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

Rolf Ergon

rolf.ergon@usn.no

University of South-Eastern Norway

April 13, 2026

Short title: A critical look at random walk modeling

Postal address:
University of South-Eastern Norway (USN)
Post office box 4
3199 Borre
NORWAY

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, but even here 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 collapses into deterministic walk processes plus sampling errors. As a result of poor step variance estimates, the directional evolution may be both under- and overestimated as compared with generalized least squares (GLS) results, which gives the best linear unbiased estimator (BLUE) of the evolutionary slope. Here, I study this problem through simulations as well as in four real data cases. Based on weighted mean square error (WMSE) comparisons, my conclusion is that GLS in cases with large measurement errors is the best method for inference of directional evolution. When the step variances must be set to zero, GLS is simplified into weighted least squares (WLS) estimation.

Keywords: generalized least squares estimation; 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_{step}^2. 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)

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

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_{step}^2 by summing the log-likelihoods according to Eq. (1) over the transitions, i.e., by use of $\ell_{total}\left(\mu_{step}, \sigma_{step}^2\right) = \sum_{i=1}^{N-1} \ell_i\left(\mu_{step}, \sigma_{step}^2\right)$. 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 $\ell\left(\mu_{step}, \sigma_{step}^2\right)$ is maximized when $\hat{\mu}_{step} = \Delta\mathrm{Y}/T$, 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_{step}^2 in cases where $\mathrm{T}\sigma_{step}^2$ 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 $\hat{\sigma}_{step}^2$ 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_{step}^2 were estimated well, except for small values of σ_{step}^2 as compared with V_{p_A} and V_{p_D}. The population phenotypic variances in these simulations appear

to be unrealistically small, as compared with 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 $\hat{\sigma}^2_{step} = 0$. In such cases the GRW model collapses into a deterministic directional walk model plus sampling errors.

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 $\hat{a}_{GRW}$ to the data by a generalized least squares approach. Here, c is a scaling constant that can be chosen freely (Hunt, 2006), and throughout the paper I will set $c = 1$.

In my simulations I also analyze data for directional evolution by means of the generalized least squares (GLS) method, which gives the best linear unbiased estimator (BLUE) of the evolutionary slope. I then find that GRW may both underestimate and overestimate prediction slopes by up to 50%. For comparisons of GRW and GLS slope estimates, the simulation results include weighted mean square errors (WMSE).

In Section 4 I analyze data from four real data cases, a bryozoan case, two ostracod cases, and a stickleback fish case. In all these cases parameter estimation with σ^2_{step} as a free variable resulted in $\hat{\sigma}^2_{step} < 0$, which means that $\hat{\sigma}^2_{step} = 0$ must be used in a realistic model, and as a result of this GLS is simplified into weighted least squares (WLS).

In the first three cases there exists a dominating evolutionary driver, and a tracking model as discussed in Ergon (2025a) is then an attractive alternative. The applicability of such a model will, however, depend on the quality of the evolutionary data, and only in the bryozoan case is the tracking model better than the WLS model. As shown in Ergon (2025b) it is also in the stickleback fish case possible to find a tracking model that is clearly better than the WLS model, but then by use of an association with rather than a direct measurement of an

environmental proxy. Ergon (2025b) also includes improved tracking models for the two ostracod cases, using a different temperature proxy.

My results are summarized and discussed in Section 5, and my conclusion is that GLS or WLS are the best alternatives for analyses of directional evolution, and that dependent on the quality of the data a tracking model may give even better prediction results. Appendix A gives some details regarding the tracking model for the stickleback fish case. Some detailed simulation results are given in Appendix B, results for various values of σ_{step}^2 are given in Appendix C, while Appendix D 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 generalized least squares approach. Here, c is a scaling constant that I consistently will set to $c = 1$.

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, $\bar{y}_{true}(50)$, $\bar{y}_{true}(150)$, ..., $\bar{y}_{true}(950)$. The initial value is chosen as $\bar{y}_{true}(1) = 0$. 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 $\bar{y}_{true}(t)$. To each of the ten samples I add the mean values of $n = 30$ individual measurement errors drawn from normally distributed populations with mean $\mu_p = 0$ and variance $V_p = 1$, and thus form ten

samples $\bar{y}(t)$. Each of these samples has 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 $\hat{\sigma}^2_{step}$ that maximize the log-likelihood function $\ell_{total} = \sum_{i=1}^{N-1} \ell_i(\mu_{step}, \sigma^2_{step})$, where ℓ_i is given by Eq. (1). For this purpose, I use the *fmincon* function in MATLAB, i.e., I minimize $-\ell_{total}(\mu_{step}, \sigma^2_{step})$.

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 Covariance matrix for random walk process

In order to find the covariance matrix $\boldsymbol{C}$ of a random walk process we may for simplicity assume $\mu_{step} = 0$, such that the expected value at all time steps is zero. For an element $c_{t,s}$ in $\boldsymbol{C}$ with $s > t$ we will then find

$$c_{t,s} = E[\bar{y}_t \bar{y}_s] = E[\bar{y}_t(\bar{y}_t + \triangle\, \bar{y}_{t+1} + \cdots + \triangle\, \bar{y}_s)] = E[\bar{y}_t^2] = t\sigma^2_{step}, \tag{3}$$

where $E[\cdot]$ is the expectation operator, and where the incremental changes $\triangle\, \bar{y}_{t+1}$ etc. are independent of $\bar{y}_t$. The result from Eq. (3) is the well-known variance of integrated discrete white noise. With N samples we thus find

$$\boldsymbol{C} = \begin{bmatrix} t_1 & t_{t_1} & t_1 & t_1 & t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & t_2 & t_2 & t_2 & t_2 & \cdots & t_2 \\ t_1 & t_2 & t_3 & t_3 & t_3 & t_3 & \cdots & t_3 \\ t_1 & t_2 & t_3 & t_4 & t_4 & t_4 & \cdots & t_4 \\ t_1 & t_2 & t_3 & t_4 & t_5 & t_5 & \cdots & t_5 \\ t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & \cdots & t_6 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & \cdots & t_N \end{bmatrix} \sigma^2_{step}. \tag{4}$$

With $\mu_{step} \neq 0$ we will have a combination of directional evolution and integrated white noise, which is a clear parallel to the combination of natural selection and genetic drift as discussed in Walsh and Lynch (2018), Ch. 18.

2.4 Generalized and weighted least squares estimation

In the simulations, predictions $\hat{\bar{y}}_{GRW}(t)$ according to Eq. (2) will be compared with generalized least squares (GLS) predictions found as $\hat{\bar{y}}_{GLS}(t) = \hat{a}_{GLS} + \hat{b}_{GLS}t$. Here, $\hat{a}_{GLS}$ and $\hat{b}_{GLS}$ are found as

$$\begin{bmatrix} \hat{a}_{GLS} \\ \hat{b}_{GLS} \end{bmatrix} = \left(\boldsymbol{X}^T \boldsymbol{\Omega}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{\Omega}^{-1} \bar{\boldsymbol{y}}, \tag{5}$$

where $\boldsymbol{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_N \end{bmatrix}^{\boldsymbol{T}}$ and $\bar{\boldsymbol{y}} = [\bar{y}(t_1) \quad \bar{y}(t_2) \quad \cdots \quad \bar{y}(t_N)]^{\boldsymbol{T}}$, while $\boldsymbol{\Omega} = \boldsymbol{C} + \boldsymbol{V}$, where $\boldsymbol{V}$ is a diagonal matrix with elements $v_t = V_p/n_t$. Note that the prediction slope parameter $\hat{b}_{GLS}$ is the best linear unbiased estimator (BLUE).

Comparisons of responses $\hat{\bar{y}}_{GRW}(t)$ and $\hat{\bar{y}}_{GLS}(t)$ can be done by use of generalized mean squared errors

$$WMSE_{GRW} = (\bar{\boldsymbol{y}}_{true}^T - \hat{\bar{\boldsymbol{y}}}_{GRW}^T)\boldsymbol{\Omega}^{-1}(\bar{\boldsymbol{y}}_{true} - \hat{\bar{\boldsymbol{y}}}_{GRW})/trace(\boldsymbol{\Omega}^{-1}), \tag{6}$$

and

$$WMSE_{GLS} = (\bar{\boldsymbol{y}}_{true}^T - \hat{\bar{\boldsymbol{y}}}_{GLS}^T)\boldsymbol{\Omega}^{-1}(\bar{\boldsymbol{y}}_{true} - \hat{\bar{\boldsymbol{y}}}_{GLS})/trace(\boldsymbol{\Omega}^{-1}), \tag{7}$$

where $\bar{\boldsymbol{y}}_{true}$, $\hat{\bar{\boldsymbol{y}}}_{GRW}$ and $\hat{\bar{\boldsymbol{y}}}_{GLS}$ are $N \times 1$ data vectors. Note that the division by $trace(\boldsymbol{\Omega}^{-1})$ makes the WMSE results consistent with MSE for ordinary least squares (OLS), where $\boldsymbol{\Omega} = \boldsymbol{I}$ and $trace(\boldsymbol{\Omega}^{-1}) = N$, but this normalization is of no importance for comparisons of GRW and GLS results. Also note that in real data cases $\bar{\boldsymbol{y}}_{true}$ must be replaced by the observation vector $\bar{\boldsymbol{y}}$.

In all four real data cases in Section 4 the realistic step variances are found to be zero, such that $\boldsymbol{C} = 0$. From Eq. (5) then follows the weighted least squares (WLS) result

$$\begin{bmatrix} \hat{a}_{WLS} \\ \hat{b}_{WLS} \end{bmatrix} = (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{y}, \tag{8}$$

where $\boldsymbol{W} = \boldsymbol{V}^{-1}$. From this follows that $\boldsymbol{\Omega}^{-1}$ in Eqs. (6) and (7) is replaced by $\boldsymbol{W}$, while $trace(\boldsymbol{\Omega}^{-1})$ becomes $\sum_{i=1}^{N} w_i$.

3. Simulation results

As in Hunt (2006), GRW parameter estimation results with phenotypic variance $V_p = 1$ were found from 1,000 repeated simulations, with typical responses from a system with $\mu_{step} = 0.1$ and $\sigma_{step}^2 = 0.1$ as shown in Fig. 1, left panel. In the same simulations, GRW and GLS prediction parameters $\hat{a}_{GRW}$ and $\hat{b}_{GRW} = \hat{\mu}_{step}$ as well as $\hat{a}_{GLS}$ and $\hat{b}_{GLS}$ were found, which gave typical responses as also shown in Fig. 1, left panel. The results are summarized in Table 1, including *WMSE* values based on normalized weights. The $\hat{\mu}_{step}$ and $\hat{\sigma}_{step}^2$ results are very much the same as found by Hunt (2006), and also here $\hat{\sigma}_{step}^2$ consistently approaches $0.1(N-1)/N = 0.09$. Note that the measurement standard errors as shown by error bars are quite small. Also note, however, that the SE value for $\hat{\sigma}_{step}^2$ 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 comparable to the ones 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 for all measurements are also shown in Table 1. Note that the SE value for $\hat{\sigma}_{step}^2$ now is much larger than the mean value, and that 397 of the 1,000 realizations resulted in $\hat{\sigma}_{step}^2 < 0$. In such realizations, $WMSE_{GWR}$ was computed by use of $\sigma_{step}^2 = 0$. Results with irregular sampling and variable numbers of individual samples, as described in Section 2, are included in Table 1, and typical responses are shown in Fig. 1, right panel. Note that the estimates of μ_{step} are very much the same when V_p is increased from 1 to 400, but that $\hat{\sigma}_{step}^2$ is very much affected. More detailed parameter estimation results are given in Appendix B.

The $WMSE$ 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 50% under- or overestimated, as compared with $\hat{b}_{GLS}$, see histogram in Fig. 2 (lower panel). Note that the GRW estimation error $\Delta\hat{b} = \hat{b}_{GRW} - \hat{b}_{GLS}$ with $\sigma^2_{step} = 0.1$ has approximately the same variance as the estimated slope $\hat{b}_{GLS}$ (upper panel). As shown in Appendix C, the variance of $\hat{b}_{GLS}$ increases with σ^2_{step}, while the variance of the GRW estimation error decreases. Three typical realizations are shown in Fig. 3.

As an illustration of the difficulties involved, Appendix C also shows responses for two quite different evolutionary realizations of a system with $\mu_{step} = 0.1$ and $\sigma^2_{step} = 10$, both apparently directional, but with different signs. Also in these cases, however, $WMSE_{GLS}$ is clearly lower than $WMSE_{GRW}$.

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
$\hat{\mu}_{step} = \hat{b}_{GRW}$	0.1002 ± 0.0106	0.1001 ± 0.0121	0.1003 ± 0.0171
$\hat{\sigma}^2_{step}$	0.0866 ± 0.0432	0.0835 ± 0.1884*	0.0780 ± 0.2270 **
$\hat{b}_{GLS}$	0.1002 ± 0.0106	0.1000 ± 0.0115	0.1001 ± 0.0118
$WMSE_{GWR}$	4.76 ± 10.21	8.54 ± 5.19	17.11 ± 32.05
$WMSE_{GLS}$	4.77 ± 10.16	6.92 ± 3.69	6.78 ± 4.50

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

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

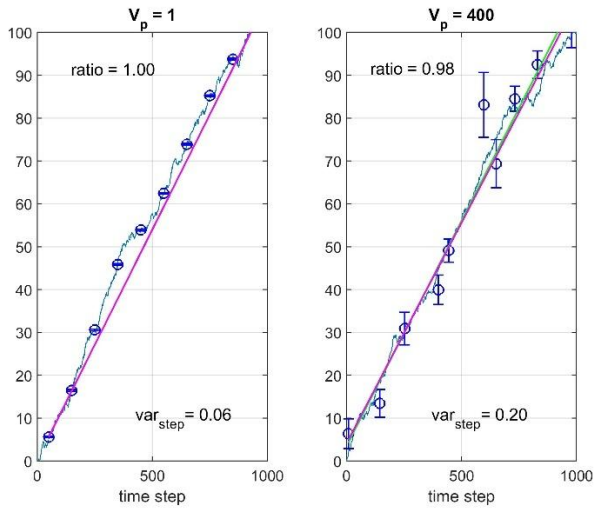

Figure 1. Typical results for simulations with $\mu_{step} = 0.1$, $\sigma_{step}^2 = 0.1$, $V_p = 1$ and regular sampling (left panel) and with $V_p = 400$ and irregular sampling (right panel). Values of $\bar{y}_{true}(t)$ are shown by solid blue lines, while $\bar{y}(t)$ are shown by circles with error bars. GLS predictions are shown by green lines, and GRW predictions are shown by red lines. Estimated step variances $\hat{\sigma}_{step}^2$ and slope ratios $\hat{b}_{GRW}/\hat{b}_{GLS}$ are shown in the plots.

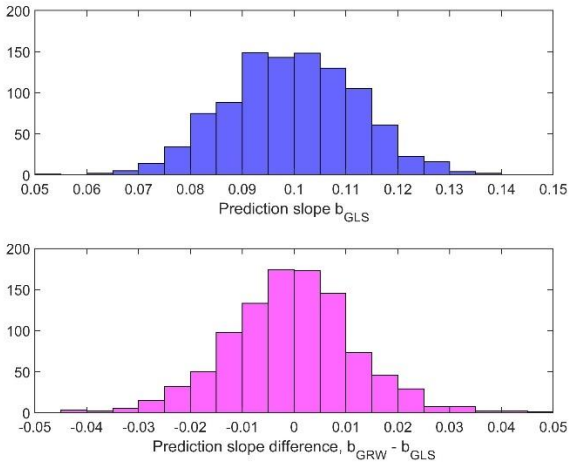

Figure 2. Typical histograms for estimated prediction slope $\hat{b}_{GLS}$ (upper panel) and for the GRW estimation error $\Delta\hat{b} = \hat{b}_{GRW} - \hat{b}_{GLS}$ (lower panel), based on 1,000 realizations with $\mu_{step} = 0.1$, $\sigma_{step}^2 = 0.1$, $V_p = 400$, irregular sampling and variable numbers of samples as described in text.

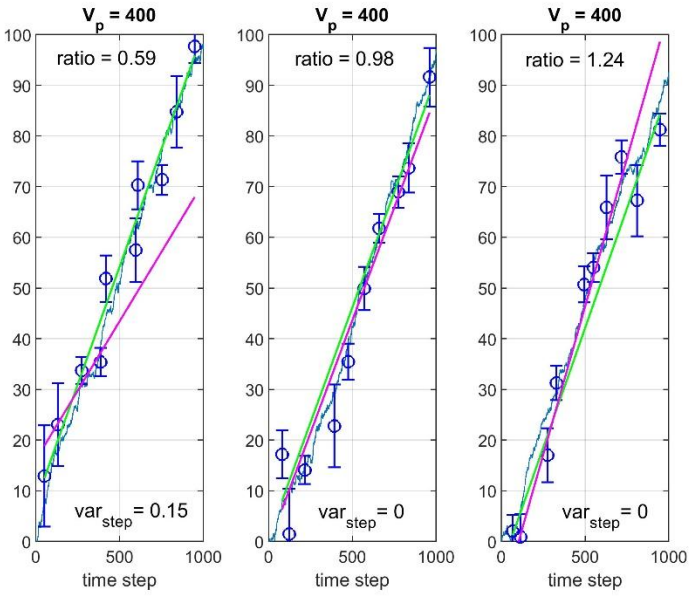

Figure 3. Prediction results for three realizations with $\mu_{step} = 0.1$, $\sigma^2_{step} = 0.1$, $V_p = 400$, irregular sampling and variable numbers of samples as described in text. Values of $\bar{y}_{true}(t)$ are shown by solid blue lines, while $\bar{y}(t)$ are shown by circles with error bars. GLS and GRW predictions are shown by green and red lines, respectively. Estimated step variances $\hat{\sigma}^2_{step}$ and slope ratios $\hat{b}_{GRW}/\hat{b}_{GLS}$ 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 (2025a) 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 (2025a), good predictions can be found using an adaptive peak tracking model based on a moving average smoothed version of the $\partial^{18}O(t)$ environmental proxy, with

window size 100. An essential feature of this model is that the mean trait values as function of $\partial^{18}O_{smooth}$ can be approximated by a straight line, see Fig. 9 in Ergon (2025a). When this line is found by WLS the result is a tracking model with known WMSE values. The $\partial^{18}O(t)$ data are found in Lisiecki and Raymo (2005), while other data are collected in Table D1 in Appendix D.

Since GWR without parameter limitations resulted in $\hat{\sigma}^2_{step} < 0$, I used $\hat{\sigma}^2_{step} = 0$, such that GLS is simplified into WLS. Fig. 4 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 $c = 1$. Prediction slope and WMSE results for the three models are given in Table 2. Note that the WLS results are better than the GRW results, but that the tracking results are even better.

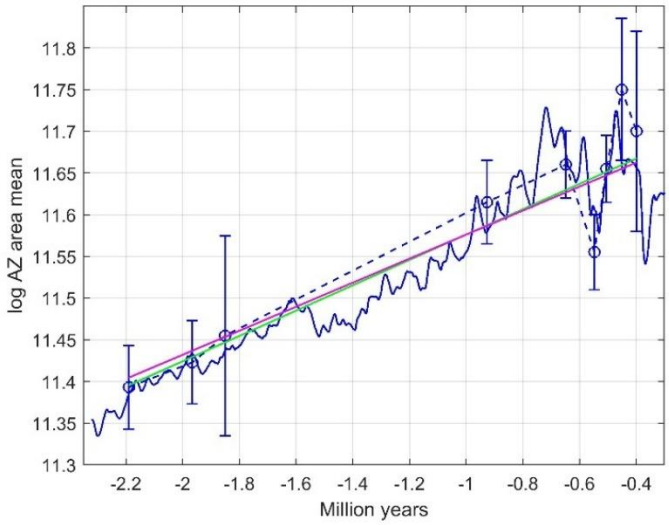

Figure 4. Tracking, WLS and GRW results for Bryozoan case, with mean trait observations (dashed blue line with circles and error bars), predictions based on a centered moving average smoothed version of the $\partial^{18}O(t)$ temperature proxy (solid blue line), WLS predictions (green line), and GRW predictions with $\hat{\sigma}^2_{step} = 0$ (red line).

2 Ostracod case 1

Hunt and Roy (2006) presented time series data for body size of the ostracod genus *Poseidonamicus*, and in this case I use data for *P. major*, extracted as shown in Table D2 in Appendix D. The data include long-term smoothed Ma/Ca 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.

As in the Bryozoan case, the mean trait as function of the temperature proxy could be approximated by a straight line, such that a tracking model could be used. Also in this case the realistic step variance is found to be $\hat{\sigma}^2_{step} = 0$, such that GLS is simplified into WLS.

Fig. 5 shows predictions by means of the Mg/Ca tracking model as well as of WLS and GRW models. Prediction slope and WMSE results are given in Table 2. Note that the WLS model is better than the tracking model, presumably owing to age uncertainty in the Mg/Ca data, and that the GRW model underestimates the prediction slope by 21%. Verification of the search results are shown in the grid contour plot in Fig. 6, where it should be noted that $\hat{\sigma}^2_{step}$ is allowed to be negative.

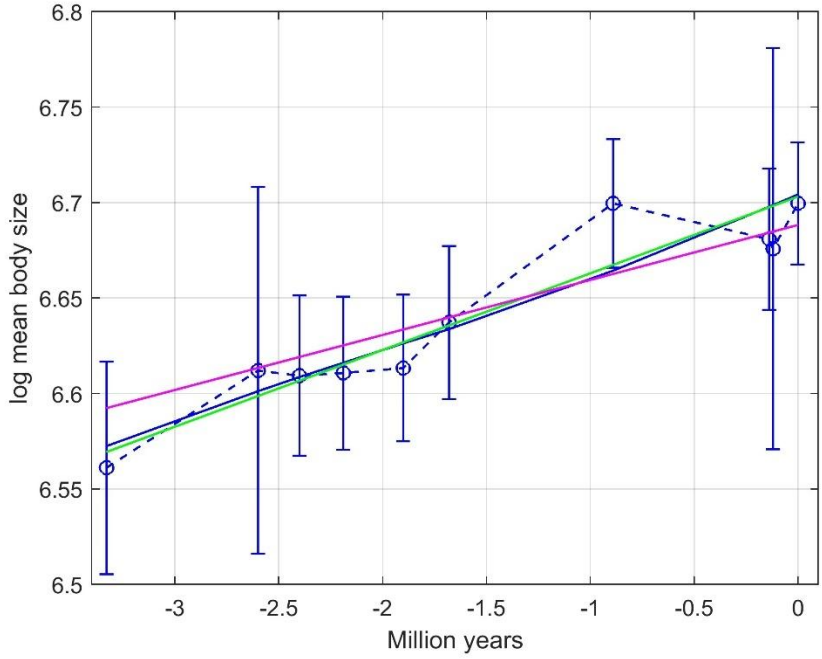

Figure 5. Results for Ostracod case 1, with mean trait observations (dashed blue line with circles and error bars), predictions based on the Mg/Ca temperature proxy (solid blue line that is not quite a straight line), WLS predictions (green line, almost identical to the solid blue line), and GRW predictions with $\hat{\sigma}^2_{step} = 0$ (red line).

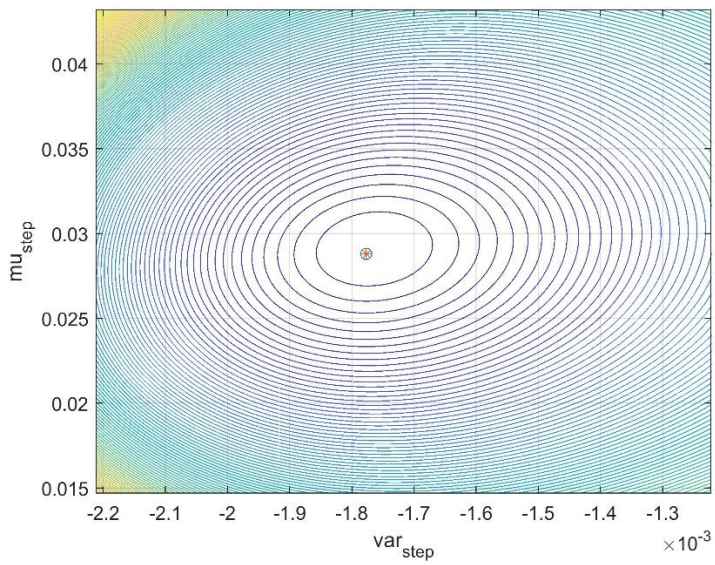

Figure 6. Grid contour results for Ostracod case 1, with the point $[\hat{\sigma}^2_{step}, \hat{\mu}_{step}]$ added as circle with star. Note that $\hat{\sigma}^2_{step} < 0$ is accepted.

4.3 Ostracod case 2

In this case I use data for *P. pintoi* in Hunt and Roy (2006), as shown in Table D3 in Appendix D. This is another case where the realistic step variance is found to be $\hat{\sigma}^2_{step} = 0$. As in the Ostracod case 1, the mean trait as function of the temperature proxy could be approximated by a straight line, such that a tracking model could be used.

Fig. 7 shows predictions by means of a Mg/Ca tracking model as well as of the WLS and GRW models. Prediction slope and WMSE results are given in Table 2. Note that also here the WLS model is better than the tracking model, presumably owing to age uncertainty in the Mg/Ca data, and that the GRW model underestimates the prediction slope by 41%. As a verification of the search results, Fig. 8 shows the log likelihood as function of μ_{step} with $\hat{\sigma}^2_{step} = -0.0164$ (the result with σ^2_{step} as free variable, upper panel) and $\hat{\sigma}^2_{step} = 0$ (lower panel). As seen in Fig. 8, the maximum log likelihood value is reduced when the search is limited to positive values of σ^2_{step}, i.e., when $\hat{\sigma}^2_{step}$ is non-optimal.

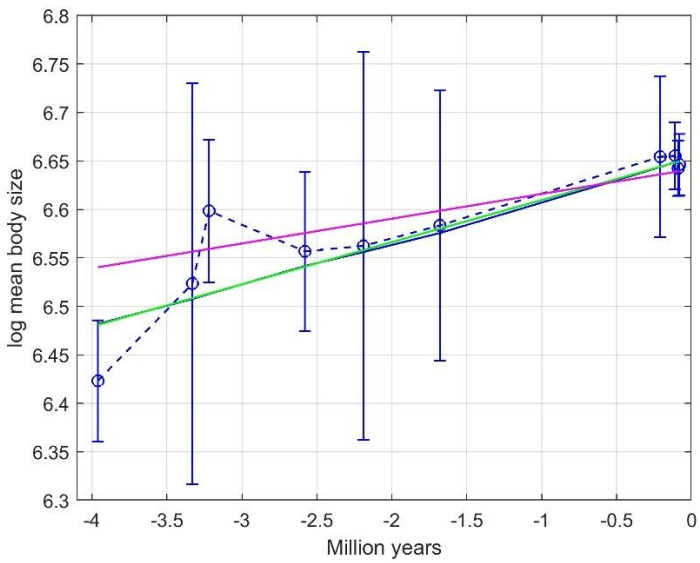

Figure 7. Results for Ostracod case 2, with mean trait observations (dashed blue line with error bars), predictions based on the Mg/Ca temperature proxy (solid blue line that

is not quite a straight line), WLS predictions (green line, almost identical to the solid blue line), and GRW predictions with $\hat{\sigma}^2_{step} = 0$ (red line).

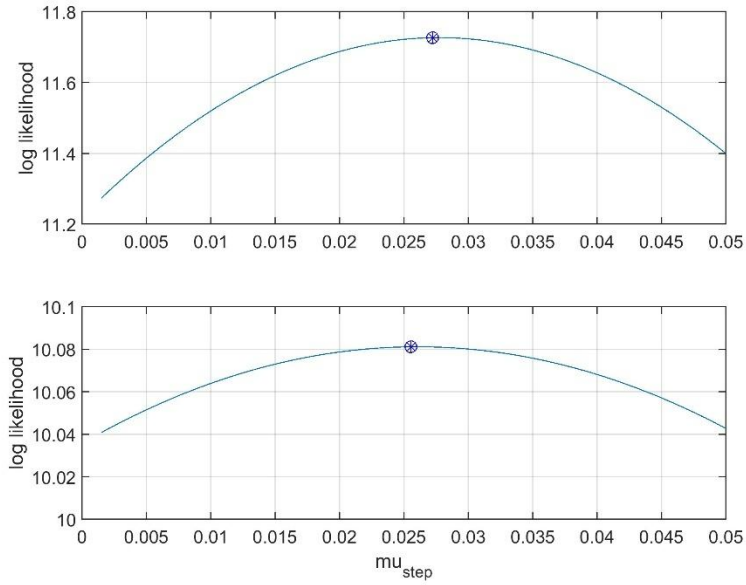

Figure 8. The log likelihood as function of μ_{step} with $\hat{\sigma}^2_{step} = -0.0164$ (upper panel) and $\hat{\sigma}^2_{step} = 0$ (lower panel), with maxima marked by circles with stars.

4.4 Stickleback fish case

In this final case I use data that show how the dorsal fin ray number evolved in the stickleback fish species *Gasterosteus doryssus* (Bell et al., 1985), see Table D4 in Appendix D for data. The zero point on the time scale is approximately -10 million years (personal information). In the data analysis the mean trait values are log-transformed, with the measurement errors transformed accordingly.

This is another example where the realistic step variance is found to be $\hat{\sigma}^2_{step} = 0$, such that GLS is simplified into WLS. See results in Fig. 9 and Table 2.

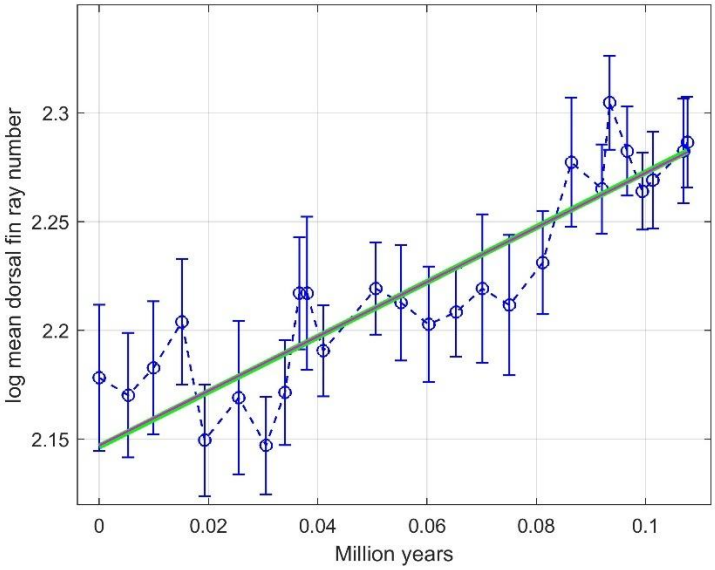

Figure 9. Results for the Stickleback fish case, with mean trait observations (dashed blue line with error bars), WLS predictions (green line), and GRW predictions with $\hat{\sigma}^2_{step} = 0$ (red line, very much on top of the green line).

As shown in Ergon (2025b), it is also in this case possible to find a tracking model. The essential elements and results for such a model are given in Appendix A.

Table 2. Results for real data cases.

	Bryozoan	Ostracod 1	Ostracod 2	Stickleback
$\hat{b}_{WLS}$	0.1519	0.0402	0.0434	1.2605
$\hat{b}_{GRW}$	0.1441	0.0316	0.0255	1.2531
$WMSE_{Tracking}$	0.00100	0.000279	0.000613	–
$WMSE_{WLS}$	0.00190	0.000241	0.000606	0.00040126
$WMSE_{GRW}$	0.00193	0.000319	0.0011	0.00040132

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 generalized least squares (GLS) estimation is a better alternative. A second aim is to show that realistically large measurement 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 collapses into a deterministic directional walk model. A final aim is to show that adaptive peak tracking in some cases can be a good modeling alternative.

In the simulations, Hunt (2006) used regular sampling as explained in Section 2, and what appears to be the unrealistically low phenotypic variance $V_p = 1$, and one of the findings was that with $\sigma^2_{step} = 0.001$ "some simulated sequences" yielded $\hat{\sigma}^2_{step} < 0$ (Fig. 4 in Hunt, 2006). This is not surprising considering the term in Eq. 1 (with Hunt's numbers) $T\sigma^2_{step} + \frac{V_{p_A}}{n_A} + \frac{V_{p_D}}{n_D} = T\left(\sigma^2_{step} + 0.0013\right)$, which obviously makes it rather difficult to estimate $\sigma^2_{step} = 0.001$ from data. Repeating the simulations in Hunt (2006) I found that "some" is around 10%, which with my simulations with 10 samples instead of 20 increases to 18%. With the more realistic phenotypic variance $V_p = 400$ and regular sampling I used $T\sigma^2_{step} + \frac{V_{p_A}}{n_A} + \frac{V_{p_D}}{n_D} = T\left(\sigma^2_{step} + 0.27\right)$, and for $\sigma^2_{step} = 0.1$ I found $\hat{\sigma}^2_{step} < 0$ in 40% of the realizations (Table 1). The probability of obtaining $\hat{\sigma}^2_{step} < 0$ will thus increase when the number of samples decreases and the phenotypic variance increases. 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 C).

In addition to simulations, different models were compared in four real data cases. In all these cases the estimated step variances were negative if they were allowed to be so, such that $\hat{\sigma}^2_{step} = 0$ must be used in the GRW models, while GLS was simplified into WLS. In the Bryozoan

and Stickleback fish cases I found that the GRW and WLS predictions were quite similar (Figures 4 and 9, and Table 2). In the Ostracod case 1 GRW underestimated the prediction slope by 21% (Fig. 5 and Table 2), while GRW in the Ostracod case 2 underestimated the prediction slope by 41% (Fig. 7 and Table 2).

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 generalized least squares approach. This makes it possible to compare GRW and GLS/WLS results using weighted mean squared errors (*WMSE*), and for all four of the real data cases with $\hat{\sigma}^2_{step} = 0$ I found $WMSE_{WLS} < WMSE_{GRW}$, although the difference is small in the Bryozoan case and very small in the Stickleback fish case (Table 2). This is not surprising, of course, since with zero step variance the GRW model collapses into 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.

In three of the real data cases, I include an adaptive peak tracking model based on a dominating and well-known evolutionary driver (Ergon (2025a), and the tracking models gave better prediction results than GRW (Table 2). In the two Ostracod cases, however, the tracking results were poorer than the WLS results, which may possibly be explained by how the Mg/Ca temperature proxy was found. As shown in Ergon (2025b), a tracking model for Ostracod case 1 using the $\partial^{18}O(t)$ temperature proxy, gives clearly improved results. Ergon (2025b) also includes a tracking model for the stickleback fish case, but then by use of an association with rather than a direct measurement of an environmental proxy. Essential elements and results for such a model are given in Appendix A.

As shown in Table 1, $WMSE_{GRW} \approx WMSE_{GLS}$ in simulations with $V_p = 1$, while simulations with the more realistic phenotypic variance $V_p = 400$ show that $WMSE_{GLS} < WMSE_{GRW}$. This is also the result in each of the four real data cases with $\hat{\sigma}^2_{step} = 0$, where GLS is replaced by WLS (Table 2), although the difference is marginal in two of the cases. The main conclusion is thus

that GLS/WLS are the best methods for comparison of directional evolution with other evolutionary modes. As discussed in Hunt (2006), such comparisons are preferably performed using the Akaike Information Criterion (AIC), which raises questions of AIC for GLS and WLS. AIC for WLS is discussed in Ergon (2025b), based on Banks and Joyner (2017), but methods for finding AIC for GLS appear to be missing in the literature.

What this conclusion 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, however, natural to ask how many of the time series in these studies that are examples of adaptive peak tracking as in the four real data cases in this paper. In such examples the different categories of random walk models may be of limited interest, since the evolutionary responses primarily are determined by the time series of the environmental drivers. The positive directional evolution in the Stickleback fish case, for example, would most likely have been negative if the fossils had been 0.2 million years younger (see Fig. B1 in Ergon, 2025b).

An interesting statistical issue for future research is the strong dependence on accurate estimates of sampling times. These are certainly measured with substantial errors, and the incremental change from one sample to another is a function of the time T between them. Hence, both the GRW and GLS/WLS solutions are influenced by errors in estimation of T, which possibly could cause large errors in the estimation results. This problem is discussed in Ergon (2025a) in relation to the Bryozoan case, but a more theoretical approach supported by a simulation study might be clarifying.

Acknowledgments

I thank one of the reviewers for corrections regarding the covariance structure used in the simulations and for an interesting literature reference. I also thank University of South-Eastern Norway for support and funding.

Conflicts of Interest

There are no conflicts of interest.

Data Availability Statement

MATLAB code is archived on biorxiv https://doi.org/10.1101/2025.11.11.687792

Oxygen isotope data for Fig. 4 are archived as Raw data for simulations on bioRxiv, https://doi.org/10.1101/2024.10.30.621046.

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. (2025a). 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

Ergon, R. (2025b). Examples of Adaptive Peak Tracking as Found in the Fossil Record, Including Obliquity Cycles Tracking. *Modeling, Identification and Control*, doi:10.4173/mic.2025.4.2.

Hendry, A.P., Peichel, C.L, Matthews, B., Boughman, J.W., and Nosil, P. (2013). Stickleback research: the now and the next. Evolutionary Ecology Research, 15: 111–141.

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

Walsh, B., and Lynch, M. (2018). Evolution and selection of quantitative traits. Oxford University Press

Westerhold, T. et al. (2020). An astronomically dated record of earth's climate and its predictability over the last 66 million years. *Science*, 369(6509):1383–1387. doi:10.1126/science.aba6853.

Appendix A. Tracking model for stickleback fish case

As shown in Ergon (2025b) it is possible to find a tracking model also in the stickleback fish case, with use of $\partial^{18}O(t)$ data as found from deep sea drilling samples (Westerhold et al., 2020). These data include obliquity cycles with a period of around 41,000 years, as found by power spectral density analysis (Fig. A1), and this information can be utilized by use of centered

moving average smoothing as a feature extraction method. The result is shown in Fig. A2, from which follows $WMSE_{WLS} = 0.000401$ and $WMSE_{tracking} = 0.000211$.

The directional trend in Fig. A2 appears to be caused by an overall temperature trend, while the cycles apparently are caused by obliquity cycles. The model makes use of an association between cycles in the $\partial^{18}O(t)$ data and in the stickleback mean trait value, but the detailed mechanisms remain to be explained. It is for example well known that obliquity affects the intensity of seasons, and one mechanism may therefore be that this affects the balance between summer survival (declining with number of spines) and winter survival (increasing with number of spines), see Hendry et al. (2013).

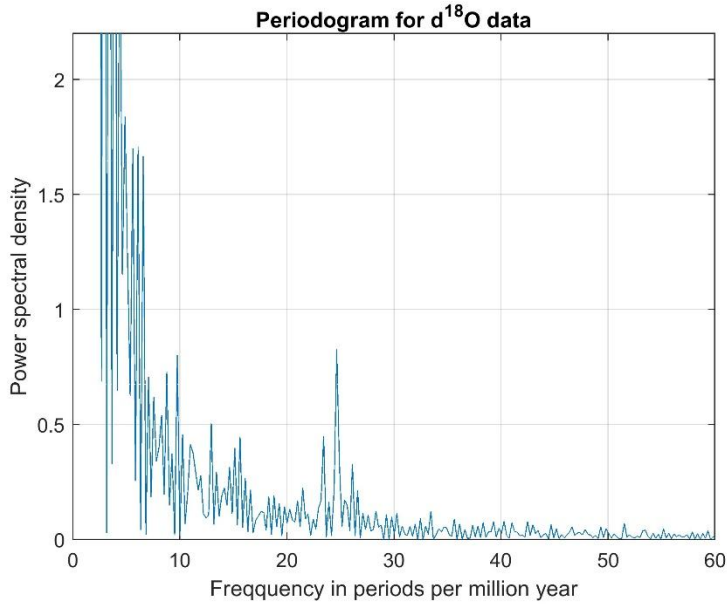

Figure A1. Periodogram for $\partial^{18}O(t)$, with obliquity cycles seen as a peak at 24.7 periods per million years.

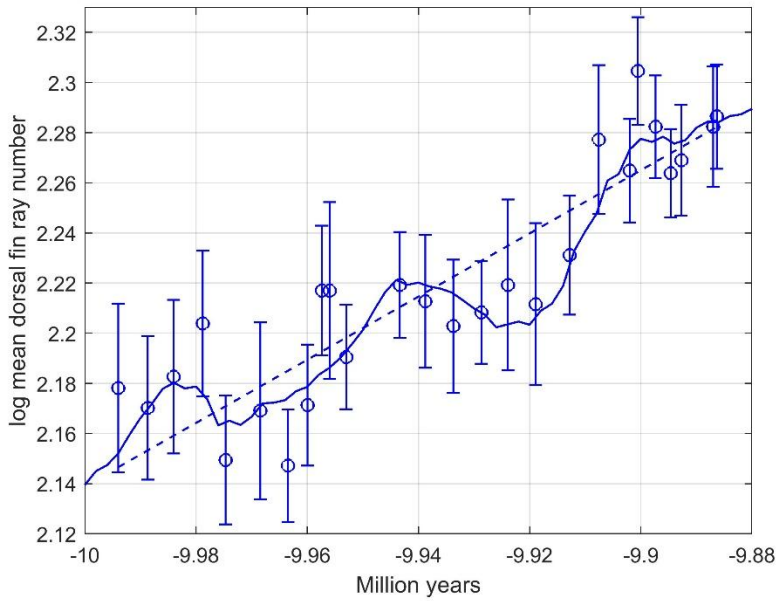

Figure A2. Predictions of stickleback fish mean trait by use of true data with obliquity cycles (solid line). Dashed line shows WLS predictions.

Appendix B. Detailed simulation results

Table B1. Simulation $\hat{\sigma}^2_{step}$ 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 $\hat{\sigma}^2_{step} < 0$.

$\hat{\mu}_{step}$	$\sigma^2_{step} = 0.001$	$\sigma^2_{step} = 0.01$	$\sigma^2_{step} = 0.1$	$\sigma^2_{step} = 1$	$\sigma^2_{step} = 10$
0	0.00091 ± 0.00082 (120)	0.0088 ± 0.0047 (0)	0.090 ± 0.045 (0)	0.88 ± 0.45 (0)	8.7 ± 4.4 (0)
0.01	0.00095 ± 0.00080 (91)	0.0090 ± 0.0048 (0)	0.088 ± 0.044 (0)	0.87 ± 0.42 (0)	9.0 ± 4.3 (0)
0.1	0.00086 ± 0.00075 (102)	0.0087 ± 0.0047 (1)	0.090 ± 0.045 (0)	0.91 ± 0.45 (0)	9.1 ± 4.5 (0)

Table B2. Simulation results with $V_p = 400$, based on 1,000 realizations with irregular 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	0.062 ± 0.153 (658)	0.06 ± 0.13 (657)	0.11 ± 0.18 (487)	0.85 ± 0.66 (41)	8.8 ± 4.6 (0)
0.01	0.052 ± 0.114 (667)	0.06 ± 0.15 (625)	0.12 ± 0.19 (453)	0.87 ± 0.69 (56)	9.1 ± 4.9 (0)
0.1	0.061 ± 0.147 (651)	0.06 ± 0.15 (654	0.11 ± 0.18 (468)	0.88 ± 0.65 (50	8.9 ± 4.8 (1)

Appendix C. Results for various values of σ^2_{step}

Fig. C1 shows histograms as in Fig. 2, but with $\sigma^2_{step} = 0.01$, 0.1 and 1. Note that the variance of $\hat{b}_{GLS}$ increases with σ^2_{step}, while the variance of the GRW estimation error $\Delta\hat{b} = \hat{b}_{GRW} - \hat{b}_{GLS}$ decreases.

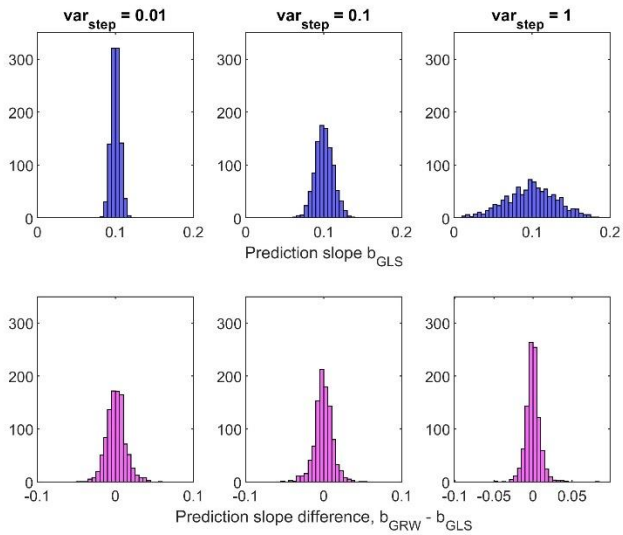

Figure C1. Typical histograms as in Fig. 2, for three values of σ^2_{step}.

As illustrations of the difficulties involved, Fig. C2 and Table C show responses and numerical results for two realizations of an evolutionary system with $\mu_{step} = 0.1$ and $\sigma^2_{step} = 10$, when the phenotypic variance is $V_p = 400$. The sampling was irregular. Note that the estimated slopes have different signs, and that the deviations from the GLS prediction lines are large compared to the standard errors. Also in these cases, however, $WMSE_{GLS}$ is clearly lower than $WMSE_{GRW}$.

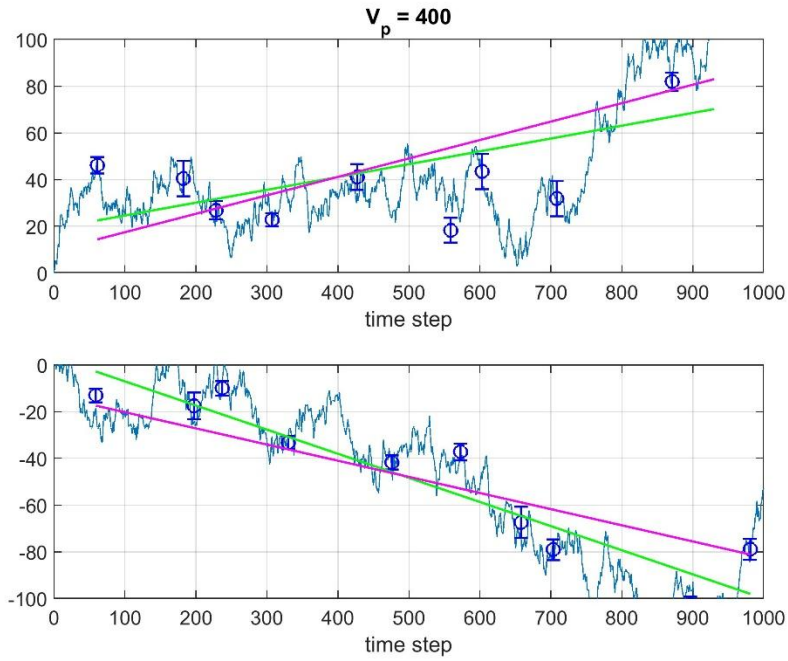

Figure C2. Responses for two realizations with $\mu_{step} = 0.1$, $\sigma^2_{step} = 10$, $V_p = 400$ and irregular sampling, with GLS (green lines) and GRW (red lines) predictions.

Table C. Estimated parameter values and WMSE results for responses in Fig. C2, i.e., for $\mu_{step} = 0.1$, $\sigma^2_{step} = 10$, $V_p = 400$ and irregular sampling.

	Upper panel	Lower panel
$\hat{\sigma}^2_{step}$	5.3	2.4
$\hat{b}_{GRW} = \hat{\mu}_{step}$	0.08	−0.07
$\hat{b}_{WLS}$	0.05	−0.10
$WMSE_{GRW}$	309	194
$WMSE_{GLS}$	271	124

Appendix D Data for real data cases

Table D1. 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 D2. 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 D3. 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 D4. 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