities of discrete character mappings and continuous characters are correlated. Thus, high probability mappings can be expected to produce high probability continuous characters when the two are coevolving. These findings suggest, that although there will be slight differences between the joint estimate developed in hOUwie and methods currently being used (described by Revell 2013), most previously published results should remain relatively robust. This also suggests a diminishing return on increasing the number of stochastic maps when conducting a study. Depending on the parameterization of the discrete model, increasing the number of stochastic maps may mean several repetitions of high probability mappings. We suspect this is why increasing the number of stochastic maps in the hOUwie simulations had little influence on the accuracy of the model. Of course, this is not always the case and there are certain discrete parameters (such as high transition rates) where there are many equally probably maps. Increasing the number of stochastic mappings would be useful in these cases.