Using the Ornstein-Uhlenbeck process to model the evolution of interacting populations

Krzysztof Bartoszek, Sylvain Glemin, Ingemar Kaj and Martin Lascoux

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-140499

N.B.: When citing this work, cite the original publication.
https://doi.org/10.1016/j.jtbi.2017.06.011

Original publication available at:
https://doi.org/10.1016/j.jtbi.2017.06.011

Copyright: Elsevier
http://www.elsevier.com/
Using the Ornstein–Uhlenbeck process to model the evolution of interacting populations

Krzysztof Bartoszek, Sylvain Glémin, Ingemar Kaj, and Martin Lascoux

September 25, 2018

Abstract

The Ornstein–Uhlenbeck (OU) process plays a major role in the analysis of the evolution of phenotypic traits along phylogenies. The standard OU process includes random perturbations and stabilizing selection and assumes that species evolve independently. However, evolving species may interact through various ecological processes and also exchange genes especially in plants. This is particularly true if we want to study phenotypic evolution among diverging populations within species. In this work we present a straightforward statistical approach with analytical solutions that allows for the inclusion of adaptation and migration in a common phylogenetic framework, which can also be useful for studying local adaptation among populations within the same species. We furthermore present a detailed simulation study that clearly indicates the adverse effects of ignoring migration. Similarity between species due to migration could be misinterpreted as very strong convergent evolution without proper correction for these additional dependencies. Finally, we show that our model can be interpreted in terms of ecological interactions between species, providing a general framework for the evolution of traits between “interacting” species or populations.

Keywords : Migration; Ornstein–Uhlenbeck process; Phylogenetic comparative methods; Species interactions; Trait evolution
1 Introduction

Comparing phenotypic traits or gene expression patterns across species or populations is intrinsically difficult since species or populations are inherently non-independent entities: species are related through phylogenies and populations through gene genealogies (Felsenstein, 1985). This historical component of the variation within and between species will also need to be accounted for if our aim is to understand trait evolution and in particular test whether traits evolved mainly neutrally or are under natural selection.

The first proposed model of continuous trait evolution was the Brownian motion (BM) model (Edwards, 1970; Felsenstein, 1985). This is essentially a neutral model of entirely random evolution with the trait fluctuating around the ancestral value with linearly increasing variance over time. The lack of any stabilization in the process was quickly noticed and hence the Ornstein–Uhlenbeck model was proposed as an alternative (Felsenstein, 1988; Hansen, 1997). Since the seminal papers of Edwards and Felsenstein (Edwards, 1970; Felsenstein, 1985), various forms of the Ornstein-Uhlenbeck process have been used to assess the relative importance of genetic drift and selection in phenotypic trait evolution across species. Initial versions incorporated only random perturbations and stabilizing selection, but recent forms of the OU process now consider simultaneously genetic drift and both stabilizing and directional selection (Beaulieu et al., 2012; Held et al., 2014). Until recently, however, species or populations were assumed to evolve independently, neither interacting nor exchanging genetic material. Since migration between populations will usually tend to homogenize phenotypes directly by sharing the underlying genetic information, the application of such models at the species level (i.e. between populations within species, see discussion by Stone et al., 2011) could pose a serious problem. More generally, hybridization between closely related species or divergence with gene flow during species formation is likely more frequent than initially thought (Abbott et al., 2013), leading to the same problem at the phylogenetic scale where BM and OU models where classically applied. Sometimes, hybridization can even generate new phenotypes not present in the parents. Ecological interactions between species can also affect phenotype evolution if the strength and/or direction of selection on the phenotype of a focal species is affected by the phenotypes of other species. For instance, competition can lead to character displacement and diversification. On the contrary, mimicry, as migration, will tend to homogenize phenotypes. Two recent studies combined
phylogenetic structure and interactions during trait evolution. Nuismer and Harmon (2015) described an interaction framework combined with an OU evolution process. However the interactions between the species occur only through “the covariance among population mean phenotypes for species $i$ and $j$”. Drury et al. (2016) worked further with this model and included species interaction in the “deterministic part” of the stochastic dynamics in their framework. In our work we focused on the migration problem and considered a selection—migration model where different species or populations were allowed to exchange genes thus affecting each others’ primary optima (in the spirit of Hansen, 1997). We showed that Drury et al. (2016)’s model is a special subcase of ours, which can thus also be applied to the species interaction problem. We then used simulations to assess the consequences of ignoring migration when testing for the presence of selection.

Very recently Manceau et al. (2016) proposed a general framework for working with interacting species evolving on a phylogeny. Our proposed here model is a subcase of their general setup. However, our aim is to understand what effects such interactions have and this cannot be readily seen from a general setup. We choose working with the OU framework as currently it is the gold standard in the phylogenetic comparative methods community. Most contemporary analyses are focused around OU–type trait evolution. Therefore, directly expanding it for between–species interactions should be of interest. In our work we explicitly show what effects including and ignoring migration has on estimation under an OU model. For network/interaction based models to become useful in the evolutionary biology community one needs to understand their properties and consequences of assuming/ignoring interactions.

2 Methods

Phylogenetic models of selection with migration

2.1 Selection model

The Ornstein-Uhlenbeck process is defined by the following stochastic differential equation (SDE) for the trait,
\[ dX(t) = -\alpha (X(t) - \theta) \, dt + \sigma dW(t), \quad (1) \]

where \( W(t) \) is the standard Wiener process and \( \alpha \geq 0, \theta \) and \( \sigma > 0 \) are parameters.

In the phylogenetic comparative methods (PCMs) context \( X(t) \in \mathbb{R} \) is usually considered to be the average value of a one-dimensional trait for a species. For each \( t \) the values \( X(t) \) follow a normal distribution supported on the real line. Since natural traits are typically positive and bounded, the Gaussian property raises some concerns about the applicability of this class of random processes to modelling traits. The complementary approach, however, of letting \( X(t) \) represent logarithmic traits rather than traits presents a method to deal with the positivity. Moreover, for \( \alpha > 0 \) the stationary behavior of the process detailed below means that \( X(t) \) with high probability is confined to a bounded interval centered at \( \theta \). This leaves open the issue whether the normality assumption is at all appropriate, but on the other hand the OU model was never intended to perfectly describe the evolution of the traits average but rather capture the most important first-order characteristics (Felsenstein, 1988; Hansen, 1997).

The OU model of Eq. (1) is by now a standard process for PCMs and the interpretation of the individual model parameters has been discussed at length (e.g. Butler and King, 2004; Hansen, 1997; Hansen et al., 2008). Briefly, \( \theta \) is the optimal value for the trait, \( \sigma \) determines the magnitude of genetic drift affecting the trait, and \( \alpha \) controls the speed of the approach to the optimum, i.e. the strength of selection. The neutral case in the sense of absence of selection, \( \alpha = 0 \), corresponds to the Brownian motion model. The expected value (mean) and variance of the process are

\[
\begin{align*}
E[X(t)] &= \theta + e^{-\alpha t} (X_0 - \theta) \\
\text{Var}[X(t)] &= \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha t}) .
\end{align*}
\quad (2)
\]

Letting \( t \) tend to infinity, it follows that over time the trait will stabilize around the optimal value \( \theta \) with stationary oscillations of variance \( \sigma^2/(2\alpha) \). The half–life (Hansen, 1997), \( t_{0.5} = \ln 2/\alpha \), which quantifies how much time is needed to lose half of the ancestral effect, adds further to the interpretation of \( \alpha \).

The trait follows the speciation pattern described by the phylogeny. Traditionally in phylogenetic comparative studies the phylogeny is pre–estimated from molecular sequence data that are assumed to be independent of the
trait. The trait changes along each branch according to the diffusion process in (1) and splits at speciation into two processes. Nearly all current phylogenetic comparative methods assume that the two daughter lineages evolve independently, by making a “no exchange or interaction between species” assumption. However, one can easily argue that such independence is biologically unrealistic: species interact with each other due to e.g. ecological interactions, gene flow or hybridization. In Fig. 1 we present an example of simulation (done by mvSLOUCH Bartoszek et al., 2012) of a BM and OU process on a 50 tip phylogeny.

2.2 Including migration

Assume we have \( n \) species or populations under selective pressure present in the time interval \([0, T]\) and that there is gene flow between some of them. For simplicity and parallelism with other OU models, we will use the term species in the rest of the manuscript, though gene exchange is more likely to occur between populations of the same species. The mathematics of the model is the same whether one studies evolving species or populations, whereas the parameter interpretations and model implications can be different.

If the state of the process at time \( t \) is a vector \( \vec{X}_t \), consisting of the \( n \)
current trait-values, then the “deterministic part” of the state $\tilde{X}_{t+\Delta t}$ after an additional time $\Delta t$ can be described as

$$
\tilde{X}_{t+\Delta t} = \tilde{X}_t - \alpha I(\tilde{X}_t - \tilde{\theta})\Delta t + M \tilde{X}_t \Delta t. \tag{3}
$$

The matrix $M$ describes the gene flow between the $n$ species while the other parameters are the standard OU process parameters (it is worth pointing out now that in this work $M$ cannot be completely arbitrary, see discussion around Eq. 9). From the point of view of the trait the matrix $M$ describes the interactions of the populations’ means for the trait. Taking $\Delta \to 0$, we obtain the ordinary differential equation (ODE)

$$
d\tilde{X}_t = -(\alpha I - M) \tilde{X}_t + \alpha I \tilde{\theta}. \tag{4}
$$

Equivalently, letting $H$ denote the $n \times n$-matrix

$$
H = \alpha I - M, \tag{5}
$$
in which all interactions between the species are coded, we have

$$
d\tilde{X}_t = -H \tilde{X}_t + \alpha I \tilde{\theta}. \tag{6}
$$

Combined with the random noise we obtain our complete model—an OU with migration stochastic process

$$
d\tilde{X}(t) = ( -H \tilde{X}(t) + \alpha I \tilde{\theta})dt + \sigma_x I d\tilde{W}(t), \tag{7}
$$

where $\tilde{W}$ is a vector of $n$ independent Wiener processes. In Eq. (7) we recognize $\tilde{X}(t)$ as a homogeneous linear SDE whose solution is (p. 99 Evans, 2013)

$$
\tilde{X}(t) = e^{-Ht} \tilde{X}(0) + e^{-Ht} \int_0^t e^{Hs} ds \alpha I \tilde{\theta} + e^{-Ht} \sigma_x \int_0^t e^{Hs} d\tilde{W}(s) \tag{8}
$$

From this equation one can immediately see that the process is a normal one and directly obtain the mean and variance–covariance functions which fully characterize its law.

**Remark 2.1** We call the model in (7) “complete” as it is complete in the usual phylogenetic comparative methods sense. Two components of the phylogenetic tree are not covered however. Firstly, the phylogeny is assumed
known, or rather conditioned on. As mentioned already, this is standard in phylogenetic comparative studies. Secondly, the layout of regimes (different levels of $\vec{\theta}$) is pre–estimated. This is also rather standard as methods that estimate the parameters of the OU process and regime layout jointly are only starting to appear (to the best of our knowledge, only Ingram and Mahler, 2013b; Khabbazian et al., 2016, allow for such joint estimation).

We now state explicitly our assumptions about $H$, which in turn restrict $M$. We assume in this work that $H$ has $n$ linearly independent eigenvectors and let $P$ denote an $n \times n$–matrix with these eigenvectors as columns. Then the matrix $H$ is eigendecomposable with representation

$$H = \alpha I - M \equiv P\Lambda P^{-1}, \quad \Lambda = \begin{pmatrix} \Lambda_1 & \cdots & 0 \\ 0 & \ddots & \vdots \\ 0 & \cdots & \Lambda_k \end{pmatrix},$$

(9)

where $\Lambda_i$ is the $n_i \times n_i$ diagonal matrix with eigenvalue $\lambda_i$ of multiplicity $n_i$ as diagonal element, $1 \leq i \leq k$, with $k \leq n$ the number of distinct eigenvalues and $\sum_{i=1}^{k} n_i = n$. It is important to point out that we do not restrict ourselves to real eigenvalues or eigenvectors, we consider them in their full complex generality. This assumption on diagonalizability of $H$ immediately imposes very mild restrictions on $M$. Should $H$ be a defective matrix (i.e. not possess $n$ linearly independent vectors) one can study the model using Jordan block decomposition. However, we do not go in this direction in this work as it would not yield any new insights only clouding the straightforward results provided by the study of the eigenvalues. Furthermore, Jordan block decomposition is numerically unstable and is not recommended in implementations. Therefore alternative numerical procedures should be used for defective matrices. However, then the closed form variance formula of Eq. (11) will not hold.

If one makes a further assumption that $H$ is non–singular, then the inverse matrix $H^{-1} = P\Lambda^{-1}P^{-1}$ exists and Eq. (7) can be re-written as

$$d\vec{X}(t) = -H(\vec{X}(t) - \alpha H^{-1}\vec{\theta})dt + \sigma_I d\vec{W}(t).$$

(10)

In the non–singular case we recognize $\vec{X}(t)$ as a multivariate, $n$–dimensional, OU process whose "optimum value" equals $\alpha H^{-1}\vec{\theta}$. Now, the mean vector and covariance matrix are given by (notice that the variance does not depend on the state value nor optimum Bartoszek et al., 2012)
\[
E[X(t)] = Pe^{-\Lambda t}P^{-1}X_0 + \alpha PA^{-1}(I - e^{-\Lambda t}) P^{-1}\theta
\]
\[
\text{Var}[X(t)] = \frac{\sigma^2}{2} PA^{-1}(I - e^{-2\Lambda t}) P^{-1}.
\] (11)

The remaining case, when \( H \) is singular, amounts to some of its eigenvalues being 0. Assuming that eigenvalue \( \lambda_i = 0 \), we need to consider in Eq. (11) the limit
\[
\Lambda_i^{-1}(1 - e^{-x\Lambda_i t}) \xrightarrow{\Lambda_i \to 0} xt.
\] (12)

Then the mean and variance formulae of Eq. (11) hold with \( xt \) at the corresponding \( n_i \) diagonal entries of the diagonal matrix \( \Lambda^{-1}(I - e^{-x\Lambda t}) \). In fact we will discuss an important submodel where this is the case, namely the model of Drury et al. (2016).

2.3 A symmetric submodel

The model setting laid out above is quite general. Each species tends towards its own optimum (which in particular may be the same as the optimum of all/some other species) with \( \alpha \) controlling the speed of the approach. Along its route it exchanges migrants (hence shares trait values) or more generally interacts with all/some other species as controlled by \( M \), hence \( H \). As a result the model is high–dimensional: with \( n \) species the matrix \( H \) is parametrized by \( n^2 + 1 \) parameters. This number increases further if \( M \) changes over the phylogeny. Usually in a phylogenetic comparative method setting there are only \( n \) measurements – all from contemporary species. Therefore, one needs to add further structure on \( M \) to obtain a feasible framework for analysis, and furthermore make the parameters interpretable.

Let us write out in detail the migration component for species (remembering that in Eq. (7) \( \Delta t \to 0 \))
\[
X_{t+\Delta t,i} = X_{t,i} + \text{selection + noise} - (-m_{i,i}X_{t,i} + m_{i,1}X_{t,1} + \ldots + m_{i,i-1}X_{t,i-1} + m_{i,i+1}X_{t,i+1} + \ldots + m_{i,n}X_{t,n}) \Delta t.
\] (13)

Now let us assume that migration is conservative, i.e. that the number of immigrants and emigrants are equal. This translates to the sum of entries in each row of \( M \) equalling 0, i.e.
\[
m_{i,i} = m_{i,1} + \ldots + m_{i,i-1} + m_{i,i+1} + \ldots + m_{i,n}.
\] (14)
Let us assume furthermore that the exchange of information is symmetric between all species, i.e. \( m_{i,i} \equiv m \) and \( m_{i,j} \equiv m/(n-1) \) for \( i \neq j \). This implies that the elements of \( H \) are given by

\[
[H]_{ij} = \begin{cases} 
\alpha + m & i = j \\
\alpha - \frac{m}{n-1} & i \neq j 
\end{cases}
\]

Such a matrix \( H \) is eigen decomposable – it has the simple eigenvalue \( \alpha \) and an eigenvalue of multiplicity \( n-1 \) equalling \( \alpha + (1 + 1/(n-1))m \). The eigenvectors matrix and its inverse are

\[
P = \begin{bmatrix}
1 & -1 & -1 & -1 & \ldots & -1 \\
1 & 1 & 0 & 0 & \ldots & 0 \\
1 & 0 & 1 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & \ldots & \ldots & \ldots & 1
\end{bmatrix}, \quad
P^{-1} = \begin{bmatrix}
\frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \\
\frac{-1}{n} & \frac{n-1}{n} & \frac{-1}{n} & \frac{-1}{n} & \ldots & \frac{-1}{n} \\
\frac{-1}{n} & \frac{-1}{n} & \frac{n-1}{n} & \frac{-1}{n} & \ldots & \frac{-1}{n} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{-1}{n} & \frac{-1}{n} & \ldots & \ldots & \frac{n-1}{n} & \frac{-1}{n}
\end{bmatrix}.
\]

From these representations one easily obtains the moments of the process, Eq. (11).

The above assumptions on the process may seem restrictive (esp. compared to the Manceau et al., 2016, setup). In particular, the whole migration part is described by a single parameter \( m \). However, as mentioned in the beginning of this section the parameter space has to be significantly reduced to allow for estimability and interpretability. The model described in this Section is certainly interpretable. It perfectly describes the “mean field” corresponding to the Wright island model in population genetics or to a situation where a number of exchangeable species constantly interacts with each other. The magnitude of migration is measured by \( m \). At the species level, this corresponds to a model of local adaptation as analyzed by Hendry et al. (2001). However, they only considered two populations.

### 2.4 Including the phylogeny

So far we have considered the evolution of traits without any branching structure – a fixed number of species evolve and interact within a given time frame. Ideally we would like to model the phylogenetic or populations history setup – where the trait evolves with time and species not only interact but also speciate (or populations split) and new lineages, that also may interact,
arise. A simple example of such a splitting with migration structure is given in Fig. 2.

In order to trace the evolution of traits in the species tree and to derive the law of the contemporary sample we introduce some additional notation. By \( t_k \) we denote the time (the origin of the tree is \( t_0 = 0 \)) of the \( k \)th speciation event, i.e. at time \( t_k \) the number of species changes from \( k \) to \( k + 1 \). Hence \( T_k := t_k - t_{k-1} \) is the duration of time when there were \( k \) lineages present. In particular if the tree has no root branch but starts at the first bifurcation event we will have \( T_1 = t_0 = t_1 = 0 \). One could argue that the speciation events should be related to the trait values, in contrast to our approach conditioned on the phylogeny. Models that allow for trait dependent speciation and extinction rates have been proposed, e.g. QuaSSE (FitzJohn, 2010), but are avoided here to not risk lengthy computation times.

We assume that in each of the \((t_{k-1}, t_k)\) time intervals all parameters are constant. A parameter with subscript \( k \), e.g. \( H_k \), indicates the value of the parameter at the time when there are \( k \) lineages present in the phylogeny. After each speciation event the two daughter lineages inherit their ancestral trait value and the traits in both lineages begin to evolve over time following the assumed OU dynamics with possible interactions. The subscripted trait vector \( \vec{X}_k(t) \) will be used to underline that at time \( t \) there are \( k \) lineages present. Just after speciation the two daughter lineages are identical, hence they have to share the same mean, variance and covariance with other lineages. To formalize the mean and covariance of the phylogeny traits process at this instance, as \( \vec{X}_k(t) \) changes dimension to \( \vec{X}_{k+1}(t) \), we introduce a branching operator \( \upharpoonright \), which increases the size of a vector by one element and the size of a square matrix by adding a row and a column (compare with the notation of Manceau et al., 2016). This operator copies the appropriate entry of the traits’ mean vector and the appropriate row and column of the variance–covariance matrix when speciation happens, resulting in a recursive (in terms of speciation events) formula for the mean and variance.
\[
E \left[ \vec{X}_k(t_k) \right] = e^{-H_k T_k} E \left[ \vec{X}_{k-1}(t_{k-1}) \right] + \left( \alpha H_k^{-1} - P_k (\alpha \Lambda_k^{-1} e^{-\Lambda_k T_k}) P_k^{-1} \right) \vec{\theta}_k \\
= e^{-H_k T_k} E \left[ \vec{X}_{k-1}(t_{k-1}) \right] + \alpha P_k \Lambda_k^{-1} (I_k - e^{-\Lambda_k T_k}) P_k^{-1} \vec{\theta}_k,
\]

(17)

\[
\text{Var} \left[ \vec{X}_k(t_k) \right] = \frac{\sigma^2}{2} H_k^{-1} (I_k - e^{-2H_k T_k}) + e^{-H_k T_k} \text{Var} \left[ \vec{X}_{k-1}(t_{k-1}) \right] e^{-H_k^2 T_k} \\
= \frac{\sigma^2}{2} P_k \Lambda_k^{-1} (I_k - e^{-2\Lambda_k T_k}) P_k^{-1} + e^{-H_k T_k} \text{Var} \left[ \vec{X}_{k-1}(t_{k-1}) \right] e^{-H_k^2 T_k}.
\]

(18)

In principle the operator should be indexed by the phylogeny and speciation number but we suppress this notation. Even though we do not focus on this here the model can easily accommodate extinction events. Upon extinction of a lineage the \( \gamma \) branching operator decreases the dimension by deleting an appropriate value instead of increasing the vector dimension and copying an appropriate value. The index \( k \) would then no longer give the number of lineages in the system but merely indicate the number of completed speciation/extinction events.

### 2.5 Including measurement error and missing observations

Our setting is targeted to maximum likelihood optimizations, which allow for analyzing measurement errors and missing observations.

Trait measurements at the tips of the phylogeny are typically species’ means obtained as averages over many individuals. The resulting intra–species variation adds a sample variance attached to each observation, which is straightforward to include but could be highly misleading to ignore (e.g. Hansen and Bartoszek, 2012). From the vector of observations, \( x_{i,1}, \ldots, x_{i,n_i} \) for species \( i \) one estimates the variance for the \( i \)–th species. However, how exactly one computes such a variance estimate can be analysis specific and we refer the reader for examples of such calculations and related discussions to e.g. Garamszegi (2014); Ives et al. (2007); Martins and Lamont (1998); Rohlfs et al. (2014); Silvestro et al. (2015). These variances are then
Figure 2: An evolving trait on a phylogenetic tree with migrations indicated, by horizontal lines.

added to the diagonal of the variance–covariance matrix of the \( n \) contemporary species, \( \text{Var} \left[ \tilde{X}_n(t_n) \right] \), given by Eq. (18). This is a common approach used in phylogenetic comparative software that allows for measurement error (see e.g. Bartoszek et al., 2012). It is worth reminding the reader that the variance–covariance matrix given by Eq. (18) is not directly estimated from the observed data. Rather the parameters it is dependent on are estimated (by e.g. maximum likelihood) and then the matrix of Eq. (18) is calculated from these parameter estimates.

To correctly handle species with missing measurements, the rows of the expectation vector in Eq. (17), and the rows and columns of the variance–covariance matrix in Eq. (18) that correspond to measurements with missing values, should be removed. In the case of each species being described by a single trait this is the same as just removing the “missing value” species from the data set. However, in the case of a multivariate analysis, when observations are just missing from some traits, there is no need to remove a whole species from the analysis. One just removes the missing components by removing the appropriate rows and columns as described above (Bartoszek
Phylogenetic uncertainty is of course a further issue. On the one hand migration alleviates the problem. On regions where migration occurs one does not need to know the topology. On the other hand, one will need to have a good idea where the migration takes place. One way of dealing with this is to estimate parameters on a collection of trees sampled from some posterior distribution and then take a weighted average of the estimates or use model selection criteria. However, our simulation study here indicates that first one will need to derive conditions for parameter identifiability.

3 Results

3.1 Simulations — effects of migration and selection

To assess the effects of migration we simulated populations of individuals with migration between them. We considered a fixed star phylogeny of species with height equal to 5 and 300 tips. We do not have a phylogenetically structured population but rather a population descending from a common ancestor and all individuals are constantly interacting.

The simulation is described in the online Supplementary material (Section S.2) and we now present the most illustrative histograms of the effects of migration. The complete simulation results can be found in the online Supplementary material. In the simulations we can immediately see that, as expected, the migration parameter $m$ mixes the individuals. When all have the same optimum value, then increasing $m$ brings all individuals closer around the mean, by decreasing the observed trait variance. This effect is thus similar to increasing selection so that neglecting migration could affect selection estimates. In fact this increase of the selection effect is clearly visible in the simulated trajectories presented in Fig. 3. In the contrasting situation of different optima an increase of the migration parameter $m$ pushes individuals away from their optimum and towards an average of these optima. This is again visible in the simulated trajectories of Fig. 3. When migration is weak the trait values still manage to tend towards their optima. However, in the case of strong migration ($m = 10, \alpha = 1$) the traits move away from their optima.

The simulations presented in Figs. 3 and 4 suggest that not taking into consideration migration can seriously bias parameter estimation. We did a
Figure 3: Simulated trajectories on the phylogeny depicted at the top. First column: $m = 0$, second column: $m = 0.5$, third column $m = 10$, first row: all species have $\theta = 5$, second row: half of species have $\theta = -5$, other half $\theta = 5$. The solid line branches have $\theta = -5$ as the optimum while the dashed have $\theta = 5$. In all presented simulations $\alpha = 1$, $X_0 = 0$, $\sigma^2 = 1$. We can clearly observe that migration has an averaging out effect on the “optimum” value.

simulation study to test this and the resulting histograms are presented in Fig. 5, Tab. S. 2 (Supplementary Material) and Figs. S. 3–6 (Supplementary Material). For each parameter combination we simulated 1000 50–tip Yule trees with speciation rate 0.1 (actually with speciation rate 1 and rescaled all branch lengths by 10) using TreeSim. On each tree we simulated a trait
Figure 4: Histograms of estimates of $\theta$ values. First column: $m = 0$, second column: $m = 0.5$, third column $m = 10$, first row: all individuals have $\theta = 5$, second row: half of individuals have $\theta = -5$, other half $\theta = 5$. In all presented simulations $\alpha = 1, X_0 = 0, \sigma^2 = 1$.

process under the law defined by the parameter combination. If $m = 0$ (i.e. usual OU process) we used the mvSLOUCH package and when $m \neq 0$ R code that drew trait values from the law described in this paper. Then, using mvSLOUCH we estimated parameters of the OU process based on the simulated data. The mvSLOUCH package does not take migration into account, hence we can observe the consequence of ignoring migration effects. When there were two $\theta$ values, half (25) the tip species had one value and half the other. We tried to cluster this as far as possible (the trees are random). Nodes with indices 1–25 (in phylo format) had one value and those with indices 26–50 the other. We drew (with equal probability) which half had which value of $\theta$. Afterwards the regimes were painted back on the tree using the Fitch algorithm (Fitch, 1971; Swofford and Maddison, 1987) implemented in the fitch.mvsl() function in mvSLOUCH.
The results of the simulation study show that the consequences of ignoring migration in a phylogenetic comparative study are as those indicated in Fig. 4 (populations evolving under an OU with migration). The histograms clearly show two phenomena. When there is a common optimum, migration has the effect of making adaptation seem faster, i.e. the species are much closer to their optimum (as if $\alpha$ was larger). In the case of multiple optima migration has the effect of masking them. Despite the original optima being distinctly separate (in terms of the model parameters) the simulated species’s trait values are clustering around the average of the optima values. The visible slight bimodality is due to the weak but ever present ($\alpha > 0$) pull towards the true optimum. Estimates of the selection parameter $\alpha$ (Fig. S. 3), moreover, are heavily inflated by the migration parameter $m$. An increase of $m$ brings the observations closer together (compare Figs. 4 and S. 5) — indicating a stronger selection effect, but to a different optimum when there are multiple optima. The difficulty of estimating $\alpha$ for the case of a single optimum and no migration, observed in Cressler et al. (2015), essentially remains under weak and strong migration. However, the accuracy in estimating $\alpha$ is much improved when there are multiple optima, and no migration (Fig. S. 3, bottom left). Intuitively, when there is a switching of optima, especially near the tips, additional information on $\alpha$ is gained by observing how quickly the trait moves from one optimum to the other. Migratory exchange again flattens the histograms and drastically reduces the quality of the $\alpha$-estimates. A similar scenario applies for estimation of $\sigma^2$ (Fig. S. 4). The presence of migration reduces the quality of the parameter estimates whereas multiple optima improves precision.

In OU phylogenetic models it is often preferable to estimate the stationary variance of the process, $v_y := \sigma^2/(2\alpha)$, rather than $\alpha$ and $\sigma^2$ separately.

It is known that the sample mean around a single optimum $\theta$ has Gaussian fluctuations of variance $v_y$, in the usual manner of the central limit theorem (under strong selection, $\alpha > 1/2$ Bartoszek and Sagitov, 2015). Here, similarly, the parameter $v_y$ is well estimated, using $\alpha = 1$ and $m = 0$ (Fig. S. 5, left column). However when we increase migration the estimate of $v_y$ can differ significantly. The maximum likelihood estimation procedure tries to fit such ($\hat{\alpha}, \hat{\sigma}^2$) values so that $\hat{v}_y$ will correspond to the variance of the contemporary sample. As migration significantly decreases the variability around the mean (see Fig. S. 6) the estimation procedure (unaware of migration) will fixate in a parameter region where $v_y$ is small.

Estimating the value of the optimum parameter, $\theta$, was one of the original
motivations behind phylogenetic comparative methods (Hansen, 1997). This parameter (or vector of parameters if there are multiple optima) is usually straightforward to estimate (Bartoszek et al., 2012; Bartoszek and Sagitov, 2015; Cressler et al., 2015). When there is a single optimum, migration furthermore improves the precision of $\theta$-estimates by decreasing sample variance and increasing the speed of convergence of the mean. For multiple optima (Fig. 5, bottom row) it is essential to interpret the migration effect in order to correctly identify the different optima. The averages of the subsamples (i.e. those under the different regimes) are much closer together, as migration mixes the individuals of the different populations. With very strong migration (Fig. 5, bottom right) we observe an interesting effect: the ancestral and selective regimes interact with each other.

When the ancestor has the trait of $\theta_1 = -5$ then the estimates of the parameter pair $(\hat{\theta}_1, \hat{\theta}_2)$ scatter around $(-1, 0)$. Were there no ancestral effects we would expect to see the estimates around $(0, 0)$ (Fig. 4), as migration is symmetric between all species. The lagging ancestral effect results in the end from the interplay between the averaging effect of the two parameters and the restraints of initial trait value and weak (compared to migration) selection effects.

In summary, ignoring migration will certainly have a large impact on the identification of the strength of selection and the magnitude of disruptive effects. If populations are mixing and are otherwise in separate selective regimes it is impossible to estimate the optima of these regimes correctly. However, if migration is only between individuals inside the same selective regime, then identification of its optimum value is significantly easier.

It is important to notice that in the modelling setup considered here the layout of the regimes is known. Only the values of the optima are considered unknown. The interesting modeling case where the layout is not known but is estimated from the data, presents a very complex problem, with possible identifiability issues (see e.g. Ingram and Mahler, 2013b; Khabbazian et al., 2016). Allowing for modelling the regime layout is a further development step.

Our results here indicate a potential issue with methods for identifying regimes. If migration is present, but ignored, then it could mask adaptation to an optimum. As we saw in Fig. 4 the two populations are clustering around the average of the two optima. In consequence software using model selection criteria could collapse these two regimes as one. Even if it correctly identified two regimes (e.g. Fig. 4 center column, bottom row) the esti-
Figure 5: Histograms of estimates of θ. First column: m = 0, second column: m = 0.5, third column m = 10, first row: all individuals have θ = 5, second row: half of individuals have θ = −5 (black estimates), other half θ = 5 (gray estimates). In all presented histograms α = 1, X0 = 0, σ² = 1. In the second row we plot separately the θ = −5 individuals (black) and θ = 5 individuals (gray). The vertical lines represent the true values. In the case of strong migration (bottom right figure) we do not present the histogram, but provide a scatter plot where each dot represents the pair of actually estimated parameters in each of the runs. Note that the histograms in the two rows are not on the same scale. The left and right-most bars are values equal or more extreme than the appropriate x-value.

Mator of the optimum value would be biased. Regime identification (when migration is ignored) could focus on the means of species returning values that are not biologically meaningful in the sense that they are unrelated to the optimum. Rather some trade-off between selection and migration value would be estimated.

The R code used for the simulation study is available from the corresponding author (KB) on request.
3.2 Can migration be recovered?

Migration has the effect of making the sample look as if it were under very strong selection and hence we check now if including the migration parameter in the estimation procedure will allow one to disentangle the two effects. We therefore simulate 100 times, 50 independent populations of 50 tips all following the law of Eq. (7). In Fig. 6 we show the histograms of the compound parameter $m/(\alpha + m)$ estimated by maximum likelihood (ML) for $m = 0.1, 1, 10$ (the remaining parameters kept fixed at $\alpha = 1, \sigma_x = 1, \theta = 1$, populations' age: 5). However, we can see that as $m$ increases the estimate of $\alpha$ increases and this indicates that there is an intrinsic parameter identifiability problem for population structured data.

In principle we should not expect easy estimation, since obtaining estimates of $\alpha$ is difficult even without the parameter $m$ (Cressler et al., 2015). To obtain consistent ML estimates the model generating the phylogeny has to satisfy specific conditions (Ané et al., 2016). As software for models with interactions between species is becoming available (e.g. Drury et al., 2016; Jhwueng and O’Meara, 2015; Manceau et al., 2016) such identifiability issues will be more common and finding conditions for the phylogeny that allow for separating selection and migration will become even more crucial. Providing these conditions is beyond the scope of our work but intuition from our population model study indicates that the user should have a good a priori idea about which parts of the tree include migration and which ones have migration–free clades (to estimate $\alpha$).

3.3 Comparison with ecological interaction between species

So far we built our model by considering migration between species or populations. However, the structure of the model allows for a wider interpretation as that of interaction between species mediated through their phenotypes, especially ecological interactions. Our model cannot describe all kind of ecological interactions but applies as far as these interactions are described by linear combinations of the phenotype of the different species, which then correspond to a particular matrix $H$.

For instance, Drury et al. (2016) used a special submodel of Eq. (7) to describe competition between evolving species and study Anoles morphology. Their model is the special case $\alpha = 0$, with their $S \equiv m$. However, as they focused on the effect of competition, $S < 0$ in their analysis. This essentially
Figure 6: Histograms of estimates of \( m/(\alpha + m) \) (first row) and of \( \sigma^2/(2\alpha) = 0.5 \) (second row) for the independently evolving population reestimation study. First column: \( m = 0.1 \), second column: \( m = 1 \), third column \( m = 10 \). All species have \( \theta = 0 \), \( \alpha = 1 \), \( \sigma^2 = 1 \) and \( X(0) = 0 \). The gray line is the true parameter value.

is a Brownian motion with migration model (without selection). The process does not have a stationary distribution as one of the eigenvalues of \( H \) is \( \alpha = 0 \). In fact its variance tends to infinity. A slight difference between the model of Eq. (11) and Drury et al. (2016)’s is that they did not assume that all species interact with each other. Instead they considered subsets of species interacting with each other, as in their Fig. 1 where the interaction is between \textit{Anoles} species living on the same island. However this setup means that \( M \) and \( H \) are block–diagonal with different “\( n \)” in different blocks, but same \( \alpha \) and \( m \). Hence the eigen decomposition is the same as that of “our” \( H \) except that it is done independently for each block.

If we assume on the contrary that \( S > 0 \), the model in Drury et al. (2016) can be interpreted as a mimicry model where selection favours a common phenotype among different species, used for example as a warning message
against predators (Mallet and Joron, 1999). For each species $i$, the equivalent ODE can be written as:

$$dX_i(t) = -S \left( X_i(t) - \frac{1}{n} \sum_{j=1}^{n} X_j(t) \right) dt$$  \hspace{1cm} (19)

which corresponds to our equation (6) with $\alpha = 0$ and an $M$ matrix with elements $m_{i,i} = S(1 - \frac{1}{n})$ and $m_{i,j} = -\frac{S}{n}$. We remark that this can be generalized to a weighted mean of phenotypes, hence species-dependent values of $m_{i,j}$, if the different species do not contribute equally to the definition of the optimum (for example as a function of their abundance).

As for competition, this mimicry process does not have a stationary distribution as one of the eigenvalues is zero. However, contrary to the situation in Drury et al. (2016), we should note that as the remaining eigenvalues are positive in this case, there is a stationary component in the mimicry process. If we condition on one species, then the conditional distribution of the remaining ones will tend to stationarity (compare with the mvSLOUCH model Bartoszek et al., 2012). This can be biologically interpreted that if one arbitrarily chooses one species then the remaining ones will “shadow” it with stable fluctuations. This one species defines a randomly changing primary optimum (Hansen et al., 2008) for the others. It is important to notice that this does not imply that the remaining species will be in any way more similar to each other. Rather it implies that the deviations of these species from the arbitrarily chosen focal species will be exhibiting stationary stochastic behaviour. Furthermore, it is important to remember that this limiting distribution may depend on the choice of the focal species, unless all the species are exchangeable in the statistical sense. We are considering fluctuations relative to one of the species.

The mimicry trait can randomly diverge but the variation among species for this trait will reach a stationary distribution. Biologically what is selected for is not the absolute value of the trait but the fact that all species are sufficiently similar.

Our model generalizes these special cases (where $\alpha = 0$) to more realistic conditions where there is stabilizing selection around fixed optima in addition to selection (stabilizing or disruptive) due to species interactions. Moreover, both migration and ecological interaction effects could be encoded in the $H$ matrix, although the two components would not be necessarily identifiable without additional biological information. Our model thus provides
a general framework to study the evolution of traits under selection, random perturbations and species interactions *sensu lato*.

Finally, we also note that Drury et al. (2016) numerically solve a system of ODEs, their Eqs. (3a) and (3b), in order to obtain the mean and covariance structure of the system and in effect are able to calculate the likelihood. However the solution of their ODE system is exactly our Eq. (11). Hence an analytical solution is possible instead of a numerical one. As long as $\mathbf{M}$ has a nice enough structure (e.g. is block diagonal) the solution can be written in closed form. For a general $\mathbf{M}$ the solution will be in closed form up to the eigen decomposition of $\mathbf{H}$ (for which there are numerous numerical libraries).

4 Discussion

4.1 Extending the OU process for interacting species

The Ornstein–Uhlenbeck process has been the model of choice to assess the relative importance of genetic drift and selection in phenotypic trait evolution across species. The availability of large phenotypic datasets – in particular, genomewide transcriptomic datasets, that inherently are equivalent to a myriad of phenotypic traits – has led to a renewed interest in the OU process and its application to phenotypic trait evolution (e.g. Rohlfs et al., 2014; Rohlfs and Nielsen, 2015; Nourmohammad et al., 2015).

Until recently species were assumed to evolve independently of each other in OU models. However, this is not always realistic and new models have started to appear where interactions between species are permitted (Nuismer and Harmon, 2015; Drury et al., 2016; Manceau et al., 2016). These studies were motivated by the presence of ecological interaction between species, while our motivation in the present work stemmed from the occurrence of demographic interactions, mainly the exchange of migrants between species. Hybridization and introgression are now well documented in both plants and animals (Abbott et al., 2013) and have played an important role in speciation and in the resulting species trees. As pointed out recently by Solís-Lemus et al. (2016) gene trees discordances have often been solely attributed to incomplete lineage sorting but gene flow is another possibility that has received less attention. However, ignoring gene flow can lead to the reconstruction of the wrong evolutionary history. This, of course, will also apply when reconstructing phenotypic trait evolution and it is therefore important to integrate
gene flow in addition to ecological interactions in OU models.

An interesting finding of our study is that including ecological interactions or migration among species lead to the same modified OU process (compare our Eq. (7) with see Eq. 2 of Drury et al., 2016). In addition to modeling migration, our model also extends the previous model of species interaction proposed by Drury et al. (2016). Moreover, it also has implications for our ability to estimate the importance of stabilizing selection. Our simulations showed that ignoring migration when it is actually present can lead to severely overestimated selection. Below we discuss some consequences and possible extensions of our results.

The analytical derivations and the simulation study presented here clarify the effect of gene flow in the OU model. The results confirm that the mixing of traits reduces the diversity of species in the sense that species trait mean values tend to a weighted average of the optima of individual species, while the variance of the species traits decays with increasing gene flow. We further demonstrate that it is intrinsically difficult to distinguish the effect of the pull towards the environmental optimum due to selection for a given species from the push effect of the interactions with different species. Hence, analyzing data where gene flow does occur between species with a OU process without a migration estimation procedure could naturally lead to the conclusion that adaptation is rapidly moving to an optimum that is common for all species. In consequence this could be understood as evidence of convergent evolution. For instance, the program SURFACE that explicitly aims at identifying cases of convergent evolution by fitting OU models does not incorporate gene flow in the model (Ingram and Mahler, 2013a). Even other approaches to assess convergence in phenotypic evolution do not explicitly incorporate gene flow or refer to it as a possible contributing factor (Moen et al., 2016). Although, the observation that convergence, for instance in Anolis, is spectacular at a regional scale but is not observed at a larger scale would fit well with an effect of gene flow rather than convergence (references in Moen et al., 2016). In many cases the assumption that gene flow did not occur between species may be reasonable, but the evidence in favor of the existence of gene exchange with congeners in many species, especially in plants, suggests that care is required when drawing inferences based on an OU model that does not consider gene flow. It should also be noted that the trees used in the phylogenetic reconstruction are often based on the assumption that gene flow is absent and on a limited number of markers, often on solely cytoplasmic markers, and that they may therefore simply miss it. The availability of
genomic data and of multiple individuals within species together with an increasing awareness of gene exchange among species and of the limits of the OU model (Cooper et al., 2016) should, however, help alleviate this issue.

4.2 Biological applications

Adaptive migration models should be widely applicable in evolutionary biology. One immediate application is to include population structure in phylogenetic comparative studies. The tips of the tree are usually species but inside them there may be distinct subpopulations that can have diverged significantly if local adaptation is strong (Savolainen et al., 2013). Neither treating them as one (including intra–species variance Hansen and Bartoszek, 2012) or separate entities is fully satisfactory. In the first case one ignores the sub–population structure and assumes common adaptive regimes for all. In the second approach the problem is taken care of, but one ignores the fact that one is looking at a single species. The different sub–populations may mix, reproduce and generally interact. The migration model allows for correct treatment of such situations, with the size of $m$ controlling how much flow there is between the different sub–populations. More generally, the migration matrix could be parametrized using additional information on gene flow among diverging populations, typically using molecular markers. The OU model with migration is also a generalization of the two-population model proposed by Hendry et al. (2001) to analyze local adaptation. First, our model generalizes Hendry et al. (2001) to any number of populations. Second, it allows to include a branching structure and not only a single stationary migration matrix structure. The OU model with migration thus provides an interesting basis for further local adaptation models.

Beyond the effect of migration on trait evolution, we have showed the equivalence with species interaction models and how it can be applied to competition or mimicry. We may also consider the use of an OU with migration model to study e.g. host–parasite or plant–pollinator systems (combining interactions with phylogenies in such systems has been recently considered, e.g. Minoarivelvo et al., 2014; Poisot and Stouffer, 2016). With the existing phylogenetic comparative methods one would need to treat the two clades separately, e.g. modelling the evolution of the host and parasite trait on either of the trees. In a migration setup the two phylogenies can be included together, with a very distant time of coalescence. Then the appropriate host–parasite pairs interact through the $M$ matrix.
4.3 Limits and possible extensions

Currently evolutionary models assume that speciation is an instantaneous event. This might be the case on long–time scales but with recent clades the time during which species split could be comparable with the tree height. Furthermore, phylogenetic tree estimation software, can result in non–binary splitting. Such multiple radiations are commonly interpreted not as true multi–furcations but rather as resulting from a lack of information. In both cases the migration framework allows one to use the tree in a phylogenetic comparative analysis. Around the point of multi–furcation/species splitting then one would model that all involved lineages interact with each other.

All of the above setups require the availability of software to estimate parameters of migration models. However, their inference power still remains to be assessed. With a contemporary sample only, it is already difficult to make inference on the value of $\alpha$ in the absence of migration. One would not expect the addition of migration to make estimation any easier. It is a question whether without fossil measurements it will be possible to separately estimate the effects of $\alpha$ and $m$ or only a combination of them. This immediately leads to the question whether different regimes optima are estimable or whether we can only estimate their linear combinations (species averages). There might be some hope however, as our simulation study indicated that (in the situation of $m = 0$) with multiple regimes $\alpha$ is well identifiable. Also the linear transformation of the optima vector ($\vec{\theta}$) depends only on $\alpha$ and $m$ — hence if one is able to identify them, or rather appropriate ratios of them (see online Supplementary material) then adaptive regimes should be identifiable. All of this, however, requires a careful study alongside inference software development.

Acknowledgments

KB was supported by the Knut and Alice Wallenbergs Foundation. SG was supported by the French CNRS and the Marie Curie IEF Grant “SELF-ADAPT” 623486. ML acknowledges support from the Swedish Research Council: VR grants 2012–04999 and 2015–03797 to which this work is related. The authors thank Dwueng–Chwuan Jhhueng for many helpful comments and suggestions.
References


S.1 Worked 3 species example

Let us consider the simple model as in Fig. S.1.

We will go through all the calculations for this toy 3 species model in detail. Let us write that the length of time interval $i$ is $T_i$. Assume that the gene flow rates are the same everywhere and also let us assume that on each branch we have a constant $\theta$ (see notation in Fig. S.1). We consider the process time interval by time interval.

- On time interval $T_1$ we have only one species so no migration can take place. Therefore the system is described by the SDE...
\[ dX_t^1 = -\alpha(X_t^1 - \theta_1)dt + \sigma dB_t^1 \]
and then at time \( T_1 \)
\[ X_{T_1}^1 = e^{-\alpha T_1} X(0) + (1 - e^{-\alpha T_1})\theta_1 + \sigma e^{-\alpha T_1} \int_0^{T_1} e^{\alpha s} dB_s^1. \]

- On the time interval \( T_2 \) we have two lineages between which there is gene flow. Therefore the system has to be described by a bivariate process

\[ d \begin{bmatrix} X_t^1 \\ X_t^2 \end{bmatrix} = \left( -\begin{bmatrix} \alpha + m & -m \\ -m & \alpha + m \end{bmatrix} \begin{bmatrix} X_t^1 \\ X_t^2 \end{bmatrix} + \begin{bmatrix} \alpha \theta_1 \\ \alpha \theta_2 \end{bmatrix} \right) dt + \sigma d \begin{bmatrix} B_t^1 \\ B_t^2 \end{bmatrix}. \]

To solve it we recall the following eigen decomposition of the “deterministic part” matrix
\[ \begin{bmatrix} \alpha + m & -m \\ -m & \alpha + m \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha & 0 \\ 0 & \alpha + 2m \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}. \]

Then we have at the end of the interval
\[ \begin{bmatrix} X_{T_2}^1 \\ X_{T_2}^2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{-\alpha T_2} & 0 \\ 0 & e^{-\alpha T_2} \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix} \begin{bmatrix} X_{T_1}^1 \\ X_{T_1}^2 \end{bmatrix} 
+ \begin{bmatrix} 1 - e^{-\alpha T_2} \\ 0 \end{bmatrix} \frac{\alpha}{\alpha + 2m} \begin{bmatrix} 0 \\ -e^{-\alpha (\alpha + 2m) T_2} \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} 
+ \sigma \begin{bmatrix} e^{-\alpha T_2} \\ 0 \end{bmatrix} \int_0^{T_2} \begin{bmatrix} 0 \\ e^{\alpha s} \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix} d \begin{bmatrix} B_s^1 \\ B_s^2 \end{bmatrix} \]

\[ = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{-\alpha T_2} & 0 \\ 0 & e^{-\alpha T_2} \end{bmatrix} \begin{bmatrix} X_{T_1}^1 \\ 0 \end{bmatrix} 
+ \frac{1}{2} \begin{bmatrix} 1 - e^{-\alpha T_2} \\ 0 \end{bmatrix} \frac{\alpha}{\alpha + 2m} \begin{bmatrix} 0 \\ -e^{-\alpha (\alpha + 2m) T_2} \end{bmatrix} \begin{bmatrix} \theta_1 + \theta_2 \\ \theta_2 - \theta_1 \end{bmatrix} 
+ \frac{\sigma}{2} \begin{bmatrix} e^{-\alpha T_2} \\ 0 \end{bmatrix} \int_0^{T_2} \begin{bmatrix} 0 \\ e^{\alpha s} \end{bmatrix} \begin{bmatrix} 0 \\ e^{\alpha s} \end{bmatrix} d \begin{bmatrix} B_s^1 + B_s^2 \\ B_s^2 - B_s^1 \end{bmatrix}. \]
\[
\begin{bmatrix}
1 & -1 \\
1 & 1
\end{bmatrix}
\left(\begin{bmatrix}
e^{-\alpha T_2}X_1^1 \\
0
\end{bmatrix} + \frac{1}{2} \begin{bmatrix}
\frac{\alpha}{\alpha + 2m} (1 - e^{-\alpha T_2})(\theta_1 + \theta_2) \\
0
\end{bmatrix}
\right) + \frac{\sigma^2}{2}
\begin{bmatrix}
e^{-\alpha T_2} \\
e^{-(\alpha + 2m)T_2}
\end{bmatrix}
\begin{bmatrix}
\int_0^{T_2} e^{\alpha s}dB_s^1 + \int_0^{T_2} e^{\alpha s}dB_s^2 \\
\int_0^{T_2} e^{(\alpha + 2m)s}dB_s^1 - \int_0^{T_2} e^{(\alpha + 2m)s}dB_s^1
\end{bmatrix}
\]

\[
= \left(\begin{bmatrix}
e^{-\alpha T_2}X_1^1 \\
e^{-\alpha T_2}X_2^1
\end{bmatrix} + \frac{1}{2}
\begin{bmatrix}
(1 - e^{-\alpha T_2})(\theta_1 + \theta_2) \\
(1 - e^{-\alpha T_2})(\theta_1 + \theta_2)
\end{bmatrix}
\right) + \frac{\sigma^2}{2}
\begin{bmatrix}
e^{-\alpha T_2} \\
e^{-(\alpha + 2m)T_2}
\end{bmatrix}
\begin{bmatrix}
\int_0^{T_2} e^{\alpha s}dB_s^1 + \int_0^{T_2} e^{\alpha s}dB_s^2 \\
\int_0^{T_2} e^{(\alpha + 2m)s}dB_s^1 - \int_0^{T_2} e^{(\alpha + 2m)s}dB_s^1
\end{bmatrix}
\]

\[
= \left(\begin{bmatrix}
e^{-\alpha T_1 + \alpha T_2}X_1(0) \\
e^{-\alpha T_1 + \alpha T_2}X_2(0)
\end{bmatrix} + \begin{bmatrix}
(1 - e^{-\alpha T_2})(\theta_1 + \theta_2) \\
(1 - e^{-\alpha T_2})(\theta_1 + \theta_2)
\end{bmatrix}
\right) + \frac{\sigma^2}{2}
\begin{bmatrix}
e^{-\alpha T_1 + \alpha T_2} \\
e^{-\alpha T_1 + \alpha T_2}
\end{bmatrix}
\begin{bmatrix}
\int_0^{T_1} e^{\alpha s}dB_s^1 \\
\int_0^{T_1} e^{\alpha s}dB_s^1
\end{bmatrix}
\begin{bmatrix}
\int_0^{T_2} e^{\alpha s}dB_s^1 + \int_0^{T_2} e^{\alpha s}dB_s^2 \\
\int_0^{T_2} e^{(\alpha + 2m)s}dB_s^1 - \int_0^{T_2} e^{(\alpha + 2m)s}dB_s^1
\end{bmatrix}
\]

• On the time interval \(T_3\) all three lineages are interacting hence the system has to be described by a trivariate process.

\[
d\begin{bmatrix}
X_1^1 \\
X_2^2 \\
X_3^3
\end{bmatrix} = \left(\begin{bmatrix}
\alpha + m & -m/2 & -m/2 \\
-m/2 & \alpha + m & -m/2 \\
-m/2 & -m/2 & \alpha + m
\end{bmatrix}
\right)
\begin{bmatrix}
X_1^1 \\
X_2^2 \\
X_3^3
\end{bmatrix} + \begin{bmatrix}
\alpha \theta_1 \\
\alpha \theta_2 \\
\alpha \theta_3
\end{bmatrix}
\]

To solve it we recall the following eigen decomposition of the “deterministic part” matrix.
and the solution at the end of the interval is ($\beta := \alpha + 3m/2$)

$$
\begin{bmatrix}
X^1_{T_3} \\
X^2_{T_3} \\
X^3_{T_3}
\end{bmatrix} = 
\begin{bmatrix}
1 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
-\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\
-\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3}
\end{bmatrix}
\begin{bmatrix}
X^1_{T_2} \\
X^2_{T_2} \\
X^3_{T_2}
\end{bmatrix}
$$

We can see that in the same fashion as above we can plug in the ancestral values to obtain the solution at the tips. We have not done this here as the formula would be too long to
be readable. After this obtaining the mean, variance and covariance will be immediate as the Brownian motions are independent on disjoint intervals. In particular we may write the expectation as

\[
\begin{align*}
E \left[ \begin{bmatrix} X^1_{T_2} \\ X^2_{T_2} \\ X^3_{T_3} \end{bmatrix} \right] &= \frac{1}{3} \left( \begin{bmatrix}
(e^{-\alpha T_3} + 2e^{-\beta T_3}) E \left[ X^1_{T_2} \right] + 2(e^{-\alpha T_3} - e^{-\beta T_3}) E \left[ X^2_{T_2} \right] \\
(e^{-\alpha T_3} - e^{-\beta T_3}) E \left[ X^1_{T_2} \right] + (2e^{-\alpha T_3} + e^{-\beta T_3}) E \left[ X^2_{T_2} \right] \\
(e^{-\alpha T_3} - e^{-\beta T_3}) E \left[ X^1_{T_2} \right] + (2e^{-\alpha T_3} + e^{-\beta T_3}) E \left[ X^2_{T_2} \right] 
\end{bmatrix} \right) \\
&+ \left[ \begin{bmatrix}
(1 - e^{-\alpha T_3} + 2\frac{\alpha}{\alpha+2m}(1 - e^{-\beta T_3}))\theta_1 + (1 - e^{-\alpha T_3} - \frac{\alpha}{\alpha+2m}(1 - e^{-\beta T_3}))\theta_2 + (1 - e^{-\alpha T_3} - \frac{\alpha}{\alpha+2m}(1 - e^{-\beta T_3}))\theta_3 \\
(1 - e^{-\alpha T_3} - \frac{\alpha}{\alpha+2m}(1 - e^{-\beta T_3}))\theta_1 + (1 - e^{-\alpha T_3} + 2\frac{\alpha}{\alpha+2m}(1 - e^{-\beta T_3}))\theta_2 + (1 - e^{-\alpha T_3} - \frac{\alpha}{\alpha+2m}(1 - e^{-\beta T_3}))\theta_3 \\
(1 - e^{-\alpha T_3} - \frac{\alpha}{\alpha+2m}(1 - e^{-\beta T_3}))\theta_1 + (1 - e^{-\alpha T_3} - \frac{\alpha}{\alpha+2m}(1 - e^{-\beta T_3}))\theta_2 + (1 - e^{-\alpha T_3} + 2\frac{\alpha}{\alpha+2m}(1 - e^{-\beta T_3}))\theta_3 
\end{bmatrix} \right]
\end{align*}
\]

where

\[
E \left[ \begin{bmatrix} X^1_{T_2} \\ X^2_{T_2} \end{bmatrix} \right] = \begin{bmatrix} e^{-(\alpha T_2 + \alpha T_2)} X(0) \\ e^{-(\alpha T_2 + \alpha T_2)} X(0) \end{bmatrix} + \begin{bmatrix} (e^{-\alpha T_2} - e^{-\alpha (T_2 + T_2)})\theta_1 \\ (e^{-\alpha T_2} - e^{-\alpha (T_2 + T_2)})\theta_1 \end{bmatrix}
\]

\[
+ \frac{1}{2} \begin{bmatrix} (1 - e^{-\alpha T_2} + \frac{\alpha}{\alpha+2m}(1 - e^{-\beta (T_2 + T_2)})\theta_1 + (1 - e^{-\alpha T_2} - \frac{\alpha}{\alpha+2m}(1 - e^{-\beta (T_2 + T_2)})\theta_2 \\
(1 - e^{-\alpha T_2} - \frac{\alpha}{\alpha+2m}(1 - e^{-\beta (T_2 + T_2)})\theta_1 + (1 - e^{-\alpha T_2} + 2\frac{\alpha}{\alpha+2m}(1 - e^{-\beta (T_2 + T_2)})\theta_2 \\
(1 - e^{-\alpha T_2} - \frac{\alpha}{\alpha+2m}(1 - e^{-\beta (T_2 + T_2)})\theta_1 + (1 - e^{-\alpha T_2} - \frac{\alpha}{\alpha+2m}(1 - e^{-\beta (T_2 + T_2)})\theta_2 
\end{bmatrix}
\]

If we take $T_2 \to \infty$, i.e. we assume that we have a pair of species exchanging gene flow forever without speciating we obtain the stationary expectation

\[
E \left[ \begin{bmatrix} X^1_\infty \\ X^2_\infty \end{bmatrix} \right] = \begin{bmatrix} (1 + \frac{\alpha}{\alpha+2m})\theta_1 + (1 - \frac{\alpha}{\alpha+2m})\theta_2 \\ (1 - \frac{\alpha}{\alpha+2m})\theta_1 + (1 + \frac{\alpha}{\alpha+2m})\theta_2 \end{bmatrix}
\]

If $m = 0$ then we recover $E \left[ \begin{bmatrix} X^1_\infty \\ X^2_\infty \end{bmatrix} \right] = [\theta_1 \theta_2]^T$ as expected. The variance $\text{Var} \left[ \begin{bmatrix} X^1_{T_3} \\ X^2_{T_3} \end{bmatrix} \right]$ can be written as,

\[
\frac{1}{3} \left( \begin{bmatrix}
(e^{-\alpha T_3} + 2e^{-\beta T_3}) \\
(e^{-\alpha T_3} - e^{-\beta T_3}) \\
(e^{-\alpha T_3} - e^{-\beta T_3}) 
\end{bmatrix} \text{Var} \left[ \begin{bmatrix} X^1_{T_2} \\ X^2_{T_2} \end{bmatrix} \right] \left( \begin{bmatrix}
(1 + \frac{\alpha}{\alpha+2m})\theta_1 + (1 - \frac{\alpha}{\alpha+2m})\theta_2 \\
(1 - \frac{\alpha}{\alpha+2m})\theta_1 + (1 + \frac{\alpha}{\alpha+2m})\theta_2 \end{bmatrix} \right) \right)
\]

\[
+ \sigma^2 \text{Var} \left[ \begin{bmatrix}
\int_0^{T_3} e^{-\alpha (T_3 - s)} - e^{\alpha (T_3 - s)} dB^1_s + \int_0^{T_3} e^{\alpha (T_3 - s)} - e^{-\alpha (T_3 - s)} dB^2_s \\
\int_0^{T_3} e^{-\alpha (T_3 - s)} - e^{\alpha (T_3 - s)} dB^1_s + \int_0^{T_3} e^{\alpha (T_3 - s)} - e^{-\alpha (T_3 - s)} dB^2_s \\
\int_0^{T_3} e^{-\alpha (T_3 - s)} - e^{\alpha (T_3 - s)} dB^1_s + \int_0^{T_3} e^{\alpha (T_3 - s)} - e^{-\alpha (T_3 - s)} dB^2_s 
\end{bmatrix} \right]
\]

where

\[
\text{Var} \left[ \begin{bmatrix} X^1_{T_2} \\ X^2_{T_2} \end{bmatrix} \right] = \sigma^2 \text{Var} \left[ \begin{bmatrix}
\int_0^{T_3} e^{\alpha (s-T_2)} dB^1_s + \int_0^{T_3} e^{\alpha (s-T_2)} dB^2_s \\
\int_0^{T_3} e^{-\alpha (s-T_2)} dB^1_s + \int_0^{T_3} e^{-\alpha (s-T_2)} dB^2_s 
\end{bmatrix} \right]
\]

\[
+ \sigma^2 \text{Var} \left[ \begin{bmatrix}
\int_0^{T_2} e^{\alpha (s-T_2)} dB^1_s + \int_0^{T_2} e^{\alpha (s-T_2)} dB^2_s \\
\int_0^{T_2} e^{-\alpha (s-T_2)} dB^1_s + \int_0^{T_2} e^{-\alpha (s-T_2)} dB^2_s 
\end{bmatrix} \right]
\]

The above can be calculated explicitly. First
\[
\sigma^2 \text{Var} \left[ e^{-\alpha(T_1+T_2)} \int_0^{T_1} e^{\alpha s} dB_s^T \right] = \frac{\sigma^2}{2\alpha} \left( e^{-2\alpha T_2} - e^{-2\alpha(T_1+T_2)} \right) \xrightarrow{T_2 \to \infty} 0.
\]

Now we use the Itô isometry for the second part of the variance.

\[
A_{T_2} = \text{Var} \left[ \int_0^{T_2} e^{\alpha(s-T_2)} + e^{(\alpha+2m)(s-T_2)} dB_s^T \right] = e^{-2\alpha T_2} \text{Var} \left[ \int_0^{T_2} e^{\alpha s} + e^{-2mT_2} e^{(\alpha+2m)s} dB_s^T \right]
\]
\[
= e^{-2\alpha T_2} T_2 \int_0^{T_2} (e^{\alpha s} + e^{-2mT_2} e^{(\alpha+2m)s})^2 ds = e^{-2\alpha T_2} T_2 \int_0^{T_2} e^{2\alpha s} + 2e^{-2mT_2} e^{2(\alpha+m)s} + e^{-4mT_2} e^{2(\alpha+2m)s} ds
\]
\[
= e^{-2\alpha T_2} \left( \frac{1}{2\alpha} (e^{2\alpha T_2} - 1) + \frac{e^{-2\alpha T_2}}{\alpha + m} (e^{2(\alpha+m)T_2} - 1) + \frac{e^{-4mT_2}}{2(\alpha+2m)} (e^{2(\alpha+2m)T_2} - 1) \right)
\]
\[
\xrightarrow{T_2 \to \infty} \frac{1}{2\alpha} (1 - e^{-2\alpha T_2}) + \frac{1}{\alpha + m} (1 - e^{-2(\alpha+m)T_2}) + \frac{1}{2(\alpha+2m)} (1 - e^{-2(\alpha+2m)T_2})
\]

Then

\[
B_{T_2} = \text{Var} \left[ \int_0^{T_2} e^{\alpha(s-T_2)} - e^{(\alpha+2m)(s-T_2)} dB_s^T \right]
\]
\[
= \frac{1}{2\alpha} (1 - e^{-2\alpha T_2}) - \frac{1}{\alpha + m} (1 - e^{-2(\alpha+m)T_2}) + \frac{1}{2(\alpha+2m)} (1 - e^{-2(\alpha+2m)T_2})
\]
\[
\xrightarrow{T_2 \to \infty} \frac{1}{2\alpha} - \frac{1}{\alpha + m} + \frac{1}{2(\alpha+2m)}
\]

\[
C_{T_2} = \text{Cov} \left[ \int_0^{T_2} e^{\alpha(s-T_2)} + e^{(\alpha+2m)(s-T_2)} dB_s^T, \int_0^{T_2} e^{\alpha(s-T_2)} - e^{(\alpha+2m)(s-T_2)} dB_s^T \right]
\]
\[
= e^{-2\alpha T_2} T_2 \int_0^{T_2} e^{2\alpha s} - e^{-4mT_2} e^{2(\alpha+2m)s} ds = e^{-2\alpha T_2} \left( \frac{1}{2\alpha} (e^{2\alpha T_2} - 1) - \frac{e^{-4mT_2}}{2(\alpha+2m)} (e^{2(\alpha+2m)T_2} - 1) \right)
\]
\[
= \frac{1}{2\alpha} (1 - e^{-2\alpha T_2}) - \frac{1}{2(\alpha+2m)} (1 - e^{-2(\alpha+2m)T_2}) \xrightarrow{T_2 \to \infty} \frac{1}{2\alpha} - \frac{1}{2(\alpha+2m)}
\]

We can see with no migration we recover that both species are independent with stationary variance \( \sigma^2/(2\alpha) \) as expected. The covariance between the two species depends on the values of \( \alpha \) and \( m \) and as expected is always positive. It increases with the increase of the migration rate.

In the three dimensional case as \( T_3 \to \infty \) we can also calculate the limits. We can immediately see that the mean value will be in the limit

\[
\text{E} \left[ \begin{bmatrix} X_{T_2}^1 \\ X_{T_2}^2 \\ X_{T_2}^3 \\ X_{T_3}^1 \\ X_{T_3}^2 \\ X_{T_3}^3 \end{bmatrix} \right] \xrightarrow{T_3 \to \infty} \begin{bmatrix} (1 + \frac{2\alpha}{\beta}) \theta_1 + (1 - \frac{\alpha}{\beta}) \theta_2 + (1 - \frac{\alpha}{\beta}) \theta_3 \\ (1 - \frac{\alpha}{\beta}) \theta_1 + (1 + \frac{2\alpha}{\beta}) \theta_2 + (1 - \frac{\alpha}{\beta}) \theta_3 \\ (1 - \frac{\alpha}{\beta}) \theta_1 + (1 - \frac{\alpha}{\beta}) \theta_2 + (1 + \frac{2\alpha}{\beta}) \theta_3 \end{bmatrix}.
\]

Using the same techniques as in the two dimensional case we calculate that the non-vanishing part of the covariance matrix will equal

\[
\begin{bmatrix}
\frac{1}{\alpha} + \frac{1}{\alpha + 2m} & -\frac{1}{\alpha + \alpha + 2m} & \frac{1}{\alpha + m} \\
\frac{1}{\alpha + \alpha + 2m} & \frac{1}{\alpha} & \frac{1}{\alpha + \alpha + 2m} \\
\frac{1}{\alpha + m} & \frac{1}{\alpha + \alpha + 2m} & \frac{1}{\alpha + m + \alpha + 2m} \\
\end{bmatrix}
\]
\[
\sigma^2 = \frac{1}{9} \begin{bmatrix}
\frac{3}{\beta^2} (1 - e^{-\alpha T_3}) + \frac{3}{\beta^2} (1 - e^{-\theta T_3}) \\
\frac{2}{\beta^2} (1 - e^{-\alpha T_3}) - \frac{2}{\beta^2} (1 - e^{-\theta T_3}) \\
\frac{1}{\beta^2} (1 - e^{-\theta T_3}) + \frac{1}{\beta^2} (1 - e^{-\alpha T_3})
\end{bmatrix}
\]

becoming in the limit

\[
\sigma^2 = \left[ \frac{1}{3\cdot2\alpha} \begin{bmatrix}
\frac{1}{3\cdot2\alpha} + \frac{2}{3\cdot2\alpha} & \frac{1}{3\cdot2\alpha} & \frac{1}{3\cdot2\alpha} - \frac{1}{3\cdot2\alpha} \\
\frac{1}{3\cdot2\alpha} & -\frac{2}{3\cdot2\alpha} & \frac{1}{3\cdot2\alpha} + \frac{2}{3\cdot2\alpha} \\
\frac{1}{3\cdot2\alpha} & \frac{2}{3\cdot2\alpha} & \frac{1}{3\cdot2\alpha} - \frac{1}{3\cdot2\alpha}
\end{bmatrix} \right].
\]

As \( \beta = \alpha + 3m/2 \) then when \( m = 0 \) one obtains the limiting variance as \( \sigma^2/(2\alpha) \) and the limiting covariance 0. Also for \( m > 0 \) the covariance is positive as expected.

Figure S.2: Simulation of migration process on a three species phylogeny, see Fig. S.1. First column: \( m = 0 \), second column: \( m = 0.5 \), third column \( m = 10 \) In all presented histograms \( \alpha = 1, X_0 = 0, \sigma^2 = 1 \) and \( \theta_1 = -5, \theta_2 = 0, \theta_3 = 5 \). The tree has height 3 as \( T_1 = T_2 = T_3 = 1 \).
S.2 Simulation description

We draw \( n = 300 \) trait values from a normal distribution with mean vector and variance–covariance matrix defined by Eqs. (7) and (8) from the main paper, for different parameter values. Time was always assumed as 5, the initial value 0, \( \sigma^2 = 1 \). The optimum \( \theta \) was either constant at 0 or 5, or half the population had \( \theta = 0 \) or \( \theta = -5 \) while the other half had \( \theta = 5 \). The adaptation parameter varied as \( \alpha \in \{0.5, 1, 10\} \) and the migration parameter \( m \in \{0, 0.1, 0.5, 1, 10\} \). Here, \( m \) is an arbitrary value on the trait scale. For the special case \( 0 \leq m \leq 1 \) this is consistent with the common interpretation of \( m \) as a proportion of individuals coming from abroad.

<table>
<thead>
<tr>
<th>Setup id</th>
<th>( \alpha )</th>
<th>( m )</th>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
<th>Setup id</th>
<th>( \alpha )</th>
<th>( m )</th>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0.5</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>10</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0.5</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>0.5</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>18</td>
<td>10</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>19</td>
<td>0.5</td>
<td>0.1</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>1</td>
<td>0.1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>21</td>
<td>10</td>
<td>0.1</td>
<td>5</td>
<td>5</td>
<td>22</td>
<td>0.5</td>
<td>0.5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>0.5</td>
<td>5</td>
<td>5</td>
<td>24</td>
<td>10</td>
<td>0.5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>25</td>
<td>0.5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>26</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>27</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>28</td>
<td>0.5</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>30</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>31</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>32</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>33</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>34</td>
<td>0.5</td>
<td>0.1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>0.1</td>
<td>0</td>
<td>5</td>
<td>36</td>
<td>10</td>
<td>0.1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>37</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>5</td>
<td>38</td>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>39</td>
<td>10</td>
<td>0.5</td>
<td>0</td>
<td>5</td>
<td>40</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>41</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>42</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>43</td>
<td>0.5</td>
<td>10</td>
<td>0</td>
<td>5</td>
<td>44</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>45</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>5</td>
<td>46</td>
<td>0.5</td>
<td>0</td>
<td>-5</td>
<td>5</td>
</tr>
<tr>
<td>47</td>
<td>1</td>
<td>0</td>
<td>-5</td>
<td>5</td>
<td>48</td>
<td>10</td>
<td>0</td>
<td>-5</td>
<td>5</td>
</tr>
<tr>
<td>49</td>
<td>0.5</td>
<td>0.1</td>
<td>-5</td>
<td>5</td>
<td>50</td>
<td>1</td>
<td>0.1</td>
<td>-5</td>
<td>5</td>
</tr>
<tr>
<td>51</td>
<td>10</td>
<td>0.1</td>
<td>-5</td>
<td>5</td>
<td>52</td>
<td>0.5</td>
<td>0.5</td>
<td>-5</td>
<td>5</td>
</tr>
<tr>
<td>53</td>
<td>1</td>
<td>0.5</td>
<td>-5</td>
<td>5</td>
<td>54</td>
<td>10</td>
<td>0.5</td>
<td>-5</td>
<td>5</td>
</tr>
<tr>
<td>55</td>
<td>0.5</td>
<td>1</td>
<td>-5</td>
<td>5</td>
<td>56</td>
<td>1</td>
<td>1</td>
<td>-5</td>
<td>5</td>
</tr>
<tr>
<td>57</td>
<td>10</td>
<td>1</td>
<td>-5</td>
<td>5</td>
<td>58</td>
<td>0.5</td>
<td>10</td>
<td>-5</td>
<td>5</td>
</tr>
<tr>
<td>59</td>
<td>1</td>
<td>10</td>
<td>-5</td>
<td>5</td>
<td>60</td>
<td>10</td>
<td>10</td>
<td>-5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table S.1: Simulation setups, population size \( n = 300 \). \( \theta_1 \) is the optimum for individuals \( 1, \ldots, 150 \) and \( \theta_2 \) is the optimum for individuals \( 151, \ldots, 300 \). In all simulations \( X_0 = 0 \) and \( \sigma^2 = 1 \). In the following histograms we present the trait values inside the simulated population for all the setups.
<table>
<thead>
<tr>
<th>Setup</th>
<th>Parameter</th>
<th>Mean</th>
<th>Variance</th>
<th>MSE</th>
<th>rbias</th>
<th>rMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\alpha = 1$</td>
<td>5.154</td>
<td>557.168</td>
<td>574.425</td>
<td>4.154</td>
<td>574.425</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2 = 1$</td>
<td>4.421</td>
<td>260.531</td>
<td>272.235</td>
<td>3.421</td>
<td>272.235</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2/(2\alpha) = 0.5$</td>
<td>0.486</td>
<td>0.011</td>
<td>0.011</td>
<td>−0.028</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta} = 5$</td>
<td>5.004</td>
<td>0.01</td>
<td>0.01</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>0.497</td>
<td>0.012</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$\alpha = 1$</td>
<td>101.076</td>
<td>258167.8</td>
<td>268182.9</td>
<td>100.076</td>
<td>268182.9</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2 = 1$</td>
<td>64.541</td>
<td>103060.5</td>
<td>107097.9</td>
<td>63.541</td>
<td>107097.9</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2/(2\alpha) = 0.5$</td>
<td>0.324</td>
<td>0.004</td>
<td>0.035</td>
<td>−0.352</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta} = 5$</td>
<td>5.004</td>
<td>0.011</td>
<td>0.011</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>0.33</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\alpha = 1$</td>
<td>141.391</td>
<td>841122.1</td>
<td>860831.8</td>
<td>140.391</td>
<td>860831.8</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2 = 1$</td>
<td>11.674</td>
<td>4906.333</td>
<td>5020.262</td>
<td>10.674</td>
<td>5020.262</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2/(2\alpha) = 0.5$</td>
<td>0.044</td>
<td>0.208</td>
<td>−0.912</td>
<td>0.416</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta} = 5$</td>
<td>5.004</td>
<td>0.011</td>
<td>0.011</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>0.045</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$\alpha = 1$</td>
<td>0.99</td>
<td>0.022</td>
<td>0.022</td>
<td>−0.01</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2 = 1$</td>
<td>0.939</td>
<td>0.049</td>
<td>0.053</td>
<td>−0.061</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2/(2\alpha) = 0.5$</td>
<td>0.523</td>
<td>0.452</td>
<td>0.452</td>
<td>0.046</td>
<td>0.905</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_1 = 5$</td>
<td>−5.22</td>
<td>7.223</td>
<td>7.271</td>
<td>0.044</td>
<td>−1.454</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_2 = 5$</td>
<td>5.02</td>
<td>0.214</td>
<td>0.215</td>
<td>0.004</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>2.622</td>
<td>4.934</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\sigma}^2$</td>
<td>2.674</td>
<td>4.928</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>$\alpha = 1$</td>
<td>998.951</td>
<td>5.915 \cdot 10^8</td>
<td>5.925 \cdot 10^8</td>
<td>997.951</td>
<td>5.925 \cdot 10^8</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2 = 1$</td>
<td>2167.767</td>
<td>2.199 \cdot 10^9</td>
<td>2.203 \cdot 10^9</td>
<td>2166.767</td>
<td>2.203 \cdot 10^9</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2/(2\alpha) = 0.5$</td>
<td>0.994</td>
<td>0.338</td>
<td>0.582</td>
<td>0.988</td>
<td>1.164</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_1 = 5$</td>
<td>−3.061</td>
<td>0.253</td>
<td>4.011</td>
<td>−0.388</td>
<td>−0.802</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_2 = 5$</td>
<td>2.996</td>
<td>0.257</td>
<td>4.271</td>
<td>−0.401</td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>1.036</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\sigma}^2$</td>
<td>1.099</td>
<td>0.831</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>$\alpha = 1$</td>
<td>62750.66</td>
<td>3.244 \cdot 10^{12}</td>
<td>3.248 \cdot 10^{12}</td>
<td>62749.66</td>
<td>3.248 \cdot 10^{12}</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2 = 1$</td>
<td>6703.932</td>
<td>3.814 \cdot 10^{10}</td>
<td>3.819 \cdot 10^{10}</td>
<td>6702.932</td>
<td>3.819 \cdot 10^{10}</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2/(2\alpha) = 0.5$</td>
<td>0.044</td>
<td>0</td>
<td>0.208</td>
<td>−0.911</td>
<td>0.415</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_1 = 5$</td>
<td>−0.459</td>
<td>0.248</td>
<td>20.873</td>
<td>−0.908</td>
<td>−4.175</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_2 = 5$</td>
<td>0.43</td>
<td>0.248</td>
<td>21.137</td>
<td>−0.914</td>
<td>4.227</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>0.047</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\sigma}^2$</td>
<td>0.047</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table S.2: Summary of simulation with reestimation by mvSLOUCH results. For each setup and parameter we report the mean, variance, mean square error (MSE), relative bias (rbias) and relative MSE (rMSE) of the estimates. The means are visualized on histograms. The parameter $\sigma^2$ is the sample variance, $\hat{\sigma}^2$ refers to the sample variance of $\theta = 5$ individuals, $\hat{\sigma}^2_2$ to $\theta = 5$ individuals. Migration is set to $m = 0$ (setup 1 and 4), $m = 0.5$ (setup 2 and 5), or $m = 10$ (setup 3 and 6).
Figure S.3: Histograms of estimates of $\alpha$. First column: $m = 0$, second column: $m = 0.5$, third column $m = 10$, first row: all individuals have $\theta = 5$, second row: half of individuals have $\theta = -5$, other half $\theta = 5$. In all presented histograms $\alpha = 1$, $X_0 = 0$, $\sigma^2 = 1$. The vertical line represents the true value. Note that the histograms are not all on the same scale. The left and right–most bars are values equal or more extreme than the appropriate $x$–value.
Figure S.4: Histograms of estimates of $\sigma^2$. First column: $m = 0$, second column: $m = 0.5$, third column $m = 10$, first row: all individuals have $\theta = 5$, second row: half of individuals have $\theta = -5$, other half $\theta = 5$. In all presented histograms $\alpha = 1$, $X_0 = 0$, $\sigma^2 = 1$. The vertical line represents the true value. Note that the histograms are not all on the same scale. The left and right–most bars are values equal or more extreme than the appropriate $x$–value.
Figure S.5: Histograms of estimates of the stationary variance $\sigma^2/(2\alpha)$ of the OU model. First column: $m = 0$, second column: $m = 0.5$, third column $m = 10$, first row: all individuals have $\theta = 5$, second row: half of individuals have $\theta = -5$, other half $\theta = 5$. In all presented histograms $\alpha = 1$, $X_0 = 0$, $\sigma^2 = 1$. Hence the stationary variance is $\sigma^2/(2\alpha) = 0.5$. The vertical line represents the true value. Note that the histograms are not all on the same scale. The left and right–most bars are values equal or more extreme than the appropriate $x$–value.
Figure S.6: Histograms of sample variances of data simulated under the migration model. First column: $m = 0$, second column: $m = 0.5$, third column $m = 10$, first row: all individuals have $\theta = 5$, second row: half of individuals have $\theta = -5$, other half $\theta = 5$. In all presented histograms $\alpha = 1$, $X_0 = 0$, $\sigma^2 = 1$. In the second row we plot separately the sample variances of $\theta = -5$ individuals (black) and $\theta = 5$ individuals (gray). The vertical line is the OU process stationary variance $\sigma^2/(2\alpha) = 0.5$. Note that the histograms are not all on the same scale. The left and right–most bars are values equal or more extreme than the appropriate $x$–value.