Paleobiology

www.cambridge.org/pab

Using state space models to understand trait evolution in fossil lineages

Gene Hunt[1], Wilmer Martínez-Rivera[2], Melanie Jane Hopkins[3], John Fricks[4] and Beckett Sterner[5]

[1]National Museum of Natural History, Smithsonian Institution, 10th Street and Constitution Avenue NW, Washington, D.C. 20560, U.S.A.
[2]Banco de la República de Colombia, Cra. 7 #14-78, Santa Fé, Bogotá, Columbia
[3]American Museum of Natural History, 200 Central Park West, New York, New York 10024, U.S.A.
[4]School of Mathematical and Statistical Sciences, Arizona State University, 901 South Palm Walk, Tempe, Arizona 85287-1804, U.S.A.
[5]School of Life Sciences, Arizona State University, 427 East Tyler Mall, Tempe, Arizona 85287-1804, U.S.A.

Article

Cite this article: Hunt, G., W. Martínez-Rivera, M. J. Hopkins, J. Fricks, and B. Sterner (2026). Using state space models to understand trait evolution in fossil lineages. *Paleobiology* **52**, 89–103.
https://doi.org/10.1017/pab.2025.10075

Received: 05 August 2024
Revised: 25 August 2025
Accepted: 27 August 2025

Handling Editor:
Rachel Warnock

Corresponding author:
Beckett Sterner;
Email: beckett.sterner@asu.edu

Abstract

Linear state space models provide a useful framework for investigating phenotypic evolution in fossil lineages for a wide variety of models representing Brownian motion, Ornstein-Uhlenbeck processes, and the potential influence of environmental covariates. A state space framework also provides access to residuals for the predicted and observed values at each time point as well as improved numerical stability. We illustrate the value of the state space approach by reanalyzing a classic dataset of trait evolution in the diatom lineage *Stephanodiscus yellowstonensis*. A series of increasingly complex models were fit to these data, including a novel modification of an Ornstein-Uhlenbeck model in which a trait tracks an exogenous covariate. These model results suggest that the number of spines on the periphery of the diatom is best explained by adaptation to changing solar insolation over time.

Non-technical Summary

Paleontologists have long used repeated observations from fossil species to document and understand patterns of trait evolution. Here we describe a flexible framework for modeling such data called linear state space models. After summarizing this approach and its properties, we apply it to a classic dataset of trait evolution in species of diatom, a kind of unicellular algae. A set of models were fit to these diatom data using the state space approach, the best supported of which involved a novel model in which the focal trait tracks variations in solar insolation over time. Overall, state space models offer a useful framework for paleontologists to robustly develop, fit, and evaluate models of trait evolution.

Introduction

The project of documenting and explaining historical patterns of evolutionary change has had enduring significance since George Gaylord Simpson introduced the ideas of evolutionary tempo and mode as a tool for bridging observation and theory across micro- and macroevolutionary timescales (Simpson 1944). Fossil trait series provide a sequence of phenotypic measurements drawn from multiple organisms in the same lineage over a period of time, typically on the scale of tens to hundreds of thousands of years. In contrast to comparative methods, which rely on measurements of different species, fossil trait series document historical patterns of evolutionary change in a single species over much longer durations than we can access in the present. This makes fossil trait series a uniquely valuable data source for addressing fundamental questions in evolutionary biology occurring at those scales, including why evolutionary rates of change appear to show a robust scaling with time (Harmon et al. 2021) and whether evolutionary divergence is driven by the gradual accumulation of small changes or short, rapid bursts of change that punctuate long periods of stasis (Hunt et al. 2015). In addition, paleobiologists have long looked at these data for signatures of other evolutionary and ecological processes, seeking to gain insights into processes of adaptation (Voje 2020; Kearns et al. 2021), extinction (Brombacher et al. 2017, 2018), stasis (Antell et al. 2021), and parallel evolution (Stuart et al. 2020).

Because evolution is stochastic and involves the interaction of multiple processes such as selection and drift, statistical modeling is essential to reliably estimate evolutionary rates and classify fossil trait series into biologically significant patterns or modes of change (Hunt 2012; Gingerich 2019). In addition, the sample sizes of fossil lineage data are often small relative to the potential complexity of the system, so it is especially important to make effective use of the available information. The appropriate interpretation of evolutionary rates in a system, for

© The Author(s), 2025. Published by Cambridge University Press on behalf of Paleontological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

PALEOBIOLOGY A PUBLICATION OF THE Paleontological SOCIETY

CAMBRIDGE UNIVERSITY PRESS

example, is sensitive to the underlying mode of change: for a lineage evolving according to classical Brownian motion, we can understand its rate of change as the variance of random fluctuations it undergoes between points of time, but for a lineage exhibiting a linear directional trend in addition to Brownian motion, we also need to incorporate the magnitude of that linear tendency (Hunt 2012). Numerical estimation of evolutionary rates is also sensitive to model choice, and using an inadequate model for the data can generate scaling artifacts in the magnitude of rates in relation to the absolute time duration of the trait series (Hunt 2012; Gingerich 2019; Harmon et al. 2021).

In this paper, we present some practical tools and background theory for the use of linear state space models, also called dynamic linear models, in the analysis of phenotypic time series from fossil lineages. This approach provides some key advances over previous methodological approaches. First, state space models, which apply to time series with observations at discrete time points, are grounded in continuous-time models, which are important for knowing how to handle time-varying parameters when there are uneven time steps between observations. Second, the state space framework allows for the ready evaluation of exogenous environmental covariates. Third, the framework provides access to residuals for the predicted and observed values at each time point, a powerful model diagnostic tool that is in standard use across other scientific disciplines. We illustrate the value of these residuals to propose a new interpretation of a classic dataset of diatom evolution in Yellowstone Lake (Theriot et al. 2006). We also present some simulation studies emphasizing the quality of parameter estimation and model selection within this framework in comparison to previous frameworks for estimation.

We focus on univariate evolutionary models from a single species, because this foundation is essential for numerically efficient and stable parameter estimation as paleontologists start to increasingly take advantage of high-throughput specimen processing, including machine-learning methods for image segmentation and trait extraction (Porto and Voje 2020; Goswami and Clavel 2024; He et al. 2024). However, the modeling framework we develop here is readily generalized to analyzing multivariate measurements from variable numbers of individual specimens at each time point. This approach also demonstrates the broad relevance of stochastic integral models to analyzing fossil lineage datasets, building on prior work by Reitan et al. (2012), and is complementary to stochastic integral models for species richness (Hannisdal and Liow 2018; Reitan and Liow 2019) and comparative phylogenetic datasets (Blomberg et al. 2020; Turley 2020).

Methods

In this section, we briefly present the definition of a linear state space model, sometimes known as a dynamic linear model. We introduce the general structure of the model along with some adaptations and quantities that this formulation facilitates. Then, we show how observing an Ornstein-Uhlenbeck (OU) process at discrete time points can be formulated as a linear state space model. In particular, we make clear the relationship between the continuously defined model and the discretely sampled observations. We also discuss the relationship of other familiar models, such as an unbiased random walk (RW), random walk with a directional trend (DT), and stasis (ST) to the linear state space models in Appendix 2.

Definition of a Linear State Space Model

The linear state space model consists of two recursively defined equations: the system equation and the observation equation. We will focus on the unidimensional case for each equation, although multivariate extension is straightforward. The system equation is a linear stochastic equation describing the underlying dynamics of a phenomenon, such as the change of the level of a trait measurement for a population at a given time. The system equation can be written as

$$X_t = \Phi X_{t-1} + W_t \tag{1}$$

for $t = 1, \ldots, T$, where $W_1, \ldots, W_T$ are a sequence of independent and identically distributed normal random variables with zero-mean and variance σ_W^2. This is an autoregressive model with coefficient Φ. The model is called autoregressive because the sequence could be viewed as having the current time observation depend on the previous observation as a covariate. The observation equation is then defined as

$$Y_t = AX_t + V_t \tag{2}$$

for $t = 1, \ldots, T$, where $V_1, \ldots, V_T$ is a sequence of independent and identically distributed normal random variables with zero-mean and variance σ_V^2, which is also independent of the sequence $W_1, \ldots, W_T$. This sequence of $Y_1, \ldots Y_T$ will model the observed data and is a linear transformation by A of the system process, X_t, plus an additional observation error, V_t. Note that the notation here follows the book by Shumway and Stoffer, where detailed calculations associated with this model can be found (Shumway and Stoffer 2000).

In the subsection that follows and in Appendix 1, we will show how a number of familiar trait evolution models fit into this framework, especially when a few features are added. This linear state space model is the framework for calculating model likelihoods using the renowned Kalman filter. The Kalman filter recursively calculates the conditional distribution of X_t given observations $Y_1, \ldots, Y_t$; this means that if you have the Kalman filter calculated up to time $t-1$ and receive a new observation Y_t, then the conditional distribution of X_t given the data up to time t can be calculated without passing back through all of the previous data. A by-product of calculating the Kalman filter is that the likelihood function can be calculated with one pass through the data, so that the computational efficiency of this calculation is of order T. The likelihood function for a linear state space model can also be calculated using a multivariate normal distribution of size T, which has been the standard approach in previous paleontological research (Hunt et al. 2008; Voje 2020); however, that approach requires calculating the inverse of a $T \times T$ matrix. Hence there can be a significant speed up and improved stability using the Kalman filter algorithm to calculate the relevant likelihood function, especially for time series with many samples.

The Kalman filter approach also makes prediction relatively straightforward. The filter calculates the conditional mean and variance of X_t given $Y_1, \ldots, Y_t$, giving us access to a "best guess" for X_t with the data up to that time. This also allows us to predict, that is, to find the conditional mean and variance, of the next observation Y_{t+1} given the observations up to time t (i.e. $Y_1, \ldots Y_t$), which is known as the one-step ahead predictor of the data and also allows us to construct residuals for our data after having fit it to a specific state space model. Again, using the notation of Shumway and Stoffer, the standardized residuals would be defined as

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

$$\frac{Y_{t+1} - AX_{t+1}^{t}}{\sqrt{\Sigma_{t+1}}} \tag{3}$$

for $t = 1, \ldots, T-1$, where Σ_{t+1} is the conditional variance of Y_{t+1} given $Y_1, \ldots, Y_t$, which is calculated as part of the Kalman filter algorithm. Residuals in a time-series context are an important tool for discerning the quality of fit of a model, in a manner similar to their use in regression analysis. It is a way to "approximate" the sequence V_t, and we therefore expect them to be approximately independent and identically distributed.

In addition to facilitating likelihood function calculations and predictions of both the system equation model and future observations, the linear state space model facilitates regression with exogenous variables, which is especially important when considering the effects of environmental variables on trait dynamics. These can enter either through the system equation or the observation equations. In addition, the parameters defining the linear state space model can be time-varying. These modifications give

$$X_t = \Phi_t X_{t-1} + \Upsilon u_t + W_t \tag{4}$$

and

$$Y_t = A_t X_t + \Gamma u_t + V_t \tag{5}$$

where u_t is an $r \times 1$ column vector of exogenous variables, typically what we think of as covariates, for each t, and Υ and Γ are the coefficients that convert changes in the input variables into changes in X_t and Y_t. In this context, we can still use the Kalman filter to calculate the likelihood and to create predictions, and thus residuals, to evaluate model performance. Visually examining the residuals is a useful complement to the formal model misspecification tests for common fossil lineage models that were introduced by Voje (2018).

Unbiased Random Walk as a Linear State Space Model

Many commonly used trait evolution models can be represented in this state space framework. Perhaps the simplest case is that of an RW, which we develop here as an example. Under this model, the trait value at one time step (X_t) is equal to the value at the previous time step (X_{t-1}), plus an evolutionary perturbation (W_t). The autoregressive coefficient is unity ($\Phi = 1$), and the variance of the W_t is usually called the step variance in paleontology. The state equation represents the true trait mean of the population at each time step. Because sample sizes are finite, however, we can never know the true means of our samples. The calculated trait means that we observe, Y_t, reflects the true population mean (X_t), plus sampling noise (V_t). In most paleontological cases, our measurements of traits are noisy but unbiased, and so $A = 1$. Biased sampling from, for example, size-selective preservation, could be accounted for in principle by appropriate values of A, although the nature of bias would need to be independently estimated. For trait means, the observational variance of V_t for each sample will be the within-sample trait variance divided by the number of individuals measured in that sample.

Ornstein-Uhlenbeck as a Linear State Space Model

State space models are applied to discretely observed time series, and so an important step in analysis is to understand how continuous-time evolutionary models manifest in discrete time. This section presents this topic for OU models, with additional models considered in Appendix 1. Readers interested mostly in practical application of state space models may wish to proceed directly to the next section.

The OU process is used in evolutionary studies to model adaptation (Hansen 1997; Hunt et al. 2008). It is often defined by its stochastic differential equation (SDE) formulation.

$$dX(t) = -\alpha X(t)\,dt + \sigma\,dW(t) \tag{6}$$

This type of differential is a shorthand to define an integral equation. Such an equation is defined as a solution to sequences of difference equations:

$$\begin{aligned} X(n\Delta) - X((n-1)\Delta) = &-\alpha X((n-1)\Delta)\Delta \\ &+ \sigma W(n\Delta) - W((n-1)\Delta), \end{aligned} \tag{7}$$

as Δ goes to zero and $n\Delta$ converges to some terminal time T and with $X(0)$ defining an initial value for the equations. (This sequence is on an evenly spaced grid, which is not necessary in general as long as the grid spacing distance shrinks to zero.) The OU process has a solution (Oksendal 2013) in the form of a stochastic integral with deterministic integrand,

$$X(t) = e^{-\alpha t}X(0) + \int_0^t e^{-\alpha(t-s)}\sigma\,dW(s) \tag{8}$$

The definition of these SDEs as a limit of solutions to certain difference equations points to one possible discretization. Namely, we could rearrange the difference equation to arrive at

$$X(n\Delta) = (1-\alpha\Delta)X((n-1)\Delta) + \sigma W(n\Delta) - W((n-1)\Delta) \tag{9}$$

but this is only an approximate solution that accurately represents the model definition when Δ is small. However, if we take the integral solution of equation (6) and manipulate it appropriately, we arrive at

$$X_n = e^{-\alpha\Delta}X_{n-1} + \int_{(n-1)\Delta}^{n\Delta} e^{-\alpha(n\Delta - s)}\sigma\,dW(s) \tag{10}$$

We can use Ito's isometry to calculate the variance of the last term, which is Gaussian and has a zero mean. The variance is then

$$Var\left(\int_{(n-1)\Delta}^{n\Delta} e^{-\alpha(n\Delta-s)}\sigma\,dW(s)\right) = \sigma^2 e^{-2\alpha n\Delta}\int_{(n-1)\Delta}^{n\Delta} e^{2\alpha s}\,ds \tag{11}$$

We solve the integral to arrive at the variance of the last term

$$\sigma_\Delta^2 = \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha\Delta}\right) \tag{12}$$

Note that when Δ is small, then a Taylor approximation will verify that this expression is approximately equal to $\sigma^2\Delta$ corresponding to equation (7). So, we can express this discretization as

$$X_n = e^{-\alpha\Delta}X_{n-1} + \sigma_\Delta\varepsilon_n \tag{13}$$

where ε_n is a sequence of independent standard normal random variables.

The OU model will eventually settle around zero regardless of the initial condition. We can modify this part of the model for the OU process to be centered around another constant, θ. The exact discrete version would be

$$X_n = c + e^{-\alpha\Delta}(X_{n-1} - \theta) + \sigma_\Delta\varepsilon_n \tag{14}$$

This formulation allows for one way to introduce exogenous variables by replacing θ with a linear combination of covariates, as demonstrated in "Application: Stephanodiscus yellowstonensis Trait Evolution.".

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

Model Estimation and Selection Simulations

We used simulations to validate model fitting in the state space framework using the Kalman filter compared to prior work. We also explored parameter estimation and model selection performance in the state space framework. Details are provided in Appendix 2, but in general, performance was as expected for likelihood-based methods. Corrected Akaike information criterion (AICc) scores generally favored the correct, generating models, increasingly so with increasing sequence length. Parameter estimates were unbiased, or nearly so, although convergence could depend on two inherent observational timescales in trait series: the time step between observed samples (Δ in equation 7, for example) and the total duration between the first and last observations. Parameter convergence may therefore depend on how increasing the sample size alters these timescales. For example, subdividing a fixed interval of time with more observations does not lead the linear trend parameter, μ, to converge asymptotically in the DT model. In contrast, smaller time steps are valuable for estimating parameters of the OU and decelerating evolution models.

Application: Stephanodiscus yellowstonensis *Trait Evolution*

To illustrate the linear state space framework, we re-analyzed the *Stephanodiscus yellowstonensis* fossil trait series published in Voje (2020), originally created by Theriot and colleagues (2006), using the discretized models described in "Methods" and implemented in the state space framework. Script and data files to perform this analysis have been uploaded to Zenodo (see "Data Availability Statement").

Stephanodiscus yellowstonensis is a species of diatom endemic to Yellowstone Lake, Wyoming, USA, and likely descended from the preexisting species *S. niagarae*, which is still extant throughout the region (Theriot et al. 2006). The fossil trait series is derived from 63 samples from a sediment core collected from the lake's central basin, and it covers approximately 14,000 years ago until the present. For each sample, Theriot and co-workers measured 50 individuals, occasionally fewer if this number of specimens was not available. They measured three traits on each individual: valve diameter, the number of costae per valve, and the number of spines per valve. All three traits show a relatively rapid increase in values from about 12,000 to 10,000 years ago, with a slower and fluctuating decrease thereafter. To illustrate the model-fitting methods, we focus on just one variable, spine count (Fig. 1). As noted by Voje (2020), all three traits are considered to be ecologically important for diatoms; valve spines, in particular, may enhance nutrient uptake and photosynthetic rate by affecting how diatoms sink through the water column. Spine counts were log-transformed before analysis, because we consider a proportional scale to be more appropriate for the evolution of this trait. This transformation also has the effect of removing a strong correlation between the mean and variance among samples ($r = 0.70$, $P < 10^{-4}$ for untransformed spine counts, $r = 0.04$, $P = 0.77$ after log-transformation).

Theriot et al. (2006) posit that morphological changes in the *S. yellowstonensis* lineage track environmental changes, and these authors synthesize the available records of regional change through the study interval. Here, we quantitatively analyze two of these records. The first is the proportion of the dominant pollen type, attributable to *Pinus contortus*, as reflecting floral change in the area. See Figure 2, digitized from figure 10 in Theriot et al. (2006). The second environmental record we analyzed is solar insolation, which peaked around 11,000 years ago and decreased to the present day. Insolation values in W/m^2 were taken from model output from Lasker et al. (2004), using the web interface http://vo.imcce.fr/insola/earth/online/earth/online/index.php for the latitude and longitude of Lake Yellowstone. As *S. yellowstonensis* blooms in summer (Theriot et al. 2006), we used insolation values for the month of June (Fig. 2).

The sampling times of the environmental records did not precisely match those of the trait time series. We used linear interpolation to produce time series of the environmental records, sampled at the same times as the traits. Both environmental records were mean-centered before analysis to facilitate model fitting.

Our overall model-fitting strategy started with the five models considered by Voje (2020): stasis (ST), random walk (RW), directional trend (DT), Ornstein-Uhlenbeck (OU), and decelerated evolution (DE). To test Theriot et al.'s (2006) suggestion that environmental changes influenced morphological evolution, we added OU models in which the trait optimum linearly tracks the two environmental covariates described earlier, June solar insolation (OU_{insol}) and the proportional abundance of *Pinus contortus* pollen (OU_{pollen}). Because model fits and residuals indicated a decrease in stochastic evolutionary change through the core (see "Results"), we considered additional models that allowed for a one-time decrease in the step variance. We specified this time to be 10,000 years ago, following Theriot et al.'s observation (2006: p. 45) that environmental conditions were much more stable after this date.

All these models were fit using functions in the R package paleoTS, the recent update of which (version 0.6.1) allows for fitting models via the Kalman filter and a state space model approach. Approximate profile confidence intervals on parameter estimates were generated using the dentist package (Boyko and O'Meara 2024).Of the five models fit by Voje, we found that DE was best supported by AICc (Table 1, models 1–5), consistent with Voje's findings. The maximum-likelihood parameter estimate for the decay parameter of the DE model implies a roughly sevenfold decrease in the step variance over the course of the 14 kyr sequence ($\hat{r} = -0.00014$; Fig. 3). Examination of residuals indicates that decrease in the stochastic component over time is an important signal in this dataset. Models without this dynamic show residuals with elevated spread early in the sequence (Fig. 4 as an example from the RW model). In contrast, residuals from the DE and best-fitting models are not structured in this way, showing a pattern closer to the ideal even spread (Figs. 5, 6). Also consistent with these data is a model with a single step-down in variance at 10 ka, with separate stochastic variances before ($\sigma^2_{W_1}$) and after ($\sigma^2_{W_2}$) this shift point. This model (Table 1, model 6) fits very slightly better than the DE model. Parameter estimates from this model indicate that the step variance decreases from 2.77×10^{-5} to 1.40×10^{-5}, about a twofold drop (Fig. 3).

Comparing the two covariate tracking models, it is more plausible that spine counts tracked June solar insolation rather than the pollen data (Table 1, models 8 and 7, AICc difference of 4.9) (Fig. 7). The OU_{insol} model shares the features of the OU model, except that the fitness optimum varies with solar insolation instead of being constant over time. The large increase in support between the OU to OU_{insol} (Table 1, ΔAICc = 6.1) is therefore a measure of the importance of solar insolation in accounting for these observations. Combining this insolation-tracking dynamic with a step decrease in stochastic variance results in a model that is best supported overall (Table 1, model 9; Table 2). This model implies a dynamic where spine counts deterministically follow summer insolation, with

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

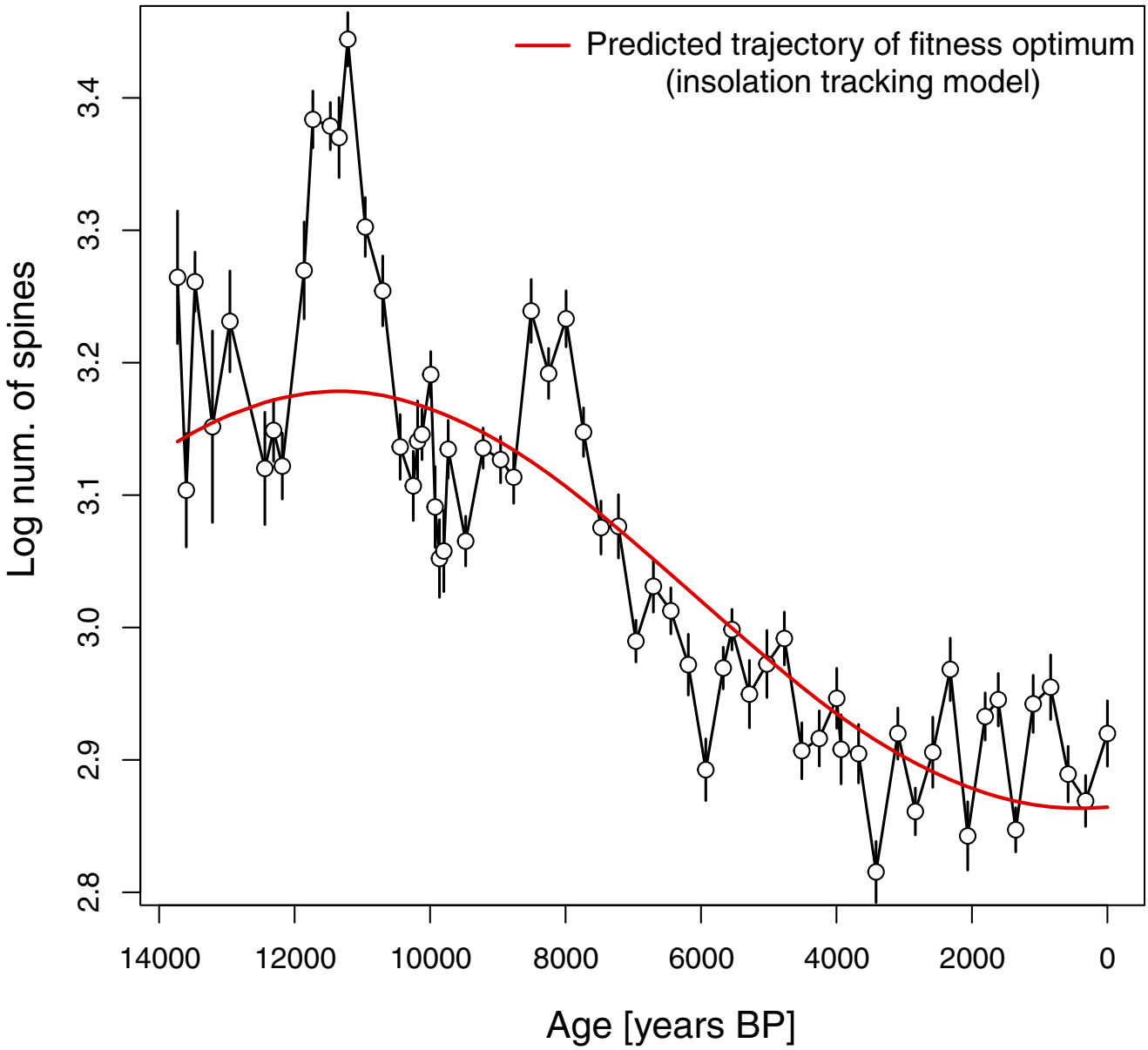


Figure 1. Evolutionary time series of log spine counts for *Stephanodiscus yellowstonensis.* Open circles and vertical bars show mean values ±1 SE on those means. The red line is the model-predicted trajectory of the fitness optimum for the best-supported model (see text). Age is in years before present day.

overlaid stochastic evolution that is initially high, but then decreases later on.

Discussion

Using the State Space Modeling Framework

We have presented a novel framework based on stochastic integrals and linear state space models for describing and analyzing five models for univariate trait evolution in fossil lineages. We have shown how the stochastic integral approach provides a clearer conceptual basis for relating underlying parameters stated in continuous time to models incorporating discretized sampling and observational error. In particular, we showed how a property of the sampling regime, the duration between observed time points (Δ), enters into the system equations of the OU model. Looking forward, the framework is naturally generalizable to multivariate systems.

The five base models considered here have all been implemented before for paleontological time series, and they have been fit using maximum likelihood (Hunt 2006; Voje 2020). The present approach, using state space models and the Kalman filter, offers an alternative means to compute these same model likelihoods. The two approaches will yield log-likelihoods that are the same (within a constant) and the resulting maximum-likelihood parameter estimates are equivalent, within precision of the hill-climbing algorithm used to search for the best parameter values. Two practical benefits of using the state space approach are that the Kalman filter calculations (1) naturally produce predicted values and residuals useful for assessing model adequacy and (2) do not require inverting a large T by T matrix, where T is the number of samples in the time series. The second benefit applies mostly to rather long time

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

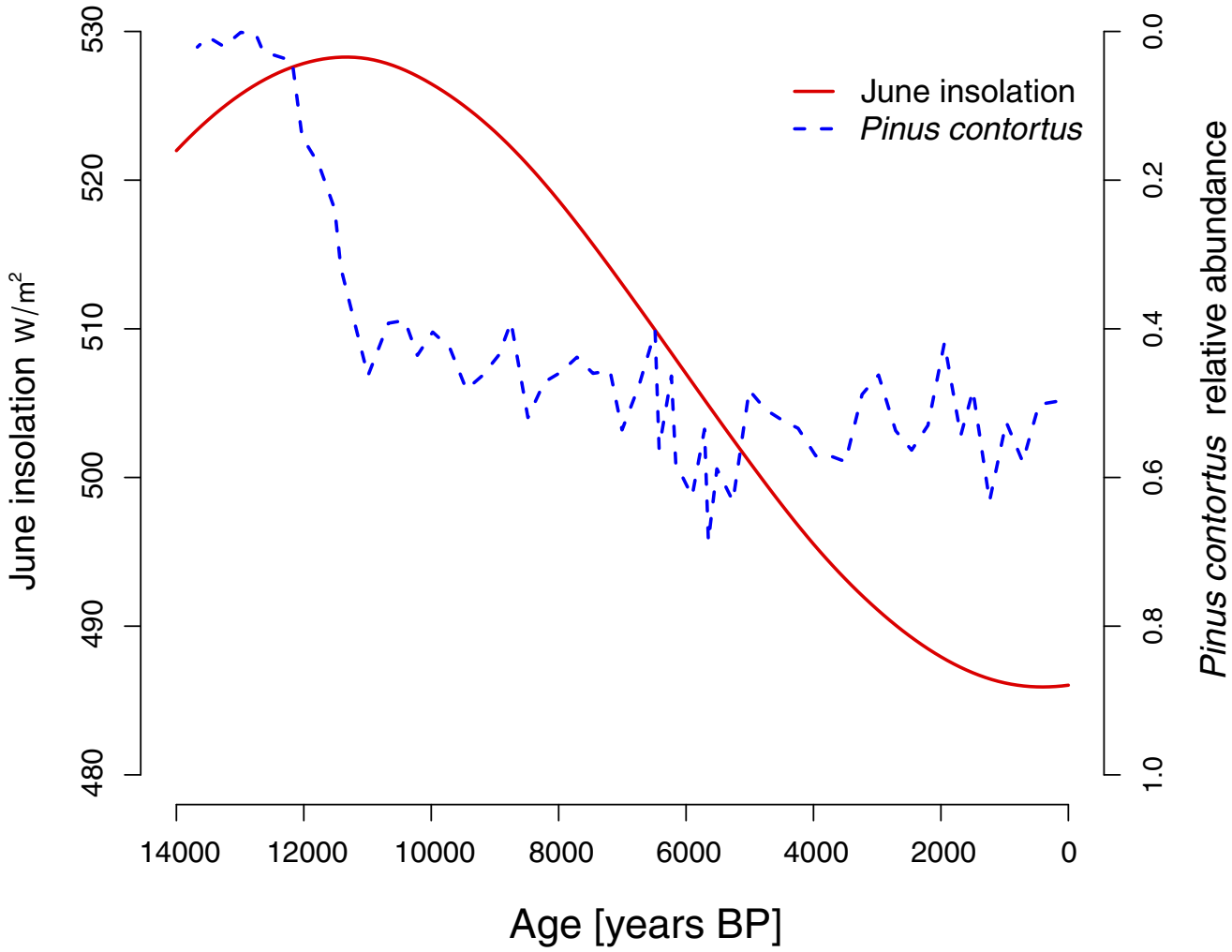


Figure 2. Measured environmental covariates, including June solar insolation (red solid line) and proportion of pollen attributable to *Pinus contortus* (blue dashed line; note reversed axis).

Table 1. Model fits to spine counts of the *Stephanodiscus yellowstonensis* lineage. From left to right, columns give model abbreviations, log-likelihoods, number of parameters, corrected Akaike information criterion (AICc) scores, ΔAICc scores, and Akaike weights. Model abbreviations: RW, random walk; DT, directional trend; OU, Ornstein-Uhlenbeck; DE, decelerating evolution; RW_{shift}, random walk with a shift in the step variance parameter at 10 ka; OU_{pollen}, OU model in which the trait optimum tracks the proportion of the pollen comprised of *Pinus contortus*; OU_{insol}, OU model in which the trait optimum tracks solar insolation; $OU_{insol-shift}$, OU model in which the trait optimum tracks solar insolation with a shift in the step variance at 10 ka

	Model	logL	K	AICc	ΔAICc	Akaike weight
1	Stasis	87.474	2	−170.748	105.625	0
2	RW	137.003	2	−269.806	6.567	0.023
3	DT	137.222	3	−268.037	8.336	0.010
4	OU	138.536	4	−268.383	7.990	0.012
5	DE	138.322	3	−270.237	6.136	0.029
6	RW_{shift}	138.454	3	−270.500	5.872	0.033
7	OU_{pollen}	140.355	5	−269.658	6.714	0.022
8	OU_{insol}	142.783	5	−274.515	1.858	0.247
9	$OU_{insol-shift}$	144.936	6	−276.373	0	0.625

series ($T > 100$), as this inversion becomes slow and is prone to fail for very large matrices.

The analysis of spine counts in *Stephanodiscus yellowstonensis* is a good example of a typical workflow with the state space approach. An initial set of models were considered, drawn from existing theory and prior interpretations of the system under study. Model fits, as well as examination of residuals, suggest that there are two important signals in the data captured by these models: (1) a decrease over time in the stochastic component of evolutionary change and (2) a correlation between diatom phenotypes and summer solar insolation. The Kalman filter calculations allowed us to quickly implement a model with both of these components, which turned out to be the best supported among those considered. The modular nature of the Kalman filter can thus facilitate model development, as it is allows users to easily combine evolutionary components into new composite models of trait evolution.

We added a single step-down in variance, rather than the exponential decrease of the DE model, to the OU covariate-tracking model. The step decrease in variance is slightly favored over the DE model, but the decision was also a practical one; its incorporation into the OU with covariate tracking is more straightforward. Given the near equivalance in model support between the DE and this discrete shift (Table 1), it is unlikely that these data could

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

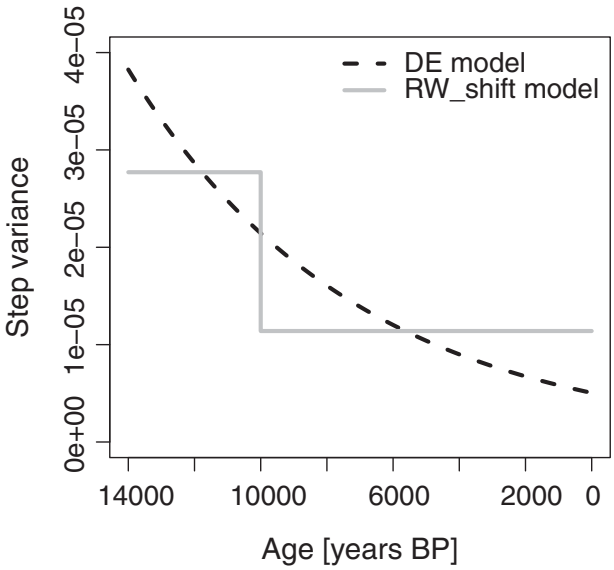


Figure 3. Modeled changes in the step variance predicted by the two models for which this parameter varies over time. The decelerated evolution (DE) model has a step variance that exponentially decreases, whereas the random walk with a shift in the step variance parameter at 10 ka (RW_{shift}) model posits a single decrease in step variance that occurs at 10 ka.

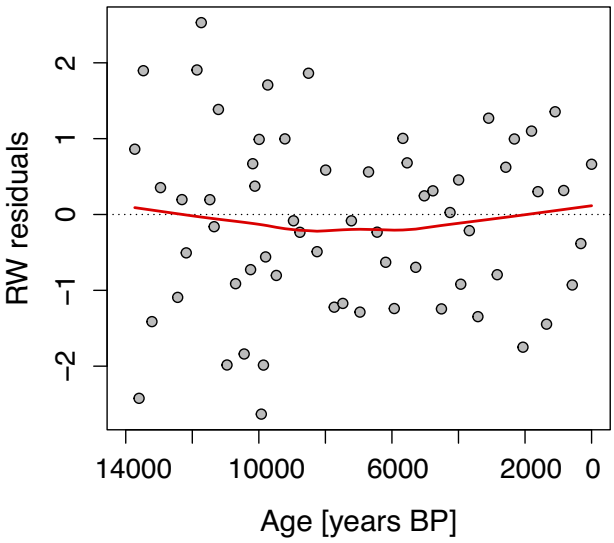


Figure 4. Residuals from the random walk (RW) model. Note the greater spread of residuals early in the sequence.

discriminate between the two different ways of modeling a reduction in the step variance over time.

Microevolution in Stephanodiscus yellowstonensis

The set of evolutionary models considered here are usually interpreted as phenomenological, not mechanistic. Phenomenological models are useful for representing, using just a few parameters, qualitatively different kinds of dynamics, such as meandering change (RW), fluctuations around a stable mean (ST), or directional trends. In some cases, these models can be shown to be the expected outcome of specific microevolutionary scenarios (e.g., neutral genetic drift will produce a random walk). But these models

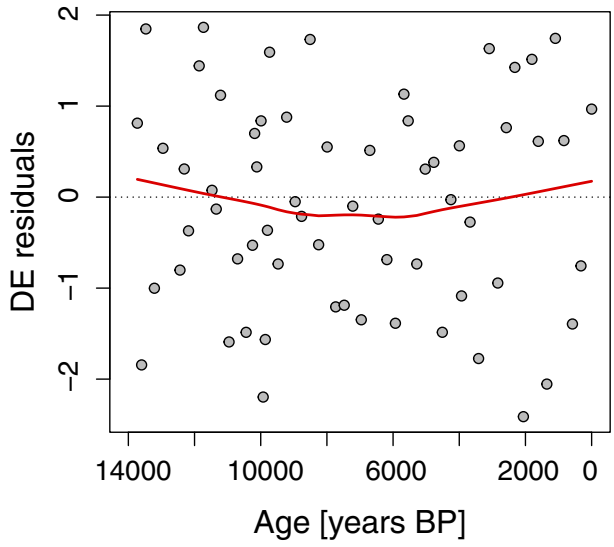


Figure 5. Residuals from the decelerated evolution (DE) model. Note they are much less structured over time compared with those from the random walk (RW) model.

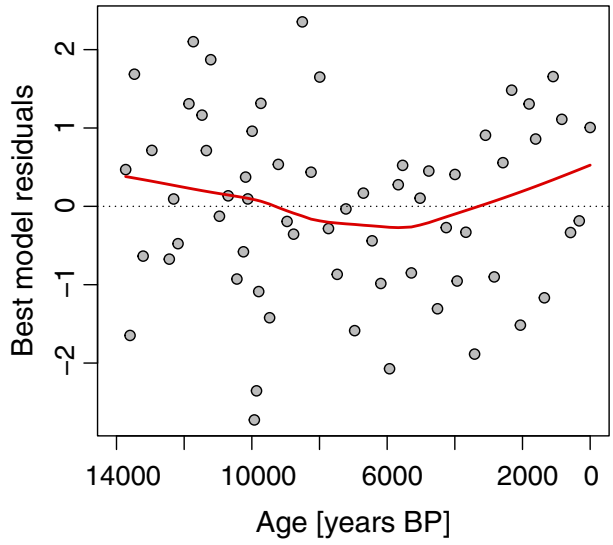


Figure 6. Residuals from the best-fitting model (model 9 in Table 1). Note that the spread of residuals does not markedly change over time as it does in the random walk (RW) model.

are usually interpreted descriptively, rather than as the outcome of specific microevolutionary mechanisms.

One potential exception is the OU model. Under a set of simplified but reasonable assumptions, this model describes the expected trajectory of a population evolving in the vicinity of a peak in the adaptive landscape (Lande 1976). The peak corresponds to the trait value that results in highest mean population fitness. This peak is stable in the OU model, but changes with extrinsic variables in the covariate-tracking versions implemented here. If the best-fitting of these (OU_{insol}) is in fact a complete description of the microevolutionary process, its parameters can be related to population genetic parameters related to the strength of natural selection (from the α parameter) and the effective population size (N_e, from σ^2), as described by Hunt et al. (2008).

Under this set of assumptions, natural selection is inferred to be rather weak. Mean fitness decreases only 1% or less for individuals

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

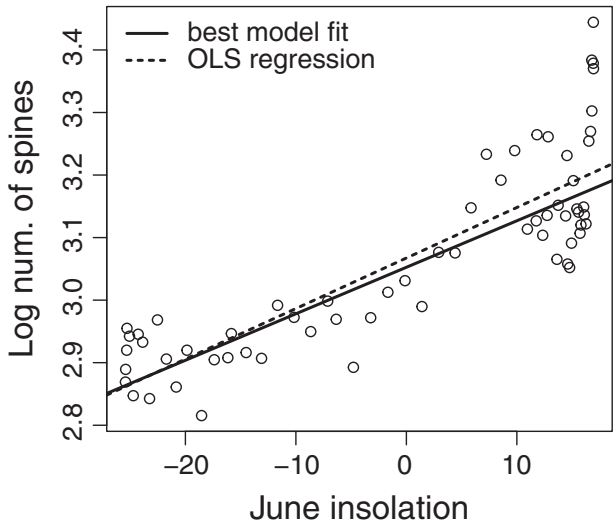


Figure 7. Log of spine count plotted with respect to June solar insolation in W/m^2. Solid line shows the relationship between the two variables predicted by the parameter estimates (Table 2) of the best-fitting model (model 9 in Table 1). Dotted line shows the fit from ordinary least-squares (OLS) regression. The close similarity between the two curves is expected with rapid adaptation, as found here.

Table 2. Maximum-likelihood estimates (MLE) and confidence intervals (CI) for the best-fitting model: an Ornstein-Uhlenbeck (OU) process in which the position of the optimum depends linearly on the value of summer insolation and the step variance is estimated separately before (σ_1^2) and after (σ_2^2) 10,000 years ago. b_0 and b_1 are the intercept and slope of the relationship between solar insolation and the trait optimum, α represents the force of attraction to that optimum, and *anc* is the estimated trait value at the start of the time series

	anc	σ_1^2	σ_2^2	α	b_0	b_1
MLE	3.24	0.0325	0.0113	0.00247	3.05	0.00744
Lower 95% CI	3.05	0.0135	0.0237	0.00025	2.98	0.00305
Upper 95% CI	3.43	0.0559	0.0451	0.00767	3.11	0.0141

Table 3. Estimates of microevolutionary parameters calculated from the Ornstein-Uhlenbeck (OU) model in which the trait optimum follows summer insolation, with a step-down in step variance at 10 ka (model 9 in Table 1). Shown are calculations assuming low (0.1) and high (0.7) plausible values of trait heritability and 10 generations per year. ω is the computed variance of the population fitness function; larger values indicate broader fitness curves and therefore weaker stabilizing selection. Fitness reduction is the resulting decrease in population mean fitness between the optimal trait value and the trait values corresponding to three population standard deviations away from the optimum. Effective population size (N_e) is computed separately before and after 10 ka

h^2	ω	Fitness reduction	Half-life of adaptation (gen)	N_e (before 10 ka)	N_e (after 10 ka)
0.1	9.37	1.1%	2802	482	1384
0.7	65.7	0.16%	2802	3375	9686

3 SD away from the optimum (Table 3). Caution should be exercised here, because the timescale of adaptation is inferred to be rapid relative to temporal sampling resolution. Its half-life—the amount of time it would take the population to progress halfway to the optimum—is only about 280 years (Table 3), which is close to the median spacing between samples (258 years). As a result, selection could be much stronger than what is estimated, but we would not be able to detect it without finer temporal resolution.

The other population genetic parameter that can be calculated is the effective population size, N_e. This parameter determines the magnitudes of change due to genetic drift in these models. Drift is more potent in smaller populations, and thus lower N_e values correspond to larger stochastic changes (= higher step variances) around the adaptive trajectory of OU models. The fit of the best model implies about a threefold increase in effective population size at 10 ka. This is consistent with a stepwise increase in the abundance of *S. yellowstonensis* observed in the core at this time (Theriot et al. 2006).

Although the direction of this change is consistent with an increase in the observed absolute abundances of this lineage, the magnitudes of estimated effective population sizes, ranging from 10^2 to 10^4 (Table 3), seem rather low for these unicellular algae, which can be found living at abundances high enough to produce that many individuals in just 100 liters of water, or fewer (see Interlandi et al. 1999: p. 679). However, it is important to note that effective population size is generally much lower than census population size, with the discrepancy between the two increasing with fluctuations in population size, differences in fitness across individuals, inbreeding, and other factors. The two literature estimates of N_e for diatoms are for widespread marine species and are about 10^7, which although very high, is still orders of magnitude lower than their peak census abundances (Krasovec et al. 2019). It is likely that our N_e estimates for *S. yellowstonensis* are unrealistically low, but the population genetics of lake diatoms is not well enough investigated to be sure of this.

Assuming that the N_e values computed from the best model are indeed too low, then genetic drift would not be sufficient to account for the stochastic component of spine count evolution. Therefore, other factors, in addition to June insolation, will have caused changes in the position of the adaptive optimum for spine counts, and hints of non-randomness in the residuals of the best-fitting model (Fig. 6) may reflect these unmeasured factors. Theriot et al.'s (2006) presentation of the paleoenvironmental record provides a detailed account of environmental variation that might contribute to these insolation-independent evolutionary change. In particular, the decrease in the stochastic evolutionary component after 10 ka may be explained by the shift to more stable conditions at this time, leading to more modest selective fluctuations in diatom morphology. Voje (2020) offers an alternative explanation in which ecological opportunity is initially high, perhaps because of the phenotypic changes associated with the origin of the new species *S. yellowstonensis*. With high ecological opportunity, stabilizing selection may be weakened, permitting greater variation and larger evolutionary steps. This interpretation is consistent with the observation that standing variation in spine counts is initially high and decreases steadily for the first 3 or 4 kyr of the sequence (Fig. 8). These two explanations for the reduction in stochastic evolutionary change—decreasing environmental variation and decreasing ecological opportunity—are not mutually exclusive.

Analysis of stratigraphic trait changes can be complicated by hiatuses, variation in sedimentation rate, and other aspects of stratigraphic architecture (Patzkowsky and Holland 2012; Hohmann et al. 2024). Thanks to the continuous sedimentation in Lake Yellowstone over the study interval (Theriot et al. 2006), however, such complications are largely avoided in the present case.

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

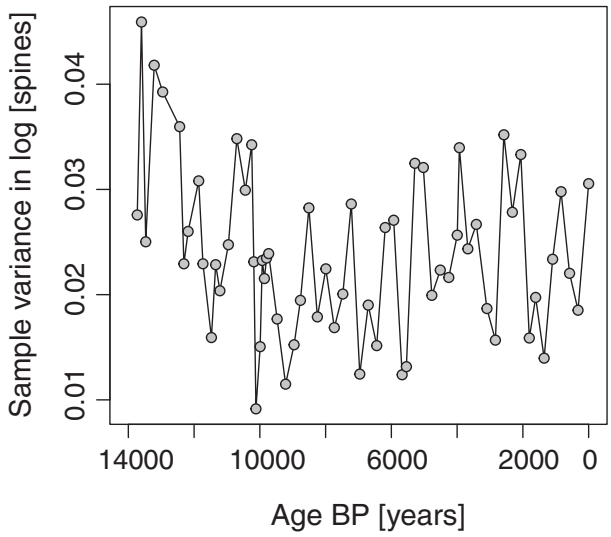


Figure 8. Within-sample variance in spine count decreases in the initial part of the *Stephanodiscus yellowstonensis* sequence.

Handling Covariates in Models of Trait Evolution

Although a model in which spine counts follow solar insolation as an OU process is the best supported among those considered, we caution that aspects of this model may make it less suitable for some situations. Our implementation requires an assumption that the position of the trait optimum is constant in between the time points at which we have observations. This is reasonable when, as is the case here, the covariates show point-to-point changes that are small compared with the total range of the time series. This assumption will be less appropriate for covariates that fluctuate widely on short timescales. In addition, the microevolutionary intepretation of the OU dynamics is only tenable if the sampling resolution is fine enough to potentially capture the adaptive dynamics of the population chasing the moving fitness peak. Even with the exceptional temporal resolution of the *S. yellowstonensis* data, the evolutionary dynamics may be too rapid to well constrain estimates for the strength of stabilizing selection. In addition, this modeling approach may be prone to receiving spuriously strong support when applied to traits that show clearly directional change if analyzed with covariates that are also trended. Including the simpler DT model of a trend (without covariates), as is done here, may protect against this effect, as the fewer parameters of the trend model give it an AICc advantage.

State space models are flexible enough to allow for other approaches to incorporate the effects of exogenous covariates that may be more suitable in other circumstances. For example, one could model a trait as an RW, with an additional pulse of change that is proportional to changes in a covariate, implemented through the Υ term of the state equation. Such a modeling approach does not attempt to capture the dynamics of a population climbing an adaptive peak and instead would be consistent with an assumption that enough time has elapsed between samples for the population to have reached the adaptive optimum. This approach would therefore be more appropriate for trait time series at more typical paleontological resolutions. And the use of *changes* in covariates as input variables, rather than covariates themselves, would render this approach less susceptible to trended sequences as described earlier.

Conclusion

The state space framework provides a practical approach for analyzing phenotypic evolution in fossil lineages that facilitates model development, including those that incorporate exogenous environmental variables. We highlighted some additional useful features, especially the ease of accessing time-series residuals and enhanced numerical stability and efficiency. Our analysis suggested a novel biological interpretation of evolution in spine count for *S. yellowstonensis* based on stabilizing selection to changing solar insolation levels. Our focus on univariate trait models in this paper provides a foundation for expanding into more complex, multivariate models that, for example, allow for estimation of trait covariances in a time-series setting. Such a multivariate approach will be essential to understand how solar insolation may jointly affect all three traits measured by Theriot et al. (2006) and to better understand in general the evolution of integrated phenotypes over time.

Acknowledgments. This research was supported by the Templeton Foundation, award number 62220, and the National Museum of Natural History. J.F. also received financial support from the French government in the framework of the France 2030 program IdEx Université de Bordeaux. We are grateful to K. Voje and an anonymous reviewer for their helpful and careful reading of the manuscript and to E. Theriot for providing raw data and insight about this system.

Competing Interests. The authors declare no competing interests.

Data Availability Statement. Data files and analysis scripts have been uploaded to Zenodo (https://zenodo.org) and can be accessed at https://doi.org/10.5281/zenodo.14984759.

Literature Cited

Antell, G. S., I. S. Fenton, P. J. Valdes, and E. E. Saupe. 2021. Thermal niches of planktonic foraminifera are static throughout glacial-interglacial climate change. *Proceedings of the National Academy of Sciences* **118**:e2017105118.

Blomberg, S. P., S. I. Rathnayake, and C. M. Moreau. 2020. Beyond Brownian motion and the Ornstein-Uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. *American Naturalist* **195**:145–165.

Boyko, J. D., and B. C. O'Meara. 2024. dentist: quantifying uncertainty by sampling points around maximum likelihood estimates. *Methods in Ecology and Evolution* **15**:628–638.

Brombacher, A., P. A. Wilson, I. Bailey, and T. H. G. Ezard. 2017. The breakdown of static and evolutionary allometries during climatic upheaval. *American Naturalist* **190**:350–362.

Brombacher, A., P. A. Wilson, I. Bailey, and T. H. G. Ezard. 2018. Temperature is a poor proxy for synergistic climate forcing of plankton evolution. *Proceedings of the Royal Society B* **285**:20180665.

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

Gingerich, P. D. 2019. *Rates of evolution: a quantitative synthesis*. Cambridge University Press, Cambridge.

Goswami, A., and J. Clavel. 2024. Morphological evolution in a time of phenomics. *Paleobiology* **51**:195–213.

Hannisdal, B., and L. H. Liow. 2018. Causality from palaeontological time series. *Palaeontology* **61**:495–509.

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

Harmon, L. J., M. W. Pennell, L. F. Henao-Diaz, J. Rolland, B. N. Sipley, and J. C. Uyeda. 2021. Causes and consequences of apparent timescaling across all estimated evolutionary rates. *Annual Review of Ecology, Evolution, and Systematics* **52**:587–609.

He, Y., J. M. Mulqueeney, E. C. Watt, A. Salili-James, N. S. Barber, M. Camaiti, E. S. E. Hunt, et al. 2024. Challenges and opportunities in applying

AI to evolutionary morphology. *EcoEvoRxiv*, https://ecoevorxiv.org/repository/view/6525/.

Hohmann, N., J. R. Koelewijn, P. Burgess, and E. Jarochowska. 2024. Identification of the mode of evolution in incomplete carbonate successions. *bioRxiv*, https://doi.org/10.1101/2023.12.18.572098.

Hunt, G. 2006. Fitting and comparing models of phyletic evolution: random walks and beyond. *Paleobiology* **32**:578–601.

Hunt, G. 2008. Evolutionary patterns within fossil lineages: model-based assessment of modes, rates, punctuations and process. *Paleontological Society Papers* **14**:117–131.

Hunt, G. 2012. Measuring rates of phenotypic evolution and the inseparability of tempo and mode. *Paleobiology* **38**:351–373.

Hunt, G., M. A. Bell, and M. P. Travis. 2008. Evolution toward a new adaptive optimum: phenotypic evolution in a fossil stickleback lineage. *Evolution* **62**: 700–710.

Hunt, G., M. J. Hopkins, and S. Lidgard. 2015. Simple versus complex models of trait evolution and stasis as a response to environmental change. *Proceedings of the National Academy of Sciences USA* **112**:4885–4890.

Interlandi, S. J., S. S. Kilham, and E. C. Theriot. 1999. Responses of phytoplankton to varied resource availability in large lakes of the Greater Yellowstone Ecosystem. *Limnology and Oceanography* **44**:668–682.

Kearns, L. E., S. M. Bohaty, K. M. Edgar, S. Nogué, and T. H. G. Ezard. 2021. Searching for function: reconstructing adaptive niche changes using geochemical and morphological data in planktonic foraminifera. *Frontiers in Ecology and Evolution* **9**:679722.

Krasovec, M., S. Sanchez-Brosseau, and G. Piganeau. 2019. First estimation of the spontaneous mutation rate in diatoms. *Genome Biology and Evolution* **11**: 1829–1837.

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

Oksendal, B. 2013. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, New York.

Patzkowsky, M. E., and S. M Holland. 2012. *Stratigraphic paleobiology: understanding the distribution of fossil taxa in space and time*. University of Chicago Press, Chicago.

Porto, A., and K. L. Voje. 2020. ML-morph: a fast, accurate and general approach for automated detection and landmarking of biological structures in images. *Methods in Ecology and Evolution* **11**:500–512.

Reitan, T., and L. H. Liow. 2019. layeranalyzer: inferring correlative and causal connections from time series data in R. *Methods in Ecology and Evolution* **10**: 2183–2188.

Reitan, T., T. Schweder, and J. Henderiks. 2012. Phenotypic evolution studied by layered stochastic differential equations. *Annals of Applied Statistics* **6**: 1531–1551.

Shumway, R. H, and D. S Stoffer. 2000. *Time series analysis and its applications*. Vol. **3**. Springer, New York.

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

Stuart, Y. E., M. P. Travis, and M. A. Bell. 2020. Inferred genetic architecture underlying evolution in a fossil stickleback lineage. *Nature Ecology & Evolution* **4**:1549–1557.

Theriot, E. C., S. C. Fritz, C. Whitlock, and D. J. Conley. 2006. Late Quaternary rapid morphological evolution of an endemic diatom in Yellowstone Lake, Wyoming. *Paleobiology* **32**:38–54.

Turley, L. M. 2020. Ornstein Uhlenbeck–Jump Models of Evolution. M.S. thesis. University of Otago, Dunedin, New Zealand. https://ourarchive.otago.ac.nz/handle/10523/10401.

Voje, K. L. 2018. Assessing adequacy of models of phyletic evolution in the fossil record. *Methods in Ecology and Evolution* **9**:2402–2413.

Voje, K. L. 2020. Testing eco-evolutionary predictions using fossil data: phyletic evolution following ecological opportunity. *Evolution* **74**:188–200.

Wagenmakers, E.-J., and S. Farrell. 2004. AIC model selection using Akaike weights. *Psychonomic Bulletin & Review* **11**:192–196.

Appendix 1. Common Models as Linear State Space Models

Stochastic Integral Models and Their Discretizations

In this section, we will discuss a number of well-known models of trait evolution in fossil lineages and give a corresponding continuous time equivalent. We will then show that each of these models can be exactly discretized to correspond to discretely observed data.

Each of these standard models may be expressed as an Ito integral with a deterministic integrand. We can then look at how each can then be expressed as a linear state space model when observed at discrete time points.

First, the definition of a Ito integral for a deterministic integrand is:

$$X(t) = \int_0^t f(t,s)\,dW(s) \tag{A.1.1}$$

where $f(t,s)$ is a deterministic function and $W(s)$ is a standard Weiner process (Oksendal 2013). In other words, for each s, $W(s)$ is a normal random variable with zero mean and variance equal to s. This process also has the property of independent increments, implying that $W(t)-W(s)$ is independent of $W(v)-W(u)$ as long as $(s,t]$ and $(u,v]$ do not overlap. The fact that $W(s)$ has zero mean for each s implies that $X(t)$ does also. Each of the models that we examine in this paper is a Gaussian process, and using this integral representation allows us to express the models in a unified way.

An important property of this integral, especially when calculating variances, is the Ito isometry (Oksendal 2013).

$$E\left[X^2(t)\right] = E\left(\int_0^t f(t,s)\,dW(s)\right)^2 = \int_0^t \left(f(t,s)\right)^2 ds \tag{A.1.2}$$

Now, we write some basic models in the form of such an integral along with their exact discretizations and approximate discretizations where appropriate.

Random Walk (RW)

For the RW model, the deterministic function is simply a constant, σ, and

$$X(t) = \sigma W(t) \tag{A.1.3}$$

Note then, that a discretely observed version of this model, assuming equal spacing in time, would be

$$X(n\Delta) = X((n-1)\Delta) + \sigma(W(n\Delta) - W((n-1)\Delta)) \tag{A.1.4}$$

which we could we write as follows by using the fact of independent increments of Brownian motion:

$$X_n = X_{n-1} + \sigma\sqrt{\Delta}\varepsilon_n \tag{A.1.5}$$

where ε_n is a sequence of independent standard normal random variables, and Δ is the amount of time between observations. This can be derived directly from the earlier integral definition with the $\sqrt{\Delta}\varepsilon_n$ corresponding to $W(n\Delta) - W((n-1)\Delta)$. In this discrete version, X_n corresponds to $X(n\Delta)$.

Directed Random Walk

For this model, we add a deterministic linear directionality to the RW model.

$$X(t) = \mu t + \sigma W(t) \tag{A.1.6}$$

In a similar way then, a discretely observed version of this model, again assuming equal spacing in time, would be

$$X_n = X_{n-1} + \mu\Delta + \sigma\sqrt{\Delta}\varepsilon_n \tag{A.1.7}$$

with similar interpretations as for the RW.

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

Decelerated Evolution (DE)

Voje (2020) considered a model for the evolution of a trait where the step variance of a RW declines exponentially. In other words, a model that could be describe with the recursion

$$X_n = X_{n-1} + \sigma_V e^{\frac{-r}{2}n}\varepsilon_n \tag{A.1.8}$$

where σ_V and r are positive parameters and ε_n is a sequence of standard normal random variables. A natural way to write this as a stochastic integral would be

$$X(t) = X(0) + \int_0^t e^{\frac{-r}{2}s}\sigma dW(s) \tag{A.1.9}$$

An Euler approximation of this integral would then be defined by the recursion

$$X_n = X_{n-1} + \sigma e^{\frac{-r}{2}n\Delta}\sqrt{\Delta}\varepsilon_n, \tag{A.1.10}$$

and if we identify $\sigma_V = \sigma\sqrt{\Delta}$, we notice that this approximation corresponds to Voje's original definition. However, we can write down an exact discretization of the stochastic integral.

$$X_n = X_{n-1} + \int_{(n-1)\Delta}^{n\Delta} e^{-\frac{r}{2}s}\sigma dW(s) \tag{A.1.11}$$

We can again look at the second term on the right and calculate the variance using Ito's isometry

$$Var\left(\int_{(n-1)\Delta}^{n\Delta} e^{-\frac{r}{2}s}\sigma dW(s)\right) = \sigma^2\int_{(n-1)\Delta}^{n\Delta} e^{-rs}ds \tag{A.1.12}$$

So, solving the integral on the right-hand side of the above equation, we find

$$\sigma_\Delta^2(n) = \sigma^2\frac{1}{r}e^{-rn\Delta}\left(e^{r\Delta}-1\right) \tag{A.1.13}$$

Applying a Taylor expansion with Δ small, we see that

$$\sigma_\Delta^2(n) \approx \sigma^2 e^{-rn\Delta}\Delta \tag{A.1.14}$$

and this corresponds to the Euler approximation.

Stasis (ST)

The ST model assumes that each observation is independent and identically distributed, typically with a normal distribution. So, X_n is normal with mean c and variance σ_S^2, which we could write as

$$X_n = c + \sigma_S\varepsilon_n \tag{A.1.15}$$

where the ε_n is a sequence of independent and identically distributed random variables. Effectively, there is no continuous version of this model. One way to think of this model, however, is as a discretely observed Ornstein-Uhlenbeck (OU) process with sufficient spacing between the observations relative to the parameter α. The OU process is stationary, meaning that after a sufficient time, the initial condition is trivially relevant and the relationship between observations at two times depends only on the distance between those observations. Because the OU process is Gaussian, the variance and covariances define the process. The covariance between two observations of a (zero-mean) OU process is given by

$$Cov(X(t),X(s)) = E[X(t)X(s)] = \frac{\sigma^2}{2\alpha}e^{-\alpha|t-s|} \tag{A.1.16}$$

So, if the observations are sufficiently spaced, this covariance is effectively zero (as long as α is not too small). Therefore, the ST model could be viewed as an OU process that is approximately stationary and sufficiently spaced with the relationship between the variances being $\sigma_S^2 = \sigma^2/(2\alpha)$, where σ^2 is the infinitesimal variance of the OU model. If we look at equation (14), with $\alpha\Delta$ being large, then the we see that the OU model is effectively the ST model.

Appendix 2. Simulations and Model Selection

To ensure comparability of results, we re-used code from previous studies to simulate data from the five models, including the paleoTS R package maintained by Hunt and supplementary materials from Voje (2020). We used paleoTS to simulate data for the stasis (ST), random walk (RW), directional trend (DT), and Ornstein-Uhlenbeck (OU) models and code from Voje (2020) to simulate data for the decelerated evolution (DE) model. However, we found that prior simulation studies explored parameter values and data-sampling regimes in a way that varied both observational and biological timescales simultaneously (Hunt 2006, 2008; Voje 2020), making their separate effects on performance difficult to disentangle. We chose simulation parameters to feature substantial levels of model uncertainty in order to illustrate dependence on sampling scales.

For model selection, we calculated model goodness-of-fit using the corrected Akaike information criterion (AICc), which is modified for better performance in small sample sizes (Hunt 2008). In general, the AICc provides an unbiased estimator of a model's expected likelihood for small sample sizes, and picking the model with the best (i.e., lowest) AIC score will asymptotically converge on the true distribution when it is unique to a single model (Burnham and Anderson 2002). We also calculated Akaike weights for each model using the AICc scores (Wagenmakers and Farrell 2004), which approximate the probability that a model is the best out of the candidates considered. For model estimation, we used the linear Kalman filter in the state space models and compared the results of our procedures to those in paleoTS and Voje's code.

Parameter Estimation

We find that estimation accuracy and uncertainty are not uniformly influenced by sampling timescales. We show simulation results for the RW and DT models (Fig. A2.1), OU model (Fig. A2.2), and DE model (Fig. A2.3). We explored three scenarios for modifying sample timescales: first, increasing sampling with constant total duration (shrinking Δ); second, increasing duration and increasing sample size (uniform Δ); and third, increasing duration while holding sample size constant (increasing Δ). The ST model is not shown but is effectively an independent, identically distributed (i.i.d.) process with time-indexed observations, and so estimation of $\widehat{\sigma}^2$ will depend on sample size but not total duration or Δ.

For the RW and DT models, the results in Figure A2.1 show that increasing total time while keeping the sample size fixed does not affect $\widehat{\sigma}^2$ but does improve $\widehat{\mu}$. Note that the sufficient statistic for $\widehat{\mu}$ is the value of the process at the terminal point of the time series, so intermediate values do not matter for estimation, only the endpoint. Increasing duration with constant Δ improves both $\widehat{\sigma}^2$ and $\widehat{mu}$, but for different reasons: $\widehat{\mu}$ is improving because the total time observed is increasing, but $\widehat{\sigma}^2$ is improving because there are more steps observed.

In contrast, the OU model parameters in Figure A2.2 show several different types of response to timescales. Both $\widehat{\alpha}$ and $\widehat{\sigma}^2$ improve for both the scenarios, with increasing sampling with constant duration or increasing duration. As Δ becomes larger in row 3, however, estimation gets worse, because $\exp(-\Delta)$ goes to zero and the process starts to look i.i.d., so that $\widehat{\alpha}$ and $\widehat{\sigma}^2$ are both being fit to a normal distribution with mean θ and variance $\sigma^2/(2\alpha)$. The initial value parameter, $\widehat{Z}$, is unaffected in all three scenarios, because better estimation of the restoring force, α, can only improve precision for the initial condition up to a point. Similar to $\widehat{\mu}$'s behavior in the RW and DT models, the OU equilibrium value parameter, $\widehat{\theta}$, does not converge under the increased sampling intensity. Neighboring observations are positively correlated, so adding more time points within a fixed interval gives diminishing returns for estimating the mean, but if T is increasing, the observations are spaced farther apart and so are more independent.

For the DE model, the variance decay parameter $\widehat{r}$ in Figure A2.3 shows improved precision in the increasing sampling intensity and increasing duration, constant-sampling scenarios but remains biased below the true value for the simulation setups we examined. The DE model shows phenomenologically the same behavior in $\widehat{Z}_0$ and $\widehat{\sigma}^2$ as the OU model but for different reasons. As the step variance of the process decays exponentially to zero with time, observing the process over a longer duration provides progressively less information.

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

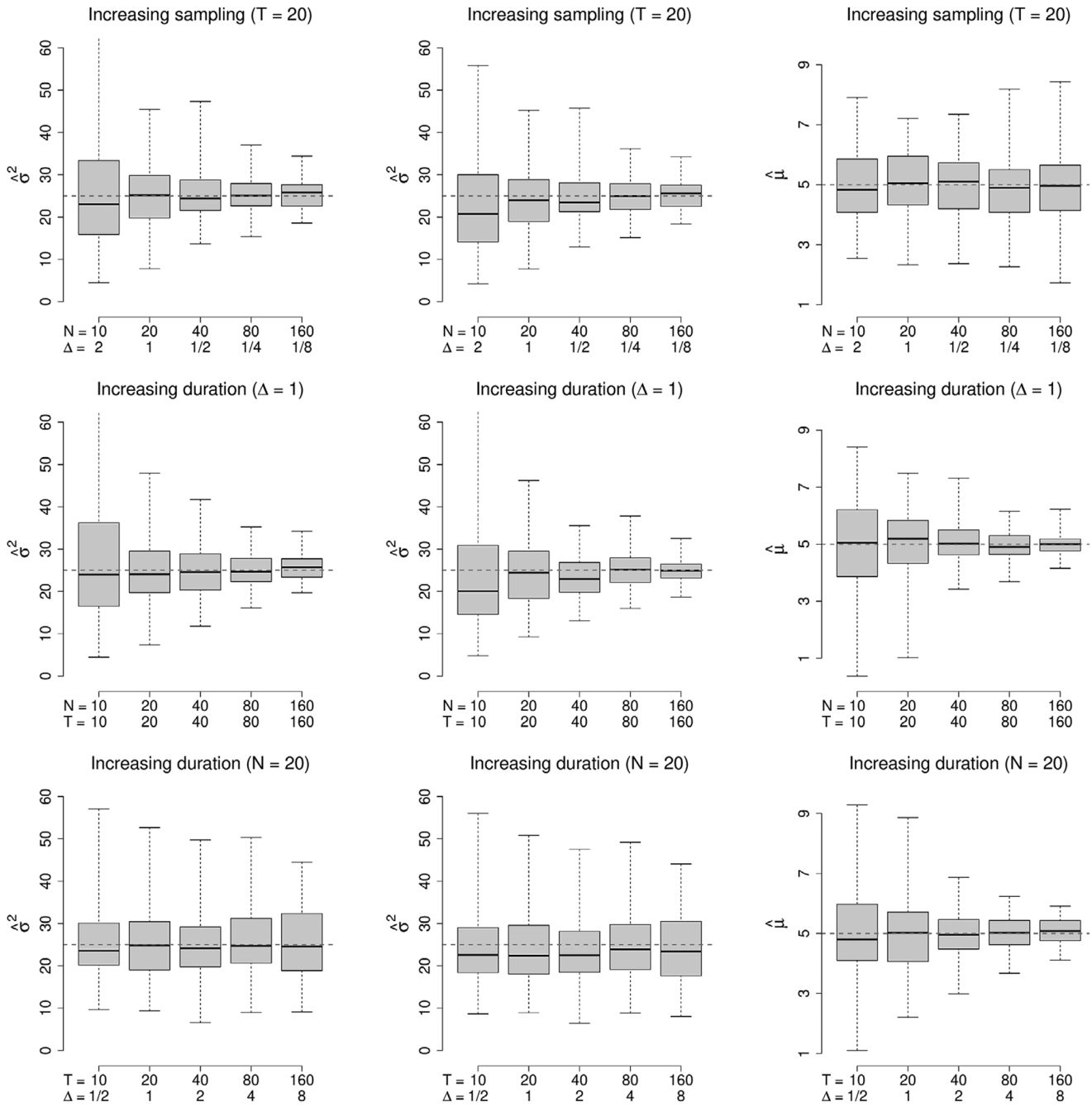


Figure A2.1. Parameter estimation for random walk (RW) and directional trend (DT) models using varying ratios of stepwise and total observational timescales. The left column shows estimation of the diffusion parameter for the RW model. The middle and right columns show estimation of the diffusion and the drift parameters, respectively, for the DT model. The true parameter values are $\sigma^2 = 25$ and $\mu = 5$. We use different values for the sample size $N = \{20,40,80,160\}$, the size of the increments $\Delta = \{1/8, 1/4, 1/2, 1\}$, and the terminal time $T = \{2.5,5,10,20\}$. Box plots show 100 replicates.

Model Selection Performance

Figure A2.4 shows how model selection performance, measured in terms of the average Akaike weight of the true model, varies with sampling. Columns in the figure show increasing total duration. Rows show denser sampling as total duration shrinks. Diagonals from top left to bottom right show increasingly dense sampling within a fixed total duration.

ST is almost exclusively conflated with OU. Increased data appear to slightly worsen false positives for OU when ST is true, likely because the AICc has a bias for nested models toward the model with more parameters. RW is most frequently confused with DT and OU models. The average Akaike weight does not vary significantly with sampling timescales, again likely because of the AICc's bias toward complex models. As expected, the evidence for DT improves significantly as total duration grows. OU shows improvement with greater duration and sampling density. The same is true for DE, which is mainly competitive with ST for small sample sizes.

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

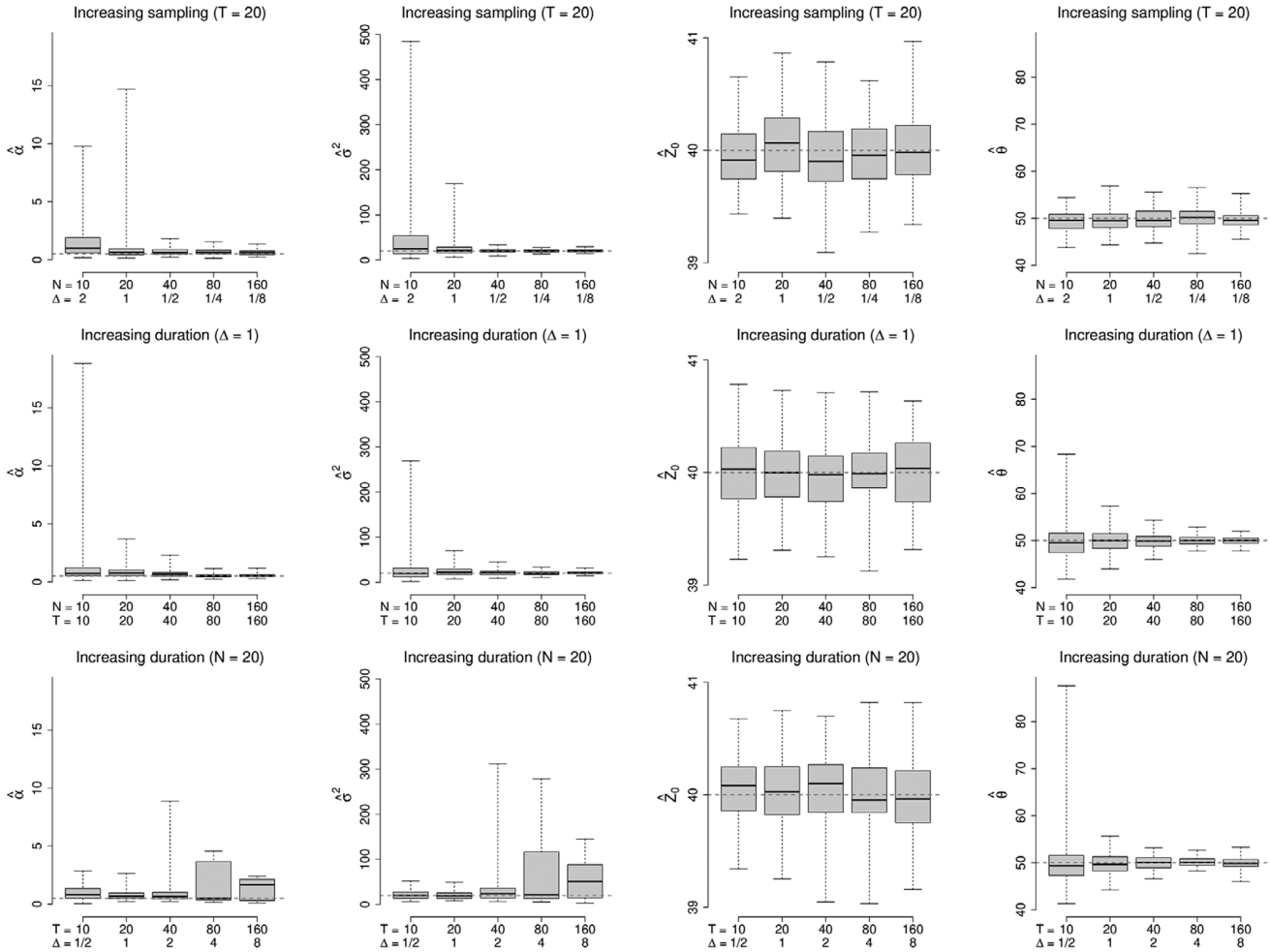


Figure A2.2. Parameter estimation on simulated data from the Ornstein-Uhlenbeck (OU) model for different combinations of sample size, time step, and total duration. Each of the columns shows parameter estimation for α, σ, Z_0, and θ, respectively. True values are $\alpha = 0.50$, $\sigma^2 = 20$, $Z_0 = 40$, and $\theta = 50$, as represented with dashed red lines. We varied the sample size $N = \{20,40,80,160\}$, the size of the increments $\Delta = \{1/8, 1/4, 1/2, 1\}$, and the terminal time $T = \{2.5,5,10,20\}$. Box plots show 100 replicates.

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

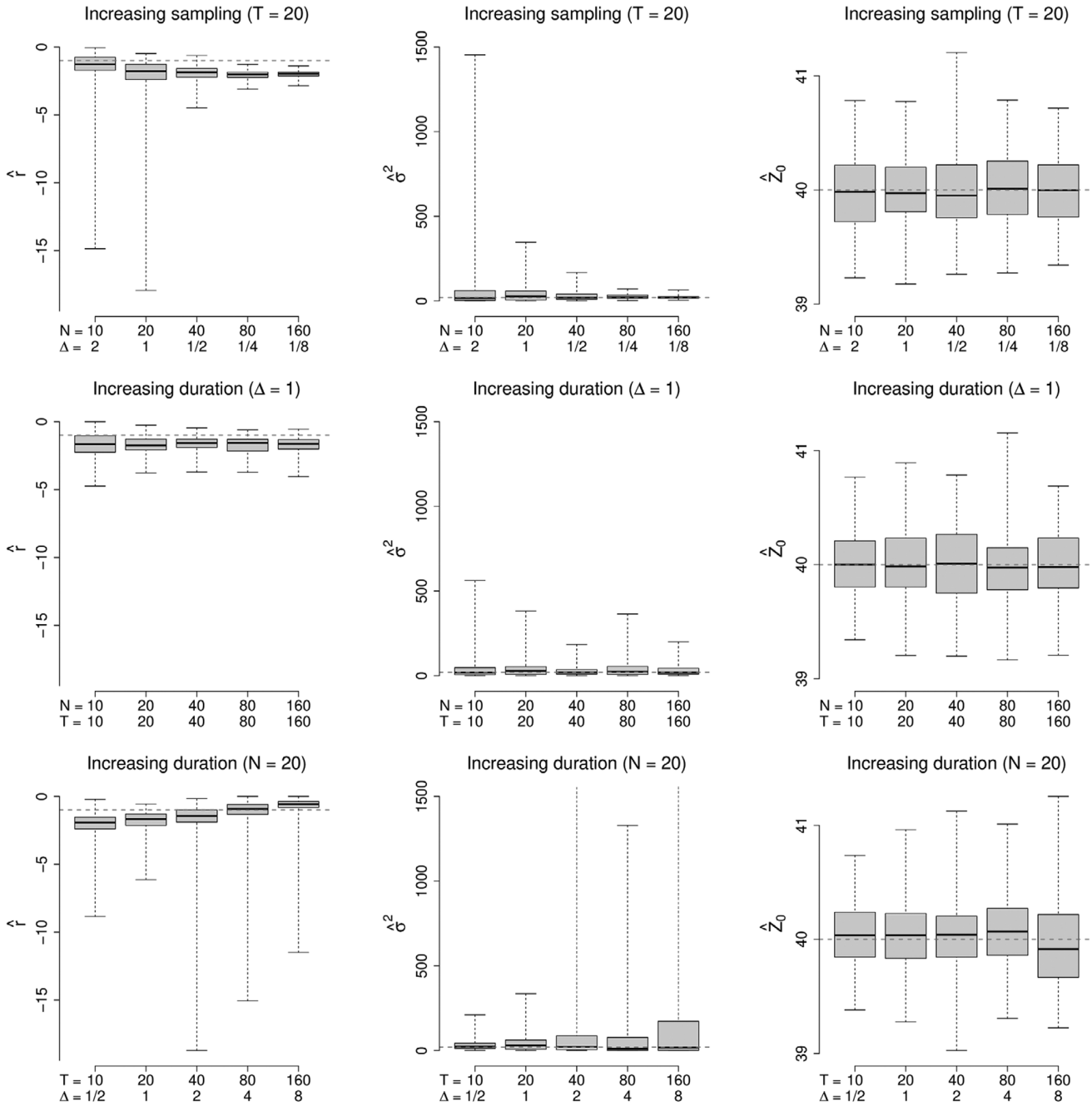


Figure A2.3. Parameter estimation for decelerated evolution (DE) model on simulated data using varying combinations of sample size, time step, and total duration. The columns show parameter estimation of r, σ, and Z_0, respectively. Dashed red lines show the true values of $r = -1$, $\sigma^2 = 20$, and $Z_0 = 40$. We used sample sizes $N = \{20,40,80,160\}$, time increments of $\Delta = \{1/8, 1/4, 1/2, 1\}$, and terminal times $T = \{2.5,5,10,20\}$. Box plots show 100 replicates.

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press

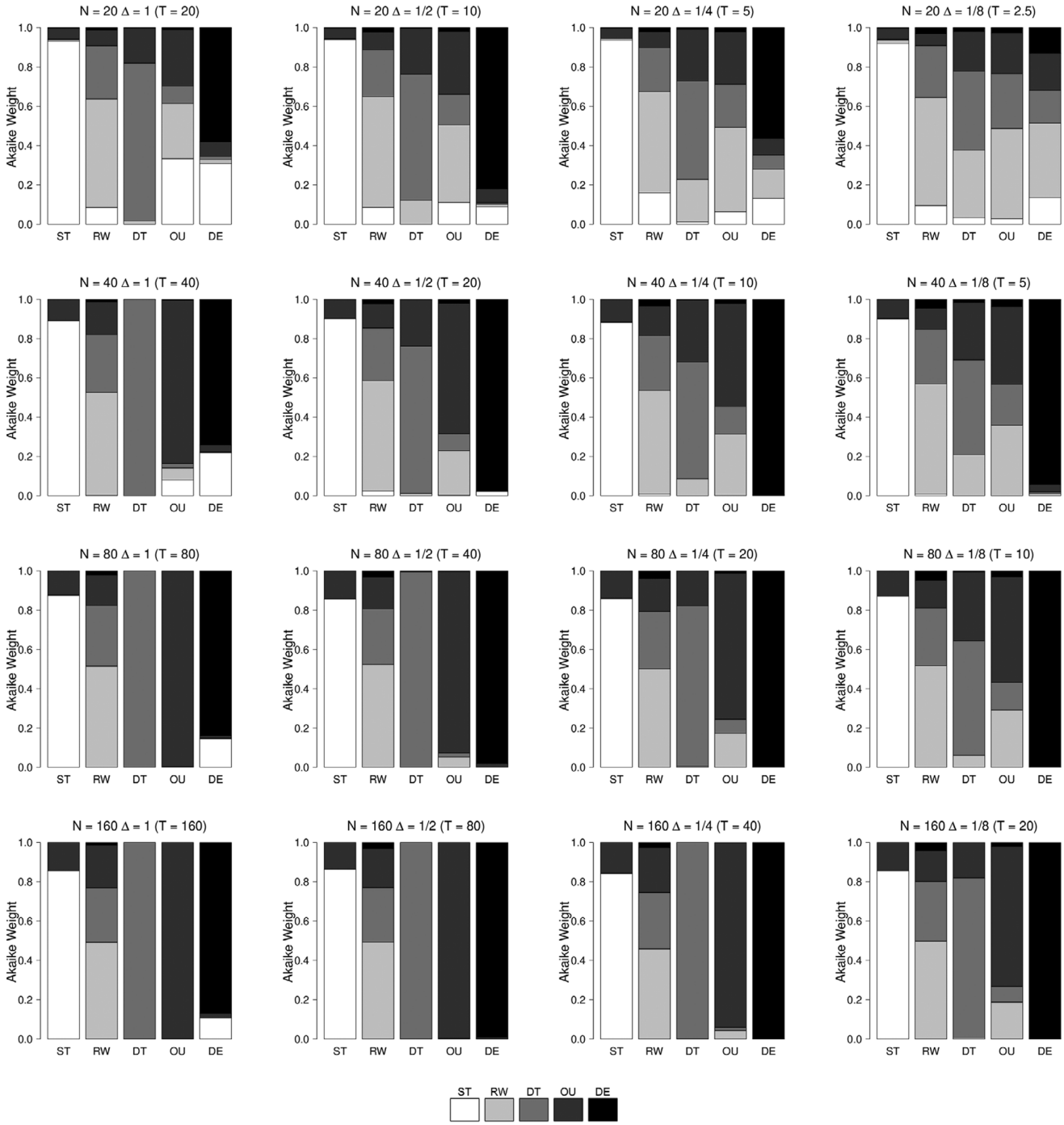


Figure A2.4. Model selection performance using the Akaike information criterion (AIC) on simulated data for varying ratios of stepwise and total observational timescales. In each panel, the true model is labeled on the x-axis, and the stacked histogram shows the average Akaike weight for each model. Perfect model performance would show each bar as completely filled by the corresponding true model's shade on the legend (ST as white, RW as light gray, DT as medium gray, etc.). The true model parameters are ST: $\theta = 50$, $\omega = 20$; RW: $\sigma^2_{step} = 20$; DT: $Drift = 5$, $\sigma^2_{step} = 25$; OU: $\theta = 50$, $\alpha = 20$, $\sigma^2_{step} = 20$; and DE: $r = -1$, $\sigma^2_{step} = 20$. The initial condition for all the models is $Z_0 = 40$, the variance of the evolutionary step is $V_p = 5$, and the vector of population sample size is $m = 50$. We varied sample size $N = \{20,40,80,160\}$, the size of the increments $\Delta = \{1/8, 1/4, 1/2, 1\}$, and the terminal time $T = \{2.5,5,10,20\}$. ST, stasis; RW, random walk; DT, directional trend; OU, Ornstein-Uhlenbeck; DE, decelerated evolution.

https://doi.org/10.1017/pab.2025.10075 Published online by Cambridge University Press