A General Model for Estimating Macroevolutionary Landscapes

Florian C. Boucher[1,2,*], Vincent Démery[3], Elena Conti[1], Luke J. Harmon[4,5], and Josef Uyeda[4]

[1]*Department of Systematic and Evolutionary Botany (ISEB), University of Zurich, Zurich, Switzerland;* [2]*Department of Botany and Zoology, University of Stellenbosch, Stellenbosch, South Africa;* [3]*Gulliver, CNRS, ESPCI Paris, PSL Research University, 10 rue Vauquelin, Paris, France;*
[4]*Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, Moscow, ID, USA; and*
[5]*Department of Fish Ecology and Evolution, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Center for Ecology, Evolution, and Biogeochemistry, 6047 Kastanienbaum, Switzerland*
Correspondence to be sent to: Department of Botany and Zoology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa;
E-mail: floflobucher@gmail.com.

Abstract.—The evolution of quantitative characters over long timescales is often studied using stochastic diffusion models. The current toolbox available to students of macroevolution is however limited to two main models: Brownian motion and the Ornstein–Uhlenbeck process, plus some of their extensions. Here, we present a very general model for inferring the dynamics of quantitative characters evolving under both random diffusion and deterministic forces of any possible shape and strength, which can accommodate interesting evolutionary scenarios like directional trends, disruptive selection, or macroevolutionary landscapes with multiple peaks. This model is based on a general partial differential equation widely used in statistical mechanics: the Fokker–Planck equation, also known in population genetics as the Kolmogorov forward equation. We thus call the model FPK, for Fokker–Planck–Kolmogorov. We first explain how this model can be used to describe macroevolutionary landscapes over which quantitative traits evolve and, more importantly, we detail how it can be fitted to empirical data. Using simulations, we show that the model has good behavior both in terms of discrimination from alternative models and in terms of parameter inference. We provide R code to fit the model to empirical data using either maximum-likelihood or Bayesian estimation, and illustrate the use of this code with two empirical examples of body mass evolution in mammals. FPK should greatly expand the set of macroevolutionary scenarios that can be studied since it opens the way to estimating macroevolutionary landscapes of any conceivable shape. [Adaptation; bounds; diffusion; FPK model; macroevolution; maximum-likelihood estimation; MCMC methods; phylogenetic comparative data; selection.]

Understanding the evolution of phenotypes over geological timescales is one of the fundamental goals of macroevolution (Simpson, 1944). Phenotypic evolution is typically inferred either from time series of measurements obtained in the fossil record (Hunt, 2007) or from the distribution of phenotypic characters at the tips of a phylogenetic tree (O'Meara, 2012). In both cases, one then fits stochastic models for the evolution of traits on single lineages or on phylogenies, all of which treat trait evolution as a diffusion process that may or may not be influenced by deterministic forces (O'Meara, 2012).

Many approaches attempt to bridge the gap between microevolutionary process and macroevolutionary pattern by interpreting model parameters in the context of the dynamics of evolution on adaptive landscapes (Wright, 1932; Simpson, 1944; Arnold et al., 2001; Arnold, 2014). Recent years have revitalized this connection with the development of numerous methodological tools specifically aimed at inferring "macroevolutionary landscapes" (Hansen et al., 2008; Eastman et al., 2013; Ingram and Mahler, 2013; Uyeda and Harmon, 2014). Such landscapes almost certainly do not reflect static landscapes upon which populations evolve over long time scales; instead, these landscapes are most productively described as representing the movements of adaptive peaks over million-year time scales (Hansen, 1997; Uyeda et al., 2011; Uyeda and Harmon, 2014). In particular, a peak on such a landscape might not be a phenotypic optimum in any particular generation of evolution of a lineage, but instead might represent a long-term average

peak location on a dynamic landscape. Throughout this article, we will refer to these as "macroevolutionary landscapes," which summarize patterns of trait evolution averaged over many generations.

Comparative methods to infer macroevolutionary landscapes are all based on the Ornstein–Uhlenbeck (OU) process (Hansen, 1997), which was itself strongly inspired by the original concept of adaptive landscape in population genetics (Lande, 1976). Under the OU process, a continuous trait evolves under both random diffusion (i.e., Brownian motion, Edwards and Cavalli-Sforza, 1964), and a force that brings back the trait close to an optimal value. Following models from quantitative genetics (e.g., Lande, 1976), these two components of the macroevolutionary OU process are sometimes interpreted as genetic drift and stabilizing selection. However, such an interpretation is almost always overly simplistic. First, many other processes can generate evolution following an OU model. For example, the shape of the peak and, in turn, the dynamics of selection and drift within populations may be less important for long-term patterns than the movement of the peak itself. Under such a scenario, both the diffusion and deterministic components of OU reflect peak movement, and both are strongly dependent on the dynamics of both selection and drift. Second, the actual parameters of OU models are almost always incompatible with Lande's model of evolution on a static adaptive landscape (Estes and Arnold, 2007; Uyeda and Harmon, 2014). However, even if we do not interpret

diffusion as *drift* and *determinism* as *selection*, it is still useful to divide macroevolutionary dynamics into these two components. Any factor leading to trait change that is random in direction from one generation to the next (e.g., drift, randomly varying selection, plasticity due to random environmental noise) will affect the diffusion component, and any factor that leads to predictable change towards some particular value (e.g., selection, predictable patterns of peak movement over time, developmental constraints towards certain values) will be seen in the deterministic component. Various extensions of the OU model have been proposed in recent years, including different optima in different clades, either determined *a priori* (Hansen, 1997; Butler and King, 2004) or not (Ingram and Mahler, 2013; Uyeda and Harmon, 2014), varying rates of diffusion and intensities of attraction towards optima in different clades (Beaulieu et al., 2012), evolution of the optimum itself (Hansen et al., 2008), or the possibility to study multivariate evolution (Bartoszek et al., 2012). The complexity of all of these extensions, however, leads to difficulties in model identifiability and parameter estimation (e.g., Khabbazian et al., 2016). While these models cover a wide range of possible scenarios, they are restricted to two main kinds of macroevolutionary landscapes: (i) macroevolutionary landscapes with a single peak continuously moving in time (Hansen et al., 2008) and (ii) macroevolutionary landscapes with one or several peaks, eventually of varying heights (varying attraction strengths) and widths (varying ratios of diffusion rate to attraction strength). Importantly, in the second case the different peaks are experienced at different epochs or by different lineages, so that no single lineage evolves in an macroevolutionary landscape with multiple peaks. In addition, another type of macroevolutionary landscape might be described by the bounded Brownian motion model (BBM, Boucher and Démery, 2016). BBM was developed as a model of neutral evolution between bounds, but it could also describe macroevolutionary landscapes in which one part of phenotypic space (i.e., traits values between the bounds) has high but constant fitness while other phenotypes have null fitness, a scenario related to holey adaptive landscapes (Gavrilets, 1997).

As can be seen from this short overview of existing methods, the types of macroevolutionary landscapes that can be estimated from comparative data are still rather limited. For example, disruptive selection is central to the theories of ecological speciation (Doebeli, 1996; Nosil, 2012) and adaptive radiation (Schluter, 2000), in which diverging lineages adapt to different ecological niches. These theories are being increasingly appreciated in the macroevolution community and ecological speciation and/or adaptive radiation are frequently invoked as explanations for the diversity of extant clades (Soulebeau et al., 2015). However, despite recent theoretical advances in modeling interspecific interactions over macroevolutionary timescales (e.g., Nuismer and Harmon, 2015; Drury et al., 2016), we still lack proper tools to infer disruptive selection on

phylogenies since macroevolutionary landscapes that contain multiple local optima cannot be inferred from phylogenetic comparative data. The current alternative is to model phenotypic evolution in multimodal macroevolutionary landscapes using OU models with multiple optima (Butler and King, 2004; Uyeda and Harmon, 2014). In this framework, cases in which transitions between peaks are frequent can be interpreted as evidence of a change in the adaptive environment, rather than the existence of multiple, simultaneously existing peaks among which lineages alternate. Such a model in which several adaptive peaks are simultaneously present in the adaptive landscape experienced by all species in the clade would be a step towards a more explicit model for disruptive selection.

In this article, we introduce a general model for the evolution of continuous characters on phylogenies that can accommodate macroevolutionary landscapes of any shape and thus attempts to provide solutions to the limitations mentioned above. In this model, the continuous trait of interest evolves under random diffusion but is also subject to deterministic change following a macroevolutionary landscape of any possible shape and strength. This model will be especially useful in situations where one thinks that all members of a clade have experienced the same macroevolutionary landscape throughout their history, in contrast to situations in which lineages shift between alternative macroevolutionary landscapes. We present algorithms for both maximum-likelihood estimation (MLE) and Bayesian estimation of model parameters, that is the value of the trait at the root of the tree, the diffusion rate and the shape of the macroevolutionary landscape. Using simulations, we show that this model is easily distinguishable from other models of trait evolution like BM, OU, and BBM. Parameter estimation is also generally reliable, and in particular the shape of the macroevolutionary landscape can be accurately estimated as long as it has been fully explored by the clade evolving on it. We also show how alternative hypotheses can be statistically tested in empirical data sets. Our approach opens new avenues for macroevolutionary research: it renders possible the detection of evolutionary trends from neontological data only, but also inference of disruptive selection or of even more complex scenarios in which the macroevolutionary landscape contains multiple peaks.

DEVELOPMENT AND IMPLEMENTATION OF THE MODEL

General Presentation of the Fokker–Planck–Kolmogorov Model

We introduce a general model for the evolution of continuous traits on phylogenies, in which a trait x undergoes a random walk (i.e., BM) that is biased by a deterministic force that can be of any shape and strength.

The force biasing trait evolution derives from a potential $V(x)$: differences in the values of the potential generate a force that attracts the trait towards regions of trait space in which the potential is the lowest and at each point x, the BM process is biased by a force proportional to $-V'(x)$. This process can be modeled using the Fokker–Planck equation, a partial differential equation widely used in statistical mechanics to describe the time evolution of the probability density of an observable under the influence of both random and deterministic forces (Risken, 1984). In population genetics, the Fokker–Planck equation has been used to model changes in allele frequencies and is better known as the Kolmogorov forward equation (Wright, 1945). This is why we label the present model Fokker—Planck—Kolmogorov (FPK). Under FPK, the probability density $p(x, x_0, t)$ of the position of a trait x with initial value x_0 after time t has elapsed follows

$$\frac{\partial p}{\partial t}(x, x_0, t) = \frac{\sigma^2}{2} \frac{\partial}{\partial x}\left[\frac{\partial p}{\partial x}(x, x_0, t) + p(x, x_0, t)\frac{\partial V}{\partial x}(x)\right]. \quad (1)$$

In this equation, the evolution of the probability density of the trait (left hand side) is determined by both random diffusion (i.e., BM; first term on the right hand side) and a deterministic force set by the derivative of the potential (second term on the right hand side). The factor $\sigma^2/2$ on the front of the second term in the right hand side is chosen so that the stationary distribution for the probability density is

$$p^*(x) = \lim_{t \to \infty} p(x, x_0, t) = \mathcal{N}\exp(-V(x)), \quad (2)$$

where $\mathcal{N}$ is a normalization factor. Finally, the initial position of the trait, x_0, gives the initial condition for this partial differential equation: $p(x, x_0, 0) = \delta(x - x_0)$ where $\delta(x)$ is the Dirac delta function.

In summary, the potential $V(x)$ determines the force $-\sigma^2 V'(x)/2$ that is exerted on the trait over the interval, and the process has a stationary distribution, which is proportional to $\exp(-V(x))$ but does not depend on σ^2. The force represents the deterministic component of trait evolution, since it pulls traits towards specific values and the stationary distribution can be interpreted as a macroevolutionary landscape because trait values are attracted towards regions of trait space with the lowest potential, which themselves corresponds to peaks of $\exp(-V(x))$. Figure 1 shows how the potential, the deterministic force, and the macroevolutionary landscape are related. In the remainder of this article, we will use the term *macroevolutionary landscape* to refer to $\mathcal{N}\exp(-V(x))$ and will avoid mentioning the potential ($V(x)$) as much as possible. It is important to note here that the evolutionary rate, σ^2, is not a measure of the strength of the random component of the process. Indeed, σ^2 determines both the intensity of the random diffusion component (first term in the right side of Eq. 1) and of the deterministic force exerted on the trait (second term in the right side of Eq. 1). The relative strengths of the random and deterministic components of the process are better

captured by the variations in the stationary distribution of the process, $p^*_{\max}/p^*_{\min} = \exp(V_{\max} - V_{\min})$: a ratio close to one means that diffusion dominates, while deterministic forces are important if this ratio is large.

We can also define a characteristic time T_c representing the time it takes for the trait to explore the macroevolutionary landscape. Supplementary Appendix I available on Dryad at http://dx.doi.org/10.5061/dryad.28g37 shows how this characteristic time can be calculated. Over short time periods, that is for $t \ll T_c$, the trait has no time to fully explore the macroevolutionary landscape. Over long time periods, that is for $t \gg T_c$, the trait explores the whole macroevolutionary landscape; the probability density thus reaches its stationary distribution (Eq. 2). An important feature of the characteristic time is that a low value $p^*_{\min}$ of the macroevolutionary landscape between two peaks ($p^*_{\max} \gg p^*_{\min}$) can considerably slow down the exploration of the landscape; in this situation the characteristic time follows the Arrhenius law: $T_c \sim p^*_{\max}/p^*_{\min}$ (Gardiner, 1985). If there is a single peak in the macroevolutionary landscape, like in the OU process, the characteristic time is proportional to the phylogenetic half-life of the process (Supplementary Appendix I available on Dryad).

Calculation of the Likelihood

The likelihood of the FPK model given a phylogenetic tree and observed values of the trait at the tips of the tree is obtained by multiplying the probability densities along each branch of the tree and integrating over all possible trait values at the internal nodes:

$$\mathcal{L} = \int\left(\prod_{i \in I \cup T} p(x_i, x_{\text{parent}(i)}, t_i - t_{\text{parent}(i)})\right)\prod_{i \in I} dx_i, \quad (3)$$

where I is the set of internal nodes (excluding the root), T is the set of tips, x_i is the value of the trait at the node i, t_i is the time at node i, and parent(i) is the parent of the node i. Computing the likelihood numerically thus requires integrating over all possible values of the trait at interior nodes of the tree, which makes it computationally challenging. Since the distribution of the trait at the tips of the tree is often not multivariate normal, fast methods like Generalized Least-Squares (Grafen, 1989), phylogenetic independent contrasts (Felsenstein, 1985; Freckleton, 2012), or the 3-point algorithm (Ho and Ané, 2014a) cannot be used either.

To compute the likelihood of FPK, we instead discretize the trait interval by considering only a set of n points equally spaced between two extreme values, $B_{\min}$ and $B_{\max}$, a procedure already used for the BBM model (Boucher and Démery, 2016). In the following of this article, we call this regular set of points the *grid*. Supplementary Appendix II available on Dryad shows how the continuous evolution equation for the probability density (1) can then be cast in a matrix form and that these discretized equations
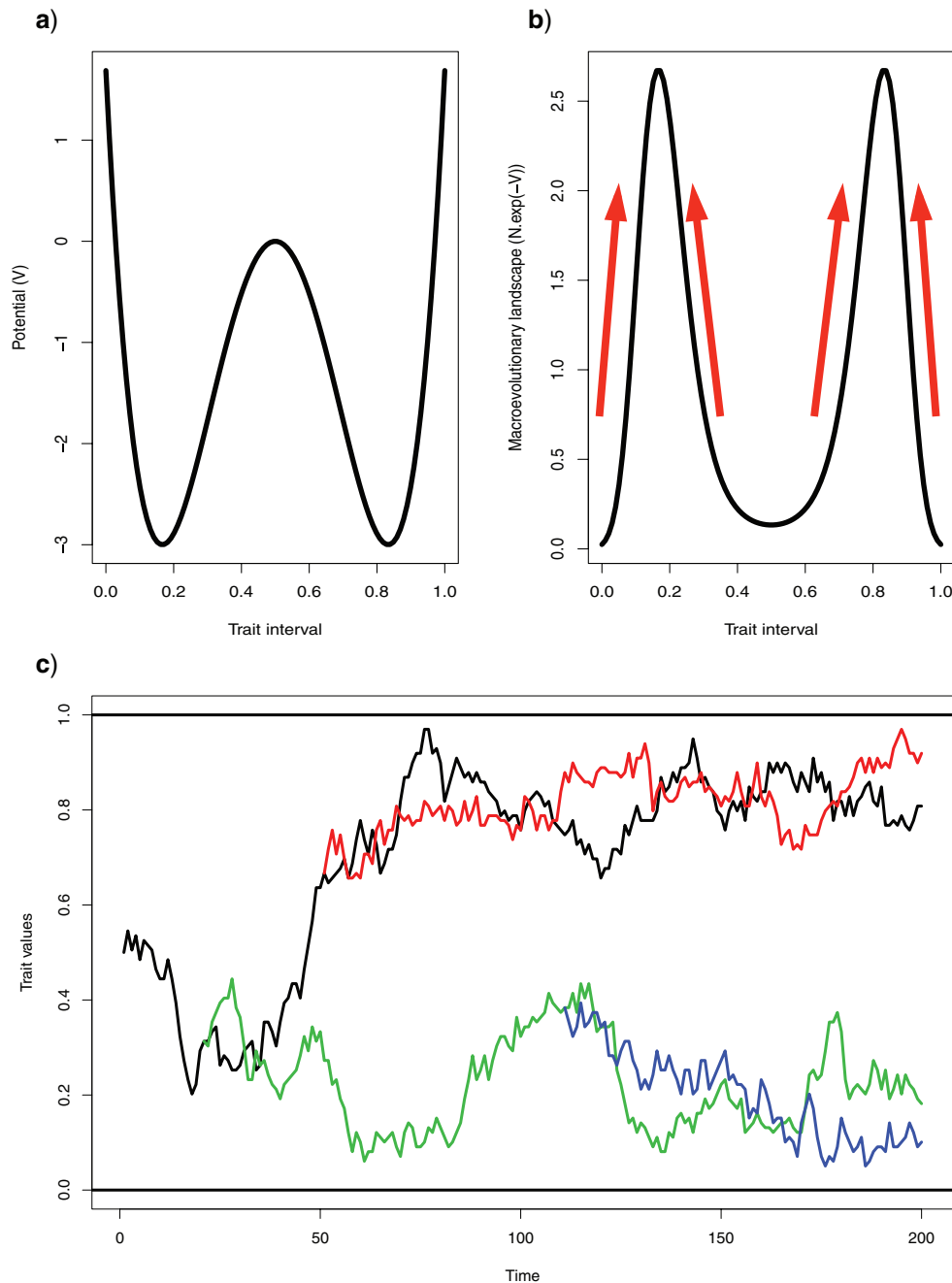
a)

b)

c)

FIGURE 1.   Behaviour of the FPK model. a) Example of a potential with $V(x) = 3x^4 - 6x^2$. b) The stationary distribution of the process, N.exp($-V(x)$), can be interpreted as a macroevolutionary landscape. Wells of the potential become peaks in the macroevolutionary landscape. Differences in the potential generate a force $-V'(x)$, which attracts trait values towards the two peaks of the macroevolutionary landscape, as indicated by the arrows. c) One simulation of the evolution of a trait in a clade of four species in this macroevolutionary landscape: the x-axis shows time and the y-axis the trait value of each species, in different colors. The ancestral trait value lies between the two peaks, and species' traits get attracted towards one of the peaks, which have equal heights. The bounds of the trait interval, represented by thick horizontal lines, are not reached during the process and thus do not influence the evolution of the trait in this case.

converge to the continuous one (1) as we increase the number of points used to discretize the trait interval (i.e., $n \to \infty$).

The use of this discretization procedure imposes that bounds on the trait interval exist: these two bounds are denoted $B_{\min}$ and $B_{\max}$. As done for the BBM model, we make the hypothesis that these two bounds are reflective and calculate the probability density of a trait evolving under FPK using the method of images (Jackson, 1998), that is by cutting and reflecting the probability density

of the unbounded model an infinite number of times at each one of the two bounds (Boucher and Démery, 2016). However, these bounds need not be reached by the trait and need not even influence the evolutionary process. We thus distinguish two different cases:

1. In situations where the stationary distribution of the FPK model converges to zero when trait values tend to $+\infty$ or $-\infty$, bounds located far apart from the observed trait interval will most likely never be reached: there is a strong force opposing trait evolution towards low-lying regions of the macroevolutionary landscape. Such scenarios can be seen on the top part of Figure 2 (scenarios a–d). In this case, the discretization procedure only introduces a very slight approximation to the likelihood function.

2. While we need to introduce bounds for technical reasons, we can also actually make use of them. The same model can thus be used to model situations in which reflective bounds on each side of the trait interval are actually reached during trait evolution. Such scenarios can be seen on the bottom part of Figure 2 (scenarios e–h). For clarity's sake, in these situations, we will call the model BBMV, for bounded BM with an evolutionary potential. BBMV is actually the most general model, since FPK is a special case of it with bounds set to $-\infty$ and $+\infty$.

The discretization procedure that we use enables calculating the transition matrix between different points on the grid. Once this matrix is obtained, we calculate the likelihood of the model as is done for the evolution of discrete characters (i.e., the Mk model, Lewis, 2001). We use the pruning algorithm (Felsenstein, 1973), starting from the probability density of the trait at the tips, and propagating it down to the root of the tree. Finally, we treat the value of the trait at the root of the phylogenetic tree, x_0, as a parameter of the FPK model. This makes comparison with other models of evolution possible, since most implementations include the root value as a parameter.

The precision of the discretization procedure (i.e., the number of points used to discretize the trait interval) naturally influences the precision of the numerical calculation of the likelihood (Boucher and Démery, 2016) and the accuracy in the estimation of the shape of the macroevolutionary landscape, but on the other hand calculation quickly grows with the number of points (the transition matrix M, which needs to be diagonalized, has n^2 terms). In the rest of this article, all simulations have been run with $n=50$ points on the trait grid. This value was used because of the large number of simulations we ran, but we generally recommend people working on a single empirical case to increase this number.

Shape of the Macroevolutionary Landscape

The model we have presented above can accommodate any shape of the macroevolutionary landscape. However, in order to infer macroevolutionary landscapes from empirical data we need to specify a parametric shape for it and optimize its parameters. One possibility to do so would have been to use step functions with different values at each point in the trait grid, but this would have led to a very large number of parameters to estimate. Using combinations of sine functions of various periods and amplitudes would also have been possible, but their periodicity renders optimization difficult. Instead, we chose a polynomial function with only three terms, parametrized as $V(x)=ax^4+bx^2+cx$, with x taken on the interval $[-1.5,+1.5]$. This interval was chosen because it is symmetric, but it is then transformed to the actual trait interval observed in the data set following an affine transformation (see Supplementary Appendix III available on Dryad). We discard the term x^3 because any function of the form $f(x)=ax^4+bx^3+cx^2+dx$ can be written as $f(x)=a(x-x_0)^4+b'(x-x_0)^2+c'(x-x_0)+d'$ with $x_0=-b/(4a)$. This means that adding a term proportional to x^3 amounts to a translation of the potential. No constant term needs to be added to this polynomial function either, since it is the derivative of $V(x)$ that controls the dynamics of the model (Eq. 1). This shape of the potential can approximate a variety of scenarios, including flat landscapes (i.e., BBM, $V(x)=0$), linear trends (e.g., $V(x)=x$), domed (e.g., $V(x)=-x^2$) or U-shaped (e.g., $V(x)=x^2$) macroevolutionary landscapes, but also macroevolutionary landscapes with two central peaks of equal (e.g., $V(x)=x^4-x^2$) or different heights (e.g., $V(x)=x^4-x^2+x$). Figure 2 shows a variety of shapes of the macroevolutionary landscape that can be obtained with this parametric function, either under the pure FPK (no bounds in practice) or the BBMV model (bounds are actually reachable). Finally, note that both BM and the OU model are special cases of the FPK model: BM corresponds to $V(x)=0$ and OU to $V(x)=(\alpha/\sigma^2)x^2-(2\alpha\theta/\sigma^2)x$, where α and θ are the attraction strength and optimum of the OU model, respectively (Hansen, 1997).

Maximum-likelihood Inference of Model Parameters

The FPK model has five parameters: the value of the trait at the root of the tree x_0, the evolutionary rate σ^2, and the three coefficients determining the shape of the macroevolutionary landscape. We have implemented MLE of the FPK model in the R statistical environment (R Core Team, 2016). All functions needed to fit the model to empirical data are freely available from the following Github repository: https://github.com/fcboucher/BBMV, and only depend on functions from the ape package (Paradis et al., 2004). We have verified that the likelihoods obtained from our code are compatible with likelihoods
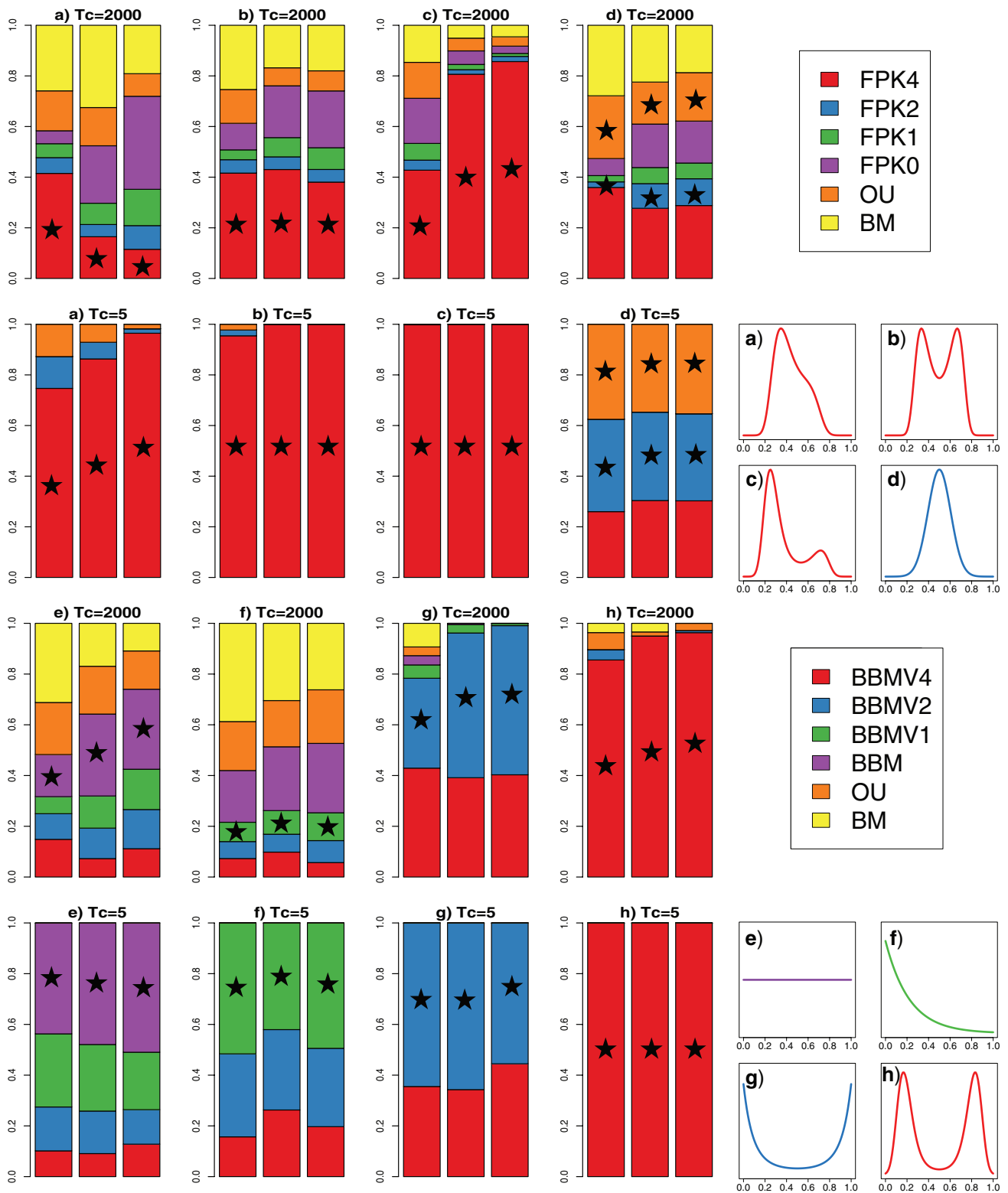
FIGURE 2.   Model discrimination in simulations. Each barplot shows AIC weights of each of the six models fitted to the simulated data: 4 versions of the FPK or BBMV models, plus BM and OU fitted using the package *geiger*, averaged over 20 simulations. In each case, black stars indicate the model that was used for simulations and blocks of three columns show results for trees of 50, 100 and 200 tips, from left to right. The top height panels show simulations of the pure FPK model, the height bottom ones simulations of the BBMV model. Numbers after FPK or BBMV in the legends indicate the degree of the leading polynomial term that was used for fitting: for example, FPK4 stands for the FPK model with $V(x) = ax^4 + bx^2 + cx$, and BBMV1 for the BBMV model with $V(x) = cx$. In case d) there are two stars since FPK2 and OU are identical models. Small panels on the right show the shape of the macroevolutionary landscape that was simulated in each case, with colors corresponding to the rest of the figure.

for other models of trait evolution implemented in the *fitContinuous* function of package geiger (Pennell et al., 2014). This makes comparison between FPK and other evolutionary models like BM or OU possible, using the Akaike information criterion (AIC) for example. Maximum-likelihood inference of model parameters is conducted using the *optim* function in R with the Nelder–Mead optimization routine (i.e., the simplex method, although our code also allows for optimization using the BFGS method with box constraints), which was found to perform better than other optimization routines following preliminary tests. For better numerical precision, we do not directly optimize the evolutionary rate, σ^2, but rather $\log(\sigma^2/2)$. In cases where obvious bounds on the trait values exist, for example if the trait under study is a proportion or a probability, the user can specify the values of the bounds on the trait interval and fit the BBMV model. In cases where no actual bounds are suspected to exist and the user wants to fit the pure FPK model, we place artificial bounds far away from the observed trait interval (i.e., $B_{min} = x_{min} - (x_{max} - x_{min})/2$ and $B_{max} = x_{max} + (x_{max} - x_{min})/2$).

In order to test complex macroevolutionary landscapes against simpler ones, we have also written R functions for fitting the FPK and BBMV models with simpler macroevolutionary landscapes, that is $V(x) = bx^2 + cx$, $V(x) = cx$, and $V(x) = 0$. Alternatives shapes of the macroevolutionary landscape can then be statistically compared based on their likelihoods, using likelihood ratio tests or any information criterion. Note that the pure FPK model does not make sense in cases where $\exp(-V(x))$ does not converge to 0 when x tends to $+\infty$ or $-\infty$: this is the case for $V(x) = cx$ and $V(x) = 0$, as well as for both $V(x) = ax^4 + bx^2 + cx$ and $V(x) = bx^2 + cx$ when the dominant polynomial coefficient is negative. This does not mean that evolutionary trends cannot be estimated from comparative data when the trait interval is not bounded: a macroevolutionary landscape which quickly raises from low to high probabilities, then decreases slowly until a given trait value, and finally quickly drops to low probabilities again would fit a scenario with a trend towards small trait values with soft bounds on trait values (scenario a in Fig. 2).

Finally, for all versions of the macroevolutionary landscape, confidence intervals (CIs) containing the 95% highest probability density around parameter estimates while fixing other parameters to their maximum likelihood estimate can be calculated. This is technically done by removing the lowest 2.5% density regions on each side of the MLE for σ^2, the parameters describing the shape of the macroevolutionary landscape, and the root value when its MLE does not lie in one of the bounds of the trait interval. If the MLE of the root value lies in one of the bounds of the trait interval, then the lowest 5% density region on the other side is removed. These CIs can be returned along with likelihood profile plots around parameter estimates.

MCMC Algorithm

In addition to maximum-likelihood optimization, we present a MCMC algorithm to estimate parameters of the FPK model, which is also written in R. Since the aim of this MCMC algorithm will often be to get an idea of the distribution of parameter estimates, we have focused on the full model with three polynomial terms. However, nested models with simpler macroevolutionary landscapes can also be fit by setting the probability of update for unnecessary parameters to zero. Numerical calculation of the likelihood of FPK in our MCMC implementation is done as for the maximum-likelihood case, and we use the Metropolis–Hastings algorithm to create a Markov chain of parameter estimates.

Parameters of the FPK model have different natures: the three coefficients determining the shape of the macroevolutionary landscape (a, b, and c) as well as the diffusion coefficient $\log(\sigma^2/2)$ are continuous variables, while the root value of the trait, x_0, is only allowed to vary on a regular grid of points (see above). These different parameters thus have different kinds of prior and proposal functions.

We have implemented two prior distributions for continuously varying parameters: either normal or uniform ones. However, one should keep in mind that very large values of a and b in particular (i.e., the coefficients of the x^4 and x^2 terms) can lead to extremely steep macroevolutionary landscapes, which will be unrealistic in most cases. For these two parameters at least, a normal prior centered on zero thus seems to be the most sensible choice. For x_0, we have only implemented a discrete uniform prior on all points of the trait grid.

As for proposal functions, both normal deviates and sliding windows are available for continuously distributed parameters (but other proposals could easily be implemented by modifying our R code). Note that since we actually update $\log(\sigma^2/2)$, this corresponds to a multiplier proposal for σ^2. Only a discrete sliding windows is possible for x_0 and a move on the trait grid is forced to occur each time this parameter is updated. The sensitivity of these proposal functions can be set by the user, but we recommend that the discrete sliding window for x_0 only allows for moves of one step on the grid. Parameters of the model are updated independently, but in order to speed up convergence of the Markov chain to its stationary distribution the relative frequencies of update of the different parameters can be modified.

In practice, we have observed two contrasting behaviors for convergence: (i) when the characteristic time T_c is larger than the depth of the phylogenetic tree, σ^2 will converge rapidly in the MCMC chain while parameters setting the shape of the macroevolutionary landscape will not; (ii) on the opposite, when T_c is smaller than the depth of the phylogenetic tree, parameters setting the shape of the macroevolutionary landscape will converge rapidly while σ^2 will be slow to converge. These two behaviors simply reflect the fact that when

T_c is large relative to the total time span of trait evolution, the distribution of traits at the tips of the phylogeny will not have converged to the stationary distribution of the FPK model: the macroevolutionary landscape is poorly explored and tip values retain very little information regarding its shape. On the contrary, when T_c is small, the macroevolutionary landscape will have been thoroughly explored by the clade, but it is difficult to determine the value of the evolutionary rate with precision (see below). Manipulating the relative frequencies at which these different parameters are updated can have dramatic effects on the speed of convergence of the MCMC chain. A good way to set this tuning parameter would be either to first run a MLE of the model or to do an initial quick MCMC run (e.g., with a very low number of points to discretize the trait interval) in order to get an idea of the values of σ^2 and V. From our experience, convergence of our MCMC algorithm takes a long time, even on rather small data sets (c. 200,000 to 1 million steps for trees with less than 50 tips). We recommend running at least two independent chains in order to make sure that they have converged to the same stationary distribution.

PERFORMANCE OF THE FPK MODEL

In order to assess the performance of FPK in terms of parameter inference and model discrimination, we have conducted a large number of simulations. Our focus was on the likelihood of FPK, thus we restricted our simulations to the maximum-likelihood optimization procedure since it is much faster than MCMC estimation. We ran simulations under height contrasted scenarios, four of them corresponding to the pure FPK model and four to the BBMV model. The macroevolutionary landscapes corresponding to these height scenarios are pictured on Figure 2.

The four scenarios of the FPK model that we simulated were the following: (a) a directional trend limited to a portion of trait space ($V(x) = 5x^4 - x^2 + x$), (b) a macroevolutionary landscape with two peaks of equal height ($V(x) = 10x^4 - 5x^2$), (c) a macroevolutionary landscape with two peaks of different heights ($V(x) = 5x^4 - 5x^2 + x$), and (d) a single peak (i.e., an OU model, $V(x) = 5x^2$). The four scenarios simulated under the BBMV model, that is in which trait evolution was actually bounded, cover a broad range of interesting cases: (e) a flat macroevolutionary landscape (BBM, $V(x) = 0$), (f) a directional trend ($V(x) = 1.5x$), (g) disruptive selection (i.e., a U-shaped macroevolutionary landscape with extreme trait values being favored, $V(x) = -1.2x^2$), and (h) two peaks of the same height ($V(x) = 3x^4 - 6x^2$).

For each one of these height scenarios, we fit four different versions of FPK: the full model ($V(x) = ax^4 + bx^2 + cx$), a model with only quadratic and linear terms ($V(x) = bx^2 + cx$), a model with only a linear term ($V(x) = cx$), and a model with a flat macroevolutionary landscape

(i.e., BBM, $V(x) = 0$). In simulations of the FPK model (scenarios a–d), bounds were placed far apart from the observed trait interval for inference, while in simulations of the BBMV model (scenarios e–h) the true bounds used in simulations were specified. In addition, we also fit BM and an OU model with a single optimum to each simulated data set using the *fitContinuous* function in the geiger package (Pennell et al., 2014).

Phylogenetic trees were simulated under a pure birth model, with unit birth rate. Trees were grown until the desired number of tips plus one was obtained and one of the two sister tips originating from the last speciation event was trimmed. Trees were then rescaled to a total depth of 100 arbitrary time units in order to enable comparison between simulations. All simulations were done with $B_{\min} = 0$, $B_{\max} = 1$, and $x_0 = 0.5$. For each of the height scenarios described above (and thus for each value of V) we used two different values of σ^2 which were calculated so that: (i) $T_c = 5$ (i.e., 1/20 of tree depth), which should ensure that the distribution of the trait at the tips of the tree has converged to the stationary distribution of the model, and (ii) $T_c = 2,000$ (i.e., 20 times tree depth), in which stationarity should not have been reached at the end of the simulation. In addition, we also explored the effect of tree size on parameter estimation and model discrimination using trees of 50, 100, and 200 tips. For each combination of the shape of the potential, the value of σ^2, and tree size, we conducted 20 different simulations (960 simulations in total).

Model Discrimination

We first focused on whether FPK and BBMV can be distinguished from other classic models of evolution using relative Akaike weights (Burnham and Anderson, 2002). Our simulations showed that when stationarity was reached ($T_c = 5$), all four scenarios of the FPK model that we simulated could easily be discriminated from BM, which always received less than 0.001% Akaike weight (Fig. 2). Discrimination from OU was also easily achieved, this model always receiving less than 13% Akaike weight, except in the case where it was the model which was actually simulated (scenario d, Fig. 2). Discrimination was even better under the four scenarios of the BBMV model that we simulated, with both BM and OU always receiving less than 0.01% Akaike weight (Fig. 2). In simulations where stationarity was not reached ($T_c = 2000$), only scenarios (c, g, and h) could still be discriminated from BM and OU (Fig. 2). All other five scenarios indeed gave relatively high Akaike weights to either BM or OU (Fig. 2). The number of tips in the tree did influence discriminatory power between FPK and BM or OU positively, but its effect was moderate (Fig. 2).

We then looked at whether FPK or BBMV models with different shapes of the macroevolutionary landscape can be statistically distinguished. Discrimination between alternative versions of the model with different shapes of the macroevolutionary landscape was also generally satisfactory. In cases where stationarity was
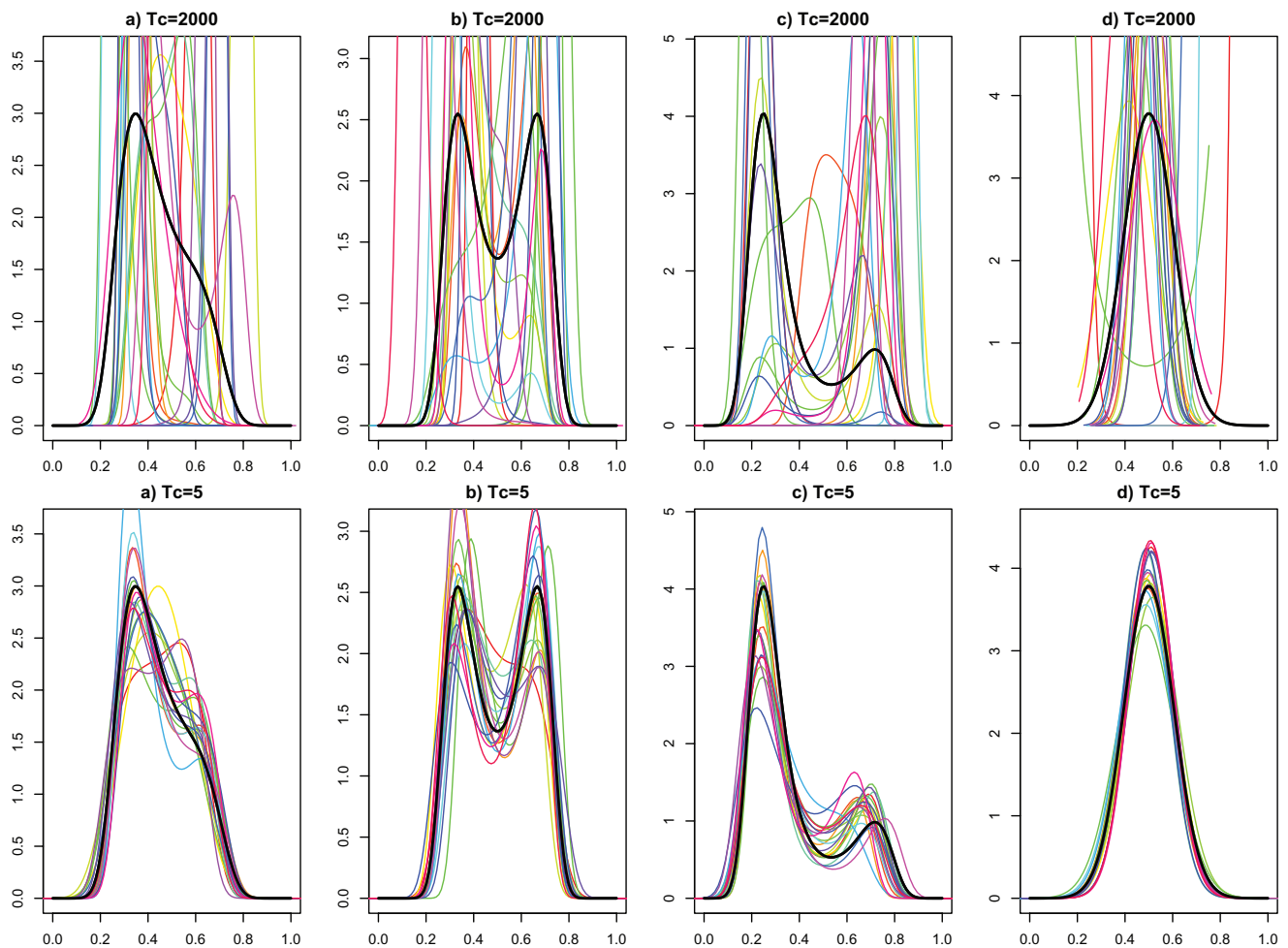
FIGURE 3.   Estimation of the macroevolutionary landscape in different versions of the FPK model. Thin lines in each plot show the macroevolutionary landscapes estimated in 20 simulations, each one in a different color, while the simulated macroevolutionary landscape is shown by the thick black line. Only results for trees with 100 tips are shown. Top row: simulations with $Tc = 2,000$, in which stationarity was not reached. Bottom row: simulations with $Tc = 5$, in which stationarity was reached. From left to right, columns show simulation for scenarios a to d.

reached ($T_c = 5$), the version of the model that was used to simulate the data was always the one to receive the highest AIC weight (Fig. 2). Increasing tree size generally lead to an increase in the Akaike weight of the generating model, the effect being substantial this time (Fig. 2). Scenarios with macroevolutionary landscapes containing two peaks (i.e., scenarios b, c, and h), which can only be accommodated by the most complex form of the macroevolutionary landscape ($V(x) = ax^4 + bx^2 + cx$) always led to more than 95% Akaike weight to this model (Fig. 2). Importantly, for all four scenarios simulated under the FPK model, models with a flat ($V(x) = 0$) and a linear potential ($V(x) = cx$) always received less than 2.3e−8 Akaike weight. We have seen above that these two shapes of the potential do not make sense in the case of the FPK model (i.e., when there are no bounds in practice): these simulations confirm that both of these models are strongly rejected statistically in these situations. In contrast, when stationarity was not reached

($T_c = 2000$) different shapes of the macroevolutionary landscape were difficult to discriminate and simpler forms were often preferred over more complex ones, especially so in trees with few tips. This is normal since in these cases, traits only had time to explore a small fraction of the macroevolutionary landscape. The only notable exception to this general observation was for scenario h (a BBMV model with two peaks), in which the full BBMV model always received over 85% Akaike weight, even when $T_c = 2000$ and with trees of 50 tips only (Fig. 2).

Parameter Inference

Accuracy of FPK models in parameter estimation was assessed by comparing the maximum-likelihood estimates of parameters with values used in simulations. We first compared the precision in the estimation of the macroevolutionary landscape as a whole, and not
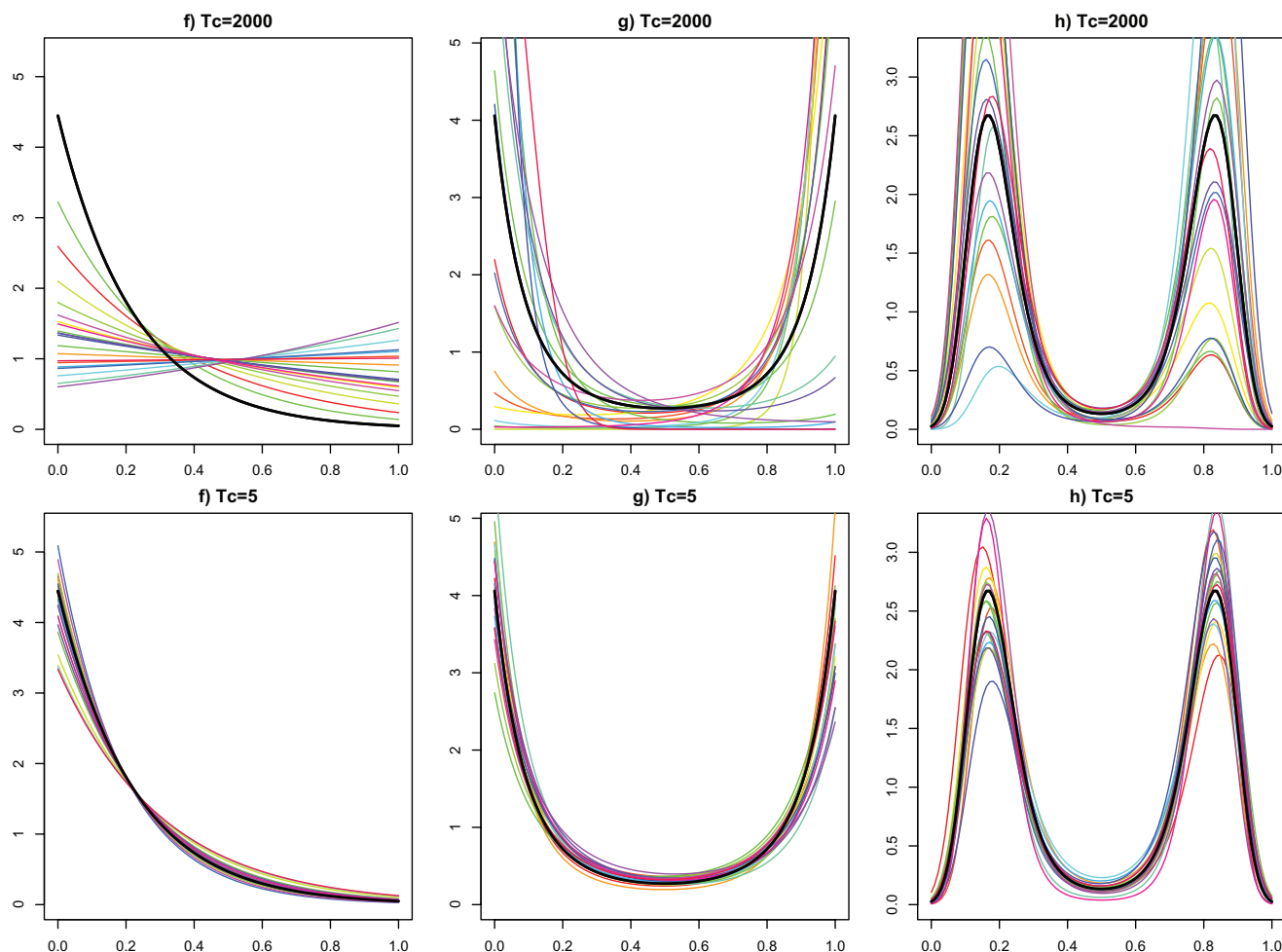
FIGURE 4.   Estimation of the macroevolutionary landscape in different versions of the BBMV model. Thin lines in each plot show the macroevolutionary landscapes estimated in 20 simulations, each one in a different color, while the simulated macroevolutionary landscape is shown by the thick black line. Only results for trees with 100 tips are shown. Top row: simulations with $Tc = 2,000$, in which stationarity was not reached. Bottom row: simulations with $Tc = 5$, in which stationarity was reached. From left to right, columns show simulation for scenarios f to h. Results for simulations of the flat landscape (BBM, scenario e) are not shown since the macroevolutionary landscape is fixed in this case.

in the estimation of a, b, and c separately. This is because these three coefficients can sometimes be redundant and lead to very similar shapes of the macroevolutionary landscape: for example, a (the coefficient of the x^4 term) and b (the coefficient of the x^2 term) are highly correlated. Simulations showed that macroevolutionary landscapes are generally accurately estimated in cases where stationarity has been reached since the actual shape that was simulated is most often recovered (Figs. 3 and 4). Estimation of the macroevolutionary landscape was even better in simulations of the BBMV model (Fig. 4) compared to simulations the pure FPK model (Fig. 3), probably because in the former case the actual bounds used in simulations were specified when inferring parameters. In all height scenarios, accuracy in the estimation of the macroevolutionary landscape increased with the number of tips in the phylogeny (Supplementary Appendix VI available on Dryad). Accuracy was much worse in simulations that had not reached stationarity (Figs. 3 and 4).

As for the other two parameters of the FPK model, accuracy in the estimation of x_0 was much less satisfactory and usually had a huge variance, especially so when T_c was small (Supplementary Appendix IV available on Dryad). The estimation of σ^2 was very accurate for large values of T_c but had much larger variance when T_c was small (Supplementary Appendix IV available on Dryad). No bias in the estimation of σ^2 was apparent for the FPK model, but it seemed that the estimation of σ^2 was slightly biased towards larger values in the four scenarios of the BBMV model that we simulated and for $T_c = 5$.

Empirical Example: Body Size Evolution in North-American Watersnakes (Tribe Thamnophiini)

We demonstrate the utility of FPK using an example of body size evolution in snakes. We decided to study North-American watersnakes (Colubridea, subfamily Natricinae, tribe Thamnophiini) because the distribution
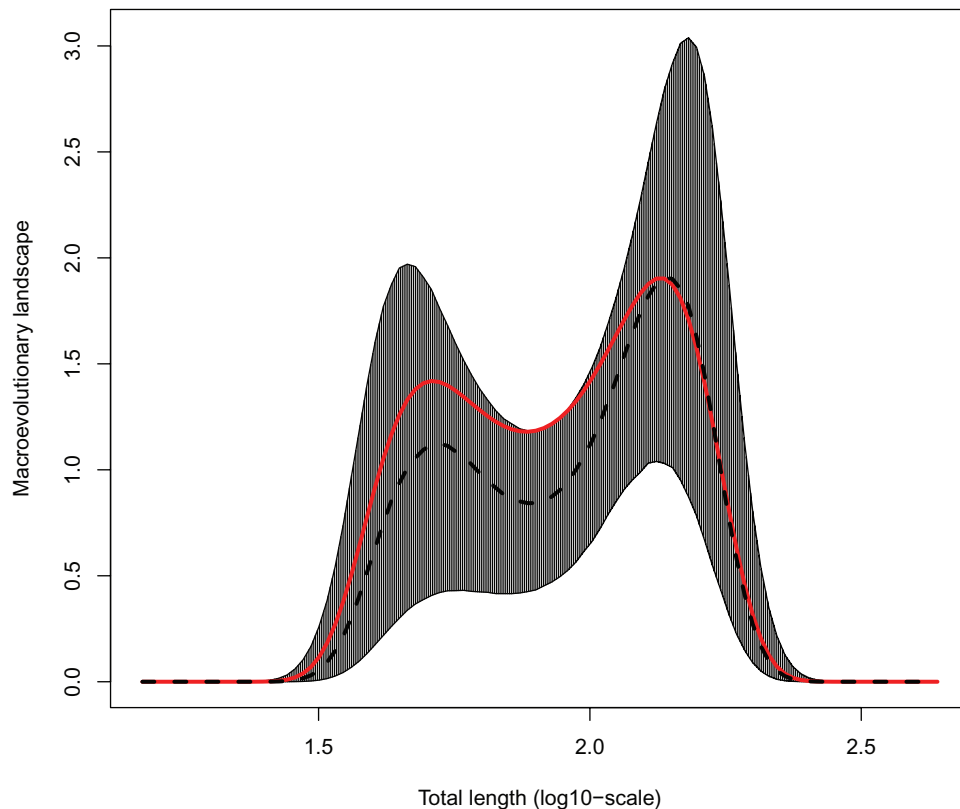
FIGURE 5.   Posterior distribution of the macroevolutionary landscape estimated for body length evolution in watersnakes (tribe Thamnophiini). This posterior distribution was obtained by concatenating the two MCMC chains after the first 10% of samples were discarded as burnin (800,000 MCMC steps in total). The figure shows the value of the macroevolutionary landscape (N.exp($-V(x)$)) on the y-axis as a function of log10(total length) measured in centimeters. The dashed black line shows the median value of the macroevolutionary landscape over the posterior, while the grey area ranges from the 25% to the 75% quantiles. The solid red line shows the maximum-likelihood estimate of the macroevolutionary landscape.

of their body length shows a slight bimodality (Burbrink and Myers, 2014). A time-calibrated phylogeny as well as measurements of total length (hereafter, TL) for 45 species included in this group were obtained from (Burbrink and Myers, 2014), and TL was log_{10}-transformed prior to analysis.

We first fit three alternative models for the evolution of TL along the watersnake phylogeny using maximum-likelihood: BM, an OU model with a single optimum, and the FPK model (with $V(x) = ax^4 + bx^2 + cx$). In addition, we used our MCMC algorithm to obtain posterior estimates of the shape of the macroevolutionary landscape in this clade (detailed methods can be found in Supplementary Appendix V available on Dryad). Convergence of MCMC chains was assessed both visually by looking at the trace plots of the parameters, likelihood, prior, and posterior, and by measuring the effective sample sizes of these different quantities using the R package coda (Plummer et al., 2006).

Among the three models compared using maximum-likelihood, the FPK model had by far the lowest AIC, followed by the OU model (ΔAIC=13.2), and finally BM (ΔAIC=15.1). The macroevolutionary landscape estimated by the FPK model contained two distinct peaks, the peak corresponding to longer TLs being the highest (Fig. 5). CIs on the maximum-likelihood estimates of model parameters confirmed that the coefficient for the x^4 term of the potential, a, was positive (ML estimate: 9.2, 95% CI: [5.8, 14.3]), while the coefficient for the x^2 term, b, was negative (ML estimate: -3.4, 95% CI: [$-4.9, -1.0$]), which is typical of macroevolutionary landscapes with two peaks. The linear coefficient of the potential, c, was not significantly different from 0 (ML estimate: -0.34, 95% CI: [$-1.9, 1.8$]). There was large uncertainty as to the value of TL for the ancestor of watersnakes, the CI spanning almost the whole distribution of TL in extent species (ML estimate: 72.9 cm, 95% CI: [$39.9, 118.1$]). We estimated a characteristic time of 22.4 Myr for the FPK process, which is slightly higher than the crown age of Thamnophiini (16.6 Myr, Burbrink and Myers, 2014). Results obtained using MLE were supported by the two MCMC chains that we ran: the posterior distribution of the macroevolutionary landscape also had two peaks of unequal heights, and the mode of

this distribution closely matched the macroevolutionary landscape estimated using maximum-likelihood (Fig. 5).

These results suggest that TL might have evolved toward two different optima in North-American watersnakes, the first one roughly corresponding to 50 cm and the second, higher optimum, to 130 cm (Fig. 5). In their original publication (Burbrink and Myers, 2014) had proposed that the skewness in the distribution of TL in watersnakes could be due to higher diversification rates for longer species. Comparing both explanations would prove especially interesting but would require extending the FPK model to account for diversification rates depending on the value of the evolutionary potential, for example using the statistical machinery already developed for state-dependent diversification models (FitzJohn et al., 2009).

DISCUSSION

In this article, we have presented equations for a very general model of evolution for continuous traits, as well as its implementation. This opens new possibilities for estimating macroevolutionary landscapes from phylogenetic comparative data. Below we discuss the strengths and weaknesses of this model. We note that (Blomberg, 2017) has recently introduced a family of essentially similar models for continuous trait evolution, but no framework exists yet to infer parameters of these models from phylogenetic comparative data.

New Avenues for Studying Phenotypic Evolution from Comparative Data

The flexibility of the new model that we propose is its greatest strength. FPK and its bounded version BBMV offer the opportunity to estimate a variety of macroevolutionary landscapes from phylogenetic comparative data. FPK can indeed model evolutionary processes like directional selection or diversifying selection leading to macroevolutionary landscapes with several peaks, eventually of varying heights. BBMV accommodates bounds on phenotypic evolution and as such can model scenarios like directional selection toward one extreme trait value or even disruptive selection. All of these scenarios lie at the core of modern (macro)evolutionary theory, but could not yet be inferred from phylogenetic comparative data (O'Meara, 2012). Fitting the FPK model to empirical data is similar in some aspects to cubic spline analysis in selection studies (Schluter, 1988): it allows inferring (and visualizing) the macroevolutionary landscape that has been experienced by species in a clade. As already noted, this model will be especially useful in situations where one has strong *a priori* expectations that all members of a clade have experienced the same macroevolutionary landscape throughout their history.

One kind of landscapes that can be inferred using FPK deserve particular mention here: macroevolutionary landscapes in which multiple peaks exist. This scenario would be especially interesting to compare to a situation in which these multiple peaks are available for different lineages, which is what is implemented in OU models with multiple optima. In this latter class of models, each lineage is indeed subject to attraction towards a single peak at a time, but lineages may shift between different peaks. These shifts are either determined *a priori* (Butler and King, 2004; Beaulieu et al., 2012) or inferred directly from phylogenetic patterns of trait evolution (Ingram and Mahler, 2013; Uyeda and Harmon, 2014; Khabbazian et al., 2016). In FPK with multiple peaks on the contrary, each lineage is always influenced by the different peaks in its macroevolutionary landscape, and transitions between peaks might be frequent if the traits of most species in the clade are located in a valley of the macroevolutionary landscape. More theoretically, these alternatives would represent two very different evolutionary scenarios: OU models with several optima might be better at describing situations in which a lineage shifts to another adaptive zone (Simpson, 1944; Landis and Schraiber, 2017), while FPK with multiple peaks might represent more genuine diversifying selection towards alternative phenotypic optima. In the later scenario, no change in the environment a lineage experiences or in other aspects of its phenotype (Simpson, 1944) are required for one lineage to shift between two phenotypic optima, and one might expect much more frequent transitions between adaptive peaks within a clade. The implementation of the FPK that we have introduced opens the way to discriminate between these two alternatives using empirical data, although the level of statistical power required to do so remains an open question.

Statistical Behavior of the FPK Model

FPK is a model that generally retains very little phylogenetic signal (hereafter, PS): over all simulations in which stationarity had been reached, the median value of the λ index of PS (Pagel, 1997) was 1.6e−109 (see Supplementary Appendix VI available on Dryad). This absence of PS stems from the strong deterministic component of the FPK model in all scenarios that we simulated, a result already known for the OU model (Münkemüller et al., 2015), which is a special case of FPK. As a result, when stationarity is reached most of the information needed to infer the shape of the macroevolutionary landscape ultimately comes from the distribution of the trait for extant species. For example, evolutionary trends can be recovered in the absence of fossil data from a highly skewed trait distribution for contemporaneous species. In the same vein, the simultaneous presence of two peaks in the macroevolutionary landscape can be inferred from a bimodal trait distribution. However, FPK can also produce trait distributions with higher levels of PS. First, high PS can be obtained when the deterministic component of FPK is small or even absent: this is the case for the BM model, also a special case of FPK. Second,

intermediate levels of PS can be obtained even under an FPK model with a strong deterministic component when stationarity is not reached. FPK should thus not be seen as a nonphylogenetic model: in the extreme case where stationarity has been reached PS will be null and the process will actually behave as if it were nonphylogenetic, but many intermediate cases exist in which the deterministic component of FPK influences trait evolution but PS is still significant (Supplementary Appendix VI available on Dryad). As already argued for the OU model (Hansen et al., 2008), fitting the FPK model is actually the best way to figure out what level of PS is present in the data and what is the relative importance of deterministic forces compared to random diffusion.

The characteristic time of the FPK process, T_c, represents the typical time needed to reach stationarity and should always be compared to the total depth of the phylogenetic tree for the clade under study, T_{tot} to get an idea of the level of PS in the data. Unsurprisingly, model performance will increase with T_{tot}/T_c and in the extreme case where $T_{tot} \ll T_c$, traits will not have explored much of the macroevolutionary landscape. T_c bears much similarity with the phylogenetic half-life of the OU process, which describes the time necessary for the trait value to move halfway from its initial position to the optimum. Supplementary Appendix I available on Dryad shows that the characteristic time of an OU process is actually directly proportional to its phylogenetic half-life: $T_c = 1/\alpha$ while the phylogenetic half-life is $\ln(2)/\alpha$. In agreement with our findings, previous studies have shown that accuracy in parameter estimation of the OU model increases with decreasing phylogenetic half-lives (Ho and Ané, 2014b; Uyeda and Harmon, 2014).

Using simulations, we have demonstrated that FPK can be distinguished from other classic models of trait evolution, and also that distinct shapes of the macroevolutionary landscape can be distinguished from each other based on their likelihoods. This means that alternative versions of the FPK model can be used for testing evolutionary hypotheses about trait evolution. Our simulations also show that the danger of overfitting is quite low with FPK since simpler models will often be preferred when stationarity has not been reached. Our focus on AIC to discriminate between alternative models was motivated by the fact that it is the most commonly used in the macroevolutionary community. However, AIC might be prone to overfitting in parameter rich macroevolutionary models and other measures that penalize more for extra parameters, like the Bayesian Information Criterion or its modified version (Zhang and Siegmund, 2007), might be preferable (Ho and Ané, 2014b). Another solution to diagnose overfitting would be to use parametric bootstrapping techniques (Boettiger et al., 2012), which is readily implementable for FPK since we provide an R function to simulate the model.

Our results also show that estimation accuracy under FPK drastically differs between parameters. Estimation of the shape of the macroevolutionary landscape is generally accurate when $T_{tot} > T_c$ and increases with T_{tot}/T_c. Estimation accuracy also increases with tree size: our results suggest that trees with 50 tips lead to reasonable estimation of the general shape of the macroevolutionary landscape (Supplementary Appendix IV available on Dryad), while trees with 100 tips should most often be large enough to obtain very reliable estimates (Figs. 3 and 4). Importantly the three coefficients that determine the shape of the macroevolutionary landscape should not be analyzed separately since numerous different combinations of a, b, and c can give similar shapes. Rather, we recommend to interpret the general shape of the macroevolutionary landscape by focusing on a few important features: (i) whether the macroevolutionary landscape is flat or not, (ii) whether it features a single trend towards one of the bounds of the trait interval, (iii) whether it contains one or several peaks, and if relevant (iv) where are these peaks located in the trait interval. In contrast, the evolutionary rate σ^2 has high estimation variance when $T_{tot} > T_c$, and is accurate when $T_{tot} < T_c$. This same effect had already been observed for BBM (Boucher and Démery, 2016) and probably reflects the fact that when the macroevolutionary landscape has been fully explored by the clade its shape and extent are easily estimated but the speed at which the landscape is travelled is not. Given this limitation, the estimate of σ^2 is difficult to interpret when fitting FPK to an empirical data set. Rather, the characteristic time of the process, T_c, should be the quantity that is interpreted in comparison with tree depth. Finally, we found that the estimation of the trait value at the root of the tree, x_0, is poor as soon as the macroevolutionary landscape has been moderately explored, a result that generalizes the one already obtained for BBM (Boucher and Démery, 2016). This stems from the fact that FPK is a model that retains low PS since it includes strong deterministic forces and wipes out any hope of confidently inferring ancestral trait values in empirical data sets.

These differences in the quality of the estimation of the shape of the macroevolutionary landscape versus σ^2 generalize results that have been obtained for the OU model. Indeed, in this model σ^2 and especially α, the attraction strength, are difficult to estimate (Butler and King, 2004; Ho and Ané, 2014b) while it seems that the stationary variance of the OU process, $\sigma^2/2\alpha$, and the trait optimum, μ, generally have higher estimation accuracy (Ho and Ané, 2014b; Münkemüller et al., 2015). This is in agreement with our results since μ and $\sigma^2/2\alpha$ respectively determine the mean and variance of the stationary distribution of the OU process, what we have called the macroevolutionary landscape in this article.

Interpretation of Macroevolutionary Landscapes Inferred from FPK

We have introduced FPK as a method for the inference of macroevolutionary landscapes from phylogenetic comparative data. The adaptive landscape is a fruitful

metaphor to understand phenotypic evolution on a variety of evolutionary scales (Wright, 1932; Simpson, 1944; Arnold et al., 2001) but care must be taken when interpreting inferences made from phylogenetic comparative data in the light of adaptive landscape theory, which was mainly developed for population genetics (Hansen and Martins, 1996; Uyeda and Harmon, 2014). Some of the models of trait evolution on phylogenies include deterministic forces that influence trait evolution in an attempt to mimic selection: the OU model includes a term that was designed to resemble stabilizing selection towards a given trait value (Hansen, 1997), and FPK can imitate a large variety of selection shapes. However, all macroevolutionary models for trait evolution, including OU and FPK, are phenomenological by nature since they rely on probabilistic diffusion equations and model evolution over long time-scales (typically thousands to million years) that are not amenable to direct observation. In other words, these models recover patterns from the data, which researchers have to interpret in terms of evolutionary processes. One example of such a confusion between long-term patterns and short-term processes has been highlighted when interpreting the good fit of an OU model to empirical data sets: several studies have indeed shown that neutral evolution between bounds produces patterns closely resembling the ones obtained under an OU process (Revell et al., 2008; Boucher et al., 2014). Development of the BBM model has now rendered possible to distinguish between these two scenarios (Boucher and Démery, 2016), but many such cases in which two different microevolutionary processes produce the same macroevolutionary pattern remain.

Macroevolutionary landscapes estimated using FPK should thus be interpreted with extreme caution: they reflect the general shape of deterministic forces that have been acting on the evolution of a continuous trait in a clade, but are agnostic regarding the nature of these deterministic forces, that is they are not actual measurements of the relation between individuals' traits and fitness. The most obvious limitation of FPK is that it makes the hypothesis that the macroevolutionary landscape is constant through time. This is likely to be wrong in a majority of cases since environmental change or interactions with other species will lead to changes in the intensity and shape of the selection gradient acting on one trait in a given clade (Simpson, 1944; Hansen, 2012). Macroevolutionary landscapes inferred using FPK will thus necessarily reflect some kind of average macroevolutionary landscape experienced by the clade throughout its evolutionary history.

Even though it is difficult to connect microevolutionary processes to macroevolutionary patterns, there is one promising way in which FPK could be used to do so. Indeed, the Bayesian implementation of the model enables the use of informative priors based on quantitative genetic parameters. (Uyeda and Harmon, 2014) have demonstrated how this could be done for the OU

model: using the quantitative genetic model of (Lande, 1976), they showed how measurements of heritability, phenotypic variance, and effective population size can inform priors on the parameters of the OU model. By connecting the parameters in Eq. 1 to quantitative genetic models, the same procedure could be carried out for FPK.

Finally, FPK need not be restricted to infer macroevolutionary landscapes. This model indeed has its roots in spatial diffusion theory and as such could be used in phylogeographic studies to model the dispersal of a set of individuals or populations for which the phylogeny is known (e.g., Grollemund et al., 2015). This field of research has indeed seen huge methodological advances in recent years (Lemey et al., 2010; Bloomquist et al., 2012). In this context, FPK could be used to infer preferred directions of dispersal (i.e., directional trends) or even to infer particular regions that act as geographic attractors for the taxon under study (i.e., one or several peaks). Hard bounds on the distribution of organisms (e.g., oceans for terrestrial organisms) could even be taken into account explicitly using the BBMV model.

Limitations of the FPK Model

Our implementation of FPK does not come without limitations. The main technical limitation is that our implementation of the model is restricted to single traits. We are fully aware that extending it to multivariate data sets would be very convenient, since multiple traits are expected to often evolve in a correlated fashion (Arnold, 1992). However, this is for the moment hampered by computational time. Indeed, the most time-consuming part in the calculation of the likelihood is to invert the instantaneous transition matrix, M. If we were to study two traits simultaneously, we would need to discretize the plane that they define into a regular grid of points, and computing time would not be multiplied by two but rather raised to the power two. The only possible solution that we can envision would be to use algorithmic tricks that avoid inverting the entire transition matrix, but rather a matrix describing transitions between a given point on the grid and its immediate neighbors, as recently proposed for inference of ancestral areas (Landis et al., 2013). This would require much development and is out of the scope of this article.

The fact that our implementation of FPK can only accommodate a maximum of two peaks in the macroevolutionary landscape might also seem frustrating for some users. Extending the model so that three peaks or more can be simultaneously present is rather straightforward: the most obvious solution would be to use polynomial functions with more terms for $V(x)$. However, this would increase computational time, and more importantly would probably yield likelihood functions that are extremely difficult to optimize. This is why we have not implemented it yet. We however note that in our code to infer the FPK model, we have left open the possibility to specify a given shape for $V(x)$: users can

thus experiment with more complex macroevolutionary landscapes if they feel this is relevant to their specific study system.

As already discussed above, the fact that the macroevolutionary landscape is constant across time and across different clades in the phylogeny is another limitation. Further developments of the FPK model could aim at extending it to cases in which the macroevolutionary landscape differs among clades or among specified time periods across the phylogeny.

The last limitation of FPK perhaps lies in its very formulation. FPK is indeed based on a constant-rate diffusion model and as such cannot model accelerating or decelerating trait evolution (Harmon et al., 2010) or sudden jumps in the value of the trait, as would be expected under quantum evolution (Simpson, 1944; Kirkpatrick, 1982) or punctuated equilibrium (Gould and Eldredge, 1977).

Conclusion

Our development and implementation of FPK generalizes the BM and OU models and greatly expands the set of models available for studying the evolution of continuous characters on phylogenies, thus enabling estimation of macroevolutionary landscapes of various shapes. We have shown that the model generally achieves good performance both in terms of parameter estimation and in terms of discrimination from alternative macroevolutionary models. R code for fitting FPK (and its special case BBMV) to empirical data is freely available from https://github.com/fcboucher/BBMV, and this repository also contains a detailed tutorial to the different functions for simulating and inferring the model.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.28g37.

Acknowledgements

Funding

References

Arnold S.J. 1992. Constraints on phenotypic evolution. Am. Nat. 140:S85–S107.

Arnold S.J. 2014. Phenotypic evolution: the ongoing synthesis. Am. Nat. 183:729–746

Arnold S.J., Pfrender M.E., Jones A.G. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica 112:9–32.

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

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

Blomberg S.P. 2017. Beyond Brownian motion and the Ornstein–Uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. bioRxiv. doi:https://doi.org/10.1101/067363.

Bloomquist E.W., Lemey P., Suchard M.A. 2012. Three roads diverged? Routes to phylogeographic inference. Trends Ecol. & Evol. 25: 626–632.

Boettiger C., Coop G., Ralph P. 2012. Is your phylogeny informative? Measuring the power of comparative methods. Evolution 66: 2240–2251.

Boucher F.C., Démery V. 2016. Inferring bounded evolution in phenotypic characters from phylogenetic comparative data. Syst. Biol. 65:651–661.

Boucher F.C., Thuiller W., Davies T.J., Lavergne S. 2014. Neutral biogeography and the evolution of climatic niches. Am. Nat. 183:573–84.

Burbrink F.T., Myers E.A. 2014. Body size distributions at community, regional or taxonomic scales do not predict the direction of trait-driven diversification in snakes in the United States. Global Ecol. Biogeogr. 23:490–503.

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

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

Doebeli M. 1996. An explicit genetic model for ecological character displacement. Ecology 77:510–520.

Drury J., Clavel J., Manceau M., Morlon H. 2016. Estimating the effect of competition on trait evolution using maximum likelihood inference. Syst. Biol. 65:700–710.

Eastman J.M., Wegmann D., Leuenberger C., Harmon L.J. 2013. Simpsonian 'Evolution by Jumps' in an adaptive radiation of anolis lizards. ArXiv e-prints. arXiv:1305.4216.

Edwards A.W.F., Cavalli-Sforza L.L. 1964. Reconstruction of evolutionary trees. In Heywood V.H., McNeill J., editors. Phenetic and phylogenetic classification. London: Systematics Association, p. 67–76.

Estes S., Arnold S.J. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. Am. Nat. 169:227–244.

Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25: 471–492.

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

FitzJohn R.G., Maddison W.P., Otto S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst. Biol. 58:595–611.

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

Gardiner C.W. 1985. Handbook of stochastic methods. 1st ed. Berlin: Springer.

Gavrilets S. 1997. Evolution and speciation on holey adaptive landscapes. Trends Ecol. Evol. 12:307–312.

Gould S.J., Eldredge N. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology 3:115–151.

Grafen A. 1989. The phylogenetic regression. Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci. 326:119–157.

Grollemund R., Branford S., Bostoen K., Meade A., Venditti C., Pagel M. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. Proc. Natl. Acad. Sci. USA 112:13296–13301.

Hansen T. 2012. Adaptive landscapes and macroevolutionary dynamics. In Svensson E., Calsbeek R., editors. The adaptive landscape in evolutionary biology. Oxford: Oxford University Press, p. 205–221.

Hansen T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

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

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

Harmon L.J., Losos J.B., Jonathan Davies T., Gillespie R.G., Gittleman J.L., Bryan Jennings W., Kozak K.H., McPeek M.a., Moreno-Roark F., Near T.J., Purvis A., Ricklefs R.E., Schluter D., Schulte Ii J.a. Seehausen O., Sidlauskas B.L., Torres-Carvajal O., Weir J.T., Mooers A.O. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–96.

Ho L.S.T., Ané C. 2014a. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. Syst. Biol. 63:397–408.

Ho L.S.T., Ané C. 2014b. Intrinsic inference difficulties for trait evolution with Ornstein–Uhlenbeck models. Methods Ecol. Evol. 5:1133–1146.

Hunt G. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. Proc. Natl. Acad. Sci. USA 104:18404–18408.

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

Jackson J.D. 1998. Classical electrodynamics, 3rd ed. New York: John Wiley & Sons.

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

Kirkpatrick M. 1982. Quantum evolution and punctuated equilibria in continuous genetic characters. Am. Nat. 119:833–848.

Lande R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

Landis M., Schraiber J.G. 2017. Punctuated evolution shaped modern vertebrate diversity. bioRxiv. doi:https://doi.org/10.1101/151175.

Landis M.J., Matzke N.J., Moore B.R., Huelsenbeck J.P. 2013. Bayesian analysis of biogeography when the number of areas is large. Syst. Biol. 62:789–804.

Lemey P., Rambaut A., Welch J.J., Suchard M.A. 2010. Phylogeography takes a relaxed random walk in continuous space and time. Mol. Biol. Evol. 27:1877–1885.

Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

Münkemüller T., Boucher F.C., Thuiller W., Lavergne S. 2015. Phylogenetic niche conservatism—common pitfalls and ways forward. Funct. Ecol. 29:627–639.

Nosil P. 2012. Ecological speciation. Oxford Series in Ecology and Evolution. Oxford: Oxford University Press.

Nuismer S.L., Harmon L.J. 2015. Predicting rates of interspecific interaction from phylogenetic trees. Ecol. Lett. 18:17–27.

O'Meara B.C. 2012. Evolutionary inferences from phylogenies: a review of methods. Annu. Rev. Ecol. Evol. Syst. 43:267–285.

Pagel M. 1997. Inferring evolutionary processes from phylogenies. Zool. Scr. 26:331–348.

Paradis E., Claude J., Strimmer K. 2004. Ape: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Pennell M.W., Eastman J.M., Slater G.J., Brown J.W., Uyeda J.C., FitzJohn R.G., Alfaro M.E., Harmon L.J. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216–2218.

Plummer M., Best N., Cowles K., Vines K. 2006. Coda: convergence diagnosis and output analysis for MCMC. R News 6:7–11.

R Core Team. 2016. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Revell L.J., Harmon L.J., Collar D.C. 2008. Phylogenetic signal, evolutionary process, and rate. Syst. Biol. 57:591–601.

Risken H. 1984. The Fokker–Planck equation. Berlin: Springer (Springer Series in Synergetics, vol. 18).

Schluter D. 1988. Estimating the form of natural selection on a quantitative trait. Evolution 42:849–861.

Schluter D. 2000. The ecology of adaptive radiation. (Oxford Series in Ecology and Evolution). Oxford: Oxford University Press.

Simpson G.G. 1944. Tempo and mode in evolution. New York: Columbia University Press.

Soulebeau A., Aubriot X., Gaudeul M., Rouhan G., Hennequin S., Haevermans T., Dubuisson J.-Y., Jabbour F. 2015. The hypothesis of adaptive radiation in evolutionary biology: hard facts about a hazy concept. Org. Divers. Evol. 15:747–761.

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

Uyeda J.C., Hansen T.F., Arnold S.J., Pienaar J. 2011. The million-year wait for macroevolutionary bursts. Proc. Natl. Acad. Sci. USA 108:15908–15913.

Wright S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proceedings of the Sixth International Congress of Genetics 1:356–366.

Wright S. 1945. The differential equation of the distribution of gene frequencies. Proc. Natl. Acad Sci. USA 31:382–389.

Zhang N.R., Siegmund D.O. 2007. A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. Biometrics 63:22–32.