Journal of Mathematical Biology (2023) 86:88
https://doi.org/10.1007/s00285-023-01922-8

Mathematical Biology

When can we reconstruct the ancestral state? Beyond Brownian motion

Nhat L. Vu[1] · Thanh P. Nguyen[2,3,4] · Binh T. Nguyen[2,3,4] · Vu Dinh[5] · Lam Si Tung Ho[1]

Received: 17 August 2022 / Revised: 18 February 2023 / Accepted: 17 April 2023

Abstract
Reconstructing the ancestral state of a group of species helps answer many important questions in evolutionary biology. Therefore, it is crucial to understand when we can estimate the ancestral state accurately. Previous works provide a necessary and sufficient condition, called the big bang condition, for the existence of an accurate reconstruction method under discrete trait evolution models and the Brownian motion model. In this paper, we extend this result to a wide range of continuous trait evolution models. In particular, we consider a general setting where continuous traits evolve along the tree according to stochastic processes that satisfy some regularity conditions. We verify these conditions for popular continuous trait evolution models including Ornstein–Uhlenbeck, reflected Brownian Motion, bounded Brownian Motion, and Cox–Ingersoll–Ross.

Keywords Ancestral state reconstruction · Consistency · Big bang condition · Ornstein–Uhlenbeck · Brownian motion · Cox–Ingersoll–Ross

Mathematics Subject Classification Primary 62F12 · Secondary 92D15 · 92B10

Nhat L. Vu and Thanh P. Nguyen have contributed equally to this work.

✉ Lam Si Tung Ho
lam.ho@dal.ca

1 Department of Mathematics and Statistics, Dalhousie University, Halifax, NS, Canada

2 AISIA Research Lab, Ho Chi Minh City, Vietnam

3 Department of Computer Science, University of Science, Ho Chi Minh City, Vietnam

4 Vietnam National University, Ho Chi Minh City, Vietnam

5 Department of Mathematical Sciences, University of Delaware, Newark, DE, USA

Published online: 04 May 2023

1 Introduction

The main task of the ancestral state reconstruction problem is to estimate the trait value of the most recent common ancestor of a group of species from their observed trait values. This problem has many applications in ecology and evolution, such as reconstructing ancestral diets (Maritz et al. 2021), studying female song in songbirds (Odom et al. 2014), and inferring the origin of infectious disease epidemics (Faria et al.. 2014; Gill et al. 2017). Moreover, ancestral state reconstruction methods have been applied to reconstruct the place of origin of a language family (Bouckaert et al. 2012; Neureiter et al. 2021). Therefore, it is important to assess whether the ancestral state can be reconstructed with high certainty. In particular, we are interested in whether reconstruction methods converge to the true ancestral trait value (in probability) as the number of sampled species increases to infinity. This property is called *consistency*, a desired characteristic for any reconstruction method.

The consistency property of ancestral state reconstruction methods has been studied previously for both discrete and continuous traits. Since species are related to each other according to an evolutionary tree, it is natural that consistency depends heavily on the structure of this tree. For continuous traits, Ané (2008) derived a necessary and sufficient condition for the consistency of the Maximum likelihood estimator (MLE) of the ancestral state under the Brownian motion (BM) model. This condition is $\mathbf{1}^\top \mathbf{V}_n^{-1}\mathbf{1} \to \infty$ where n is the number of sampled species, $\mathbf{1}$ is an $n \times 1$ vector whose elements are 1, and $\mathbf{V}_n$ is an $n \times n$ matrix whose ith-row and j-th column is the distance from the root to the most recent common ancestor of leaves i and j. Under mild assumptions, Fan and Roch (2018) provided a necessary and sufficient condition, called big bang, for the existence of a consistent ancestral state reconstruction method for discrete traits. The big bang condition is satisfied if for any positive number ϵ, the number of intersection points between the circle with radius ϵ centred at the root and the tree converges to infinity as the number of sampled species increases. In other words, the number of branches in the proximity of the root is large. A natural question is whether the condition $\mathbf{1}^\top \mathbf{V}_n^{-1}\mathbf{1} \to \infty$ and the big bang condition are equivalent. Recently, Ho and Dinh (2022) gave a positive answer to this question when the sequence of trees is nested and has bounded height as the sample size increases. In particular, they showed that both conditions are necessary and sufficient conditions for the existence of a consistent ancestral state reconstruction method under the BM model. The result unifies the theory of ancestral state reconstruction between discrete trait evolution models and the BM model. What remains unknown is whether this unified theory also holds for other evolution models of continuous traits. In this paper, we will address this open question.

Researchers often model the evolution of a continuous trait along an evolutionary tree using a diffusion process. At a node of the tree, the children lineages inherit the trait value of the parent lineage as their starting trait value and evolve independently from each other. The simplest evolution model for continuous traits is the BM model, which only considers neutral evolution (Felsenstein 1985). Later, Hansen (1997) proposed the Ornstein–Uhlenbeck (OU) model to incorporate natural selection. The BM and OU models are the most popular evolution models for continuous traits due to the existence of efficient computational methods (Ho and Ané 2014). Recently, much effort has been

made to move away from these Gaussian models (Boucher and Démery 2016; Boucher et al. 2018; Blomberg et al. 2020; Jhwueng 2020). Following this spirit, we consider a general setting where the evolution of traits follows a general stochastic process on trees. We will show that under mild assumptions, the big bang condition is a necessary and sufficient condition for the existence of a consistent ancestral state reconstruction method. Our setting includes several popular evolution models for continuous traits such as the OU model (Hansen 1997), the reflected Brownian motion (RBM) model (Boucher and Démery 2016), the bounded Brownian motion (BMM) model (Boucher and Démery 2016), and the Cox–Ingersoll–Ross (CIR) model (Lepage et al. 2006; Blomberg et al. 2020).

2 Mathematical formulation

In this paper, we consider the common setting for studying the asymptotic theory of trait evolution models where the sequence of trees $(\mathbb{T}_n)_{n=1}^{\infty}$ is nested. That is, $\mathbb{T}_n$ is a subtree of $\mathbb{T}_{n+1}$ for all n. Without loss of generality, we assume that tree $\mathbb{T}_n$ has n species and all trees in the sequence have the same root. We make a standard assumption that the tree topology and edge lengths of $\mathbb{T}_n$ are known. In this paper, we consider the scenario where the height of the sequence of trees $(\mathbb{T}_n)_{n=1}^{\infty}$ is uniformly bounded from above. Specifically, let t_i be the distance from the root to a leaf i. The height of tree $\mathbb{T}_n$ is defined by $h_n = \max\{t_i : i = 1, 2, \ldots, n\}$. Under our setting, $h^* := \sup_n h_n < +\infty$.

For a tree $\mathbb{T}$, we denote the leaf set of $\mathbb{T}$ by $\partial\mathbb{T}$ and the tree obtained by truncating $\mathbb{T}$ at distance s from the root by $\mathbb{T}(s)$. It is worth noticing that $\partial\mathbb{T}(s)$ is called a cutset corresponding to time s away from the root (Ané et al. 2017). Let $|A|$ be the number of elements of the set A. The big bang condition (Fan and Roch 2018) is defined as follows:

Definition 2.1 (big bang condition) A sequence of nested trees $(\mathbb{T}_n)_{n=1}^{\infty}$ satisfies the big bang condition if $\lim_{n\to\infty}|\partial\mathbb{T}_n(s)| = \infty$ for all $s > 0$.

A layman's explanation of the big bang condition is that the number of branches in the proximity of the root is large. We will prove that the big bang condition is the necessary and sufficient condition for the existence of a consistent ancestral state reconstruction method for a general class of continuous trait evolution models. Throughout this paper, we assume that all other parameters of the model are known. Specifically, we assume that the model satisfies the following regularity conditions:

(A_1) The trait evolves along a tree according to a time-homogeneous Markovian stochastic process $\{Y(t)\}_{t\geq 0}$ on a state-space $\mathcal{S} \subset \mathbb{R}$ and there exist functions u, v, ϕ such that

$$\mathbb{E}\{\phi[Y(t)] \mid Y(0)\} = u(t)\phi[Y(0)] + v(t), \tag{2.1}$$

and

$$\mathrm{var}\{\phi[Y(t)] \mid Y(0)\} \leq Ct, \quad \forall t \in [0, h^*], \tag{2.2}$$

where C is a constant that does not depend on t; u, v are continuous functions; ϕ is a continuous, injective function from $\mathcal{S}$ to $\mathbb{R}$ and the inverse function ϕ^{-1} is also continuous; and $u(t) > 0$ for all $t \geq 0$. Note that $u(0) = 1$ and $v(0) = 0$.

(A_2) We denote the trait values at the leaves of a tree $\mathbb{T}$ by $\mathbf{Y}_{\mathbb{T}}$ and the trait value at the root by ρ. Let $P_{\rho,\mathbb{T}}$ be the joint distribution of the observations $\mathbf{Y}_{\mathbb{T}}$ at the leaves of the tree $\mathbb{T}$ given that the ancestral state is ρ. We assume that for any tree $\mathbb{T}$, there exists $\rho_1 \neq \rho_2$ such that the overlapped support of $P_{\rho_1,\mathbb{T}}$ and $P_{\rho_2,\mathbb{T}}$ is not trivial. Here, an overlapped support $\mathcal{A}$ of $P_{\rho_1,\mathbb{T}}$ and $P_{\rho_2,\mathbb{T}}$ is trivial if $\min\{P_{\rho_1,\mathbb{T}}(\mathcal{A}), P_{\rho_2,\mathbb{T}}(\mathcal{A})\} = 0$. Furthermore, we assume that $P_{\rho,\mathbb{T}}$ admits a probability density function $f_{\rho,\mathbb{T}}$.

Condition (A_1) ensures that the observations contain sufficient information about the root. Specifically, (2.1) means that the information about the root is contained in the expected value of the observations. On the other hand, (2.2) controls the decay rate of the information about the root through time. Condition (A_2) is similar to the Downstream Disjointness condition in Fan and Roch (2018). The main purpose of this condition is to remove trivial scenarios. If (A_2) does not hold, $P_{\rho_1,\mathbb{T}}$ and $P_{\rho_2,\mathbb{T}}$ are disjoint for any ρ_1, ρ_2. In this case, it is trivial to reconstruct the ancestral state because each observed vector of values at the leaves only corresponds to one value at the root. However, that is too good to be true for practical settings.

3 A necessary and sufficient condition for the existence of a consistent estimator for the ancestral state

We recall the definition of a *consistent estimator*:

Definition 3.1 An estimator $\hat{\rho}_n$ of ρ is consistent if and only if for all $\epsilon > 0$, we have

$$P_{\rho,\mathbb{T}_n}(|\hat{\rho}_n - \rho| > \epsilon) \to 0.$$

In other words, $\hat{\rho}_n$ converges to ρ in probability ($\hat{\rho}_n \to_p \rho$).

Now, we are ready to state our main result:

Theorem 3.1 *Assume that the regularity condition (A_1) and (A_2) are satisfied. Then, the big bang condition is the necessary and sufficient condition for the existence of a consistent ancestral reconstruction method.*

First, let us review briefly about properties of the total variation distance. For any two probability measures μ_1 and μ_2, the total variation distance between them is defined as

$$\begin{aligned} d_{\mathrm{TV}}(\mu_1, \mu_2) &= \sup_{\mathcal{A}} |\mu_1(\mathcal{A}) - \mu_2(\mathcal{A})| \\ &= \frac{1}{2} \int |\mu_1(x) - \mu_2(x)| dx \\ &= \frac{1}{2} \int [\mu_1(x) \vee \mu_2(x) - \mu_1(x) \wedge \mu_2(x)] dx, \end{aligned}$$

where $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. On the other hand, since

$$\frac{1}{2}\int [\mu_1(x) \vee \mu_2(x) + \mu_1(x) \wedge \mu_2(x)]dx = 1,$$

we have

$$d_{\mathrm{TV}}(\mu_1, \mu_2) = 1 - \int \mu_1(x) \wedge \mu_2(x)dx. \tag{3.1}$$

This implies that if the overlapped support of μ_1 and μ_2 is not trivial, then their total variation distance is strictly less than one.

The proof of Theorem 3.1 is divided into two parts. In Sect. 3.1, we prove that the big bang condition is a necessary condition. The main idea is to show that if the big bang condition does not hold, the total variation distance between $P_{\rho_1,\mathbb{T}_n}$ and $P_{\rho_2,\mathbb{T}_n}$ is bounded away from 1 for some $\rho_1 \neq \rho_2$. This implies that there is no consistent estimator for the ancestral state ρ. In Sect. 3.2, we provide a simple consistent estimator for the ancestral state when the big bang condition is satisfied.

3.1 Necessary condition

Theorem 3.2 *Given that the condition (A_2) holds, then the big bang is the necessary condition for the existence of a consistent estimator for the ancestral state.*

Proof We will use the contra-positive approach: no big bang condition implies no consistent estimator for the ancestral state. That is, we need to prove that if there exists $s_0 > 0$, $n_0 \geq 1$, and $K > 0$ such that $|\partial\mathbb{T}_n(s_0)| = K$ for all $n \geq n_0$, then there is no consistent estimator for the ancestral state.

Recall that $P_{\rho,\mathbb{T}_n}$ is the joint distribution of the observations $\mathbf{Y}_{\mathbb{T}_n}$ at the leaves of tree $\mathbb{T}_n$ with the ancestral state ρ. Note that, since $|\partial\mathbb{T}_n(s_0)| = K$ for all $n \geq n_0$, $\mathbb{T}_n(s_0)$ is a fixed K-species star tree with edge lengths all equal to s_0 for all $n \geq n_0$. From (A_2) the overlapped support of $P_{\rho_1,\mathbb{T}_n(s_0)}$ and $P_{\rho_2,\mathbb{T}_n(s_0)}$ is not trivial for some ancestral state $\rho_1 \neq \rho_2$. Since this overlapped support is fixed for all $n \geq n_0$, by (3.1), we have $d_{\mathrm{TV}}(P_{\rho_1,\mathbb{T}_n(s_0)}, P_{\rho_2,\mathbb{T}_n(s_0)}) \leq C_0 < 1$ for all $n \geq n_0$.

Let $f_{\mathbb{T}|\mathbb{T}(s)}(\cdot \mid \tau)$ be the conditional probability density function of $\mathbf{Y}_{\mathbb{T}}$ given $\mathbf{Y}_{\mathbb{T}(s)} = \tau$. Note that $f_{\mathbb{T}|\mathbb{T}(s)}(\cdot \mid \tau)$ does not depend on the ancestral state ρ at the root of $\mathbb{T}$ due to Markov property. We have

$$\begin{aligned}
&d_{\mathrm{TV}}(P_{\rho_1,\mathbb{T}_n}, P_{\rho_2,\mathbb{T}_n}) = \frac{1}{2}\int |\, f_{\rho_1,\mathbb{T}_n}(y) - f_{\rho_2,\mathbb{T}_n}(y)|dy \\
&= \frac{1}{2}\int \left| \int \left[f_{\mathbb{T}_n|\mathbb{T}_n(s_0)}(y \mid \tau) f_{\rho_1,\mathbb{T}_n(s_0)}(\tau) - f_{\mathbb{T}_n|\mathbb{T}_n(s_0)}(y \mid \tau) f_{\rho_2,\mathbb{T}_n(s_0)}(\tau) \right] d\tau \right| dy \\
&= \frac{1}{2}\int \left| \int f_{\mathbb{T}_n|\mathbb{T}_n(s_0)}(y \mid \tau) \left[f_{\rho_1,\mathbb{T}_n(s_0)}(\tau) - f_{\rho_2,\mathbb{T}_n(s_0)}(\tau) \right] d\tau \right| dy \\
&\leq \frac{1}{2}\int \int f_{\mathbb{T}_n|\mathbb{T}_n(s_0)}(y \mid \tau) \left| f_{\rho_1,\mathbb{T}_n(s_0)}(\tau) - f_{\rho_2,\mathbb{T}_n(s_0)}(\tau) \right| d\tau dy
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2} \int \left| f_{\rho_1, \mathbb{T}_n(s_0)}(\tau) - f_{\rho_2, \mathbb{T}_n(s_0)}(\tau) \right| \int f_{\mathbb{T}_n | \mathbb{T}_n(s_0)}(y \mid \tau) dy d\tau \\
&= \frac{1}{2} \int \left| f_{\rho_1, \mathbb{T}_n(s_0)}(\tau) - f_{\rho_2, \mathbb{T}_n(s_0)}(\tau) \right| d\tau \\
&= d_{\mathrm{TV}}(P_{\rho_1, \mathbb{T}_n(s_0)}, P_{\rho_2, \mathbb{T}_n(s_0)}) \leq C_0 < 1. \qquad (3.2)
\end{aligned}
$$

Now suppose that there exists a consistent estimator $\hat{\rho}_n$ for the ancestral state ρ. Consider the event $\mathcal{A}_\epsilon := \{|\hat{\rho}_n - \rho_1| \leq \epsilon\}$ where ϵ is sufficiently small. Then, we have

$$
P_{\rho_1, \mathbb{T}_n}(\mathcal{A}_\epsilon) \to 1 \quad \text{and} \quad P_{\rho_2, \mathbb{T}_n}(\mathcal{A}_\epsilon) \to 0,
$$

for any $\rho_1 \neq \rho_2$. Thus,

$$
d_{\mathrm{TV}}(P_{\rho_1, \mathbb{T}_n}, P_{\rho_2, \mathbb{T}_n}) = \sup_{\mathcal{B}} \left| P_{\rho_1, \mathbb{T}_n}(\mathcal{B}) - P_{\rho_2, \mathbb{T}_n}(\mathcal{B}) \right| \geq \left| P_{\rho_1, \mathbb{T}_n}(\mathcal{A}_\epsilon) - P_{\rho_2, \mathbb{T}_n}(\mathcal{A}_\epsilon) \right| \to 1,
$$

which is a contradiction to (3.2). Therefore, there is no consistent estimator for the ancestral state in the absence of the big bang condition.

□

3.2 Sufficient condition

Theorem 3.3 *Assume that the condition* (A_1) *is satisfied. Then, the big bang condition implies that there exists a consistent estimator for the ancestral state.*

Proof We will construct a consistent estimator for the ancestral state ρ. Under big bang condition, there exists an increasing sequence $\{n_k\}_{k=1}^{\infty}$ such that $|\partial \mathbb{T}_{n_k}(1/k)| \geq k$ for all k. This implies that there exists a k-species subtree of $\mathbb{T}_{n_k}$ such that the distances from the root to all internal nodes are at most $1/k$. Let $(Y_{i,k})_{i=1}^k$ be the observations at the leaves and $(t_{i,k})_{i=1}^k$ be the distance from the root to the i-th leaf of this subtree. Denote the trait value of the most recent common ancestor of i-th, j-th leaves of the subtree by $Y_{ij,k}$, and the distance from the root to this ancestor by $t_{ij,k}$. It is worth noticing that $t_{i,k} \leq h^*$ and $t_{ij,k} \leq 1/k$. We consider the following estimator for the ancestral state ρ:

$$
\hat{\rho} = \phi^{-1}\left(\frac{1}{k} \sum_{i=1}^{k} \frac{\phi(Y_{i,k}) - v(t_{i,k})}{u(t_{i,k})} \right).
$$

Since ϕ is a continuous, injection function, the proposed estimator is well-defined. Moreover, ϕ^{-1} is a continuous function. So, to prove that $\hat{\rho}$ is consistent, it is sufficient to show

$$
\frac{1}{k} \sum_{i=1}^{k} \frac{\phi(Y_{i,k}) - v(t_{i,k})}{u(t_{i,k})} \to_p \phi(\rho).
$$

Indeed, we have

$$\begin{aligned}\mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k}\frac{\phi(Y_{i,k})-v(t_{i,k})}{u(t_{i,k})}\right) &= \frac{1}{k}\sum_{i=1}^{k}\frac{\mathbb{E}[\phi(Y_{i,k})]-v(t_{i,k})}{u(t_{i,k})}\\ &= \frac{1}{k}\sum_{i=1}^{k}\frac{[u(t_{i,k})\phi(\rho)+v(t_{i,k})]-v(t_{i,k})}{u(t_{i,k})}\\ &= \phi(\rho).\end{aligned}$$

On the other hand,

$$\begin{aligned}&\mathrm{var}\left(\frac{1}{k}\sum_{i=1}^{k}\frac{\phi(Yi,k)-v(t_{i,k})}{u(t_{i,k})}\right)\\ &\leq \frac{1}{k^2m^2}\left(\sum_{i=1}^{k}\mathrm{var}[\phi(Yi,k)]+\sum_{1\leq i<j\leq k}\mathrm{Cov}[\phi(Y_{i,k}),\phi(Y_{j,k})]\right)\end{aligned}$$

where $m := \min_{t\in[0,h^*]} u(t) > 0$. Note that $\mathrm{var}[\phi(Yi,k)] \leq Ct_{i,k} \leq Ch^*$ and

$$\begin{aligned}\mathrm{Cov}[\phi(Y_{i,k}),\phi(Y_{j,k})] &= \mathrm{Cov}\{\mathbb{E}[\phi(Y_{i,k})\mid Y_{ij,k}],\mathbb{E}[\phi(Y_{j,k})\mid Y_{ij,k}]\}\\ &\quad+\mathbb{E}\{\mathrm{Cov}[\phi(Y_{i,k}),\phi(Y_{j,k})\mid Y_{ij,k}]\}\\ &= u(t_{ij,k})^2\mathrm{var}[\phi(Yi,k)]]\leq M^2Ct_{ij,k}\leq\frac{CM^2}{k},\end{aligned}$$

where $M := \max_{t\in[0,h^*]} u(t) < \infty$. Therefore

$$\mathrm{var}\left(\frac{1}{k}\sum_{i=1}^{k}\frac{\phi(Yi,k)-v(t_{i,k})}{u(t_{i,k})}\right)\leq\frac{Ch^*}{km^2}+\frac{CM^2}{km^2}\to 0\quad\text{as } k\to\infty.$$

For any $\epsilon > 0$, by applying Chebyshev's inequality, we have

$$\Pr\left(\left|\frac{1}{k}\sum_{i=1}^{k}\frac{\phi(Y_{i,k})-v(t_{i,k})}{u(t_{i,k})}-\phi(\rho)\right|>\epsilon\right)\leq\frac{1}{\epsilon^2}\mathrm{var}\left(\frac{1}{k}\sum_{i=1}^{k}\frac{\phi(Yi,k)-v(t_{i,k})}{u(t_{i,k})}\right)\to 0.$$

Hence,

$$\frac{1}{k}\sum_{i=1}^{k}\frac{\phi(Y_{i,k})-v(t_{i,k})}{u(t_{i,k})}\to_p\phi(\rho),$$

which means $\hat{\rho}$ is consistent.

□

We note that the estimator in this proof is sub-optimal since it only uses a subset of the observations.

4 Applications

In this section, we will apply our results to several popular trait evolution models including Ornstein–Uhlenbeck (OU), reflected Brownian motion (RBM) model, bounded Brownian motion (BBM) model, and Cox–Ingersoll–Ross (CIR) models. Since we focus on the ancestral state reconstruction problem, we consider the classical setting where other parameters of the models are known.

4.1 Ornstein–Uhlenbeck (OU) model

The OU model assumes that a continuous trait evolves along a phylogeny according to an OU process. The process is equipped with a "selection optimum" parameter μ which captures the optimal trait value; a "selection strength" parameter α which represents the strength of the selection force that pulls the trait toward μ; and the variance parameter σ^2 of the neutral drift. The model has been used extensively to take into account natural selection in evolutionary studies (Beaulieu et al. 2012; Rohlfs et al. 2014; Uyeda and Harmon 2014; Bastide et al. 2021). Although the consistent property of estimators for μ and α have been studied thoroughly (Ho and Ané 2013; Bartoszek and Sagitov 2015; Ané et al. 2017), the consistency of ancestral state reconstruction methods under the OU model is not well-understood. Here, we will fill in this gap. Applying Theorem 3.1, we derive the following result:

Theorem 4.1 *Under the OU model, the big bang condition is the necessary and sufficient condition for the existence of a consistent ancestral reconstruction method.*

Proof It is sufficient to check the regularity conditions (A_1) and (A_2) for the OU model. Let $(Y_t)_{t\geq 0}$ be an OU process, we have

$$\mathbb{E}[Y(t) \mid Y(0)] = e^{-\alpha t} Y(0) + (1 - e^{-\alpha t})\mu,$$

and

$$\mathrm{Var}[Y(t) \mid Y(0)] = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}) \leq \sigma^2 t.$$

The condition (A_1) is satisfied with $C = \sigma^2; u(t) = e^{-\alpha t}; v(t) = (1 - e^{-\alpha t})\mu; \phi(y) = y$. On the other hand, $P_{\rho_1,\mathbb{T}}$ and $P_{\rho_2,\mathbb{T}}$ are multivariate normal distributions. Therefore, the condition (A_2) is trivial since the overlapped support is $\mathbb{R}^{|\partial\mathbb{T}|}$. □

4.2 Reflected Brownian motion (RBM) model

A limitation of both BM and OU models is that they cannot accommodate hard bounds on trait values. Unfortunately, hard bounds do exist in nature. For example, morphological measurements, such as body size and body mass, can only take positive values. Some trait values are proportion (e.g., allele frequencies and genomic GC content) and thus are bounded between 0 and 1. Boucher and Démery (2016) propose the Bounded Brownian motion (BBM) model, which assumes traits evolve under BM with two reflecting boundaries. A particular case of this model for traits with positive values is the RBM model (Boucher and Démery 2016). Recall that if $X(t)$ is a BM starting from $X(0) > 0$, then $Y(t) = |X(t)|$ is a RBM starting from $Y(0) = X(0)$. That is, the RBM model assumes that traits evolve according to a BM with a single reflecting boundary at zero. We have the following theorem:

Theorem 4.2 *Under the RBM model, the big bang condition is the necessary and sufficient condition for the existence of a consistent ancestral reconstruction method.*

Proof Again, we only need to verify that the RBM model satisfies the regularity conditions (A_1) and (A_2). Note that $X(t) \mid X(0)$ follows a normal distribution $\mathcal{N}(X(0), \sigma^2 t)$ and its moment generating function is $\psi(s) = \exp(X(0)s + \sigma^2 t s^2/2)$. Therefore,

$$\mathbb{E}[X(t)^2 \mid X(0)] = \frac{\partial^2 \psi}{(\partial s)^2}(0) = X(0)^2 + \sigma^2 t$$
$$\mathbb{E}[X(t)^4 \mid X(0)] = \frac{\partial^4 \psi}{(\partial s)^4}(0) = X(0)^4 + 6\sigma^2 t X(0)^2 + 3\sigma^4 t^2.$$

Since $Y^2(t) = X^2(t)$ and $Y(0) = X(0)$, we have

$$\mathbb{E}[Y(t)^2 \mid Y(0)] = Y(0)^2 + \sigma^2 t,$$

and

$$\begin{aligned}\mathrm{Var}[Y(t)^2 \mid Y(0)] &= \mathbb{E}[Y(t)^4 \mid Y(0)] - (\mathbb{E}[Y(t)^2 \mid Y(0)])^2 \\ &= 4\sigma^2 t Y(0)^2 + 2\sigma^4 t^2 \\ &\le (4\sigma^2 Y(0)^2 + 2\sigma^4 h^*)t, \quad \forall t \in [0, h^*].\end{aligned}$$

Thus, the condition (A_1) is satisfied with $C = 4\sigma^2 Y(0)^2 + 2\sigma^4 h^*$; $u(t) = 1$; $v(t) = \sigma^2 t$; $\phi(y) = y^2$. The condition (A_2) is trivial. □

4.3 Bounded Brownian motion (BBM) model

The BBM model (Boucher and Démery 2016) assumes that traits evolve under BM with two reflecting boundaries. For simplicity, we assume that the BM is bounded in [0, 1]. We have the following theorem:

Theorem 4.3 *Under the BBM model, the big bang condition is the necessary and sufficient condition for the existence of a consistent ancestral reconstruction method.*

Proof We will verify the condition (A_1) with $C = \sigma^2\pi^2$; $u(t) = \exp(-\sigma^2 t\pi^2/2)$; $v(t) = 0$; $\phi(y) = \cos(\pi y)$. First, we recall that the density function of $X(t) \mid X(0) = x_0$ is (Boucher and Démery 2016)

$$p(x, x_0, t) = \frac{1}{\sqrt{2\pi t}\sigma}\left\{\sum_{k=-\infty}^{\infty}\left[\exp\left(\frac{-(x - x_0 - 2k)^2}{2\sigma^2 t}\right) + \exp\left(\frac{-(x + x_0 - 2k)^2}{2\sigma^2 t}\right)\right]\right\}.$$

Therefore,

$$\mathbb{E}[\cos(\pi X(t)) \mid X(0)] = \int_0^1 \cos(\pi x)\, p(x, X(0), t)dx.$$

Note that

$$\int_0^1 \cos(\pi x)\exp\left(\frac{-(x - x_0 - 2k)^2}{2\sigma^2 t}\right)dx = \int_{-2k}^{-2k+1}\cos(\pi x)\exp\left(\frac{-(x - x_0)^2}{2\sigma^2 t}\right)dx$$

and

$$\begin{aligned}\int_0^1 \cos(\pi x)\exp\left(\frac{-(x + x_0 - 2k)^2}{2\sigma^2 t}\right)dx &= \int_{-2k}^{-2k+1}\cos(\pi x)\exp\left(\frac{-(x + x_0)^2}{2\sigma^2 t}\right)dx \\ &= \int_{2k-1}^{2k}\cos(\pi x)\exp\left(\frac{-(x - x_0)^2}{2\sigma^2 t}\right)dx\end{aligned}$$

Hence

$$\begin{aligned}\mathbb{E}[\cos(\pi X(t)) \mid X(0)] &= \frac{1}{\sqrt{2\pi t}\sigma}\int_{-\infty}^{\infty}\cos(\pi x)\exp\left(\frac{-(x - X(0))^2}{2\sigma^2 t}\right)dx \\ &= \Re\left(\frac{1}{\sqrt{2\pi t}\sigma}\int_{-\infty}^{\infty} e^{i\pi x}\exp\left(\frac{-(x - X(0))^2}{2\sigma^2 t}\right)dx\right) \\ &= \Re\left(e^{i\pi X(0) - \sigma^2 t\pi^2/2}\right) \\ &= \exp(-\sigma^2 t\pi^2/2)\cos(\pi X(0))\end{aligned}$$

where $\Re(\cdot)$ is the real part. Similarly, we have

$$\begin{aligned}\mathbb{E}[\cos^2(\pi X(t)) \mid X(0)] &= \mathbb{E}\left[\frac{1 + \cos(2\pi X(t))}{2} \mid X(0)\right] \\ &= \frac{1 + \exp(-2\sigma^2 t\pi^2)\cos(2\pi X(0))}{2}.\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{Var}[\cos(\pi X(t)) \mid X(0)] &= \mathbb{E}[\cos^2(\pi X(t)) \mid X(0)] - (\mathbb{E}[\cos(\pi X(t)) \mid X(0)])^2 \\
&= \frac{1+\exp(-2\sigma^2 t\pi^2)\cos(2\pi X(0))}{2} - \exp(-\sigma^2 t\pi^2)\cos^2(\pi X(0)) \\
&= \frac{1}{2}\left(1-\exp(-\sigma^2 t\pi^2)\right)\left[1-\exp(-\sigma^2 t\pi^2)\cos(2\pi X(0))\right] \\
&\le \frac{1}{2}\left(1-\exp(-\sigma^2 t\pi^2)\right)2 \le \sigma^2 t\pi^2.
\end{aligned}$$

We conclude that the condition (A_1) is satisfied with $C = \sigma^2\pi^2$; $u(t) = \exp(-\sigma^2 t\pi^2/2)$; $v(t) = 0$; $\phi(y) = \cos(\pi y)$. The condition (A_2) is trivial. □

4.4 Cox–Ingersoll–Ross (CIR) model

The CIR model is an evolution model for traits that have positive values. This model has been utilized for modelling evolutionary rate (Lepage et al. 2006), longevity of carnivores and ungulates (Blomberg et al. 2020), and rate of adaptive trait evolution (Jhwueng 2020). Similar to the OU process, the CIR process has a "selection optimum" parameter μ and a "selection strength" parameter α. Let $Y(t)$ be a CIR process, then $Y(t)$ follows the following stochastic differential equation:

$$dY(t) = \alpha(\mu - Y(t))dt + \sigma\sqrt{Y(t)}dB(t),$$

where $B(t)$ is the standard BM. We have

$$\mathbb{E}[Y(t) \mid Y(0)] = Y(0)e^{-\alpha t} + \mu(1-e^{-\alpha t}),$$

and

$$\begin{aligned}
\mathrm{Var}[Y(t) \mid Y(0)] &= Y(0)\frac{\sigma^2}{\alpha}(e^{-\alpha t} - e^{-2\alpha t}) + \frac{\mu\sigma^2}{2\alpha}(1-e^{-\alpha t})^2 \\
&\le [Y(0) + \mu/2]\frac{\sigma^2}{\alpha}(1-e^{-\alpha t}) \\
&\le [Y(0) + \mu/2]\sigma^2 t, \quad \forall t \in [0, h^*].
\end{aligned}$$

Hence, the condition (A_1) holds with $C = [Y(0) + \mu/2]\sigma^2$; $u(t) = e^{-\alpha t}$; $v(t) = \mu(1-e^{-\alpha t})$; $\phi(y) = y$. Again, the condition (A_2) is trivial. By Theorem 3.1, we have:

Theorem 4.4 *Under the CIR model, the big bang condition is the necessary and sufficient condition for the existence of a consistent ancestral reconstruction method.*

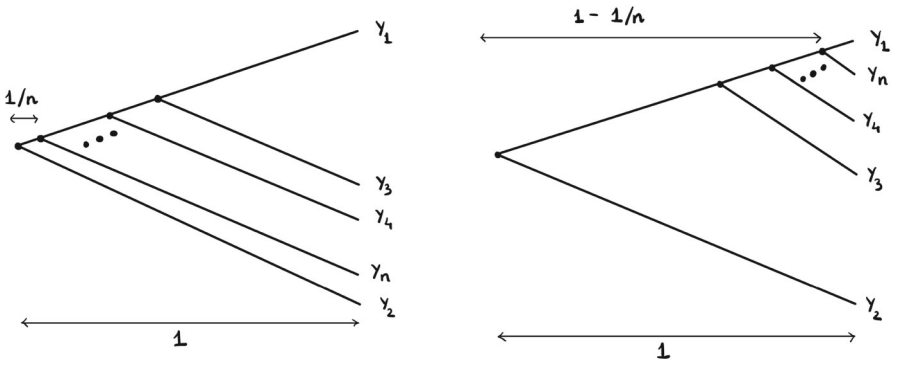

Fig. 1 Two sequences of caterpillar trees consider in the simulation. Left: the big bang condition is satisfied (new branch is closer and closer to the root). Right: the big bang condition is not satisfied (all branches are far away from the root)

4.5 Simulation under the bounded Brownian motion (BBM) model

In this section, we will use simulations to illustrate the result in Theorem 4.3. Specifically, we consider two sequences of bifurcating ultrametric caterpillar trees. Recall that the distance from the root to the leaves of an ultrametric tree is constant and a caterpillar tree has a path, called the main path, that contains all internal nodes. For both sequences, we create the n-th tree by adding a new leaf directly to the main path of the $(n-1)$-th tree such that the distance from the n-th leaf to the root is 1. The first sequence of trees satisfies the big bang condition: the n-th leaf attaches to the tree at the n-th internal node, whose distance to the root is $1/n$ (see Fig. 1 – left). On the other hand, the second sequence does not satisfy the big bang condition: the n-th leaf attaches to the tree at the n-th internal node, whose distance to the root is $1-1/n$ (see Fig. 1 – right).

In this simulation, we use the `R` functions provided in Boucher and Démery (2016) for simulating and fitting under the BBM model.[1] We focus on trees with size $n = 10, 100, 1000$. So, we have 6 trees in total (3 trees for each sequence). For each tree, we simulate the trait values at the leaves 1000 times under the BBM model with the ancestral state $\rho = 0$, variance $\sigma^2 = 1/2$, and two bounds ± 1 using the `R` function `Sim_BBM`. Then, we use the function `fit_BBM_model_uncertainty` to fit the BBM model and return the MLE of the ancestral state ρ. When the big bang condition holds, the estimate $\hat{\rho}$ is more precise as the number of species increases (see Fig. 2 – left). On the other hand, when the big bang condition is not satisfied, the precision of $\hat{\rho}$ stays the same for all trees.

5 Discussion and Conclusion

In this paper, we prove that under some regularity conditions, the big bang condition is a necessary and sufficient condition for the existence of a consistent ancestral

[1] https://github.com/fcboucher/BBM.

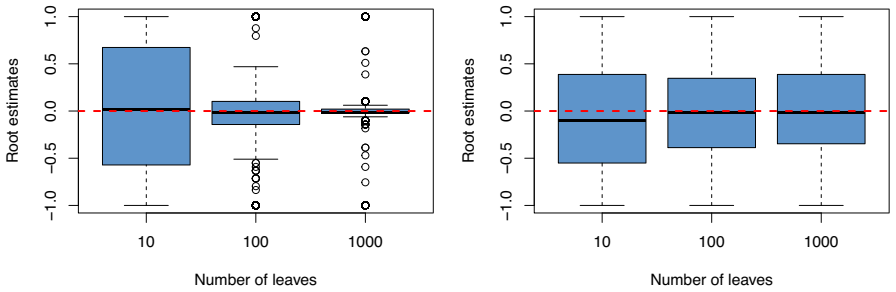

Fig. 2 Estimates of the ancestral state ρ under the BBM with $\rho = 0$, $\sigma^2 = 1/2$, and two bounds ± 1. Left (the big bang condition holds): the estimates become more accurate as the number of leaves increases. Right (the big bang condition does not hold): the accuracy is the same for all trees

state reconstruction method for continuous traits. We verify these conditions for Ornstein–Uhlenbeck, reflected Brownian motion, bounded Brownian motion and Cox–Ingersoll–Ross models.

It is worth noticing that under the BM model, the MLE for the ancestral state is consistent if and only if the big bang condition holds (Ho and Dinh 2022). However, it is unclear if this result is still true beyond BM models. Since MLE is the most popular method for reconstructing the ancestral state, studying its consistency property is of great interest.

The results in this paper assume that both the tree topology and branch lengths are known. However, we may know the tree topology but not branch lengths in practice. Many popular tree reconstruction methods such as maximum parsimony, neighbour-joining, and quartet puzzling only return the tree topology. A recent study shows that for discrete traits, the big bang condition may not guarantee the existence of a consistent ancestral state reconstruction method when branch lengths are unknown (Ho and Susko 2022). An open question is whether the big bang condition is still a sufficient condition for the existence of a consistent estimator for the ancestral state of continuous traits.

Acknowledgements LSTH was supported by the Canada Research Chairs program, the NSERC Discovery Grant RGPIN-2018-05447, and the NSERC Discovery Launch Supplement DGECR-2018-00181. VD was supported by a startup fund from the University of Delaware, a University of Delaware Research Foundation's Strategic Initiatives Grant, and National Science Foundation grant DMS- 1951474. The authors would like to thank Douglas Rizzolo for a helpful discussion that lead to the bounds for the Bounded Brownian Motion model.

Data Availability Statement The R code for the simulations is available at https://github.com/lamho86/When-can-we-reconstruct-the-ancestral-state-Beyond-Brownian-motion.

Declarations

Conflict of interest The authors have no competing interests to declare.

References

Ané C (2008) Analysis of comparative data with hierarchical autocorrelation. Ann Appl Stat 2(3):1078–1102

Ané C, Ho LST, Roch S (2017) Phase transition on the convergence rate of parameter estimation under an Ornstein–Uhlenbeck diffusion on a tree. J Math Biol 74(1):355–385

Bartoszek K, Sagitov S (2015) Phylogenetic confidence intervals for the optimal trait value. J Appl Probab 52(4):1115–1132

Bastide P, Ho LST, Baele G, Lemey P, Suchard MA (2021) Efficient Bayesian inference of general Gaussian models on large phylogenetic trees. Ann Appl Stat 15(2)

Beaulieu JM, Jhwueng DC, Boettiger C, O'Meara BC (2012) Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. Evolution 66(8):2369–2383

Blomberg SP, Rathnayake SI, Moreau CM (2020) Beyond Brownian motion and the Ornstein–Uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. Am Nat 195(2):145–165

Boucher FC, Démery V (2016) Inferring bounded evolution in phenotypic characters from phylogenetic comparative data. Syst Biol 65(4):651–661

Boucher FC, Démery V, Conti E, Harmon LJ, Uyeda J (2018) A general model for estimating macroevolutionary landscapes. Syst Biol 67(2):304–319

Bouckaert R, Lemey P, Dunn M, Greenhill SJ, Alekseyenko AV, Drummond AJ, Gray RD, Suchard MA, Atkinson QD (2012) Mapping the origins and expansion of the indo-european language family. Science 337(6097):957–960

Fan WTL, Roch S (2018) Necessary and sufficient conditions for consistent root reconstruction in markov models on trees. Electron J Probab 23

Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pépin J et al (2014) The early spread and epidemic ignition of HIV-1 in human populations. Science 346(6205):56–61

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

Gill MS, Tung Ho LS, Baele G, Lemey P, Suchard MA (2017) A relaxed directional random walk model for phylogenetic trait evolution. Syst Biol 66(3):299–319

Hansen TF (1997) Stabilizing selection and the comparative analysis of adaptation. Evolution 51(5):1341–1351

Ho LST, Ané C (2013) Asymptotic theory with hierarchical autocorrelation: Ornstein–Uhlenbeck tree models. Ann Stat 41(2):957–981

Ho LST, Ané C (2014) A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. Syst Biol 63(3):397–408

Ho LST, Dinh V (2022) When can we reconstruct the ancestral state? a unified theory. Theor Popul Biol 148:22–27

Ho LST, Susko E (2022) Ancestral state reconstruction with large numbers of sequences and edge-length estimation. J Math Biol 84(4):1–28

Jhwueng DC (2020) Modeling rate of adaptive trait evolution using Cox–Ingersoll–Ross process: an approximate Bayesian computation approach. Comput. Stat. Data Anal. 145:106924

Lepage T, Lawi S, Tupper P, Bryant D (2006) Continuous and tractable models for the variation of evolutionary rates. Math Biosci 199(2):216–233

Maritz B, Barends JM, Mohamed R, Maritz RA, Alexander GJ (2021) Repeated dietary shifts in elapid snakes (Squamata: Elapidae) revealed by ancestral state reconstruction. Biol J Lin Soc 134(4):975–986

Neureiter N, Ranacher P, van Gijn R, Bickel B, Weibel R (2021) Can Bayesian phylogeography reconstruct migrations and expansions in linguistic evolution? Royal Soc Open Sci 8(1):201079

Odom KJ, Hall ML, Riebel K, Omland KE, Langmore NE (2014) Female song is widespread and ancestral in songbirds. Nat Commun 5(1):1–6

Rohlfs RV, Harrigan P, Nielsen R (2014) Modeling gene expression evolution with an extended Ornstein–Uhlenbeck process accounting for within-species variation. Mol Biol Evol 31(1):201–211

Uyeda JC, Harmon LJ (2014) A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Syst Biol 63(6):902–918