A critical look at directional random walk modeling of sparse fossil data

Rolf Ergon

rolf.ergon@usn.no

University of South-Eastern Norway

November 11, 2025

Abstract

The general random walk model (GRW) of Hunt (2006) is used to infer directional evolution in mean trait values from sparse fossil data. Such evolutions are modeled as the accumulated result of small steps with mean step sizes and step variances. As shown in simulations and real data cases, the mean step sizes are often easy to estimate from data, except for cases where the mean step size is small compared to the step variance. The step variances, on the other hand, can be estimated reasonably well only when the mean trait values have small measurement errors, although also then the step variance estimation may be difficult. For fossil data with realistic measurement errors, the step variances appear to be extremely difficult to find, and they are often found to be negative. They must then be set to zero, such that GRW deteriorates into deterministic walk processes. As a result of poor step variance estimates, the directional evolution may be both under- and overestimated as compared with weighted least squares results, which are the best linear unbiased predictions (BLUP). Here, I study this problem through simulations, as well as in four real data cases. Based on weighted mean square error (WMSE) and Akaike information criterion (AIC) comparisons, my conclusion is that

WLS in cases with large measurement errors is the best method for inference of directional evolution.

Keywords: Akaike information criterion; general random walk model; mean step value; step variance; weighted least squares estimation

1 Introduction

Hunt (2006) developed what he called a general random walk (GRW) model for analyses of fossil time series, later referred to as a model for directional evolution (Hunt, 2012). This model assumes a random walk process where at each timestep an increment of evolutionary change in a mean trait value is drawn at random from a distribution of evolutionary steps, and that this incremental change is normally distributed with mean value μ_{step} and variance σ^2_{step}. For an observed evolutionary mean trait change ΔY over T time steps, the log-likelihood function for a GRW process is given by (Hunt, 2006)

$$\ell\left(\mu_{step}, \sigma^2_{step}\right) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\mathrm{T}\sigma^2_{step} + \frac{V_{p_A}}{n_A} + \frac{V_{p_D}}{n_D}\right) - \frac{\left(\Delta\mathrm{Y} - \mathrm{T}\mu_{step}\right)^2}{2\left(\mathrm{T}\sigma^2_{step} + \frac{V_{p_A}}{n_A} + \frac{V_{p_D}}{n_D}\right)}, \tag{1}$$

where n_A and n_D are the numbers of observed ancestors and descendants, respectively, while V_{p_A} and V_{p_D} are the corresponding population phenotypic variances. With N irregular and sparse samples of mean trait values, multiple ancestor-descendant trait differences over $N-1$ evolutionary transitions may be used jointly to estimate μ_{step} and σ^2_{step} by summing the log-likelihoods according to Eq. (1) over the transitions, i.e., by use of $\ell_{total}\left(\mu_{step}, \sigma^2_{step}\right) = \sum_{i=1}^{N-1} \ell_i\left(\mu_{step}, \sigma^2_{step}\right)$. From this expression, step parameters can be found by means of maximum likelihood estimation, and the estimated step size $\hat{\mu}_{step}$ will then be a measure of directional change over time.

For a single transition it is obvious from Eq. (1) that $\ell\left(\mu_{step}, \sigma^2_{step}\right)$ is maximized when

$\hat{\mu}_{step} = \Delta Y/T$, and as shown in simulations in Section 3, μ_{step} can be well estimated also from data over several transitions. From Eq. (1) it is also obvious that it is difficult to estimate σ^2_{step} in cases where $T\sigma^2_{step}$ is small compared to V_p/n for the samples.

When μ_{step} is fixed to zero we obtain an unbiased random walk (URW) process, and the estimated step variance $\hat{\sigma}^2_{step}$ then becomes a scaled version of the rate of directional change (Hunt, 2012). I will not make use of this property, but it should be clear that the estimation difficulties will not disappear when μ_{step} is set to zero.

Hunt (2006) presented simulation results over 19 transitions with clearly directional evolution, where both μ_{step} and σ^2_{step} were estimated well, except for very small values of σ^2_{step} as compared with μ_{step}. He did, however, use what appears to be unrealistically small values of phenotypic variances, as compared with the real data cases in Section 4. In Hunt (2012), similar small values appear in Figure 2, panel A, illustrating directional evolution. In simulations in Section 3 I essentially repeat parts of Hunt's simulations with what appears to be more realistic phenotypic variances, and in around 40% of the realizations I find that $\hat{\sigma}^2_{step} < 0$, which in practice means that $\sigma^2_{step} = 0$. In such cases the GRW model deteriorates into a deterministic directional walk model.

Eq. (1) is not primarily intended for predictions, but once $\hat{\mu}_{step}$ has been determined, a prediction model $\hat{\bar{y}}_{GRW}(t) = \hat{a}_{GRW} + \hat{b}_{GRW}t$ is found by use of $\hat{b}_{GRW} = c\hat{\mu}_{step}$, and by fitting $\hat{a}$ to the data by a least squares approach. Here, c is a scaling constant that can be chosen freely, and throughout the paper I will set $c = 1$.

In my simulations I also analyze data for directional evolution by means of the weighted least squares (WLS) method, which with weights chosen as the reciprocals of the measurement variances gives the best linear unbiased estimate (BLUE) of the evolutionary slope. I then find that GRW may both overestimate and underestimate prediction slopes $\hat{b}_{WLS}$ by up to 40%. The

simulation results include Akaike information criterion (AIC) and weighted mean square error (WMSE) values.

In Section 4 I analyze data from four real data cases, and I find similar differences between GRW and WLS results as in the simulations. In all these cases estimation with σ_{step}^2 as a free variable resulted in $\hat{\sigma}_{step}^2 < 0$, which means that $\sigma_{step}^2 = 0$ must be used in a realistic model. In three of the cases there exists a dominating evolutionary driver, and a tracking model as discussed in Ergon (2025) is then an attractive alternative. The applicability of such a model will, however, depend on the quality of the evolutionary driver data, and only in one case is the tracking model better than the WLS model.

My results are summarized and discussed in Section 5, and my conclusion is that WLS is an attractive alternative for analyses of directional evolution, and that dependent on the quality of the data, a tracking model may give even better prediction results. Detailed simulation results are given in Appendix A, an illustration of model selection difficulties are given in Appendix B, while Appendix C gives data for the four real data cases.

2 Methods

2.1 GRW predictions

Eq. (1) is not intended for predictions, but once $\hat{\mu}_{step}$ is determined, a prediction model

$$\hat{\bar{y}}_{GRW}(t) = \hat{a}_{GRW} + \hat{b}_{GRW} t \tag{2}$$

is found by use of $\hat{b}_{GRW} = c\hat{\mu}_{step}$, and by fitting a parameter a_{GWR} to the data by a least squares approach. Here, c is a scaling constant that I consistently will set to $c = 1$. Note that AIC_c values for estimates $\hat{\mu}_{step}$ and $\hat{\sigma}_{step}^2$ do not incorporate the estimation of a_{GWR}.

2.2 Methods for simulations

In the simulations in Section 3 I first repeat some of the simulations in Hunt (2006), but with the number of samples reduced from 20 to 10. I generate time series data over 1,000 time steps as $\bar{y}_{true}(t+1) = \bar{y}_{true}(t) + \mu_{step} + \sqrt{\sigma_{step}^2}\varepsilon(t)$, where $\varepsilon(t)$ is a normally distributed random number with variance one, and I then extract ten samples $\bar{y}_{true,0}(50)$, $\bar{y}_{true,0}(150)$, ..., $\bar{y}_{true,0}(950)$. The initial value is chosen as $\bar{y}_{true,0}(1) = 0$. Here I use $\mu_{step} = 0.1$ and $\sigma_{step}^2 = 0.1$, which over 1,000 timesteps results in an approximately linear change of 100 in $\bar{y}_{true,0}(t)$. To each of these samples I add the mean value of $n = 30$ individual samples drawn from a normally distributed population with mean $\mu_p = 0$ and variance $V_p = 1$, and thus form ten samples $\bar{y}(t)$. Each of these samples have a standard error $e(t) = \sqrt{V_p/n} = 0.1826$, i.e., a small error compared to the total change in $\bar{y}_{true}(t)$ and thus in $\bar{y}(t)$.

Based on the generated $\bar{y}(t)$ data with equal V_p/n values for all samples, I find estimates $\hat{\mu}_{step}$ and $\hat{\sigma}_{step}^2$ that maximize the log-likelihood function $\ell_{total} = \sum_{i=1}^{N-1} \ell_i(\mu_{step}, \sigma_{step}^2)$, where ℓ_i is given by Eq. (1). For this purpose, I use the *fmincon* function in MATLAB, i.e., I minimize $-\ell_{total}(\mu_{step}, \sigma_{step}^2)$.

In order to obtain standard errors comparable to real data, as presented in Section 4, I repeated the simulations with $V_p = 400$. In addition, I also let the sampling times be irregular as uniformly distributed random numbers in the ranges $1-100$, $101-200$, ..., $901-1000$, while the numbers of individual samples varied between 3 and 60 with uniform random distributions.

2.3 Weighted least squares estimation

In the simulations as well as in the real data cases in Section 4, predictions $\hat{\bar{y}}_{GRW}(t)$ according to Eq. (2) will be compared with weighted least squares predictions found as $\hat{\bar{y}}_{WLS}(t) = \hat{a}_{WLS} + \hat{b}_{WLS}t$. In the simulations with equal T values as described above, $\hat{a}_{WLS}$ and $\hat{b}_{WLS}$ are found as

$$\begin{bmatrix} \hat{a}_{WLS} \\ \hat{b}_{WLS} \end{bmatrix} = (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{y}, \tag{3}$$

where $\boldsymbol{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 50 & 150 & \cdots & 950 \end{bmatrix}^{\boldsymbol{T}}$ and $\boldsymbol{y} = [\bar{y}(50) \quad \bar{y}(150) \quad \cdots \quad \bar{y}(950)]^{\boldsymbol{T}}$, while $\boldsymbol{W}$ is a diagonal matrix with equal weights $w_i = \left(V_p/n\right)^{-1}$. Generally, the $\boldsymbol{X}$ and $\boldsymbol{y}$ data will of course be different, and w_i may vary from sample to sample with the variance of the observations. Comparisons of responses $\hat{\bar{y}}_{GRW}(t)$ and $\hat{\bar{y}}_{WLS}(t)$ can in any case be done by use of weighted mean square errors

$$WMSE_{WLS} = (\bar{\boldsymbol{y}}_{true}^T - \hat{\bar{\boldsymbol{y}}}_{WLS}^T)\boldsymbol{W}(\bar{\boldsymbol{y}}_{true} - \hat{\bar{\boldsymbol{y}}}_{WLS})/\sum w_i, \tag{4}$$

and

$$WMSE_{GRW} = (\bar{\boldsymbol{y}}_{true}^T - \hat{\bar{\boldsymbol{y}}}_{GRW}^T)\boldsymbol{W}(\bar{\boldsymbol{y}}_{true} - \hat{\bar{\boldsymbol{y}}}_{GRW})/\sum w_i, \tag{5}$$

where $\bar{\boldsymbol{y}}_{true}$, $\hat{\bar{\boldsymbol{y}}}_{WLS}$ and $\hat{\bar{\boldsymbol{y}}}_{GRW}$ are $N \times 1$ data vectors, and where $\boldsymbol{W}$ is a diagonal matrix with the weights w_i. I thus assume independent and normally distributed random variables $\bar{y}(t)$. In real data cases, $\bar{\boldsymbol{y}}_{true}$ must be replaced by the observation vector $\bar{\boldsymbol{y}}$. Note that when the weights are the reciprocals of the variances of the measurements, as done here, the prediction slope parameter $\hat{b}_{WLS}$ is the best linear unbiased estimate (BLUE).

2.4 Theory for the Akaike Information Criterion

For model comparisons of random walk and other models of fossil data, it is common to compute the Akaike Information Criterion with compensation for short data (Hunt, 2006),

$$AIC_{c,GRW} = 2k - 2\ell_{total}\left(\hat{\mu}_{step}, \hat{\sigma}_{step}^2\right) + 2\mathrm{k}(\mathrm{k}+1)/(\mathrm{N}-\mathrm{k}-2)\,, \tag{6}$$

where $\mathrm{k} = 2$ is the number of estimated parameters, while N is the number of samples (the number of transitions is $\mathrm{N} - 1$). In all the real data examples in Section 4, the estimated variance $\hat{\sigma}_{step}^2$ is negative, and the realistic value of $AIC_{c,GRW}$ is then found by setting $\sigma_{step}^2 = 0$. However, the better results obtained with $\hat{\sigma}_{step}^2 < 0$ are also recorded.

For comparisons with WLS models, Eq. (6) must be replaced by (Banks and Joyner, 2017)

$$AIC_{c,WLS} = 2k - 2\ln\left(\mathcal{L}\left(\widehat{\Theta}_{MLE}|\overline{\boldsymbol{y}}\right)\right) + 2\mathrm{k}(\mathrm{k}+1)/(\mathrm{N}-\mathrm{k}-1), \tag{7}$$

where $\mathcal{L}(\widehat{\Theta}_{MLE}|\overline{\boldsymbol{y}})$ is the likelihood of the estimated parameters using maximum likelihood estimation, given the vector $\overline{\boldsymbol{y}}$ of observed data. Banks and Joyner (2017) further assume a statistical model $Y_i = f(t_i, \boldsymbol{q}_0) + \omega_i \mathcal{E}_i$, where $\boldsymbol{q}_0$ is a parameter vector, and where ω_i are known weights and ε_i, $i = 1, 2, \ldots, N$, are i.i.d. $\mathcal{N}(0, \sigma^2)$, such that σ^2 must be estimated together with $\boldsymbol{q}_0$. In the present setting the weights are $\omega_i = \sqrt{1/w_i}$, and the final AIC_{WLS} expression in Banks and Joyner (2017) thus becomes

$$AIC_{WLS} = 2k + Nln(2\pi) + N + 2\sum_{i=1}^{N} \ln\left(\sqrt{1/w_i}\right) + Nln\left(\frac{\sum_{i=1}^{N} w_i(\overline{y}_i - \hat{\overline{y}}_i)^2}{N}\right), \tag{8}$$

from which $AIC_{c,WLS}$ follows by adding $2\mathrm{k}(\mathrm{k}+1)/(\mathrm{N}-\mathrm{k}-1)$. Here, $k = 3$, since also the underlying parameter σ^2 must be estimated in addition to a_{WLS} and b_{WLS}.

3. Simulation results

As in Hunt (2006), GRW parameter estimation results with $V_p = 1$ were found from 1,000 repeated simulations, with typical responses from a system with $\mu_{step} = 0.1$ and $\sigma^2_{step} = 0.1$ as shown in Fig. 1, left panel. In the same simulations, GRW and WLS prediction parameters $\hat{a}_{GRW}$ and $\hat{b}_{GRW} = \hat{\mu}_{step}$, as well as $\hat{a}_{WLS}$ and $\hat{b}_{WLS}$ were found, which gave typical responses as also shown in Fig. 1. The results are summarized in Table 1, including AIC_c and *WMSE* values. Note that $WMSE_{WLS}$ is slightly better than $WMSE_{GRW}$ (because errors in $\hat{a}_{GRW}$ are included), but that $AIC_{c,GRW}$ (where errors in $\hat{a}_{GRW}$ are not included) is lower (better) than $AIC_{c,WLS}$. The $\hat{\mu}_{step}$ and $\hat{\sigma}^2_{step}$ results are very much the same as found by Hunt (2006), and also here $\hat{\sigma}^2_{step}$ consistently approaches $0.1(N-1)/N = 0.09$. Note that the standard errors as shown by error bars are quite small. Also note, however, that the SE value for $\hat{\sigma}^2_{step}$ is 50% of the mean value, indicating the estimation difficulty also in the simple case with $V_p = 1$.

In order to obtain standard errors similar to the once in the real data examples in Section 4, we must increase V_p considerably. Simulation results with $V_p = 400$, regular sampling, and equal numbers of individuals are also shown in Table 1. Typical responses are shown in Fig. 1, right panel. Note that the SE value for $\hat{\sigma}^2_{step}$ now is much larger than the mean value, and that 376 of the 1,000 simulations resulted in $\hat{\sigma}^2_{step} < 0$. In such cases, the $AIC_{c,GWR}$ and $WMSE_{GWR}$ values were computed by use of $\sigma^2_{step} = 0$. Results with irregular sampling and variable numbers of individual samples, as described in Section 2, are included in Table 1.

Note that the estimates of μ_{step} are very much the same when V_p is increased from 1 to 400, but that $\hat{\sigma}^2_{step}$ is very much affected. Also note that $WMSE_{WLS}$ is consistently lower (better) that $WMSE_{GRW}$, and that that $AIC_{c,GRW}$ is lower (better) than $AIC_{c,WLS}$ only in the unrealistic case with $V_p = 1$. A closer look reveals that $AIC_{c,WLS} < AIC_{c,GRW}$ for $V_p > 150$, and at that limit around 18% of the realizations result in $\hat{\sigma}^2_{step} < 0$.

The $WMSE$ and AIC_c values are as seen in Table 1 increased when V_p is increased from 1 to 400, but the essential differences from the simulations in Hunt (2006) are still the poor estimation results for the step variances. As a result, the slope of directional evolution as found from $\hat{b}_{GRW}$ can be both under- and overestimated, as compared with $\hat{b}_{WLS}$, see three typical realizations in Fig. 2. This is consistent with the results that $AIC_{c,WLS} < AIC_{c,GRW}$. More detailed parameter estimation results are given in Appendix A.

As illustrations of the difficulties involved, Appendix B shows two quite different evolutionary responses for a system with $\mu_{step} = 0.1$ and $\sigma^2_{step} = 10$, both apparently directional, but with different signs. Note, however, that the deviations from the straight prediction line are much larger than the measurement standard errors, which indicates that the model is not trustworthy.

Table 1. Simulation results given as $mean \pm SE$ from *1,000* realizations.

	$V_p = 1$ regular sampling	$V_p = 400$ regular sampling	$V_p = 400$ irregular sampling
$\hat{\mu}_{step} = \hat{b}_{GRW}$	1.0004 ± 0.0106	0.0997 ± 0.0122	0.0996 ± 0.0167
$\hat{\sigma}^2_{step}$	0.0917 ± 0.0469	0.0782 ± 0.1757*	0.0732 ± 0.2412 **
$\hat{b}_{WLS}$	1.1003 ± 0.0112	0.0996 ± 0.0119	0.0997 ± 0.0117
$WMSE_{GWR}$	7.50 ± 5.25	11.36 ± 7.26	22.05 ± 26.42
$WMSE_{WLS}$	6.56 ± 4.60	9.01 ± 5.10	8.34 ± 5.03
$AIC_{c,GWR}$	50 ± 5	63 ± 4	65 ± 5
$AIC_{c,WLS}$	55 ± 6	59 ± 6	60 ± 6

* 376 of the 1,000 simulations resulted in $\hat{\sigma}^2_{step} < 0$

$**$ 465 of the 1,000 simulations resulted in $\hat{\sigma}^2_{step} < 0$

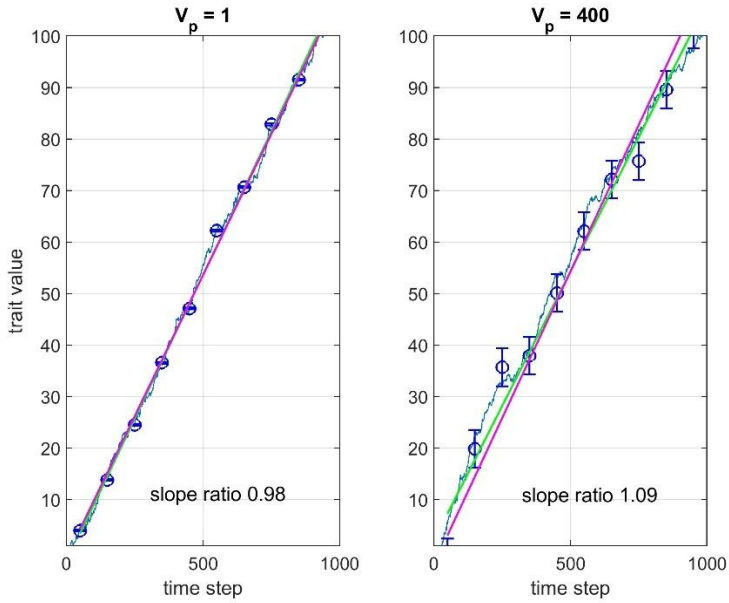

Figure 1. Typical results for simulations with $\mu_{step} = 0.1$ and $\sigma^2_{step} = 0.1$ when $V_p = 1$ (left panel) and $V_p = 400$ (right panel). Values of $\bar{y}_{true}(t)$ are shown by solid blue line, while $\bar{y}(t)$ are shown by circles with error bars. WLS predictions are shown by green

line, and GRW predictions are shown by red line. Slope ratios $\hat{b}_{GRW}/\hat{b}_{WLS}$ are shown in the plots.

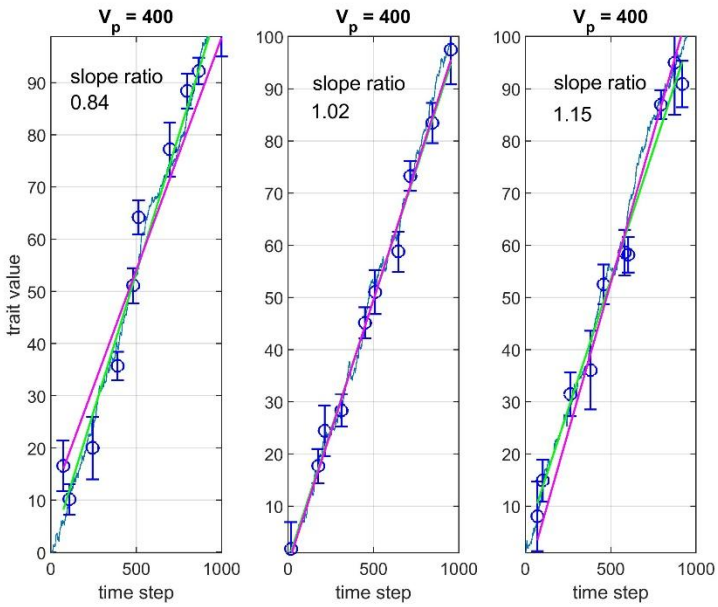

Figure 2. Typical results for simulations with $\mu_{step} = 0.1$, $\sigma^2_{step} = 0.1$, $V_p = 400$, irregular sampling and variable numbers of samples as described in text. Values of $\bar{y}_{true}(t)$ are shown by solid blue line, while $\bar{y}(t)$ are shown by circles with error bars. Typical realizations of WLS and GRW predictions are shown by green and red lines, respectively. Slope ratios $\hat{b}_{GRW}/\hat{b}_{WLS}$ are shown in the plots.

4 Real data cases

4.1 Bryozoan case

Liow et al. (2024) presented results from a field study of the bryozoan species *Microporella agonistes*, based on a fossil record over 2.3 million years. Here, I will use the time series for *log AZ area mean* as shown in Fig. S2, panel A, in Liow et al. (2024), with the two last data points omitted (because they do not fit into a linear model, see Ergon (2025) for details). In this case

there are no individual data points, but instead colony means, and the mean trait values are thus means of colony mean values.

As shown in Ergon (2025), good predictions can be found by use of a tracking model based on a centered moving average filtered version of the $\partial^{18}O(t)$ environmental driver. The $\partial^{18}O(t)$ data are found in Lisiecki and Raymo (2005), while other data are collected in Table C1 in Appendix C. Fig. 3 shows predictions by means of the tracking model, as well as WLS and GRW predictions. Here, the prediction slope $\hat{b}_{GRW} = \hat{\mu}_{step}$, i.e., I use $\hat{b}_{GRW} = c\hat{\mu}_{step}$ with $c = 1$. Ordinary least squares (OLS) results are included. Prediction slope, WMSE and AIC results for the three methods are given in Table 2, with use of the lower limit $\sigma^2_{step} = 0$ in the search for the optimal GRW model. Here, the $AIC_{c,GRW}$ value is computed for the slope value $\hat{\mu}_{step}$, not for $\bar{y}_{GRW}(t) = \hat{a}_{GRW} + \hat{\mu}_{step}t$. Note that the WLS results are better than the GRW results, but that the tracking results are even better. When σ^2_{step} was a free variable, $\hat{\sigma}^2_{step}$ became negative, and the prediction slope for the GRW model then increased from $b_{GWR} = 0.1441$ to $b_{GRW} = 0.1617$. In this unrealistic case $AIC_{c,GRW} < AIC_{WLS}$, but still $AIC_{tracking} < AIC_{c,GRW}$.

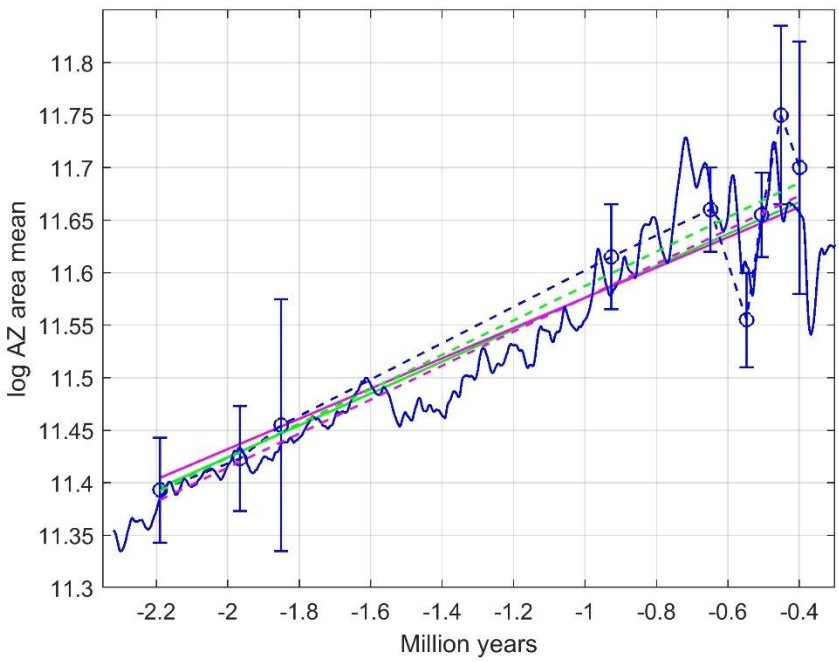

Figure 3. WLS and GRW results for Bryozoan case, with mean trait observations (dashed blue line with circles and error bars), predictions based on a centered moving average filtered version (with window size 100) of the $\partial^{18}O(t)$ environment (solid blue line), WLS predictions (green line), OLS predictions (dashed green line), GRW predictions with $\sigma_{step}^2 \geq 0$ in the search for the optimum (red line), and GRW predictions with $\hat{\sigma}_{step}^2 = -0.0147$ (dashed red line).

Table 2. Prediction slope, WMSE and AIC results for predictions in Fig. 3.

	Tracking model	WLS model	GRW model
Prediction slope with $\sigma_{step}^2 \geq 0$	$-$	0.1519	0.1441
Prediction slope with $\hat{\sigma}_{step}^2 = -0.0147$		0.1519	0.1617
WMSE	0.0010	0.00190	0.00193
WMSE with $\hat{\sigma}_{step}^2 = -0.0147$	$-$	0.00190	0.00195
AIC_c	-23	-18	-10
AIC_c with $\hat{\sigma}_{step}^2 = -0.0147$	-23	-18	-21

2 Ostracode case 1

Hunt and Roy (2006) presented time series data for body size of the ostracode genus *Poseidonamicus*, and in this case I use data for *P. major*, extracted as shown in Table C2 in Appendix C. The data include long-term smoothed magnesium–calcium temperature data and the number of samples in each population. Mean valve length is measured in μm, with measurement variance estimated as $467\ (\mu m)^2$ (Hunt et al., 2015). In the data analysis the trait values are log-transformed, with the measurement errors transformed accordingly, i.e., $e_{i,log} = e_i \cdot \bar{y}_{i,log}/\bar{y}_i$. The original and log-transformed samples thus have equal coefficients of variation.

Fig. 4 shows predictions by means of a Mg/Ca tracking model, as well as of WLS and GRW models. Here, the prediction slope $\hat{b}_{GRW} = \hat{\mu}_{step}$, i.e., I use $\hat{b}_{GRW} = c\hat{\mu}_{step}$ with $c = 1$. Ordinary least squares (OLS) results are included. Prediction slope, WMSE and AIC results with $\sigma^2_{step} \geq 0$ in the search for the optimum are given in Table 3. Here, the AIC_{GRW} value is computed for the slope value $\hat{\mu}_{step}$, not for $\bar{y}_{GRW}(t) = \hat{a}_{GRW} + \hat{\mu}_{step}t$. Note that the WLS model is better than the tracking model, presumably owing to age uncertainty, and that the GRW model underestimates the prediction slope by 21%. When σ^2_{step} was a free variable, $\hat{\sigma}^2_{step}$ was negative, and the prediction slope for the GRW model then decreased from $b_{GRW} = 0.0316$ to $b_{GRW} = 0.0288$. Verification of the search results are shown in the grid contour plot in Fig. 5, where it should be noted that $\hat{\sigma}^2_{step}$ is allowed to be negative.

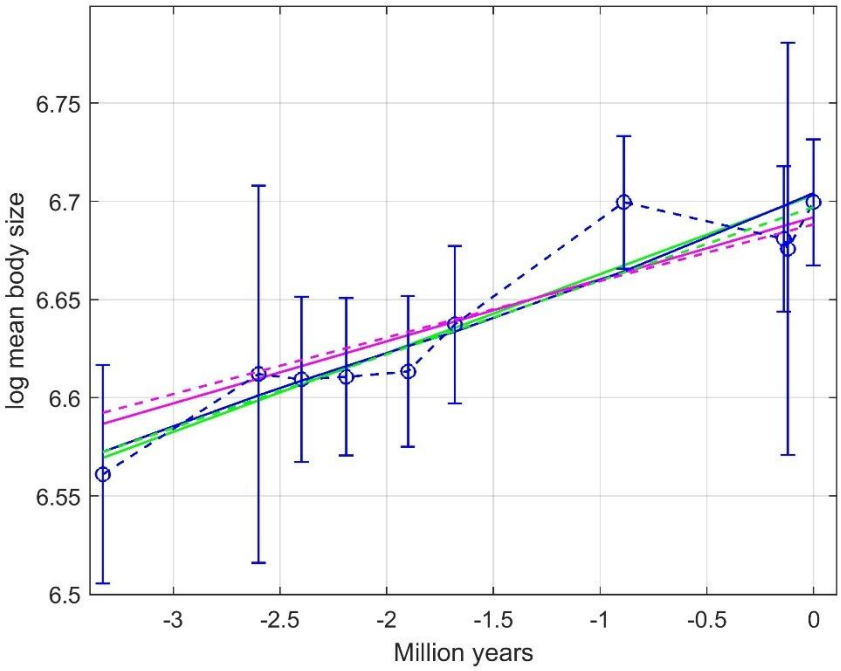

Figure 4. WLS and GRW results for Ostracode case 1, with mean trait observations (dashed blue line with circles and error bars), predictions based on the Mg/Ca temperature environment (solid blue line), WLS predictions (green line, almost identical to the solid blue line), OLS predictions (dashed green line, almost identical to the solid

blue line), GRW predictions with $\sigma_{step}^2 \geq 0$ in the search for the optimum (red line), and GRW predictions with $\hat{\sigma}_{step}^2 = -0.0018$ (dashed red line).

Table 3. Prediction slope, WMSE and AIC results for predictions in Fig. 4.

	Tracking model	WLS model	GRW model
Prediction slope with $\sigma_{step}^2 \geq 0$	$-$	0.0402	0.0316
Prediction slope with $\hat{\sigma}_{step}^2 = -0.0018$	$-$	0.0402	0.0288
WMSE	0.000279	0.000241	0.000319
WMSE with $\hat{\sigma}_{step}^2 = -0.0018$	$-$	0.000241	0.000378
AIC_c	-41	-43	-23
AIC_c with $\hat{\sigma}_{step}^2 = -0.0018$	-41	-43	-24

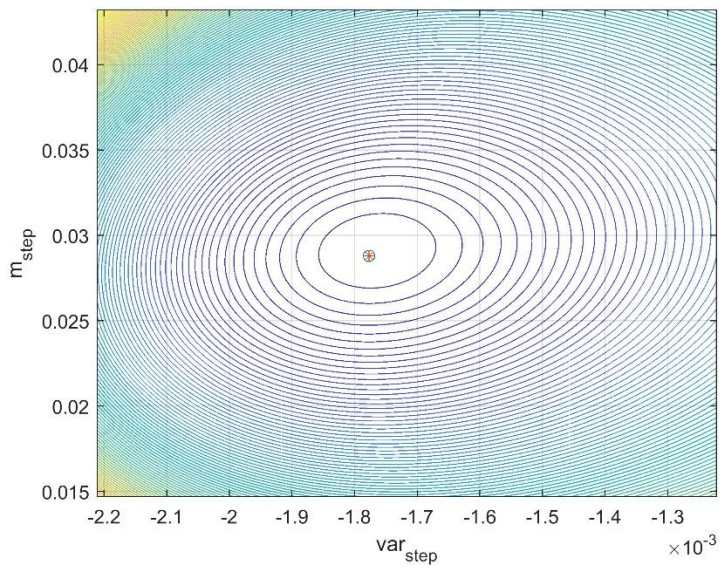

Figure 5. Grid contour results for Ostracode case 1, with the point $[\hat{\sigma}_{step}^2, \hat{\mu}_{step}]$ added as circle with star. Note that $\hat{\sigma}_{step}^2 < 0$ is accepted.

4.3 Ostracode case 2

In this case I use data for *P. pintoi* in Hunt and Roy (2006), as shown in Table C3 in Appendix C. Fig. 6 shows predictions by means of a Mg/Ca tracking model, as well as of the WLS and GRW models. Here, the prediction slope $\hat{b}_{GRW} = \hat{\mu}_{step}$, i.e., I use $\hat{b}_{GRW} = c\hat{\mu}_{step}$ with $c = 1$. Ordinary least squares (OLS) results are included. Prediction slope and WMSE results with $\sigma^2_{step} \geq 0$ in the search for the optimum are given in Table 4. Here, the AIC_{GRW} value is computed for the slope value $\hat{\mu}_{step}$, not for $\bar{y}_{GRW}(t) = \hat{a}_{GRW} + \hat{\mu}_{step}t$. Note that also here the WLS model is better than the tracking model, presumably owing to age uncertainty, and that the GRW model underestimates the prediction slope by more than 40%. When σ^2_{step} was a free variable, $\hat{\sigma}^2_{step}$ was negative, and the prediction slope for the GRW model then decreased from $b_{GRW} = 0.0272$ to $b_{GRW} = 0.0255$. As a verification of the search results, Fig. 7 shows the log likelihood as function of μ_{step} with $\hat{\sigma}^2_{step} = -0.0164$ (upper panel) and $\sigma^2_{step} = 0$ (lower panel). The maximum log likelihood value is thus reduced when the search is limited to positive values of σ^2_{step}, i.e., when σ^2_{step} is non-optimal.

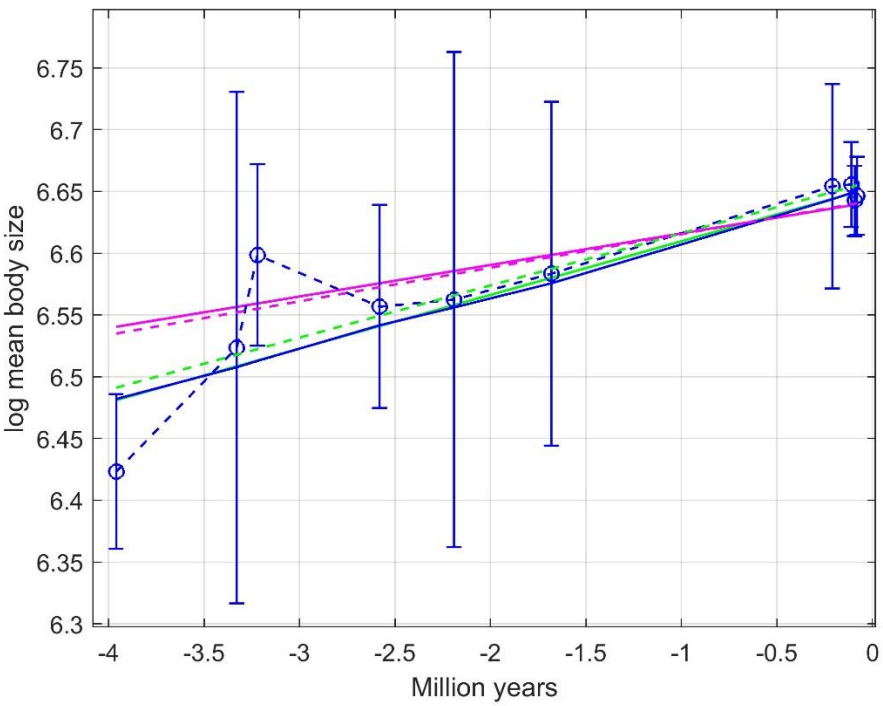

Figure 6. WLS and GRW results for Ostracode case 2, with mean trait observations (dashed blue line with error bars), predictions based on the Mg/Ca temperature environment (solid blue line), WLS predictions (green line, almost identical to the solid blue line), OLS predictions (dashed green line), GRW predictions with $\sigma_{step}^2 \geq 0$ in the search for the optimum (red line), and GRW predictions with $\hat{\sigma}_{step}^2 = -0.0164$ (dashed red line).

Table 4. Prediction slope, WMSE and AIC results for predictions in Fig. 6.

	Tracking model	WLS model	GRW model
Prediction slope with $\sigma_{step}^2 \geq 0$	$-$	0.0434	0.0255
Prediction slope with $\hat{\sigma}_{step}^2 = -0.0164$	$-$	0.0434	0.0272
WMSE	0.000613	0.000606	0.0011
WMSE with $\hat{\sigma}_{step}^2 = -0.0164$	0.000613	0.000606	0.0010
AIC_c	-28	-28	-14
AIC_c with $\hat{\sigma}_{step}^2 = -0.0164$	-28	-28	-17

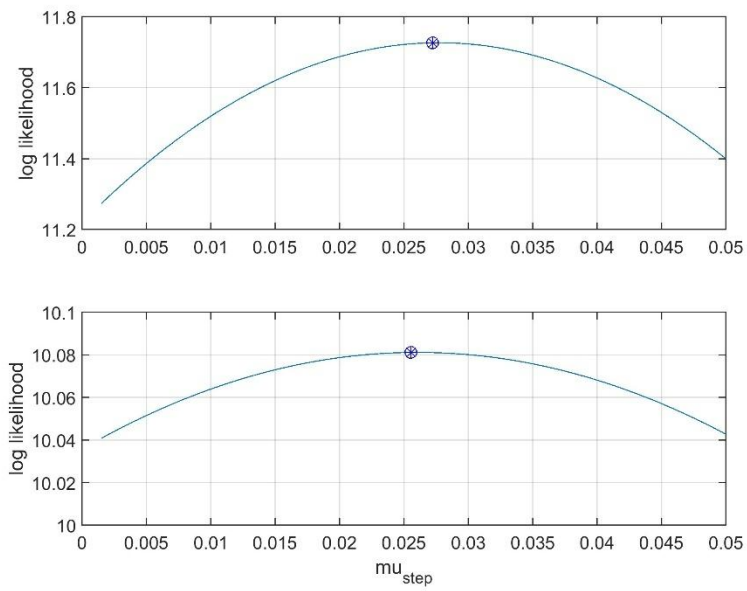

Figure 7. The log likelihood as function of μ_{step} with $\hat{\sigma}^2_{step} = -0.0164$ (upper panel) and $\sigma^2_{step} = 0$ (lower panel), with maxima marked by circles with stars.

4.4 Stickleback fish case

In this final case I use data that show how the dorsal fin ray number evolve in the stickleback fish species *Gasterosteus doryssus* (Bell et al., 1985), see Table C4 in Appendix C for data. In the data analysis the mean trait values are log-transformed, with the measurement errors transformed accordingly, i.e., $e_{i,log} = e_i \cdot \frac{\bar{y}_{i,log}}{\bar{y}_i}$, where e_i is found from the given standard deviations SD as $e_i = \sqrt{SD_i^2/n_i}$. The original and log-transformed samples thus have equal coefficients of variation.

With σ^2_{step} as a free variable, the term $\mathrm{T}\sigma^2_{step} + \frac{V_{p_A}}{n_A} + \frac{V_{p_D}}{n_D}$ in Eq. (1) for one or several of the transitions becomes negative in the course of the parameter search, such that the log-likelihood becomes a complex number. This is avoided with the lower limit $\sigma^2_{step} > -0.091$, but not for $\sigma^2_{step} > -0.092$. See results in Fig. 8 and Table 5. Note that $\sigma^2_{step} = -0.091$ is much more negative than in the other cases.

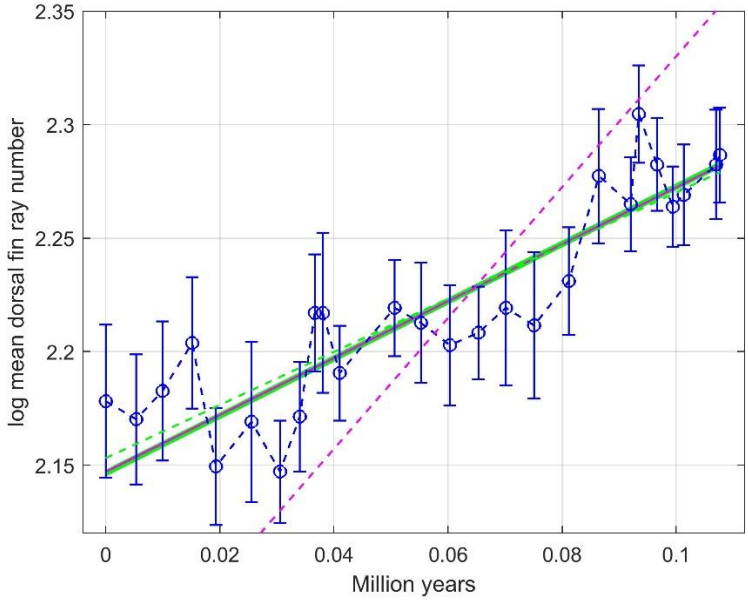

Figure 8. WLS and GRW results for Stickleback fish case, with mean trait observations (dashed blue line with error bars), WLS predictions (green line), OLS predictions (dashed green line), GRW predictions with $\sigma_{step}^2 \geq 0$ in the search for the optimum (red line, very much on top of the green line), and GRW predictions with $\sigma_{step}^2 = -0.091$ (dashed red line).

Table 5. Prediction slope, WMSE and AIC results for predictions in Fig. 8.

	WLS model	GRW model
Prediction slope with $\sigma_{step}^2 = 0$	1.2605	1.2531
Prediction slope with $\sigma_{step}^2 = -0.091$	0.0434	2.88
WMSE	0.00040126	0.00040132
WMSE with $\sigma_{step}^2 = -0.091$	0.00040126	0.0033
AIC_c	-121	-104
AIC_c with $\sigma_{step}^2 = -0.091$	-121	-111

5 Summary and discussion

The main aim of this paper is to show that the general random walk (GRW) model of Hunt (2006) with realistic measurement errors may be both under- and overestimating the degree of directional evolution, and that weighted least squares (WLS) prediction may be a better alternative. A second aim is to show that realistically large sampling errors quite often result in a negative estimated step variance, such that the step variance must be set to zero, with the consequence that the GRW model becomes a deterministic directional walk model.

In his simulations, Hunt found that very small step variances relative to the mean step value could result in negative step variance estimates, but he then used what appears to be an unrealistic low value of the population variance. With more realistic population variances I found negative step variances in up to 40% of the realizations. See Table 1 and Figs. 1 and 2 for simulation results. In the special case with step variance much larger than the step mean value, it is also difficult to estimate the step mean size, and it can in fact get the wrong sign (Appendix B).

In addition to simulations, different models were compared in four real data cases, and in all cases the estimated step variances were negative if they were allowed to be so. In three cases the step mean value estimates changed little when the step variances were set to zero (Figures 3, 4 and 6). In the Stickleback fish case with σ^2_{step} as a free variable, the total log-likelihood becomes a complex number in the course of the parameter search, but that was avoided by searching for $\sigma^2_{step} < -0.091$ (which is much more negative than $\hat{\sigma}^2_{step}$ for the other cases). In this case the difference in step mean values for $\sigma^2_{step} = 0$ and $\sigma^2_{step} = -0.091$ is large. In the Bryozoan and Stickleback fish cases I found that the GRW and WLS predictions were quite similar for $\sigma^2_{step} = 0$ (Figures 3 and 8, and Tables 2 and 5), while in the Ostracode case 2 the step mean value was underestimated by more than 40%, as compared to WLS estimates (Fig. 6 and Table 4).

The GRW model is not primarily intended for predictions, but once the mean step value is estimated, a prediction model can be found by a least squares approach. This makes it possible to compare GRW and WLS results by use of the weighted mean square error (*WMSE*), and for all four of the real data cases with $\sigma_{step}^2 = 0$, I found that $WMSE_{WLS} < WMSE_{GRW}$, although the difference is small in the Bryozoan case, and very small in the Stickleback fish case. This is not surprising, of course, since with zero step variance the GRW model becomes a deterministic walk model, and there is no reason to believe that it then should perform better than a WLS model, which after all gives the best linear unbiased estimate (BLUE) of the evolutionary slope. Ordinary least squares (OLS) models were also tested, and the differences between WLS and OLS results are in all four real data cases quite small (Figs. 3, 4, 6 and 8).

For three of the real data cases there exists a dominating and well known evolutionary driver, and use of an adaptive peak tracking model, as discussed in Ergon (2025), gave better prediction results than GRW. In the two Ostracode cases, however, the WLS predictions were even better, which may possibly be explained by errors in the sample age data. Note, however, that the WLS results do not depend on a known evolutionary driver, there may be many cases of directional evolution driven by a moving adaptive peak, where the possibly several environmental drivers are not well enough known for use in a tracking model. An example of this is the Stickleback fish case.

Models were also compared by use of the Akaike Information Criterion (AIC), and the results confirmed the WMSE results. It may be interesting to note that AIC in all cases is better (more negative) for the estimated negative step variance, than for the step variance zero.

For the realistic simulations with $V_p = 400$ (Table 1), as well as for the four real data cases with $\sigma_{step}^2 = 0$ (Tables 2, 3, 4 and 5), the results are that $AIC_{WLS} < AIC_{c,GRW}$. The main conclusion is thus that WLS is the most reliable method for comparison of directional evolution with other evolutionary modes. This is so also because $\hat{\sigma}_{step}^2$ may be so negative that the numerical search for parameters may fail, as seen in the Stickleback fish case. A WLS model

can, however, be trusted only if sample deviations from a straight prediction line are of the same order as the standard errors. This is the case in all the real data cases, but not in the simulated case in Appendix B. Large and systematic deviations from the WLS prediction line may, however, be caused by a nonlinear model, as discussed in Ergon (2025). What all this might mean for model categorization of fossil time series, as done in, e.g., Hunt et al. (2015) and Voje et al. (2018), is a question beyond the scope of this paper. It is also beyond the scope of the paper to explore non-linear WLS, where with a parameter vector $\boldsymbol{q}$ the optimal estimates $\hat{\bar{y}}_i$ for $i = 1, 2 \ldots, N$ are replaced by $f(t_i, \hat{\boldsymbol{q}})$.

Conflicts of Interest

The first author of Liow et al. (2024), which is reference, is my daughter-in-law. Otherwise, there are no conflicts of interest.

References

Banks, H.T., and Joyner, M.L. (2017). AIC under the framework of least squares estimation. Applied Mathematics Letters. http://dx.doi.org/10.1016/j.aml.2017.05.005

Bell, M. A., Baumgartner, J. V., & Olson, E. C. (1985). Patterns of temporal change in single morphological characters of a Miocene

Ergon, R. (2025). Adaptive Peak Tracking as Explanation of Sparse Fossil Data Across Fluctuating Ancient Environments. Ecology and Evolution. First published: 03 July 2025 https://doi.org/10.1002/ece3.71705

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

Hunt, G. and Roy, K. (2006). Climate change, body size evolution, and Cope's Rule in

deep-sea ostracodes. PNAS, vol. 103/5: 1347–1352. www.pnas.org/cgi/doi/10.1073/pnas.0510550103

Hunt, G. (2012). Measuring rates of phenotypic evolution and the inseparability of tempo and Mode. Paleobiology, 38(3), 2012, pp. 351–373. paleo_Hunt_2012_Pb_rates.pdf

Hunt, G., Hopkins, M.J., and Lidgard, S. (2015). Simple versus complex models of trait evolution and stasis as a response to environmental change. PNAS, vol. 112 /16: 4885–4890. www.pnas.org/cgi/doi/10.1073/pnas.1403662111

Lisiecki, L. E., and Raymo, M. E. (2005). A Pliocene–Pleistocene stack of 57 globally distributed benthic δ18O records. Paleoceanography 20, A1003. doi:10.1029/ 2004PA001071

Liow, L.H., Porto, A., and Di Martino, E. (2024). Trait-Fitness Associations via Fecundity and Competition in a Two-Million-Year-Long Fossil Record. Am. Nat. 204/3. https://doi.org/10.1086/73133

Voje, K. L, Starrfelt, J., and Liow, L. H. (2018) Model Adequacy and Microevolutionary Explanations for Stasis in the Fossil Record. The American Naturalist, 191/4

Appendix A Detailed simulation results

Table A1. Simulation results with $V_p = 1$, based on 1,000 realizations with regular sampling and a fixed number of samples, $n = 30$. Numbers in parentheses are the number of realizations with $\hat{\sigma}^2_{step} < 0$.

$\hat{\mu}_{step}$	$\sigma^2_{step} = 0.001$	$\sigma^2_{step} = 0.01$	$\sigma^2_{step} = 0.1$	$\sigma^2_{step} = 1$
-0.1	0.00088 ± 0.00076 (90)	0.0089 ± 0.0051 (0)	0.089 ± 0.046 (0)	0.90 ± 0.46 (0)
-0.01	0.00088 ± 0.00080 (101)	0.0089 ± 0.0051 (1)	0.0089 ± 0.0051 (0)	0.089 ± 0.045 (0)
0	0.00086 ± 0.00085 (119)	0.0088 ± 0.0046 (0)	0.089 ± 0.046 (0)	0.90 ± 0.46 (0)
0.01	0.00085 ± 0.00081 (100)	0.0089 ± 0.0050 (0)	0.088 ± 0.044 (0)	0.88 ± 0.45 (0)
0.1	0.00092 ± 0.00080 (96)	0.0089 ± 0.0049 (0)	0.090 ± 0.046 (0)	0.089 ± 0.043 (0)

Table A2. Simulation results with $V_p = 400$, based on 1,000 realizations with regular sampling and variable numbers of samples, $n = 3$ to $n = 60$. Numbers in parentheses are the number of realizations with $\hat{\sigma}^2_{step} < 0$.

μ_{step}	$\sigma^2_{step} = 0.001$	$\sigma^2_{step} = 0.01$	$\sigma^2_{step} = 0.1$	$\sigma^2_{step} = 1$	$\sigma^2_{step} = 10$
-0.1	-0.005 ± 0.175 (640)	-0.01 ± 0.18 (630)	0.07 ± 0.23 (460)	0.86 ± 0.67 (53)	8.9 ± 4.8 (1)
-0.01	-0.000 ± 0.188 (646)	0.01 ± 0.18 (649)	0.07 ± 0.22 (455)	0.87 ± 0.65 (47)	9.1 ± 4.8 (1)
0	-0.011 ± 0.172 (639)	-0.00 ± 0.19 (619)	0.08 ± 0.23 (437)	0.85 ± 0.63 (43)	8.9 ± 4.6 (0)
0.01	-0.004 ± 0.186 (645)	0.00 ± 0.19 (60)	0.08 ± 0.23 (465)	0.85 ± 0.65 (52)	8.9 ± 4.8 (0)
0.1	-0.011 ± 0.170 (652)	-0.00 ± 0.19 (645)	0.08 ± 0.24 (457)	0.87 ± 0.66 (44)	8.7 ± 4.6 (1)

Appendix B Difficult case

As isllustrations of the difficulties involved, Fig. B2 shows responses for an evolutionary system with $\mu_{step} = 0.1$ and $\sigma^2_{step} = 10$, when the population variance is $V_p = 400$. Note the large deviations from the WLS prediction lines compared to the standard errors.

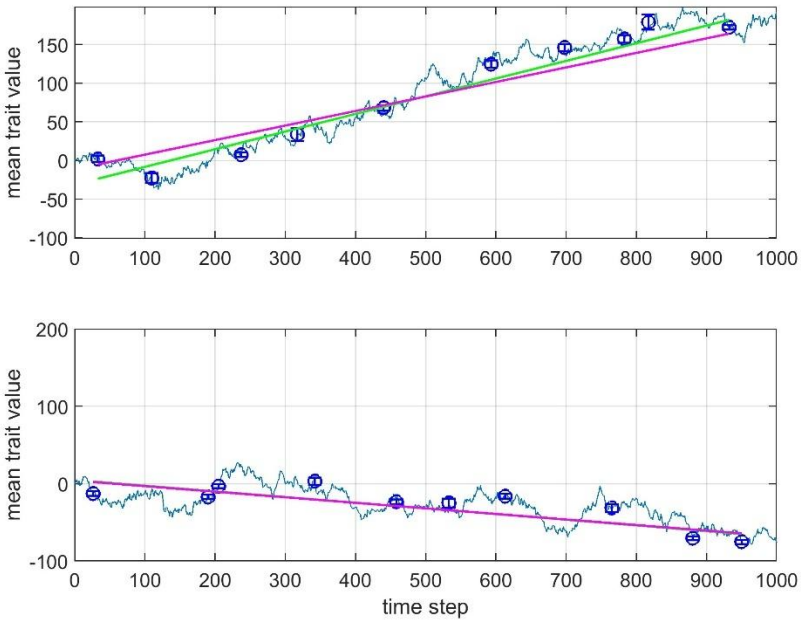

Figure B1. Responses with $\mu_{step} = 0.1$, $\sigma^2_{step} = 10$ and $V_p = 400$, with WLS (green line) and GRW (red line) predictions.

Table B1. Estimated parameter values and WMSE and AIC results for responses in Fig. B1, i.e., for $\mu_{step} = 0.1$ and $\sigma^2_{step} = 10$, and with $V_p = 400$.

	Upper panel	Lower panel
$\hat{\sigma}^2_{step}$	3.6	2.6
$\hat{b}_{GRW} = \hat{\mu}_{step}$	0.19	−0.07
$\hat{b}_{WLS}$	0.23	−0.07
$WMSE_{GRW}$	382	177
$WMSE_{WLS}$	253	178
$AIC_{c,GRW}$	86	82
$AIC_{c,WLS}$	96	91

Appendix C Data for real data cases

Table C1. Data for Bryozoan case

Age Ma	$\partial^{18}O(t)$ temperature	log AZ area mean	SE value
2.1895	3.6518	11.3930	0.0500
1.9660	3.7409	11.4230	0.0500
1.8505	3.7487	11.4550	0.1200
0.9265	4.0233	11.6150	0.0500
0.6485	4.2053	11.6600	0.0400
0.5480	4.0763	11.5550	0.0450
0.5055	4.1545	11.6550	0.0400
0.4510	4.1775	11.7500	0.0850
0.3990	4.1735	11.7000	0.1200

Table C2. Data for Ostracode case 1

Age Ma	Mg/Ca temperature	n	Mean valve length μm
3.33	2.54	13	707
2.60	2.23	4	744
2.40	2.15	21	742
2.19	2.07	23	743
1.90	1.96	25	745
1.68	1.88	22	763
0.89	1.55	28	812
0.14	1.19	24	797
0.12	1.18	3	793
0.00	1.12	31	812

Table C3. Data for Ostracode case 2

Age Ma	Mg/Ca temperature	n	Mean valve length μm
3.96	2.79	13	616
3.33	2.54	1	681
3.22	2.49	7	734
2.58	2.21	6	704
2.19	2.07	1	708
1.68	1.88	2	723
0.21	1.22	5	776
0.11	1.17	29	777
0.09	1.17	43	767
0.08	1.16	35	770

Table C4. Data for Stickleback fish case

Time	n	$\bar{y}$	SD
0	30	8.8300	0.7470
0.0053	33	8.7600	0.6630
0.0100	30	8.8700	0.6810
0.0152	36	9.0600	0.7150
0.0193	55	8.5800	0.7620
0.0256	40	8.7500	0.8990
0.0305	75	8.5600	0.7750
0.0340	52	8.7700	0.7030
0.0366	22	9.1800	0.5010
0.0381	28	9.1800	0.7720
0.0410	54	8.9400	0.6270
0.0506	45	9.2000	0.5880
0.0552	35	9.1400	0.6480
0.0603	37	9.0500	0.6640
0.0653	41	9.1000	0.5390
0.0701	25	9.2000	0.7070
0.0751	30	9.1300	0.7300
0.0811	58	9.3100	0.7540
0.0865	32	9.7500	0.7180
0.0920	84	9.6300	0.8030
0.0934	82	10.0200	0.8460
0.0967	89	9.8000	0.8280
0.0994	95	9.6200	0.7320
0.1013	57	9.6700	0.7150
0.1070	45	9.8000	0.6940
0.1077	64	9.8400	0.7180