**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**

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.

**Introduction**

The two most common approaches for applying OU process with multiple , , and/or have been to either assume selective regimes a priori (e.g. Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012) or to infer the underling regimes based on the distribution of the continuous trait often called “shift-detection” (e.g. Ingram and Mahler 2013; Uyeda and Harmon 2014; Khabbazian et al. 2016). With the former approach, regime paintings are often generated through stochastic mapping, a form of ancestral state reconstruction (ASR) (Revell 2013). The advantage of this approach is that there is an explicit model for how regimes change through time. However, the evolution of these regimes is 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 which fits the discrete data best generates stochastic maps which do not fit the continuous data. Under this approach, although we have explicitly considered how regimes change, we have not considered whether the changes are appropriate for the continuous trait. The second approach of detecting shifts in underlying regimes can consider both the discrete and continuous character by evaluating the likelihood of the continuous character given a regime painting which matches the discrete data. These methods work quite well at detecting shifts although a modified BIC should be used to avoid overfitting (Khabbazian et al. 2016). The shift detection approach can be taken a step further, as in the case of Uyeda et al. (2018), who modeled a species’ underlying regime as the weighted average of regimes informed by shifts in the discrete character and regimes informed by the continuous data alone. However, this approach will suffer from the opposite problem of the ancestral state reconstruction approach. Although we will have explicitly considered the probability of our continuous data given regimes (which may or may not be informed by discrete character data), the probability of the regimes themselves is not explicitly considered. This could lead to scenarios where regimes are well suited for the continuous trait distribution, but poorly describe the discrete character.

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.

Additionally, we have allowed for the possibility that regime changes follow a hidden Markov model (Felsenstein and Churchill 1996; Yang 1994; Beaulieu et al. 2013). 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.

Previous work discussing character dependence and independence has focused on whether a focal trait influences diversification rate. Less attention has been placed on the potential biases associated with pure phenotypic evolution.

How organisms move through space is of tremendous consequence to the fitness of individuals and the species composition of entire communities (ref.). 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. 2013). In part, this means that understanding the movement of plants through space can be achieved through a better understanding of how seed dispersal has changed through time. Furthermore, the mechanism of seed dispersal is often broadly classified into abiotic or biotic modes of seed dispersal. To some extent, these can be inferred from features of fruit morphology where fleshy fruits are biotically dispersed and dry fruits are abiotically dispersed (ref.). However, this is clearly an oversimplification of a complex phenotype that can be more aptly described as a syndrome (ref.).

Furthermore, beyond the complexity of the dispersal syndrome phenotype, how the mode of seed dispersal changes through time will depend on the environmental conditions of the lineage. 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.). Generally, the expectation is that seeds dispersed by frugivores are going to be dispersed to environments more like their parents’ environment (ref.), whereas abiotically dispersed seeds are likely to be more erratic in their dispersal patterns (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. 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.

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).

**Methods**

The hOUwie model

The hOUwie model is composed of two processes: one describes the evolution of a discrete character and the other 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*). 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 adaptive optimum. In this model, a trait () is pulled towards an optimum at a rate scaled by the parameter , while the optimum itself (which may change through time) is denoted by the parameter . is piecewise constant on intervals and takes values in a finite set. This represents the set of “selective regimes”, “regimes”, or Simpson’s “adaptive zones” (Cressler et al. 2015). Additionally, random deviations are introduced by Gaussian white noise , which is distributed as a normal random variable with mean zero and variance one. The magnitude of these deviations is scaled by the noise intensity . has been interpreted as the strength of selection (Simpson 1953, Lande 1976, 1980) and has been referred to as genetic drift (Lande 1976, Hansen 1997). However, both interpretations have been criticized. The former because may be more aptly described as a measure of phylogenetic signal (Ho and Ané 2014*a*) and the latter because the stochasticity modeled by an OU on a macroevolutionary time scale is likely not a consequence of genetic drift (Hansen 2012). Formally, the OU process is an Itô diffusion satisfying:

Here, 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 noise intensity () 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),

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 the function three.point.compute from the R-package phylolm (Ho et al. 2016) 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 () (similar to the pathway likeilhood in Steel and Penny 2000). 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 an equal rates model (where the maximum likelihood estimate is zero changes ref.), any path of 0, 2, or 100 changes has the same likelihood. 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 and the total branch length available. The convolution of these independent random variables is given by an Erlang distribution (Felsenstein 1978). Ultimately, the effect of this inaccuracy is that the number of transitions has no influence on the probability of a branch 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 Appendices A and B.

Instead, 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, , and 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 a discrete state to be defined along the entire branch, thus we place transitions halfway between nodes and internodes.

Our likelihood function for the joint probability of continuous and discrete character is:

Possible hOUwie models

Our simulation studies examine 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 model identifiability, we constrained transitions between hidden states to be the same for observed states, but this constraint can be relaxed if desired by the user. 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 OU parameter differs between observed state, while character independent models test whether observed states can be described by the same 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 (CID). 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.

Figure 1: A state-transition diagram describing the examined hOUwie model structures.

A picture containing bubble chart

Description automatically generated

Joint probability for all possible mappings

Ideally, to evaluate the likelihood of a set of parameters we would marginalize the probability of the discrete and continuous characters for all possible stochastic mappings. This is computationally infeasible for even small phylogenies and thus our approach samples stochastic mappings proportional to the discrete character probabilities. However, 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. We expect a calculation of joint probabilities to be important in cases where the continuous character and discrete character probabilities are not perfectly correlated. In these cases, the highest overall likelihood may not be the highest individual combination of the marginal probabilities of the discrete and continuous character. To examine this assumption we create a pectinate phylogeny with four branches and a height of one, is matched to all possible binary discrete variable datasets. In total, there are 16 combinations of a binary variable for a four-taxon tree, but only 14 when invariant datasets are excluded. For each dataset we assign a continuous trait value that can be considered character-dependent by assigning discrete state 0 a continuous value of 5 and discrete state 1 a continuous value of 10. This dataset should be consistent with an expectation of joint and marginal probabilities differing. We can then examine the contributions of each stochastic map to the overall likelihood, as well as the contributions of the discrete and continuous probabilities to the joint probability by plotting the correlation between the discrete and continuous probabilities.

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. 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.

Next, to evaluate model power, 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 use Akaike Information Criterion (AIC) to determine whether the best fitting model is the same as the generating model (Burnham and Anderson 2002). 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.

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. For example, there are some fruits which are technically classified as dry but are consumed and dispersed by frugivores (e.g., *Gaultheria* in Ericaceae; Stevens chapter). If the evolutionary dynamics of endozoochorous dry fruits are not adequately described by being lumped with other dry fruits, then it is important to allow for hidden variation to be incorporated into our evolutionary model.

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 in our previous study we had found two counter-intuitive results. First, we found that the phenotypic optima of abiotically dispersed seeds was more humid than biotically dispersed seeds. Second, we found that the rate of climatic evolution was greater in biotically dispersed seeds than abiotically dispersed seeds. Additionally, our previous work assumed either character dependence or simple character independence. As our simulations will show, 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 run 27 hOUwie models: 2 CID, 10 CD, 10 CID+, and 5 HYB. As we alluded to previously, *Gaultheria* is technically a dry-fruited genus within Ericaceae but has a persistent fleshy calyx that attracts frugivores. 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**

The joint probability for all possible mappings

Figure 2a presents 1 of the 14 non-invariant datasets for a binary discrete character as well as the contribution of the discrete and continuous probabilities to the overall likelihood of the mapping. The total likelihood for the set of parameters given the dataset would be found by summing the log likelihoods of all mappings. Figure 2 demonstrates a scenario where mappings 5 to 8 contribute more to the overall likelihood than 1 to 4, however this is not always the case and will depend on the distribution of the discrete data (Fig S1). For all possible discrete datasets, we find that discrete and continuous likelihoods are correlated with one another (Fig. 2b). This suggests that sampling based on the discrete weighted probabilities is an efficient way to find high probability mappings and calculate the total likelihood. However, although discrete probability is a good predictor of the overall likelihood contribution, the continuous character still contributed unique information which can be seen in mappings with lower discrete probabilities that have equal or greater overall likelihoods (Fig. 2b).

Diagram

Description automatically generated

Figure 2: A) All possible mappings for 1 of 14 discrete binary character datasets. For three internal nodes there are 8 possible mappings each of which will have a unique discrete and continuous likelihood for a set of parameters. The overall likelihood for this dataset is \_\_\_ which is found by summing the log likelihood of all possible mappings. Each bar below a mapping represents the overall likelihood of a mapping which can be decomposed into the discrete (orange) and continuous (green) contributions. B) Each point represents a single mapping for one of the 14 possible datasets. Points are numbered based on their corresponding map. A simple linear regression is plotted.

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 which parameters are 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). It is possible that increasing the number of taxa could improve inference of these models, but previous work (Beaulieu et al. 2012; Ho and Ané 2014*a*; Cressler et al. 2015) suggests that alpha estimation is often a consequence of the strength of signal relative to the noise in the data. Thus, even with more taxa, if alpha is not large enough relative to sigma, estimation may remain poor. 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 always below 0.05, 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 each other and have identical likelihoods. For example, if we were to simulate a dataset where hidden state A has a /theta value of 12, and hidden state B has a /theta value of 24, there would be no difference in likelihood if /theta\_A=12 & /theta\_B=24 or /theta\_A=24 & /theta\_B=12. This is also referred to as the \_\_\_\_ problem in hidden Markov literature and is an active area of research (ref.).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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*

Despite variable theta models having the lowest sign errors and highest successful detection of character dependent models, they also had some of the highest RMSE for parameter estimates. In part, this can be explained by the magnitude of the parameter. Because the simulated optima values range from 12 to 24, errors will tend to be higher compared to sigma or alpha which range between 0.5 and 3. Nonetheless, the high error rates are concerning and warrant further analysis. 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 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 or simply inaccurate. 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). Suggesting that hOUwie can estimate phenotypic optimum outside of the observed phenotypic distributions. This is likely to be true of most OU models, but hOUwie has the advantage of utilizing information from the discrete transition rates in addition to the distribution of regimes at the tips. We are also able to examine the expected continuous trait values according to the fitted model and see they are similar to the observed phenotypic values (Fig. 3c). This suggests that the error in the estimation in the optimum is meaningful, and likely a consequence of a lack of information around the optimum because of a lack of observations around the true optimum phenotype. ![Chart

Description automatically generated]()The power to detect the generating model

Simple character independent (CID) and character dependent (CD) models preformed the best in our power analyses. 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. This reenforces 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. 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 (Table 1).

Seed dispersal and climatic evolution

Consistent with previous findings we find little evidence of hidden Markov models, but we do find that dispersal mode influenced rates of climatic evolution. Figure 4.

**Discussion**

Phylogenetic comparative methods have been widely applied to study discrete and continuous characters separately, but there are surprisingly few options which can 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 allowing information from the discrete and continuous characters to jointly influence the regime paintings and contribute to the overall likelihood.

The importance of jointly modeling characters

An alternative conception of joint probabilities as mutual information may be useful here. The combination of two independent marginal probability will equal the joint probability if the variables contain no information about one another. If knowing the value of a particular variable (e.g., whether your roommate is wearing a raincoat or a t-shirt) provides information about another variable (e.g., whether it is raining or sunny outside) then the joint probability will differ from the combination of the marginal probabilities. For phenotypic evolution we often suspect this is the case, particularly when dealing with trait syndromes (ref.).

It is important to view the correlation between two characters outside the context of simply testing null hypotheses. If two traits are correlated that is certainty an important finding but discovering the nature of that correlation is what is most intriguing from an evolutionary perspective.

Why should we care about jointly modeling discrete and continuous traits? Firstly, this explicitly allows for the traits to talk with each other – a concept formally referred to as mutual information. The explicit modeling of two or more characters is not only biologically more relevant but can improve our modeling inferences. The quality of our modeling inferences also has a direct consequence on how accurate our hypotheses are. For example, if one is interested in testing the presence of convergent evolution of pollination syndromes. The first step to establishing a syndrome is to test for the convergence of the traits involved. This requires that we examine whether the traits have independently associated multiple times through the history of the clade. Modeling discrete characters independently can alter the certainty and the states of ancestral state reconstruction and can completely alter how we perceive the evolution of convergence. For example, if we are interested in how seed size has evolved alongside climate, reconstructing seed size independently may lead to different conclusions than if we modeled them jointly. This is because the distribution of the discrete trait alone does not utilize the information that is shared between climate and seed size. This relationship allows for a more accurate ancestral state reconstruction and thus a more accurate answer to the question of convergent evolution. The flipside of this is if the traits are perfectly correlated then there is not going to be any difference between independently modeling and jointly modeling. This is because change in one variable is matched perfectly by change in the other variable. Here the advantage of joint modeling is simply in the certainty provided in the reconstruction. If such a strong relationship was detected, the uncertainty that would normally be accumulated by independently modeling the variables will be reduced because we now have multiple sources of information suggesting a particular ancestral state.

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. 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. 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 demonstrated 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.

It is also important to discuss the OU parameters and their interpretation. This has been widely discussed other places, so we will be brief.

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 just 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 (ref.). However, if implemented in a phylogenetic context, it would completely remove the need for a simulation approach – a common feature of most current attempts at jointly modeling discrete and continuous characters (refs.).

Another potential avenue of improvement would be to take an expectation maximization (EM) approach (ref.). In this case, discrete mappings could be treated as an unobserved latent variable. In hOUwie, the distribution of the discrete mappings is assumed to follow the probability distribution of the discrete tip states given a set of rate parameters. However, the distribution of the mappings would ideally be a combination of the continuous and discrete character. Under an EM algorithm, the latent mapping variable would be drawn from a fixed number of possible mappings and then the likelihood of the hOUwie parameters would be maximized. This would be repeated iteratively until convergence is achieved. One challenge to this approach would be designing an efficient way to draw potential latent mappings such that exhaustive search is unnecessary.

It is also possible for the methods described here to be applied to State-dependent speciation and extinction (SSE) models. Currently methods available to the linking of continuous characters to speciation and extinction dynamics are dubious and slow. This extension would require a different calculation of the underlying regime mapping probability but would be relatively straightforward. A more challenging aspect of this extension would be generating high joint probability mappings. Although approaches for simulating SSE models exist (ref.), it may be best to use an approach such as EM or jointly sampling a posterior in a Bayesian setting.

Discovering hidden variation

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.

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.

Often we are biased for looking for adaptation in traits that are easily measured, but that does not mean there are not other important aspects of a species’ natural history which is the true adaptive feature. It is possible and likely that not all of the adaptive features of any particular hypothesis are going to be measured. This is where the utility of hidden Markov models is most apparent. It is possible that in some lineages color is the most important feature and in others things other than color are more important. By identifying regions where the association between color and foraging behavior do not match we are able to generate testable predictions. And highlight areas where we can readily supplement comparative analyses with experimental and observation evidence. The advantage is that we generate hypotheses and outline where in the phylogeny we can go looking.

Another advantage of modeling joint probabilities explicitly is that it allows us to put our models in the context of likelihood. With an explicit likelihood we are able to do model comparison and model averaging. Rather than testing whether our pet hypothesis is better than a trivial null model we can discover the relative power of each model to describe the variation in our dataset. Then we can use that information to inform our model averaging and examine the parameter estimates from our set of models holistically. The upside of this is that we are able to account for model structure uncertainty in addition to parameter estimation uncertainty. Ultimately the use of modeling in general is to describe our data in ways that we cannot do using summary statistics. In part, this means hypothesis testing and those are important steps. However, parameter estimates are comparable across studies and why support for hypotheses is important, it is reminiscent of vote counting in which we examine blankly whether a hypothesis was supported. Instead, parameter estimates are similar to examining effect sizes and are often directly comparable across studies since they are in the same units.

The problem of heterogeneous character independent models leading to foal 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.

Consider a situation in which one wanted to compare OU parameters across multiple studies. It would be exceedingly difficult to make parameter comparisons when the OU parameters themselves are linked to mappings of the chose author. One would have to ask the question whether it was the mapping or the focal trait that influenced parameter estimates.

Conclusion