

A TIME WARPING MODEL FOR SEASONAL DATA WITH APPLICATION TO AGE ESTIMATION FROM NARWHAL TUSKS

BY LARS N. REITER ^{1,a}, MADS PETER HEIDE-JØRGENSEN ^{2,d}, EVA GARDE ^{2,c},
ADELINE SAMSON^{3,e}, SUSANNE DITLEVSEN ^{1,b}

¹*Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark*, ^alrn@math.ku.dk;
^bsuzanne@math.ku.dk

²*Greenland Institute of Natural Resources*, ^cevga@mail.ghsdk.dk; ^dmhj@mail.ghsdk.dk

³*Jean Kuntzmann Laboratory, Université Grenoble Alpes*, ^eadeline.leclercq-samson@univ-grenoble-alpes.fr

Signals with varying periodicity frequently appear in real-world phenomena, necessitating the development of efficient modelling techniques to map the measured nonlinear timeline to linear time. Here we propose a regression model that allows for a representation of periodic and dynamic patterns observed in time series data. The model incorporates a hidden strictly positive stochastic process that represents the instantaneous frequency, allowing the model to adapt and accurately capture varying time scales. A case study focusing on age estimation of narwhal tusks is presented, where cyclic element signals associated with annual growth layer groups are analyzed. We apply the methodology to data from one such tusk collected in West Greenland and use the fitted model to estimate the age of the narwhal. The proposed method is validated using simulated signals with known cycle counts and practical considerations and modelling challenges are discussed in detail. This research contributes to the field of time series analysis, providing a tool and valuable insights for understanding and modeling complex cyclic patterns in diverse domains.

1. Introduction. Time series data with cyclic patterns are commonly encountered in real-world applications, ranging from environmental, ecological, physical and physiological studies to economic forecasting (Glass and Mackey, 1988; Shumway and Stoffer, 2017). Understanding and accurately modelling such patterns are crucial for gaining insights into underlying dynamics and making reliable conclusions.

In some applications, the measurements are taken along a segment, where each measured point at some distance corresponds to an unknown time. Key examples are tree rings, sea sediment cores, ice cores, hair, ear stones (Monserud and Marshall, 2001; Jouzel and Masson-Delmotte, 2010; Charles et al., 1996; Hüsy et al., 2021a; Mosbacher et al., 2016), or in this case, narwhal tusks. A given variable is measured along a transect line at equidistant spatial points, and the oscillation patterns in the data are hypothesized to reflect yearly variations. The measurements can then be dated, i.e., distances can be transformed to a timeline.

A significant challenge presented by such datasets relates to the unknown nonlinear relation between distance and the underlying timeline, complicating the precise mapping of the signals to their actual temporal positions. In cases where reference signals are available or when the primary objective of the study centers on signal alignment rather than precise temporal mapping, methodologies like Graphical and Dynamic Time Warping (Jeong, Jeong and Omitaomu, 2011; Salvador and Chan, 2007; Wang et al., 2016) offer effective solutions for tasks involving pattern matching and classification. However, when reference signals are absent, and the exact mapping from spatial position to real time is of particular concern, the literature on time warping frameworks targeting cyclic signals is limited. One positive example is Dahlhaus et al. (2017), who have developed a particle-EM variant that recovers the hidden phase and cycle shape jointly. Another example is Napolitano and Gardner (2016),

48 who propose two ways to estimate the time-warping function in signals with irregular cy-
 49 cles, either directly estimating the warping functions or by using a probabilistic approach to
 50 construct a suited estimator.

51 This paper presents a novel approach for modelling noisy non-stationary cyclic signals
 52 with changing periodicity and amplitude. As opposed to the traditional seasonal decomposi-
 53 tion (Dagum, 2013; Cleveland et al., 1990) where a signal is split into trend, seasonal, cyclic,
 54 and noise components, our model incorporates the seasonal and cyclic effects, and combine
 55 them into a single component. A key innovation in our model is the incorporation of a hidden
 56 strictly positive stochastic process, specifically the square-root diffusion process or CIR
 57 process as it is also known in literature (Cox et al., 1985). The integral of this process is
 58 strictly increasing and allow us to model the material growth, by modulating the frequency
 59 while also respecting the direction of time. This provides realism in capturing the varying
 60 periodicity observed in real-world data.

61 Estimation within our model is achieved using a stochastic variant of the EM algorithm,
 62 where we repeatedly filter the unobserved square-root diffusion process using sequential
 63 monte carlo (SMC) and improve our estimates of the parameters in each new run. This pro-
 64 cedure is explained in detail in section 3. While our model has applications across various
 65 domains, we focus here on its use in a ecological context. In this field, time series data are
 66 commonly associated with annual cycles, such as the growth patterns of organisms where
 67 various exogenous and endogenous factors can influence the speed of growth (Stounberg
 68 et al., 2022; Amais et al., 2021; Wooller et al., 2021). Our method can effectively capture the
 69 changing frequencies and provide more accurate predictions for these cyclic patterns. To see
 70 this, we will consider data obtained from narwhal tusks. The narwhal (*Monodon monoceros*)
 71 is a toothed whale, endemic to the Arctic regions. Growth layer groups (GLGs) are deposited
 72 in the narwhal tusk annually (Garde et al., 2024), which offers insight into the growth history
 73 of the whale and also a simple way to assess the age. Unlike tree rings, however, these growth
 74 layers are usually more faint and the transition between layers is more subtle, which makes
 75 manual counting difficult. Nonetheless, manual counting remains one of the oldest and most
 76 widely used techniques for assessing an animal’s age (Read, Hohn and Lockyer, 2018; Laws,
 77 1952). In our case study (Section 5) we demonstrate the utility of our model, by presenting
 78 an alternative age estimation technique, where we focus on the mineralized elements (specif-
 79 ically the isotope Barium-137) in the tusks, which is deposited in the same annual pattern
 80 as the growth layers. For this data, decoding the hidden stochastic process present in our
 81 model, corresponds to uncovering the process that drives the growth of the tusk. Unlocking
 82 the growth process is of scientific interest, because it reveals life chapters of accelerating or
 83 declining growth, which may reflect underlying biological or environmental changes (Dietz
 84 et al., 2021). From this process, we are then able to estimate the number of cycles, thereby
 85 also the age of the narwhal. We conclude with a discussion of the model in section 6. In the
 86 next section, we will outline the model, and in section 4 we perform a small simulation study.

87 **2. Time warping model.** Consider a regression model with additive noise,

$$(1) \quad y_i = f(x_i, \theta) + \epsilon_i, \quad i = 0, 1, \dots, n; x_i = i\Delta,$$

88 where y_i is the measured variable (e.g., a chemical element) at spatial location x_i and $\Delta =$
 89 $x_{i+1} - x_i$ is the distance between measurements. The observation interval is thus $[0, n\Delta]$. The
 90 function f is detailed below and depends on parameters θ . The error terms ϵ_i are assumed
 91 independent and normally distributed with mean 0 and variance σ^2 . The variable x is the
 92 distance along a cross section of some object of interest, where a specific position corresponds
 93 to an unknown time through a monotonic and possibly non-linear function g ,

$$(2) \quad t_i = g(x_i, \theta).$$

94 We impose that $g(x, \theta)$ is a strictly increasing function with respect to x , to ensure that time
 95 is not going backwards.

96 For the signal, we assume the form

$$(3) \quad f(x, \theta) = A \sin(g(x, \theta) + b) - B \cos(2g(x, \theta) + 2b), \quad g(0, \theta) = 0,$$

97 where $A > 0$ and $B > 0$ determine the amplitudes and b is the phase at $x = 0$. We require
 98 $A > B$ for identifiability, and also that A/B is not too small - see supplementary material for
 99 a detailed discussion of both criteria. The signal (3) is composed of a low-frequency sine wave,
 100 and a fast-frequency cosine wave: it traverses two cycles for each cycle in the sine wave. The
 101 sine wave may, for instance, represent annual cycles, while the fast cosine wave captures
 102 the summer-winter variations. See Figure 1 for a illustration. A generalization to higher fre-
 103 quency terms, for example a model with semi-annual effects, is discussed in supplementary
 104 material.

105 There is no trend in our model (1), and we assume it is removed in a prepossessing step,
 106 if necessary. For the signals used in our case study, no apparent trend was detected, and this
 107 paper will not address detrending techniques, of which there exists several choices [insert
 108 references]. We do however acknowledge the potential for varying amplitudes, and a level
 109 difference relative to $f(x, \theta) = 0$. This must be addressed before estimation, thus in section
 110 3.1 we illustrate our preferred way of bringing our data to a form compatible with the model.

111

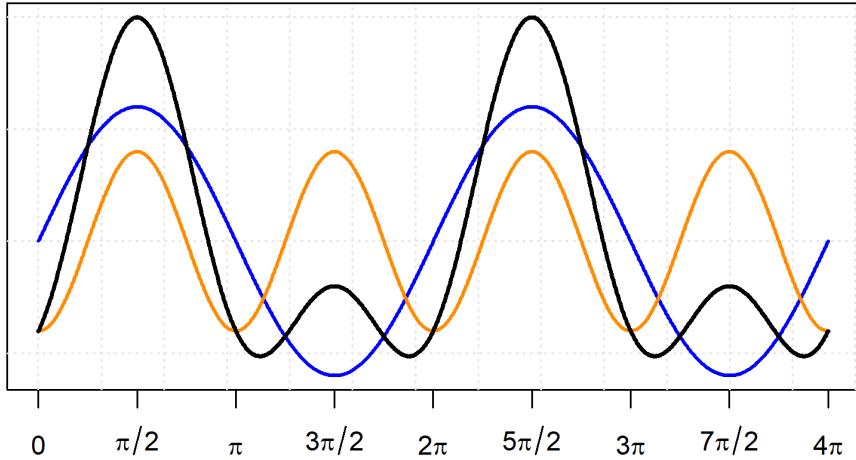


Fig 1: **Signal composition.** Slow sine wave (blue) and fast cosine wave (orange) as well as their composition $f(x, \theta)$ (black). Simulation details: $A = 0.6$ and $B = 0.4$ and no phase offset $b = 0$. Here the growth process is non-stochastic and linear $g(x) = x$.

112 In the linear and deterministic case where $g(x, \theta) = ax$ the cycles are regularly occur-
 113 ring (Figure 1). This is unrealistic in most settings, where the length of one cycle will typi-
 114 cally vary due to environmental factors, leading to quasi-periodic dynamics. To describe this,
 115 we allow the instantaneous growth rate to be stochastic. We denote the unknown stochas-
 116 tic transformation $g(x, \theta)$ as *the growth process*. Let $g(x, \theta) = \int_0^x \xi_s ds$ where $(\xi_x)_{x \geq 0}$ is a
 117 positive stochastic process described by unknown parameters θ . For $(\xi_x)_{x \geq 0}$, we propose
 118 the square root diffusion process, also known as the Cox-Ingersoll-Ross (CIR) process (Cox
 119 et al., 1985), satisfying the stochastic differential equation (SDE)

$$(4) \quad d\xi_x = -\beta(\xi_x - a)dx + \omega \sqrt{\xi_x} dW_x, \quad \xi_0 = 0,$$

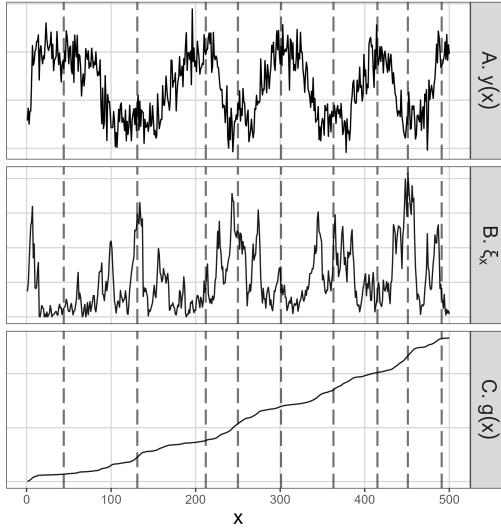


Fig 2: Example of a simulated signal.
 The dataset was simulated using $n = 500$, $\Delta = 1$, $\beta = 0.07$, $\sigma = 0.3$, $A = 0.6$, $B = 0.4$, $a = 0.05$, $b = \frac{\pi}{20}$, $\omega^2 = 0.064$.
 A. Simulated signal with measurement noise $y(x)$. B. Simulated hidden process $(\xi_x)_{x \geq 0}$. C. Simulated growth process $g(x)$. Black dashed lines are signal peaks.

where W_x is a standard Wiener process. The hidden process $(\xi_x)_{x \geq 0}$ describes instantaneous frequency deviations from the baseline frequency $a/2\pi$. The parameter β determines the rate of adjustment to the baseline and ω scales the noise level. This is an ergodic process with stationary distribution the Gamma distribution with shape parameter $2\beta a/\omega^2$, scale parameter $\omega^2/(2\beta)$, mean a and variance $a\omega^2/2\beta$, provided that $\beta, a, \omega > 0$, and $2\beta a \geq \omega^2$ (Ditlevsen, Rubio and Lansky, 2020). These parameter restrictions ensure that every realization of the stochastic process with dynamics described by eq. (4) is strictly positive (except at $x = 0$) with long-term mean a , leading to a strictly increasing process $g(x, \theta)$. Define also the one-lag autocorrelation of the growth process $\rho = \exp(-\Delta\beta)$ and the stationary variance $\gamma^2 = a\omega^2/2\beta$. The exact transition density, i.e., the distribution of $\xi_{x+\Delta}$ given ξ_x follows a non-central χ^2 distribution scaled by a constant factor. Specifically, $\xi_{x+\Delta} | \xi_x \sim Z_x/2c$, where $Z_x \sim \chi_\nu^2(\lambda_x)$ is non-central chi-squared distributed with $\nu = \frac{4a\beta}{\omega^2}$ degrees of freedom and non-centrality parameter $\lambda_{x+\Delta} = 2c\rho\xi_x$. The multiplicative constant is $\frac{1}{2c}$, where $c = \frac{2\beta}{(1-\rho)\omega^2}$ (Ditlevsen and Lansky, 2006).

Our main goal is to infer the growth process $g(x, \theta)$ in order to date the observations y_i to $t_i = g(x_i, \theta), i = 0, \dots, n$. From $g(x, \theta)$ we are also able to predict the number of cycles in the signal:

$$(5) \quad \text{Number of Cycles} = \frac{g(x_n, \theta)}{2\pi}.$$

The parameters of the model are summarized in Table 1. In Figure 2 we illustrate a signal simulated from the model (1) and (3), along with the embedded growth process (4).

The next section is devoted to explaining the method, by which we estimate within the model. The method has three phases, a pre-processing phase described in section 3.1, a initialization phase described in section 3.5 and our main algorithm which is described in sections 3.2, 3.3 and 3.4. We end with section 3.6 and 3.7 on how to validate our model and quantify uncertainty using bootstrapping.

3. Estimation.

3.1. *Preprocessing: Level adjustment and amplitude normalization.* To ensure that our signal is centered around $y = 0$ (i.e. has no level), we subtract the mean \bar{y} . We also require constant amplitudes. To this end, we will assume that the signal (roughly) concentrates

TABLE 1
Model parameters. Name, support and interpretation.

Parameter	Support	Interpretation
A	$(0, \infty)$	Amplitude for the sine wave.
B	$(0, \infty)$	Amplitude for the cosine wave.
b	$(0, 2\pi)$	Phase-offset of the signal.
σ^2	$(0, \infty)$	Variance of measurements
a	$(0, \infty)$	Mean of instantaneous growth rate
β	$(0, \infty)$	Adjustment rate of instantaneous growth rate
ρ	$(0, 1)$	One-lag autocorrelation of instantaneous growth rate, $\rho = \exp(-\Delta\beta)$
ω^2	$(0, 2a\beta)$	Infinitesimal variance of instantaneous growth rate
γ^2	$(0, a^2)$	Stationary variance of instantaneous growth rate, $\gamma^2 = a\omega^2/2\beta$

148 around a narrow band of frequencies. We will also assume that the amplitude process is
 149 slowly evolving. Under these assumptions, we will use the Hilbert Transform to compute
 150 the envelope of the signal (Osgood, 2007, p. 318–320). We will then apply a rolling average
 151 with a window size of approximately 10% to the resulting envelope to ensure a smooth en-
 152 velope, which we denote $z_i = z(x_i)$. In summary, the signal is zero adjusted and normalized
 153 as $y_i \leftarrow \frac{y_i - \bar{y}}{z_i}$. The normalization ensures that the amplitudes are not only constant, but also
 154 that the maximum peak value is unity, which in our model implies

$$(6) \quad \max(f(x, \theta)) = A + B = 1.$$

155 3.2. *Principle of the SAEM algorithm.* Denote the unknown parameters by $\theta :=$
 156 $(A, B, a, b, \sigma^2, \beta, \omega^2)$. Here we propose an algorithm for estimating θ from observations
 157 $y = (y_0, y_1, \dots, y_n)$.

158 The likelihood of the data y_i is not explicit as model (3) depends on the hidden stochastic
 159 process $(\xi_x)_{x \geq 0}$. A common approach for estimating in models with latent and unobserved
 160 data is the *expectation-maximization* (EM) algorithm (Dempster, Laird and Rubin, 1977). In
 161 this iterative procedure at iteration m , the E-step is carried out by evaluating the expected
 162 log-likelihood given data and current estimates $\hat{\theta}^{m-1}$:

$$(7) \quad \begin{aligned} Q_m(\theta) &\stackrel{\text{def}}{=} Q(\theta | \hat{\theta}^{m-1}) \\ &\stackrel{\text{def}}{=} \mathbb{E}_{\xi | y, \hat{\theta}^{m-1}} [\ell(y, \xi; \theta) | (y; \hat{\theta}^{m-1})], \end{aligned}$$

163 where ℓ is the complete log-likelihood of y and the hidden process $\xi := \xi_{0:n} = (\xi_0, \dots, \xi_n)$.
 164 The expectation is taken wrt. the probability distribution $p(\xi | y, \hat{\theta}^{m-1})$. In the M-step, $Q_m(\theta)$
 165 is then maximized in order to obtain new estimates $\hat{\theta}^m$:

$$(8) \quad \hat{\theta}^m = \arg \max_{\theta} Q(\theta | \hat{\theta}^{m-1}).$$

166 The complete log-likelihood is explicit and given by

$$\ell(y, \xi; \theta) = \sum_{i=0}^n \log p(y_i | \xi_i) + \sum_{i=1}^n \log p(\xi_i | \xi_{i-1}) + \log p(\xi_0)$$

$$(9) \quad = -\frac{1}{2\sigma^2} \sum_{i=0}^n (y_i - f(x_i, \theta))^2 - \frac{n+1}{2} \log \sigma^2 + \sum_{i=1}^n \log p(\xi_i | \xi_{i-1}, \theta) + \log p(\xi_0).$$

167 As explained in section 2, the transition density $p(\xi_i | \xi_{i-1}, \theta)$ is a non-central χ^2 -distribution.
168 However, the conditional distribution of ξ given y is not explicit due to the non-linear
169 nature of the regression function and the non-central χ^2 distribution of the transition den-
170 sity. Thus, we are unable to perform the E-step in eq. (7). Instead we use (a variant of) the
171 *Stochastic Approximation EM* (SAEM) algorithm (Delyon, Lavielle and Moulines, 1999a;
172 Jank, 2006). In this procedure, the E-step is replaced by two steps, a simulation (S) step
173 and a stochastic approximation (SA) step. In the simulation step, we use a Sequential Monte
174 Carlo (SMC) (Del Moral, Doucet and Jasra, 2006a; Gordon, Salmond and Smith, 1993) sam-
175 pler to draw the non-observed data ξ^m , conditionally on $\hat{\theta}^{m-1}$. We then apply a stochastic
176 approximation to equation (7):

$$(10) \quad \tilde{Q}_m(\theta) = \tilde{Q}_{m-1}(\theta) + \alpha_m (\ell(y, \xi^m; \hat{\theta}^{m-1}) - \tilde{Q}_{m-1}(\theta)),$$

177 where α_m is a decreasing sequence of positive numbers quantifying the memory in the ap-
178 proximation process, fulfilling $\sum_m \alpha_m = \infty$ and $\sum_m \alpha_m^2 < \infty$.

This algorithm can be simplified when the complete likelihood belongs to the curved exponential family, which then reduces to approximating the minimal sufficient statistics of the model during the E step, while the M step is explicit through the sufficient statistics. Nevertheless, the distribution $p(\xi_i | \xi_{i-1})$ does not belong to a curved exponential family and we have therefore no sufficient statistics. In that case, each iteration of the EM algorithm solves

$$\partial_\theta \tilde{Q}_m(\theta) = 0.$$

179 However, solving the score function for the CIR process can be numerically unstable
180 (Sørensen, 2012). We instead use martingale estimating functions (Bibby and Sørensen,
181 1995), leading to an approximated version of the EM algorithm.

182 *3.3. Martingale estimating function and SAEM-variant.* At iteration m of the algorithm,
183 given the current estimate ξ^{m-1} of the hidden process ξ obtained using SMC (section 3.4),
184 the M step is updated by solving the score function, i.e., by differentiating the complete log-
185 likelihood $\ell(y, \xi^m, \theta)$ with respect to θ , the score function being approximated during the SA
186 step.

187 Since the parameters a, β and ω of the hidden process ξ are difficult to estimate from the
188 likelihood, we use a martingale estimating score function (Sørensen, 2012). The estimator is
189 defined directly as the solution to equating the martingale estimation function to 0, mimicking
190 the score equation. This leads to the statistics S_1, S_2, S_3 (Sørensen, 2012, p. 21):

$$(11) \quad S_1 = \frac{\frac{1}{n} \sum_{i=1}^n \xi_{i\Delta}^m / \xi_{(i-1)\Delta}^m - (\frac{1}{n} \sum_{i=1}^n \xi_{i\Delta}^m)(\frac{1}{n} \sum_{i=1}^n (\xi_{(i-1)\Delta}^m)^{-1})}{1 - (\frac{1}{n} \sum_{i=1}^n \xi_{(i-1)\Delta}^m)(\frac{1}{n} \sum_{i=1}^n (\xi_{(i-1)\Delta}^m)^{-1})};$$

$$(12) \quad S_2 = \frac{1}{n} \sum_{i=1}^n \xi_i^m + \frac{S_1}{n(1-S_1)} (\xi_{n\Delta}^m - \xi_0^m);$$

$$(13) \quad S_3 = \frac{\sum_{i=1}^n (\xi_{i-1}^m)^{-1} (\xi_{i\Delta}^m - \xi_{(i-1)\Delta}^m) S_1 - S_2(1-S_1))^2}{\sum_{i=1}^n (\xi_{(i-1)\Delta}^m)^{-1} ((\frac{1}{2} S_2 - \xi_{(i-1)\Delta}^m) S_1^2 - (S_2 - \xi_{(i-1)\Delta}^m) S_1 + \frac{1}{2} S_2)}.$$

191 The parameter σ^2 is easier to obtain, as the derivative of the complete log-likelihood leads
192 to the following statistic:

$$(14) \quad S_4 = \frac{1}{n} \sum_{i=0}^n (y_i - f(x_i, \hat{\theta}^{m-1} | \xi^m))^2,$$

193 where conditioning on the hidden process ξ^m imply a growth process given by $g(x_i, \hat{\theta}^{m-1}) =$
 194 $\int_0^{x_i} \xi_s^m ds$. These statistics are then approximated by the stochastic approximation scheme, as
 195 in the standard SAEM. Following Ditlevsen and Samson (2014); Lavielle (2014), we choose
 196 the form

$$(15) \quad \alpha_m = \begin{cases} 1 & m \leq m_0 \\ \left(\frac{99}{100}\right)^{(m-m_0)} & m > m_0 \end{cases},$$

197 for the Stochastic Approximation sequence in eq. (10), where m_0 determines the iteration
 198 step wherafter memory from the previous step is included. We obtain approximations of the
 199 statistics, for $k = 1, \dots, 4$:

$$(16) \quad s_k^{(m)} = s_k^{(m-1)} + \alpha_m(S_k - s_k^{(m-1)}).$$

200 We then update the parameter estimates in terms of these statistics

$$\hat{\rho}^m = s_1^{(m)}; \quad \hat{\beta}^m = -\frac{1}{\Delta} \log(\hat{\rho}^m); \quad (\hat{\omega}^m)^2 = s_3^{(m)}; \quad (\hat{\sigma}^m)^2 = s_4^{(m)};$$

$$\hat{a}^m = s_2^{(m)} \mathbb{1}_{\{2\hat{a}^0 > s_2^{(m)} > \frac{1}{2}a^0\}} + \hat{a}^{m-1} \left(1 - \mathbb{1}_{\{2\hat{a}^0 > s_2^{(m)} > \frac{1}{2}a^0\}}\right).$$

201 Since the number of cycles is proportional to a , we ensure that the process does not mistake
 202 random fluctuations for cycles by restricting estimates of a to be between half and double of
 203 our initial estimate (see section 3.5). If this condition is fulfilled for the statistic associated
 204 with a we update, otherwise the estimate from the previous step is kept.

205 To update the amplitude parameters A and B , consider that after normalization, we have
 206 eq. 6, which implies $B = 1 - A$. With this substitution eq. (3) takes the form

$$(17) \quad f(x_i, \theta) = Aw_i + v_i,$$

207 where $w_i = \sin(g(x_i) + b) + \cos(2g(x_i) + 2b)$ and $v_i = -\cos(2g(x_i) + 2b)$. Equation (17)
 208 allow for a linear regression with constraints $1 = A + B = A > B = 1 - A$, which imply
 209 $A \in (\frac{1}{2}, 1)$. In conclusion

$$(18) \quad \hat{A}^m = \arg \min_{A \in (\frac{1}{2}, 1)} \sum_{i=1}^n (y_i - Aw_i - v_i)^2$$

$$(19) \quad = \min \left(1, \max \left(\frac{1}{2}, \frac{\sum_{i=1}^n w_i(y_i - v_i)}{\sum_{i=1}^n w_i^2} \right) \right),$$

210 where $\frac{\sum_{i=1}^n w_i(y_i - v_i)}{\sum_{i=1}^n w_i^2}$ is the least squares estimator in the unconstrained problem. From this
 211 we also update $\hat{B}^m = 1 - \hat{A}^m$.

212 It turns out that the intial phase b and the growth process $g(x)$, and consequently ξ , are
 213 not identifiable individually, and must be estimated jointly. In section 3.4, we go into detail
 214 on how we can modify the sequential monte carlo sampler to simultaneously sample \hat{b}^m in
 215 iteration m , by making a grid of plausible b candidates, and run a single SMC filter for each
 216 b . We call this SMC plus. We refer to supplementary material for a brief example showing
 217 why sequential estimation of $g(x)$ and b is not feasible.

218 The SAEM algorithm iterates until some stopping criteria is met. Several rules exist for de-
 219 termining when convergence can be assumed (Sammaknejad, Zhao and Huang, 2019; Dhull
 220 and Kumar, 2022; Lixoft SAS, 2024; Delyon, Lavielle and Moulines, 1999b). One common
 221 rule is to ensure that the maximum of the relative differences of all parameters between con-
 222 secutive iterations falls below some threshold to ensure that the SAEM estimates have sta-
 223 bilized. We took the threshold to be 10^{-4} following Searle, Cassella and McCullouch (1992).

224 We also made a simulation study to investigate the sensitivity to this choice of threshold,
 225 which can be found in supplementary material.

226 To summarize, our algorithm is the following. We first use SMC to filter multiple trajec-
 227 tories of ξ called particles, using the exact transition density. This is repeated over a grid of
 228 candidates for the phase parameter b . This algorithm is detailed in section 3.4. Once the the
 229 latent process is filtered, we proceed to the stochastic approximation step. In the final maxi-
 230 mization (M) step, we update the estimates $\hat{\theta}^m$. This procedure reiterates until some stopping
 231 criteria is met. The basic estimation procedure, tracing the steps above, is summarized in
 232 Algorithm 1. Later we generate confidence intervals for the parameters and other derived
 233 quantities. This subsequent step is summarized in Algorithm 4.

Algorithm 1: Parameter estimation using SAEM and martingale estimating functions

Data: $(X, Y) := \{(x_1, y_1), \dots, (x_n, y_n)\}$

Result: $\xi, \hat{\theta} := (\hat{A}, \hat{B}, \hat{b}, \hat{\sigma}^2, \hat{\rho}, \hat{\omega}^2)$

(1) Initialize parameters $\hat{\theta}^0$ and the unknown process ξ^0 (see section 3.5);

$m \leftarrow 1$;

while "Stopping criteria" **do**

234 (2) Update \hat{b}^m, ξ^m given $\hat{\theta}^{m-1}$ using SMC plus (Algorithm 2) ;

(3) Do the stochastic approximation of statistics S_1, \dots, S_4 ;

(4) Update $\hat{\sigma}^m, \hat{\rho}^m, (\hat{\omega}^2)^m, (\hat{\sigma}^2)^m$ using the statistics $s_1^{(m)}, \dots, s_4^{(m)}$;

(5) Update \hat{A}^m using eq. (19) and $\hat{B}^m = 1 - \hat{A}^m$;

$m \leftarrow m + 1$

end

235 3.4. *SMC algorithm.* Inside Algorithm 1 we apply SMC to filter out the hidden pro-
 236 cess ξ . The SMC algorithm provides a set of n_p particles $(\xi^{(j)})_{j=1,\dots,n_p}$ and weights
 237 $(W^{(j)})_{j=1,\dots,n_p}$ approximating the conditional smoothing distribution $p(\xi_0, \dots, \xi_n | y_0, \dots, y_n; \theta)$
 238 (Del Moral, Doucet and Jasra, 2006b; Doucet, De Freitas and Gordon, 2001). The SMC algo-
 239 rithm is an iterative algorithm. At each iteration in time, particles are sampled from a proposal
 240 distribution denoted $q(\xi_i | \xi_{i-1}, y_i, y_{i-1})$. To ease notation, we denote $\xi_{0:i} = (\xi_0, \dots, \xi_i)$ and
 241 likewise for $y_{0:i}$.

242 We choose the exact transition density $p(\xi_i | \xi_{i-1})$ introduced in section 2 as proposal q .
 243 This choice gives a simplified weight equal to

$$(20) \quad W = \frac{p(y_i, \xi_{0:i} | y_{0:i-1}, \theta)}{q(\xi_i | y_i, \xi_{0:i-1}, \theta)} = p(y_i | \xi_i, \theta) \sim \mathcal{N}(f(x_i, \theta), \sigma^2).$$

244 The basic SMC algorithm is presented as pseudocode in Algorithm 2 for a given value θ .
 245 It provides an empirical measure which is an approximation to the smoothing distribution
 246 $p(\xi_{0:n} | y_{0:n}, \theta)$. A draw from this empirical distribution can be obtained by sampling an index
 247 j from a multinomial distribution with probabilities $(W_n^1, \dots, W_n^{n_p})$.

248 As mentioned in the previous section, the parameter b which is the phase at $x = 0$ and the
 249 growth process $g(x)$ modeling the accumulated phase are intrinsically connected, and needs
 250 to be handled jointly. We therefore propose to use SMC with a grid search adaption. We call
 251 this SMC plus, and is described in Algorithm 3. SMC plus works by considering equally
 252 spaced candidates of b , centered around the current estimate of b , and run the basic SMC
 253 with each candidate. We will consider $G = 10$ equidistant values. We also need to specify the

254 range $2w$ around our current estimate, that we would like to search. Initially, we set half the
 255 range equal to $w = \pi$ but as the Stochastic Approximation sequence in eq. (15) decreases, so
 256 will w , resulting in a grid that collapses around the current estimate of b . Our initial estimate
 257 will be $b^0 = \pi$, and since the initial width is π we will search the entire unit circle.

Algorithm 2: Sequential Monte Carlo (SMC)

Data: $(X, Y), \theta = (A, B, a, b, \sigma^2, \beta, \rho, \omega^2)$
Result: $(\xi^{(j)} = (\xi_0^{(j)}, \dots, \xi_n^{(j)}))_{j=1, \dots, n_p}$

- (1) Initialize $n_p \times n + 1$ matrices ξ and weights W , where n_p is the number of particles ;
- (2) For $j = 1, \dots, n_p$: set $\xi_0^{(j)} = a$ and $W_0^{(j)} = 1/n_p$;
- for** $i = 1$ **to** n **do**
 - (3) Set $\lambda_i = 2\rho c \xi_{i-1}^{(j)}$ for $j = 1, \dots, n_p$;
 - (4) Update particles: $\xi_i^{(j)} \sim \frac{1}{2c} \chi_\nu'^2(\lambda_i)$ for $j = 1, \dots, n_p$;
 - (5) Compute weights: $W_i^{(j)} = p(y_j | \xi_i^{(j)}, \theta)$ for $j = 1, \dots, n_p$;
 - (6) Normalize weights: $W_i^{(j)} = W_i^{(j)} / \sum_{i=1}^{n_p} W_i^{(j)}$ for $j = 1, \dots, n_p$;
 - (7) Resample particles by drawing n_p indices from the set $\{1, \dots, n_p\}$ with corresponding probabilities $W_i^{(1)}, \dots, W_i^{(n_p)}$. Denote the realizations $\{I_1, \dots, I_{n_p}\}$ and set $\xi_{0:i}^{(j)} := \xi_{0:i}^{(I_j)}$;
- end**

Algorithm 3: SMC plus: Grid Search

Data: $(X, Y), \theta = (A, B, b, a, (\sigma^2), \beta, \rho, (\omega^2))$
Result: $\hat{b}, \hat{\xi}$

- (0) Set number of candidates $G = 10$ (grid granularity) ;
- (1) Define half-width
$$w = \begin{cases} \pi, & m \leq m_0, \\ \pi \times 0.99^{m-m_0}, & m > m_0. \end{cases}$$
- ;
- (2) Form grid of G phase zero candidates:
$$b_j = (b - w + (j-1)\frac{2w}{G-1}) \bmod 2\pi, \quad j = 1, \dots, G.$$
- ;
- for** $j = 1$ **to** G **do**
 - (3) Run SMC (Algorithm 2) with θ with b_j replacing b , to obtain $\hat{\xi}^{(j)}$. ;
 - (4) Given θ , with b_j replacing b , and $\hat{\xi}^{(j)}$ compute the log likelihood using eq. (9) and denote it L_j
 - ;
- end**
- (5) Set $(\hat{b}, \hat{\xi}) = (b_{j^*}, \xi^{(j^*)})$ where $L_{j^*} \geq L_j$ for $j = 1, \dots, G$.

258 In the next two sections we outline the initialization details of step (1) of Algorithm 1 and
 259 discuss the bootstrap procedure used to estimate the confidence intervals of the estimates. We
 260 end with a section on model validation.

261 3.5. *Parameter initialization.* Since the loss function of the non-linear model (3) will
 262 likely exhibit many local maxima due to the trigonometric functions, proper initialization
 263 will assist the algorithm in convergence towards a global optimum. Obtaining good initial es-
 264 timates is a non-trivial task, and typically involves heuristic approaches. Here we present our
 265 chosen initialization, denoted $\hat{\theta}^0$. In step (1) and (2) we will use a smoothed signal, denoted
 266 \tilde{y} . The signal was smoothed using a moving average with a window spanning 2% of the data,
 267 implemented with base R functions.

268

269 **(1) Amplitudes A, B and phase offset b :** One can show (see supplementary material),
 270 that the signal (3) attains maximum and minimum values equal to $f_{max} := A + B$ and
 271 $f_{min} := -B - A^2/(8B)$ assuming, in case of the minimum, that $\frac{A}{4B} \leq 1$. We don't know
 272 f_{max} and f_{min} but we can approximate them by considering the smoothed signal \tilde{y} , such
 273 that $f_{max} \approx \max(\tilde{y}) = \tilde{y}_{max}$ and $f_{min} \approx \min(\tilde{y}) = \tilde{y}_{min}$. Solving for A and B , we obtain
 274 $\hat{B} = \frac{1}{9} (\tilde{y}_{max} - 4\tilde{y}_{min} \pm \sqrt{(\tilde{y}_{max} - 4\tilde{y}_{min})^2 - 9\tilde{y}_{max}^2})$ and $\hat{A} = \tilde{y}_{max} - \hat{B}$ which offers two
 275 possible solutions, (\hat{A}_1, \hat{B}_1) and (\hat{A}_2, \hat{B}_2) . Define the set of amplitude candidates

$$(21) \quad S = \{(A, B) \in ((\hat{A}_1, \hat{B}_1), (\hat{A}_2, \hat{B}_2), (y_{max}, 0)) : A > B\}.$$

276 We added $(A, B) = (y_{max}, 0)$ to ensure this is non-empty. We initialize the amplitudes
 277 \hat{A}^0, \hat{B}^0 by drawing a random pair from S . The phase offset is initialized to $\hat{b}^0 = \pi$ (halfway
 278 in a cycle) to ensure that the initial grid of b candidates span the entire unit circle in the first
 279 run of the algorithm. See also section 3.4.

280

281 **(2) Measurement error:** σ^2 . The variance of the measurement noise is approximated
 282 from the smoothed signal \tilde{y} , setting $(\hat{\sigma}^2)^0 = \frac{1}{n} \sum_{i=1}^n (y(x_i) - \tilde{y}(x_i))^2$.

283

284 **(3) Hidden process parameters:** a, β, ω^2 . With initial estimates for all other parameters,
 285 we now run SMC (Algorithm 2) with a fixed number of particles n_p , but where each particle
 286 is initialized with a random pair $(a, \beta, \omega^2) \in (\frac{2\pi C_{min}}{n\Delta}, \frac{2\pi C_{max}}{n\Delta}) \times (0.01, 0.5) \times (0.01, 0.3)$
 287 obtained from rejection sampling under the feller constraint $2a\beta > \omega^2$. Here, $C_{min/max}$ is
 288 our prior belief on the lower/upper number of cycles in the signal. We resample among the
 289 particles at each step and select a final candidate based on the final weights. The parameters
 290 associated with the selected particle define our initial estimates $(\hat{a}^0, \hat{\beta}^0, (\hat{\omega}^2)^0)$.

291

292 3.6. *Bootstrapping.* Standard errors and confidence intervals for all parameters esti-
 293 mated using SAEM (Algorithm 1) are computed using residual bootstrapping. The general
 294 procedure is described in Algorithm 4.

Algorithm 4: Residual Bootstrapping (RB)

Data: Observations $y = (y_1, \dots, y_n)$ and fitted values $\hat{y} = (\hat{y}_1, \dots, \hat{y}_n)$

Result: Vector of estimates $(\hat{\theta}_j)_{j=1,\dots,M}$

$j \leftarrow 1$;

(1) Set number of bootstrap replications $M \leftarrow M_0$;

while $j < M$ **do**

(2) Compute residuals $r_i = y_i - \hat{y}_i$;

(3) Let $r_i^{(j)}$ be a random sample from the set of residuals $\{r_1, r_2, \dots, r_n\}$;

(4) Construct replicate signal as $y_i^{(j)} = \hat{y}_i + r_i^{(j)}$;

(5) Estimate $\hat{\theta}_j$ using Algorithm 1 on the replicate $(y_1^{(j)}, y_2^{(j)}, \dots, y_n^{(j)})$;

$j \leftarrow j + 1$

end

296 3.7. *Model validation.* To assess the quality of our model, we consider the raw residuals
297 defined as

$$(22) \quad \text{res}_i = y_i - \hat{y}_i,$$

298 where \hat{y}_i is the fitted value of observation i . We use the (raw) residuals to validate the mean
299 structure and homoscedasticity assumption using the residual plot. Additionally, we compare
300 the (raw) empirical residuals, equation (22), with the theoretical normal distributed errors.
301 This allows us to assess how well the model's residuals reflect the variability and distribu-
302 tional assumptions implied by the model.

303 To address the estimator variability and bias, we computed the (vector of) relative differ-
304 ence(s) (*estimated parameter – true parameter*)/*true parameter* for all parameters,
305 except for the phase b . For the phase parameter b , we instead computed the smallest abso-
306 lute difference (maximum of π) on the unit circle and divided by π . For the case-study, we
307 applied Algorithm 4 and produced several bootstrap estimates $\tilde{\theta}_j$, $j = 1, \dots, M$. We then
308 considered differences between the estimated and bootstrapped versions: $(\tilde{\theta} - \tilde{\theta}_j)/\tilde{\theta}$. In the
309 simulation study, where parameters θ_m are known apriori for each signal m , we considered
310 the difference between estimated and true parameters: $(\theta_m - \hat{\theta}_j)/\theta_m$.

311 In the next section, we simulate a collection of diverse signals - varying in parameter
312 configurations and lengths - and evaluate both the accuracy of our estimates and the precision
313 with which we recover the number of cycles described by eq. (5).

314 **4. Simulation study.** A total of 800 signals was simulated, with a fixed step size equal to
315 $\Delta x = x_i - x_{i-1} = 1$. Below, we describe how parameters were selected. For now, assume all
316 parameters are known. To obtain the hidden stochastic process, the solution to equation (4),
317 we used the exact transition density (described in section 2) to simulate a trajectory starting
318 with $\xi_0 = 0$. We then obtained $g(x_i) = \sum_{j=1}^i \xi_j \Delta$. Finally, we sampled a signal using (1) and
319 (3).

320 For each simulated signal m , we computed the relative differences between any true pa-
321 rameter θ_m and its estimate $\hat{\theta}_m$ (Figure 3) defined as $(\hat{\theta}_m - \theta_m)/\theta_m$. We included the number
322 of cycles to measure how well the model predicts the number of peaks in the signal. The vio-
323 linplots in Figure 3 show how the estimates deviate from the true parameters relative to their
324 scale. From the figure, we see that ρ and ω^2 displays some variability, but without a promi-
325 nent bias, suggesting that the rate of adjustment and infinitesimal variance of the growth

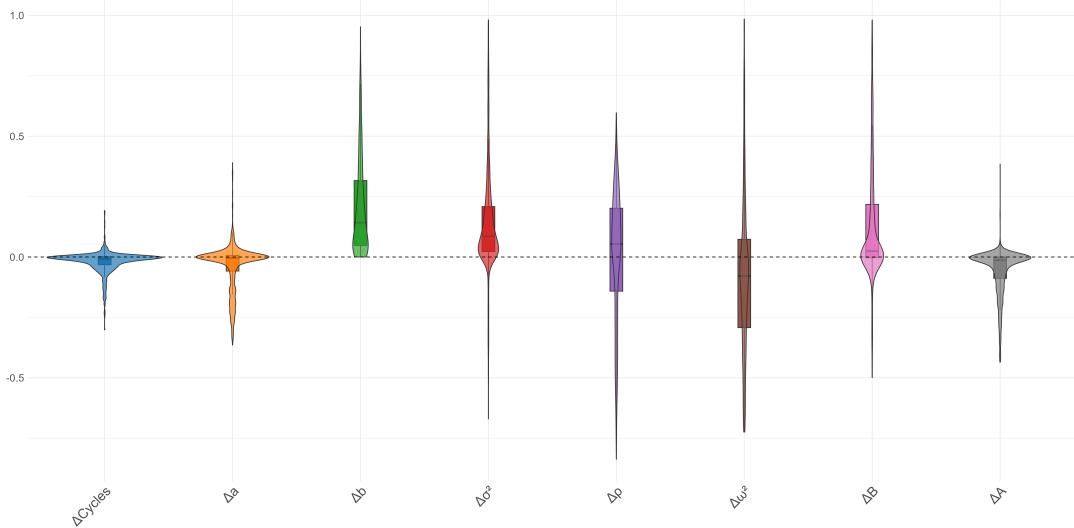


Fig 3: Estimator variability. $M = 800$ simulations equally divided into 4 groups of different length ($N = 200, N = 400, N = 600, N = 800$) and with random parameter configurations (see section 4). The relative difference $(\theta_m - \hat{\theta}_m)/\theta_m, m = 1, \dots, 800$ of parameter estimates $\hat{\theta}_m$ obtained from signals simulated with parameter θ_m . Included are all parameters, and the number of cycles are computed with formula (5). For b we computed the smallest absolute relative difference on the unit circle, since there are no principal direction. This was divided to by π to obtain a relative difference.

process could be hard to identify. The noise variance σ^2 also displays high variability possible because not all minor shifts in the underlying diffusion process ξ is properly captured in the estimation, which then overestimate the magnitude of the residuals used to calculate the estimate of the noise variance. Likewise, if ξ overfit to noise fluctuations, we risk smaller residuals, leading to underestimate of the noise variance. For the remaining parameters, there are no substantial bias, and a more dense concentration of the estimates around their true values. Most importantly, the baseline frequency a and the number of cycles (used as a proxy for how well the model describe the periodic signal) display a low variance and a low bias. One of the key questions, was how good our model was able to estimate the number of cycles defined by eq. (5). This is displayed in Figure 4 which shows that 91% of the estimates fall within 1 cycle from the true number of cycles, and less than 4% deviate by more than 3 cycles, and this mostly happening for signals with many cycles, where the peaks are more tightly concentrated.

We will now explain the simulation details. The 800 signals is divided into 4 equal sized groups, corresponding to short ($N = 200$), medium short ($N = 400$), medium long ($N = 600$) and long ($N = 800$) signals. We define $C_{\min} = 2$ and $C_{\max} = \max(6, N/100 \times 5)$ to be the minimum and maximum allowed cycles. The maximum depends on N , to ensure that we gradually allow for more cycles, as the signal gets longer. If unconstrained, peaks could potentially be so concentrated, that the entire signal cannot be distinguished from white noise. The corresponding upper and lower frequency is then now set to $a_{\min} = \frac{2\pi C_{\min}}{N\Delta}$ and $a_{\max} = \frac{2\pi C_{\max}}{N\Delta}$. The sampling distributions of almost all parameters were uniformly distributed \mathcal{U} as

$$(23) \quad \beta \sim \mathcal{U}(0.01, 3), \quad a \sim \mathcal{U}(a_{\min}, a_{\max}), \quad \omega \sim \mathcal{U}(0.01, 0.3)$$

$$(24) \quad \sigma \sim \mathcal{U}(0.2, 0.6), \quad b \sim \mathcal{U}(0, 2\pi), \quad A \sim \mathcal{U}(0.5, 1),$$

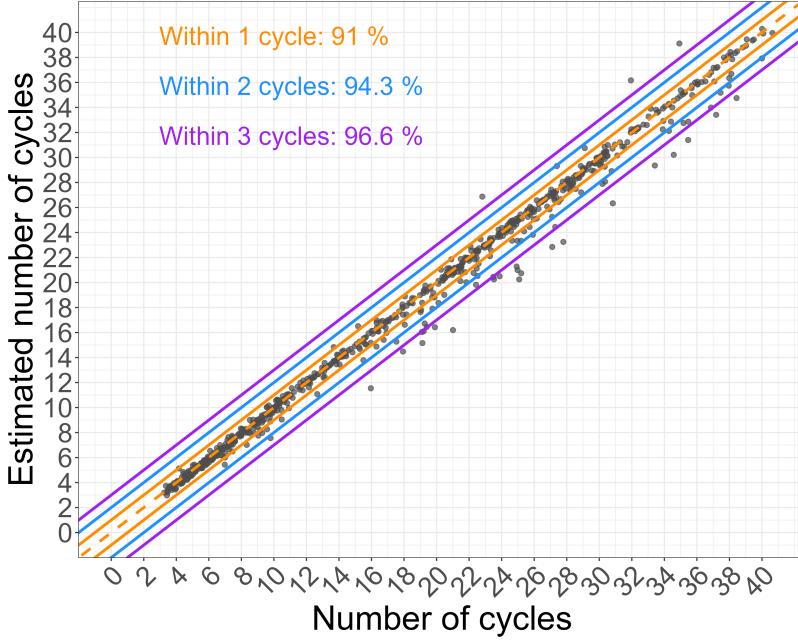


Fig 4: Estimated number of cycles against true number of cycles. Red, blue and green ribbons enclose estimates of number of cycles from simulated data sets with less than 1, 2, and 3 cycles of deviation from the true number of cycles, respectively.

(25)

and from this $B = 1 - A$ and $\rho = \exp(-\Delta\beta)$. The support of the uniform distributions was chosen to best reproduce the real signals of the case study. For the process parameters in eq. (23) we did rejection-sampling, and only accepted the triple (a, β, ω) if they fulfilled the feller condition $2a\beta > \omega^2$. This ensured that $\xi_x > 0$ for all x .

An example of a simulated signal is presented in the top plot of Figure 5, which also displays the model fit. In the second row of Figure 5 we have provided visualizations of the hidden process $(\xi_x)_{x \geq 0}$ associated with the simulated signal, both the actual process and the estimated version. The third row of Figure 5 displays the growth process $\hat{g}(x)$, along with its estimated counterpart, obtained through the integration of $(\hat{\xi}_x)_{x \geq 0}$.

In section 5 we will fit our model to real data, obtained from the measurement of isotopes within narwhal tusks. The section starts with a short introduction to the case. We will then introduce the data in section 5.1. In section 5.2 we will talk about how our model (and the data) can be used to date the narwhals. In section 5.3 we will show the estimated age, along with a confidence interval. We will end with section 5.4 concerning model validation, where we check if the data conform to our model assumptions.

5. Age determination of a narwhal tusk. Narwhals exhibit a consistent and well-documented annual migration pattern as they move between specific summer and winter habitats, which are situated in fjords and inlets, and offshore areas, respectively (Heide-Jørgensen et al., 2020). The primary prey items consist of halibut, squids, and polar cod (Laidre and Heide-Jørgensen, 2005; Watt and Ferguson, 2014). On their seasonal migration routes, the narwhals travel through the same regional water masses to and from summer- and winter grounds. The strict diet and rigid migration pattern results in annual depositions of growth layers in the dentin zone of the tusk, consisting of an opaque and a translucent layer, see Figure 6b. Continuous fluctuations in the deposition/biomineralization rates show up in the signals and likely occur due to seasonal variations driven by environmental factors, such as temperature or salinity (Hüssy et al., 2021b). In addition, the physiological characteristics of the animals, notably age and health, likely modulate metabolic processes and as a result

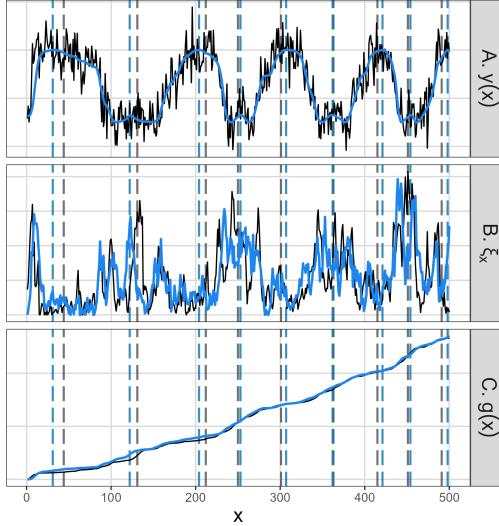


Fig 5: Simulated and fitted signal. Simulated signals are in black, fitted signals in blue. The dataset was simulated using $n = 500$, $\Delta = 1$, $\beta = 0.07$, $\sigma = 0.3$, $A = 0.6$, $B = 0.4$, $a = 0.05$, $b = \frac{\pi}{20}$, $\omega^2 = 0.064$. The algorithm used $n_p = 1500$ number of particles. A. Simulated signal and fitted signal. B. Simulated hidden process $(\xi_x)_{x \geq 0}$ and estimate $(\hat{\xi}_x)_{x \geq 0}$. C. Simulated growth process $g(x)$ and estimate $\hat{g}(x)$. Black dashed lines are (true) signal peaks. Orange dashed lines are estimated signal peaks.

potentially influence the biomineralization rate (Mosegaard, Svedäng and Taberman, 1988; Hüssy and Mosegaard, 2004). These assumptions are based on otoliths studies, and it seems reasonable to assume that the tusk growth rate exhibit similar characteristics. The recurring deposition of GLGs in conjunction with a variable deposition rate produces a cyclic profile that fits well into our working model (1). Since the hypothesis is that one GLG is deposited annually, counting these cycles will theoretically provide an age estimate of the narwhal. An approximate determination of the range of GLGs in a tusk piece can be made through visual observation, see Figure 6b. As explained in the introduction, in practice visually reading of the GLGs is both difficult and highly uncertain, and automatic robust methods based on more objective criteria and quantitative measures are needed. Thus, rather than considering the GLGs, we will focus on the mineral content deposited with the same annual pattern.

5.1. *Data acquisition and preparation.* We analyzed one tusk (ID 956) collected from the Inuit Hunt of narwhals in Niaqornat, West Greenland, in 2010. The tusk was sectioned in two halves and one was used for subsequent analysis. The half tusk were divided into several tusk pieces covering all GLGs (see Figure 6a). Few consecutive pieces had a complete overlap of GLGs, thus only one of these were used. After the exclusion, 12 pieces remained. Each tusk piece was measured along a transect (see Figure 6b) using Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS) by the Geological Survey of Denmark and Greenland (GEUS), providing the concentrations of 14 isotopes and trace elements (Sr, Ba, Zn, Mg, K, Li, Mn, Pb, Cu, P, Ti, Cr, Co, Rb) (Thomsen et al., 2022; Koch and Günther, 2011). Our analysis centers on Barium-137 normalized by the isotope Calcium-43 in high abundance, since among the set of elements, Barium displayed distinct cyclic patterns speculated to correlate with migration between summer and winter grounds (Heimbrand et al., 2020; Limburg et al., 2011). Consequently two peaks in the signal would correspond to one annual cycle. The combined signal over all pieces displaying the cyclic pattern is visualized for tusk 956 in Figure 7.

5.2. *Dating the observations.* For each piece j of the tusk, we fit the model with Algorithm 1 (subpanel A in Figure 8 and Figure 9). We fit each tusk piece individually allowing different parameter values in each, since the values will depend on the angle of the transect line (see red line in Figure 6b) across the piece. For example, a steeper line will increase the length of a year, changing the distance scale, which will change all parameters related

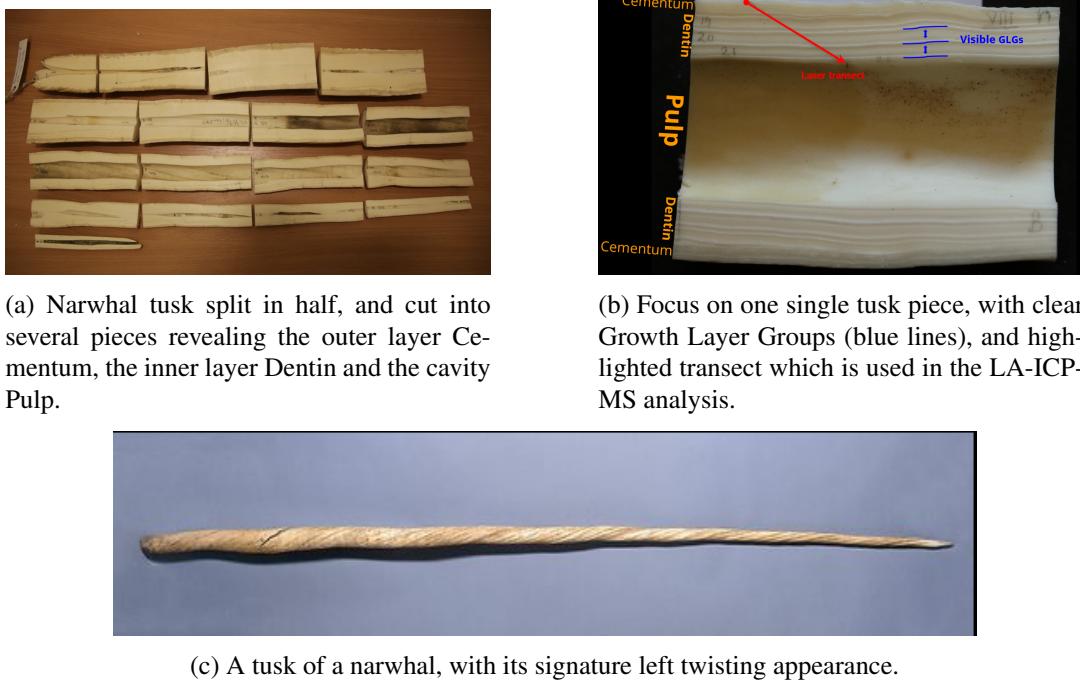
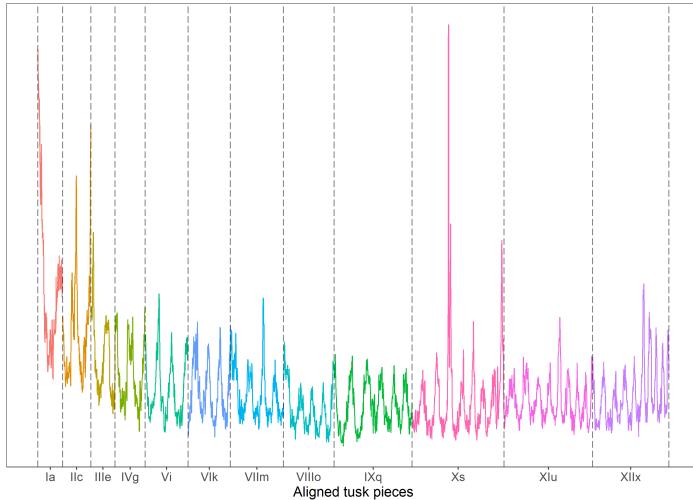
Fig 6: **Tusk and tusk pieces.**

Fig 7: Analyzed data set. Calcium-43 adjusted Barium-137 signals obtained from tusk 956. The data set consists of 12 piecewise signals merged together into one common time series in chronological order. The first part of the time series corresponds to the tip of the tusk, deposited when the whale was born, and the last part of the series corresponds to the base nearest the skull, deposited just before death.

405 to the growth process. Likewise, amplitudes are expected to change over the lifetime of the
 406 animal, and thus, change from piece to piece. We obtain estimates of the growth processes
 407 $\hat{g}_j(x_{ij})$, $i = 1, \dots, n_j$, $j = 1, \dots, 12$, where n_j are the number of observations in piece j
 408 (Figure 8B). This takes us from the original coordinates $(y_{ij}, x_{ij})_{i=1, \dots, n_j, j=1, \dots, 12}$ to a new

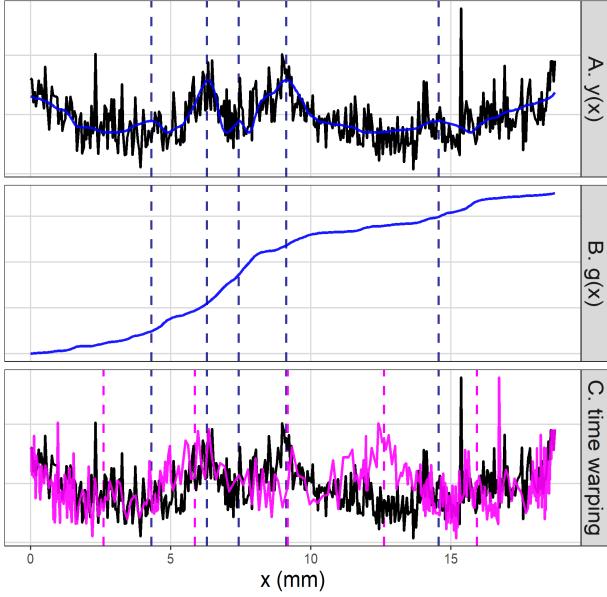


Fig 8: Time warping of barium signal in narwhal tusk for piece 956-VII. A. Fitted signal (blue curve) and true signal (black curve). B. Estimate $\hat{g}(x)$ of the growth process $g(x)$. C. Transformation from measured distance x (black curve, same as in panel A) to estimated time t (time warping, magenta curve). Blue dashed lines are estimated peak positions. Magenta dashed lines are the warped peak positions. The horizontal axis is displayed both in years (lower axis) and mm (distance on tusk; upper axis).

409 set of coordinates (y_{ij}, t_{ij}) with $t_{ij} = \hat{g}_j(x_{ij})$, where t_{ij} is a increasing sequence proportional
410 to the number of cycles. This shift in coordinates warps the signal and adjusts for any
411 non-linearity present in the original timeline (Figure 8C).

412 For each of the 12 pieces, we obtain a growth process $\hat{g}_j(x_{ij})$. The total number of obser-
413 vations is $n_{tot} = \sum_{j=1}^{12} n_j$. Re-enumerate the set of indices by concatenating chronologically

$$(26) \quad ij \rightarrow m = i + \sum_{l=1}^j n_l, \quad m = 1, \dots, n_{tot}.$$

414 Assuming each piece signal can be glued together, without any gap in the timeline, we can
415 then construct an *aggregated growth process*:

$$(27) \quad \hat{g}(x_m) = \hat{g}(x_{ij}) = \begin{cases} \hat{g}_1(x_{i1}) & i \leq n_1 \\ \hat{g}_{j-1}(x_{n_{j-1}, j-1}) + \hat{g}_j(x_{ij}) & n_{j-1} < i \leq n_j, j = 2, \dots, 12 \end{cases}.$$

416 One revolution around the unit circle corresponds to one annual cycle. Thus, the number
417 of elapsed years at observation m is $\hat{g}(x_m)/2\pi$, see eq. (5). Since we know the time of death
418 of the narwhal (in this case 2010), we date each observation m by

$$(28) \quad \text{Year}(m) = 2010 - (\hat{g}(x_n) - \hat{g}(x_m))/2\pi, \quad m = 1, \dots, n_{tot}.$$

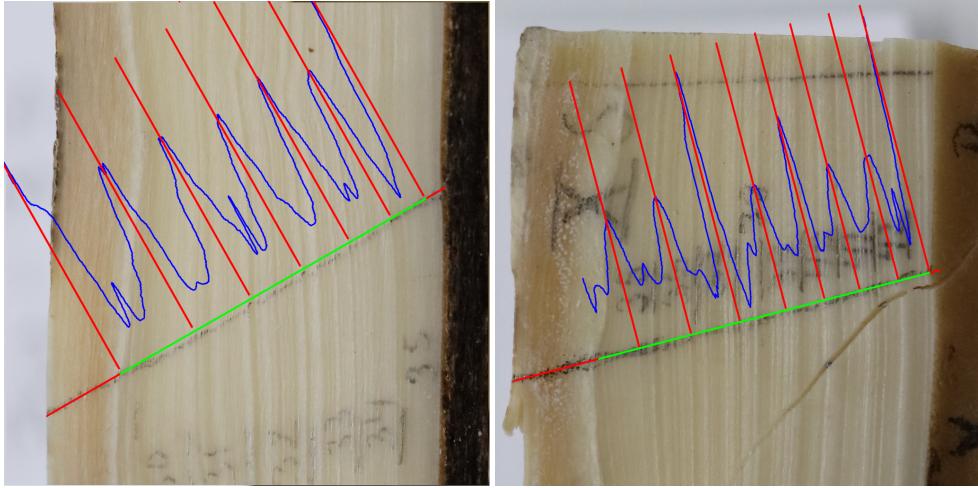
419 **5.3. Age estimation.** The quantity of interest is the age of the narwhal, which is equiva-
420 lant to the number of cycles easily obtained using formula (5) and $\hat{g}(x_m)$. A confidence inter-
421 val is obtained by bootstrapping (Algorithm 4). This is accomplished by performing paramet-
422 ric bootstrapping on each piece, and combine the bootstrapped signals randomly to produce
423 $M = 10^7$ realizations of aggregated growth processes $\hat{g}(x_m)$, of which each has its own age
424 estimate. The age estimate (number of GLGs), including upper and lower confidence bound,
425 are summarized in table 2. For benchmarking, the table also lists the most common age esti-
426 mator - manual counting of GLGs by our expert with proficiency in handling tusks - and the
427 age estimate of a newer, modern approach using Carbon-14 measurements.

428 In the supplementary material, we present a Monte Carlo analysis to assess the coverage
429 of our age estimates. While the nominal coverage is 95%, our results indicate a slight un-
430 dercoverage. This discrepancy is likely due to the residual bootstrap occasionally producing

TABLE 2

Age estimates (total GLGs) with 95% confidence interval. * Based on radiocarbon dating using the bomb pulse (Zoppi et al., 2004) and Carbon-14 measurements from the narwhal tusk (Garde et al., 2024).

Tusk ID	Age (estimate)	Age (confidence interval)	GLGs (manual count)	Age (^{14}C count)*
956	51	44 - 60	57	54



(a) Tusk piece 956-IX.

(b) Tusk piece 956-Xs.

Fig 9: Fitted signals overlain tusk pieces. Two tusk pieces with fitted signals (blue lines), estimated placement of high peaks (red vertical lines), laser line in dentin region (green segment) and pulp/cementum region (red segments).

431 multi-modal distributions of the age estimates. In such cases, the algorithm may converge to
 432 a solution that corresponds to a period doubling or halving.

433 **5.4. Model validation.** We have overlaid the fitted curves with the tusk pieces, to com-
 434 pare the identified peaks with the GLGs of the tusks. In Figure 9 we show two of such
 435 examples. The laser line is highlighted in green and red, when inside and outside the dentin
 436 area, respectively. The positions of the peaks are marked with vertical red lines, and we can
 437 discern how these peaks appear to align with consecutive white lines, which presumably rep-
 438 resent one GLG. A more formal model validation was carried out by inspecting the residuals
 439 (defined in section 3.7) visualized in Figure 10. The left plot shows a residual plot (raw resi-
 440 duals against fitted values) for each of the 12 fitted signals and the right column display a
 441 QQ-plot, where we compare the quantiles of raw residuals to those from the theoretical stan-
 442 dard normal distribution. The residual plots generally conform well to the homoscedasticity
 443 assumption and shows only a few systematic excursions, likely caused by potential rapid
 444 secondary cycles unique to our case study signals, which are not incorporated in our model
 445 (1). These cycles are likely associated with sub-annual accessory layers (Zhao, Matthews and
 446 Watt, 2025). For the QQ-plots, most residuals are symmetrically distributed around the diag-
 447 onal, indicating that the central tendencies are well-captured by the model. However, skewed
 448 behavior in the tails, such as in piece Xs, suggests that the model may not fully capture
 449 the behavior of extreme values. We refer to supplementary material, where we more closely
 450 explore this departure from normality for pieces XIIx, XIu, and Xs.

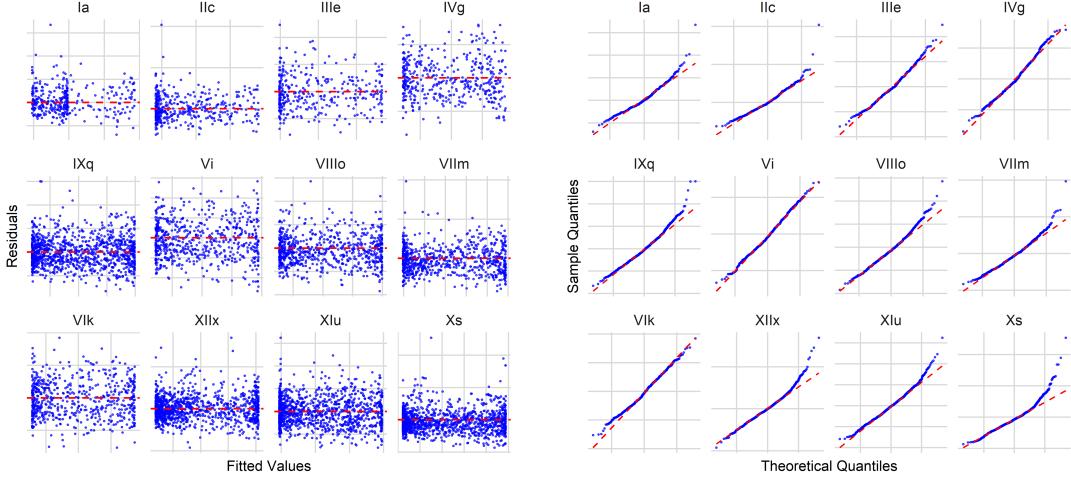


Fig 10: **Model validation plots for the case-study signals.** Residual plots (left) and QQ-plots (right) for all 12 signals of tusk 956.

451 Estimator uncertainties within each tusk piece is visualized in Figure 11 using violinplots
 452 for all parameters. The procedure is described in section 3.7. The simulated signals generally
 453 result in consistent estimates for parameters. However, among the 12 signals used in the case
 454 study, Figure 11 also reveals that several estimators display a sizeable amount of variability
 455 and bias. For example, the one-lag autocorrelation ρ is highly variable and show some un-
 456 systematic bias (i.e. bias in no particular direction) across several pieces. The infinitesimal
 457 variance ω^2 is also variable, but less prone to bias.

458 **6. Discussion and research outlook.** The main attraction of the model is arguably the
 459 identification of the growth process $g(x) = \int_{s=0}^x \xi_s ds$ which effectively translates into identifi-
 460 cation of the timeline of the underlying hidden process (4) driving the changing periodicity.
 461 While the square-root diffusion process $(\xi_x)_{x \geq 0}$ is slightly difficult to infer, likely caused by
 462 the integration of the process within our model, the integrated process $g(x)$ appear to be well
 463 identified.

464 It is important to consider the interplay between frequency and amplitude. For any periodic
 465 signal, we need to determine whether changes are primarily driven by variations in amplitude
 466 (i.e. changes in peak magnitude) or frequency (i.e. oscillation rate). In this study, we assumed
 467 that any instantaneous jumps in the signal reflects shifts in frequency, while amplitudes are
 468 slowly varying. This lets us first estimate and fix the amplitude envelope, then normalize to
 469 constant amplitudes before fitting our model. For the narwhal tusks, these assumptions imply
 470 that rapid changes in the signal occur due to changes in growth rate rather than fluctuations
 471 in the uptake of elements in the tusk.

472 While we perceive our presented model as both flexible and theoretically sound, the prac-
 473 tical implementation still requires some choices which needs a stronger theoretical founda-
 474 tion, or should possibly be tailored to the specific usecase. Some examples include selection
 475 of smoothing parameters in the initialization phase, runtime parameters (number of itera-
 476 tions inside Algorithm 1, particles in SMC and the threshold for the stopping criteria). These
 477 choices do not only assist the method in proper convergence, but are also important for con-
 478 trolling algorithmic runtime. These choices should not be perceived as weaknesses within
 479 our model but rather as strategic responses that will serve the algorithm in producing more
 480 reliable estimates and guide the optimization procedure.

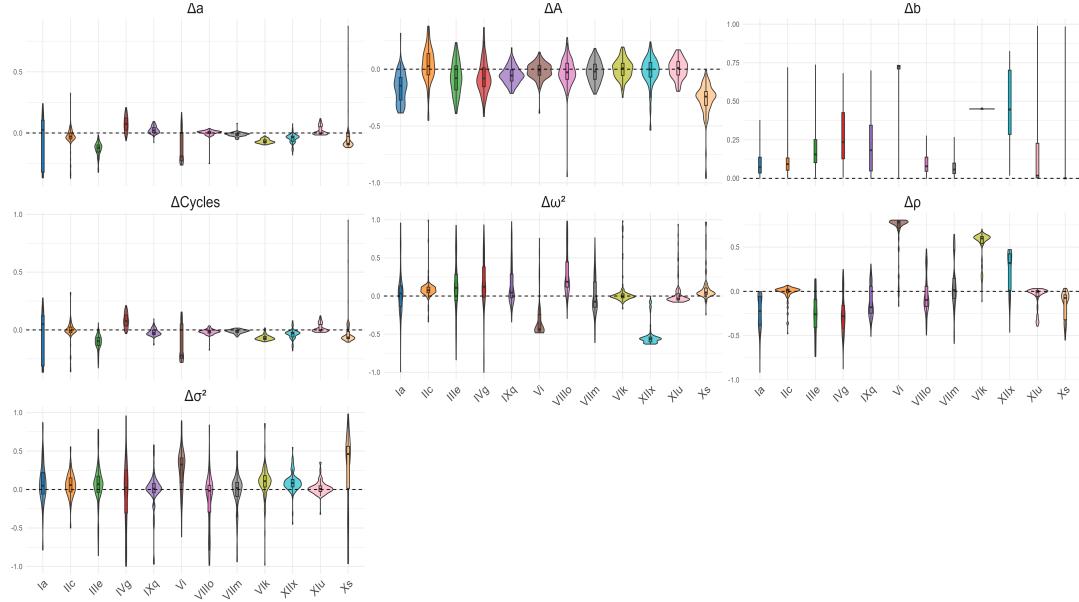


Fig 11: Estimator variability of tusk 956. The x-axis represent the different tusk pieces (12 in total). For an estimate $\hat{\theta}$ of some model parameter θ we compute 100 bootstrap estimates $\tilde{\theta}$ and plot the differences $\Delta\theta := (\tilde{\theta} - \hat{\theta})/\hat{\theta}$. The difference in phase Δb is computed as the shortest distance on the unit circle, not in absolute difference, thus all differences are non-negative.

481 **Acknowledgments.** We would like to thank the Greenlandic hunters for providing the
 482 narwhal tusk, Jeppe Møhl for sectioning the tusk, and Tonny Bernt Thomsen and Benjamin
 483 Heredia at the Geological Survey of Denmark and Greenland for performing the elemental
 484 analyses. The tusk was imported to Denmark from Greenland on CITES import permit no.
 485 IM 0822-296/18. Finally, we would like to thank Adam Gorm Hoffman, who was working
 486 with modelling and data analysis of a narwhal tusk in the early stage of the project.

487 **Funding.** This work has received funding by Novo Nordisk Foundation NNF20OC0062958;
 488 by NordForsk (Grant number 105053), and by the French-Danish CNRS International Re-
 489 search Network MaDeF IRN.

REFERENCES

- 490 AMAIS, R. S., MOREAU, P. S., FRANCISCHINI, D. S., MAGNUSSON, R., LOCOSSELLI, G. M., GODOY-
 491 VEIGA, M., CECCANTINI, G., RODRIGUEZ, D. R. O., TOMAZELLO-FILHO, M. and ARRUDA, M. A.
 492 (2021). Trace elements distribution in tropical tree rings through high-resolution imaging using LA-ICP-MS
 493 analysis. *Journal of Trace Elements in Medicine and Biology* **68** 126872.
- 494 BIBBY, B. M. and SØRENSEN, M. (1995). Martingale estimation functions for discretely observed diffusion
 495 processes. *Bernoulli* **17**–39.
- 496 CHARLES, C. D., LYNCH-STIEGLITZ, J., NINNEMANN, U. S. and FAIRBANKS, R. G. (1996). Climate connec-
 497 tions between the hemisphere revealed by deep sea sediment core/ice core correlations. *Earth and Planetary
 498 Science Letters* **142** 19–27.
- 499 CLEVELAND, R. B., CLEVELAND, W. S., MCRAE, J. E. and TERPENNING, I. (1990). STL: A seasonal-trend
 500 decomposition. *J. Off. Stat* **6** 3–73.
- 501 COX, J. C., INGERSOLL, J. E., JR. and ROSS, S. A. (1985). A Theory of the Term Structure of Interest Rates.
 502 *Econometrica* **53** 385–408.
- 503 DAGUM, E. (2013). Time Series Modelling and Decomposition. *Statistica* **70**. <https://doi.org/10.6092/issn.1973-2201/3597>

- 505 DAHLHAUS, R., DUMONT, T., LE CORFF, S. and NEDDERMEYER, J. C. (2017). Statistical inference for oscillation processes. *Statistics* **51** 61–83.
- 506
- 507 DEL MORAL, P., DOUCET, A. and JASRA, A. (2006a). Sequential monte carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **68** 411–436.
- 508
- 509 DEL MORAL, P., DOUCET, A. and JASRA, A. (2006b). Sequential monte carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **68** 411–436.
- 510
- 511 DELYON, B., LAVIELLE, M. and MOULINES, E. (1999a). Convergence of a stochastic approximation version of EM algorithm. *The Annals of Statistics* **27** 94–.
- 512
- 513 DELYON, B., LAVIELLE, M. and MOULINES, E. (1999b). Convergence of a stochastic approximation version of the EM algorithm. *Annals of statistics* 94–128.
- 514
- 515 DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **39** 1–38.
- 516
- 517 DHULL, M. S. and KUMAR, A. (2022). Expectation-Maximization Algorithm for Autoregressive Models with Cauchy Innovations. *Engineering Proceedings* **18** 21.
- 518
- 519 DIETZ, R., DESFORGES, J.-P., RIGÉT, F. F., AUBAIL, A., GARDE, E., AMBUS, P., DRIMMIE, R., HEIDE-JØRGENSEN, M. P. and SONNE, C. (2021). Analysis of narwhal tusks reveals lifelong feeding ecology and mercury exposure. *Current Biology* **31** 2012–2019.e2. <https://doi.org/10.1016/j.cub.2021.02.018>
- 520
- 521
- 522 DITLEVSEN, S. and LANSKY, P. (2006). Estimation of the input parameters in the Feller neuronal model. *Phys. Rev. E* **73** 061910. <https://doi.org/10.1103/PhysRevE.73.061910>
- 523
- 524 DITLEVSEN, S., RUBIO, A. C. and LANSKY, P. (2020). Transient dynamics of Pearson diffusions facilitates estimation of rate parameters. *Communications in nonlinear science and numerical simulation* **82**. <https://doi.org/10.1016/j.cnsns.2019.105034>
- 525
- 526 DITLEVSEN, S. and SAMSON, A. (2014). Estimation in the partially observed stochastic Morris-Lecar neuronal model with particle filter and stochastic approximation methods. *Annals of Applied Statistics* **8** 674–702. <https://doi.org/10.1214/14-AOAS729>
- 527
- 528 DOUCET, A., DE FREITAS, N. and GORDON, N. (2001). An introduction to sequential Monte Carlo methods. *Sequential Monte Carlo methods in practice* 3–14.
- 529
- 530 GARDE, E., DITLEVSEN, S., OLSEN, J. and HEIDE-JØRGENSEN, M. P. (2024). A radiocarbon bomb pulse model for estimating the age of North Atlantic cetaceans. *Biology Letters* 1–19. In press.
- 531
- 532 GLASS, L. and MACKEY, M. C. (1988). *From Clocks to Chaos: The Rhythms of Life*. Princeton University Press.
- 533
- 534 GORDON, N. J., SALMOND, D. J. and SMITH, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE proceedings F (radar and signal processing)* **140** 107–113. IET.
- 535
- 536 HEIDE-JØRGENSEN, M. P., BLACKWELL, S. B., WILLIAMS, T. M., SINDING, M. H. S., SKOVRIND, M., TERVO, O. M., GARDE, E., HANSEN, R. G., NIELSEN, N. H., NGÔ, M. C. and DITLEVSEN, S. (2020). Some like it cold: Temperature-dependent habitat selection by narwhals. *Ecology and Evolution* **10** 8073–8090. <https://doi.org/10.1002/ece3.6464>
- 537
- 538 HEIMBRAND, Y., LIMBURG, K. E., HÜSSY, K., CASINI, M., SJÖBERG, R., PALMÉN BRATT, A.-M., LEVINSKY, S.-E., KARPUSHEVSKAIA, A., RADTKE, K. and ÖHLUND, J. (2020). Seeking the true time: Exploring otolith chemistry as an age-determination tool. *Journal of Fish Biology* **97** 552–565. <https://doi.org/10.1111/jfb.14422>
- 539
- 540 HÜSSY, K. and MOSEGAARD, H. (2004). Atlantic cod (*Gadus morhua*) growth and otolith accretion characteristics modelled in a bioenergetics context. *Canadian Journal of Fisheries and Aquatic Sciences* **61** 1021–1031.
- 541
- 542 HÜSSY, K., KRÜGER-JOHNSON, M., THOMSEN, T. B., HEREDIA, B. D., NÆRAA, T., LIMBURG, K. E., HEIMBRAND, Y., MCQUEEN, K., HAASE, S., KRUMME, U. et al. (2021a). It's elemental, my dear Watson: validating seasonal patterns in otolith chemical chronologies. *Canadian Journal of Fisheries and Aquatic Sciences* **78** 551–566.
- 543
- 544 HÜSSY, K., LIMBURG, K. E., DE PONTUAL, H., THOMAS, O. R., COOK, P. K., HEIMBRAND, Y., BLASS, M. and STURROCK, A. M. (2021b). Trace element patterns in otoliths: the role of biomineralization. *Reviews in Fisheries Science & Aquaculture* **29** 445–477.
- 545
- 546 JANK, W. (2006). Implementing and Diagnosing the Stochastic Approximation EM Algorithm. *Journal of Computational and Graphical Statistics* **15** 803–829.
- 547
- 548 JEONG, Y.-S., JEONG, M. K. and OMITAO MU, O. A. (2011). Weighted dynamic time warping for time series classification. *Pattern recognition* **44** 2231–2240.
- 549
- 550 JOUZEL, J. and MASSON-DELMOTTE, V. (2010). Paleoclimates: what do we learn from deep ice cores? *Wiley Interdisciplinary Reviews: Climate Change* **1** 654–669.
- 551
- 552 KOCH, J. and GÜNTHER, D. (2011). Review of the state-of-the-art of laser ablation inductively coupled plasma mass spectrometry. *Applied Spectroscopy* **65** 155A–162A.
- 553
- 554 LAIDRE, K. L. and HEIDE-JØRGENSEN, M. P. (2005). Winter feeding intensity of narwhals (*Monodon monoceros*). *Marine Mammal Science* **21** 45–57. <https://doi.org/10.1111/j.1748-7692.2005.tb01207.x>
- 555
- 556
- 557
- 558
- 559
- 560
- 561
- 562
- 563

- 564 LAVIELLE, M. (2014). *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC
565 press.
- 566 LAWS, R. (1952). A new method of age determination for mammals. *Nature* **169** 972–973.
- 567 LIMBURG, K., OLSON, C., WALTHER, Y., DALE, D., SLOMP, C. P. and HØIE, H. (2011). Tracking Baltic
568 hypoxia and cod migration over millennia with natural tags. *Proceedings of the National Academy of Sciences
569 of the United States of America* **108** E177–82. <https://doi.org/10.1073/pnas.1100684108>
- 570 MONSERUD, R. A. and MARSHALL, J. D. (2001). Time-series analysis of $\delta^{13}\text{C}$ from tree rings. I. Time trends
571 and autocorrelation. *Tree Physiology* **21** 1087–1102.
- 572 MOSBACHER, J. B., MICHELSSEN, A., STELVIG, M., HENDRICHSEN, D. K. and SCHMIDT, N. M. (2016).
573 Show me your rump hair and I will tell you what you ate—the dietary history of muskoxen (*Ovibos moschatus*)
574 revealed by sequential stable isotope analysis of guard hairs. *PLoS One* **11** e0152874.
- 575 MOSEGAARD, H., SVEDÄNG, H. and TABERMAN, K. (1988). Uncoupling of Somatic and Otolith Growth Rates
576 in Arctic Char (*Salvelinus alpinus*) as an Effect of Differences in Temperature Response. *Canadian Journal of
577 Fisheries and Aquatic Sciences* **45** 1514–1524. <https://doi.org/10.1139/f88-180>
- 578 NAPOLITANO, A. and GARDNER, W. A. (2016). Algorithms for analysis of signals with time-warped cyclosta-
579 tionarity. In *2016 50th Asilomar Conference on Signals, Systems and Computers* 539–543. IEEE.
- 580 OSGOOD, B. (2007). *The Fourier Transform and Its Applications*. Stanford University.
- 581 READ, F. L., HOHN, A. A. and LOCKYER, C. H. (2018). A review of age estimation methods in marine mam-
582 mals with special reference to monodontids.
- 583 SALVADOR, S. and CHAN, P. (2007). Toward accurate dynamic time warping in linear time and space. *Intelligent
584 Data Analysis* **11** 561–580.
- 585 SAMMAKNEJAD, N., ZHAO, Y. and HUANG, B. (2019). A review of the Expectation Maxi-
586 mization algorithm in data-driven process identification. *Journal of Process Control* **73** 123–136.
587 <https://doi.org/10.1016/j.jprocont.2018.12.010>
- 588 LIXOFT SAS (2024). SAEM Lixoft SAS Monolix Suite 2024R1. Available at <https://monolixsuite.slp-software.com/monolix/2024R1/saem> (accessed April 16, 2025).
- 589 SEARLE, S., CASSELLA, G. and McCULLOUGH, C. (1992). Variance Components In: New York: Jonh Wiley-
590 Sons. DOI: <http://dx.doi.org/10.1002/9780470316856>.
- 591 SHUMWAY, R. H. and STOFFER, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*, 4th
592 ed. Springer.
- 593 SØRENSEN, M. (2012). *Statistical Methods for Stochastic Differential Equations* Estimating functions for diffu-
594 sion type processes. Chapman and Hall/CRC.
- 595 STOUNBERG, J., THOMSEN, T. B., HEREDIA, B. D. and HÜSSY, K. (2022). Eyes and ears: A comparative
596 approach linking the chemical composition of cod otoliths and eye lenses. *Journal of Fish Biology* **101** 985–
597 995.
- 598 THOMSEN, T., KEULEN, N., HEREDIA, B. and MALKKI, S. (2022). *Cannon balls from Bremervörde (D). A
599 characterization of historic cannon balls using SEM and LA-ICPMS microchemistry*. Danmarks og Grønlands
600 Geologiske Undersøgelse Rapport **2022**. GEUS. <https://doi.org/10.22008/gpub/34640>
- 601 WANG, Y., MILLER, D. J., POSKANZER, K., WANG, Y., TIAN, L. and YU, G. (2016). Graphical time warping
602 for joint alignment of multiple curves. *Advances in Neural Information Processing Systems* **29**.
- 603 WATT, C. and FERGUSON, S. (2014). Fatty acids and stable isotopes ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$) reveal temporal
604 changes in narwhal (*Monodon monoceros*) diet linked to migration patterns. *Marine Mammal Science* **31**.
605 <https://doi.org/10.1111/mms.12131>
- 606 WOOLLER, M. J., BATAILLE, C., DRUCKENMILLER, P., ERICKSON, G. M., GROVES, P., HAUBENSTOCK, N.,
607 HOWE, T., IRRGEHER, J., MANN, D., MOON, K. et al. (2021). Lifetime mobility of an Arctic woolly mam-
608 moth. *Science* **373** 806–808.
- 609 ZHAO, S.-T., MATTHEWS, C. J. and WATT, C. A. (2025). $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ cycles in narwhal (*Monodon mono-*
610 *ceros*) embedded teeth reveal seasonal variation in ecology and/or physiology. *Royal Society Open Science* **12**
611 242237.
- 612 ZOPPI, U., SKOPEC, Z., SKOPEC, J., JONES, G., FINK, D., HUA, Q., JACOBSEN, G., TUNIZ, C. and
613 WILLIAMS, A. (2004). Forensic applications of ^{14}C bomb-pulse dating. *Nuclear Instruments and Methods in
614 Physics Research Section B: Beam Interactions with Materials and Atoms* **223** 770–775.