A modelling approach to assessing the timescale uncertainties in proxy series with chronological errors

D. V. Divine¹, F. Godtliebsen¹,², and H. Rue³

¹Department of Mathematics and Statistics, Faculty of Science, University of Tromsø, 9037, Norway
²Norwegian Polar Institute, FRAM Centre, 9296 Tromsø, Norway
³Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, 7491, Norway

Received: 6 December 2011 – Accepted: 17 December 2011 – Published: 3 January 2012
Correspondence to: D. V. Divine (dmitry.divine@npolar.no)
Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

The paper proposes an approach to assessment of timescale errors in proxy-based series with chronological uncertainties. The method relies on approximation of the physical process(es) forming a proxy archive by a random Gamma process. Parameters of the process are partly data-driven and partly determined from prior assumptions. For a particular case of a linear accumulation model and absolutely dated tie points an analytical solution is found suggesting the Beta-distributed probability density on age estimates along the length of a proxy archive. In a general situation of uncertainties in the ages of the tie points the proposed method employs MCMC simulations of age-depth profiles yielding empirical confidence intervals on the constructed piecewise linear best guess timescale. It is suggested that the approach can be further extended to a more general case of a time-varying expected accumulation between the tie points. The approach is illustrated by using two ice and two lake/marine sediment cores representing the typical examples of paleoproxy archives with age models based on tie points of mixed origin.

1 Introduction

The skills of modern GCM models in predicting future climate scenarios are substantially constrained by the relative shortness and sparsity of available high quality instrumental climate data. Improvement of the models requires a better understanding of the interactions between the forcing factors and environmental responses occurring in the past. It conditions the use of various climate proxy archives in reconstruction of past climate evolution and then testing and refining the existing models (e.g. Hegerl et al., 2006). This procedure however requires that all possible uncertainties associated with the use of the proxy-based reconstructions are accurately taken into account.

A substantial part of proxy data used in paleoclimate research has chronological uncertainties. It implies that any constructed timescale is only an estimate of the true, but
unknown timescale. Since the timescale errors tend to map into the errors in the reconstructed variable, the realistic analysis requires the paleoenvironmental reconstruction to be considered in the probabilistic framework. For an accurate assessment of the timing and magnitude of events in the paleoproxy series and networks (e.g. Mayewski et al., 2004; Rohling and Pälike, 2005), as well as the use of proxy-based paleoclimate reconstructions in GCM model scoring experiments the effect of these errors needs to be thoroughly analyzed.

Typically, a chronological control of a proxy series stemming from all types of marine and terrestrial sedimentary archives is based on the use of $^{210}$Pb and AMS $^{14}$C dates, precisely dated reference horizons, or their combination. Depending on the prevalent origin of the available fix points (or tie points, age markers) as well as their density and relative precision a number of techniques to timescale construction were proposed. Besides a trivial curve fitting using simple mathematical functions, it includes various computational methods aimed at construction of a single best guess chronology and relevant confidence intervals (CI) (e.g. Goslar et al., 2009; Heegaard et al., 2005). Other methods, Blaauw and Christen (2005), involve a Bayesian inference to modelling the multitude of depth-age profiles.

We propose here an approach to assessment of chronological errors which is based on the modelling of the accumulation process associated with formation of a proxy archive. The value that characterizes the annual accumulation is considered an independent Gamma random variable. For a core section between the two neighboring dated horizons it allows calculation of the joint probability density of the sum of annual accumulations and simulation from this distribution conditionally on the specified sum, i.e. a precisely known depth interval. Parameters of the distribution are partly data driven and partly prescribed as a site(proxy)- dependent external parameter. In a particular case of the absolutely dated horizons the method provides an analytical solution for the CI. In a more general situation the analytical solution may not be reached and the estimates are made using the Monte Carlo technique. The approach generates a multitude of timescales creating a number of possible realizations of a proxy
series or a proxy based reconstruction in the time domain. This allows consideration of a proxy record in a probabilistic framework. The effect of accounting for uncertainties in chronology, also on a reconstructed environmental variable, is then illustrated with four case studies of ice cores and marine/lake sediment records.

2 Methods

2.1 Timescale modelling using MCMC simulation of a perturbed set of tie points

For a core with a continuous depth scale \( d \), the chronological model \( t(d) \) is typically constructed using a set of fixpoints (age markers) \( \{t_i(d_i), i = 1,\ldots,I\} \) produced by any type of dating technique, e.g. the AMS. The ages of these tie points are given as probability density functions \( p(t_i|d_i) \), with the shape of the distribution varying from point to point and depending on the type of the dating method applied. The Markov rule \( p(t_{i+1}|d_{i+1}) \propto p(t_i + c_i|d_{i+1}) \), where \( c_i = \text{const} > 0 \) and indexing runs from the top to the bottom of the core, illustrates a chronological constraint of no age reversals in the ordered set of derived fixpoints. For simplicity we also assume that the materials extracted for the dating procedure are representative of the time periods associated with the respective core slices.

Once the dates \( t_i \) are assigned to the set of age markers \( d_i \), the inferences drawn from the analysis of the reconstructed environmental variable(s) often rely on a single best guess chronology. Different approaches for constructing \( t(d) \) can be conceived, e.g. piecewise-linear least squares regression, a smooth polynomial fit or a generalized mixed-effect regression (Heegaard et al., 2005). The choice is usually rather subjective and largely depends on a time-depth distribution of age markers as well as available prior information, e.g. about past variations in accumulation rate. Note that the notion of “accumulation” used here refers to any type of process(es) responsible for formation of the considered sedimentary archive. The date \( \hat{t}_i \) maximizing the probability density function \( p(t_i|d_i) \) is typically assigned to a core slice at depth \( d_i \). However, uncertainty in
the absolute dates of age markers limits the validity of this approach as any constructed timescale is eventually only an estimate of the true, but unknown timescale.

Our modelling approach accounts for errors in the depth-age relationship via generation of the copies of \( t(d) \) yielding a “swarm” of chronologies, which is then converted into the probability density of \( p(t|d) \). This is implemented by drawing random numbers from the age distribution of each individual tie point \( p(t_i|d_i) \) for the entire set of age markers.

A possible overlap between \( p(t_i|d_i) \) of the neighboring tie points in the time domain may result in age reversals in the simulated sequence of tie points. To avoid such reversals, we used Markov-chain Monte Carlo (MCMC) implementation of the Gibbs sampler (Rue and Held, 2005). We introduce the procedure in a way similar to (Mudelsee et al., 2009) technique used in estimation of the effect of timescale errors on the spectrum of a time series.

Every of \( k = 1,...,K \) sets of fixpoints is generated sequentially progressing from the youngest to the oldest tie point. At each step for a given set \( k \) a tie point \( t_{ik} \) is simulated by drawing random numbers from the respective age distribution \( p(t_i|d_i) \). The age to the tie point \( t_{ik} \) is assigned if the condition \( t_{ik} > t_{(i-1)k} \) is met. If the condition is violated, the age-simulation is repeated until the stratigraphic constraint is fulfilled. The simulation goes over all tie point ages of the section, which completes one iteration of the Gibbs sampler and yields a strictly monotonically increasing sequence of simulated tie points. The set of \( K t(d) \) timescales can then be produced using any stipulated method. This empirical \( p(t|d) \) yields a numerical estimate of the depth evolution of timescale errors which can be mapped into the uncertainty in the space of the reconstructed variable.

### 2.2 Timescale modelling by simulation of accumulation process between the dating horizons

Reference horizons (e.g. volcanic or dust layers, T bomb peak) often provide absolutely dated fixpoints. In a more general case one can assume that for an arbitrary tie point \( t_i \):
\[ \sigma(t_i|d_i) < \{ E(t_{i+1}-t_i); E(t_i - t_{i-1}) \}, \text{ where } \sigma \text{ and } E \text{ stand for the standard deviation and the expected value, respectively.} \]

The availability of some a priori information about the features of the accumulation at the study site enables modelling of the accumulation process between the adjacent age markers.

Natural variability in accumulation rate causes the dating uncertainty in the interpolated timescale to grow together with a span to the nearest fixpoint. The two-step approach is therefore proposed:

- Step a: MCMC (see Sect. 2.1) is applied to generate \( K \) sets of perturbed fixpoints \( t_i(d_i), i = 1,...,l \).

- Step b: The accumulation between each pair \((t_i^k, t_{i+1}^k), k = 1,...,K; i = 1,...,l - 1\) is modelled \( N \) times to yield \( N \times K \) age-depths models.

Let the annual accumulation \( \lambda \) for a particular core segment \([i, i+1]\) associated with a time interval of \([t_i, t_{i+1}]\) be distributed according to \( \text{Gam}(\lambda; a, b) \) law, where \( \text{Gam}(\lambda; a, b) \) denotes the probability density function (pdf) of a gamma random variable with parameters \( a \) and \( b \):

\[
\text{Gam}(\lambda; a, b) = \frac{\lambda^{a-1} b^a e^{-b\lambda}}{\Gamma(a)} \tag{1}
\]

The Gamma distribution has been widely used in climate studies to model e.g. the amount of local precipitation in a given time period as the amount of moisture accumulated in a reservoir can be modelled by a Gamma process (e.g. Groisman and Easterling, 1994). This approximation therefore appears to be physically justified for a simplistic modelling of the annual snow accumulation on the top of an ice cap (glacier) and hence ice core formation process, provided that layer thinning due to compression is also taken into account. Parameters of the distribution are defined as follows. The average annual accumulation \( \lambda_{ik} \) for each core segment \([i, i+1]\) and realization \( k \) of the perturbed set of fixpoints is estimated directly from the data as \( \lambda_{ik} = (d_{i+1} - d_i)/(t_{i+1}^k - t_i^k) \).
For convenience a variability in the amount of annual accumulation is introduced via an additional parameter called signal to noise ratio (SNR) defined as \( \gamma = \frac{\lambda_i}{\sigma_i} \). Using the mean \( \left\{a/b\right\} \) and the variance \( \left\{a/b^2\right\} \) of the associated Gamma-distribution, \( \left\{a,b\right\} \) can then be calculated as:

\[
  a_i^k = a = \gamma^2; \quad b_i^k = \frac{\gamma^2}{\lambda_i^k} \tag{2}
\]

Note that in what follows we explicitly assume that \((a_i^k, b_i^k) = C_i^k\), i.e. the accumulation in the time interval between the layers (horizons) \(d_i\) and \(d_{i+1}\) does not experience any substantial shifts and can be approximately modeled by a random Lévy process.

The sum of gamma variables is itself a gamma-distributed random variable. Therefore, conditioning on the sum of annual accumulations \(\sum_{m=1}^{\Delta t} \lambda^k_{mi} = \Delta d_i = (d_{i+1} - d_i)\) constituting the length of the considered core segment \([i, i+1]\) is required. Here \(\Delta t^k_i = (t^k_{i+1} - t^k_i)\) is for the time interval between the tie points in the \(k\)th realization of the perturbed set of fixpoints and \(t^k_i \in \mathbb{Z}_n\) for convenience.

The conditional density of \((\lambda_1, \ldots, \lambda_{\Delta t})\) on \([i, i+1]\) given \(\sum_{m=1}^{\Delta t} \lambda_m = \Delta d\) is (Hogg and Craig, 1995)

\[
  (\lambda_1, \ldots, \lambda_{\Delta t} | \Delta d) \sim \Delta d \text{Dir}(a_1, \ldots, a_{\Delta t}) \tag{3}
\]

where Dir(\ldots) denotes the multivariate Dirichlet distribution function defined as

\[
  (\lambda_1, \ldots, \lambda_{\Delta t}; a_1, \ldots, a_{\Delta t}) \sim \prod_{m=1}^{\Delta t} \lambda_m^{a_m-1}. \tag{4}
\]

In our particular case \(\lambda_m\) are drawn from the same probability distribution, hence \(a_1 = a_k = a\). Note also that here and below in order to simplify the presentation we omitted indices \((i, k)\).

The samples corresponding to possible realizations of the annual accumulation can now be drawn from Dirichlet distribution for each core segment with a set of parameters...
according to Eq. (2). For sampling from Dirichlet distribution we used Fastfit Matlab toolbox of Thomas Minka (available at http://research.microsoft.com/en-us/um/people/minka/software/fastfit/).

In a simplified case of absolutely dated age markers (i.e. \( p(t_i|d_i) = t_i\delta(t - t_i) \) and \( K = 1 \)) the theoretical CIs for \( p(t|d) \) in \([i, i + 1]\) can be estimated using the marginalizing properties of \( \text{Dir}(a) \) pdf. Namely, for the total accumulation of \( d_M = \sum_{m=1}^{M} \lambda_m \) by year \( t_M \in [t_i; t_{i+1}] \) inclusive,

\[
\left( \sum_{m=1}^{M} \lambda_m, \lambda_{M+1}, \ldots, \lambda_{\Delta t} \right) \sim \Delta d \text{Dir}(aM, a_{M+1}, \ldots, a_{\Delta t})
\]

yielding

\[
p(d_M, (\Delta d - d_M)|t_M) \sim \Delta d \text{Be}(aM, a(\Delta t - M))
\]

where \( \text{Be}(\ldots) \) denotes the Beta distribution function defined as

\[
\text{Be}(\lambda; aM, a(\Delta t - M)) \sim \lambda^{aM-1}(1 - \lambda)^{a(\Delta t - M)-1}
\]

Using the Bayes rule and Eqs. (6, 7), the sought time vs. depth probability density function for \( t_m \) on \([i, i + 1]\) and \([t_i; t_{i+1}]\) can be derived

\[
p(t_m|d_m) = \frac{p(d_m|t_m)p(t_m)}{p(d_m)} \sim \Delta t \text{Be}(ma, (\Delta t - m)a)
\]

The inverse of the Beta cumulative density function at \( \alpha/2 \% \) and \( 100(1 - \alpha/2) \% \) yields exact \( 100(1 - \alpha) \% \) confidence limits on the estimates as

\[
\text{Cl}_{\alpha/2, (1 - \alpha/2)} = \Delta t \text{Be}^{-1}((\alpha/2, (1 - \alpha/2))|ma, (\Delta t - m)a)
\]

As a simple alternative, approximate confidence limits at every \( t_m \) can be established using the variance for \( p(t_m|d_m) \) and assuming asymptotic normality of the estimates.
The correspondent 100(1 − α) % CIs are given by $t_m \pm Z_{\alpha/2} \sqrt{V_p(t_m|d_m)}$ where $Z_{\alpha/2}$ denotes the $\alpha/2$ quantile of the standard Gaussian distribution and

$$V_p(t_m|d_m) = \frac{m(\Delta t - m)}{a\Delta t + 1}$$

is derived from the variance expression for Beta pdf. Equation (10) demonstrates that provided that the mean accumulation rate between the dating horizons is stationary, the timescale error between the tie points approximately follow the power law with the maximum error attained in the middle of the considered time interval. Inferred inverse dependence of $\sqrt{V_p(t_m|d_m)}$ on $\sqrt{a}$ and hence $\gamma$ implies an increased potential dating error for series with lower SNR, with especially steep growth in timescale uncertainty for the cases of $\gamma < 1$.

We can argue that in a more general situation of the non-fixed in time tie points the $p(t_m|d_m)$ in $[i, i + 1]$ will still be a Beta-distributed random variable. This Bayes’ rule based inference follows from the Binomial distribution as a reasonably good approximation for $p(\Delta t|t_m)$, Beta-pdf as a prior for $p(t_m)$ and the fact that Beta distribution is a conjugate prior of the Binomial distribution. It therefore suggests that the timescale error between the dating points in the general case also follows the power law.

One should also note that Eq. (10) provides very simple and efficient way of assessing potential timescale uncertainty between the dating horizons. Although some degree of subjectivity remains about the choice of the value for $\gamma$ parameter, the typical procedure of an ice core dating provides sufficient information for a reasonable estimate of $\gamma$ to be made. This is demonstrated in the next section.

3 Case studies

For the case studies we consider two ice, one lake and one marine sediment core with well established best guess depth-age relationships. For the ice and lake cores the high
resolution chronologies based on annual layer counting technique are available. We use them to estimate the values of the $\gamma$ parameter and test if the proposed timescale error model indeed comprises the actual core chronologies. For the marine core series the analysis also demonstrates the effect of timescale errors on the uncertainty in the reconstructed environmental parameter – summer sea surface temperature (SSST).

### 3.1 Lomonosovfonna ice core, Svalbard

This 121 m deep ice core was drilled at the summit of the ice field Lomonosovfonna (78°51′ N, 17°25′ E, 1250 m a.s.l.) in the spring of 1997. General information about the ice core and analyses can be found in Isaksson et al. (2001); Pohjola et al. (2002); Divine et al. (2008). The core dating is based on the updated (Divine et al., 2011) chronology using the well-known Nye (1963) relation, constrained by depths of the radioactivity peaks found in the core, which appear in 1962/1963 and 1954 (Pinglot et al., 1999), the 1903 Grímsvötn (Wastegård and Davies, 2009) and the 1783 Laki (Kekonen et al., 2005) volcanic eruptions detected by ion analyses and identification of tephra particles. Most of the core was sampled for $\delta^{18}O$ at 5 cm resolution implying subannual time resolution as far back as approximately 1456 AD. This higher resolution oxygen isotope stratigraphy made it possible, in addition to the glaciological modelling, apply summer peak counting procedure providing the annual timescale back to 1613 AD. Here we apply the proposed method for the section of the core above the Laki horizon of 1783 AD. Although an uncertainty of $\pm 1$ yr is typically assigned to even such precise age markers as volcanic ash layers, here for simplicity we treat them as absolutely dated.

Figure 1 shows the histogram of the values of annual accumulation for this period estimated from the high resolution core chronology. The mean annual accumulation rate at the core site is about 0.36 m water equivalent (w.eq.) for the 1783–1997 period and varies within $\approx 0.1–0.9$ m w.eq. The value of $\gamma$ for the whole series is estimated to be 2.2. When calculated within the sliding 50 yr long intervals (not shown) $\gamma$ slowly varies within $[1.6, 3.5]$ interval reaching the minimum during 1900–1930 and attaining
the constantly higher than average value in a post-1950 period. We note that this can be due to both natural factors, i.e. actual variability in snow accumulation plus postdepositional noise at the core site, as well as possible errors in the layer counting procedure.

Results of the MC modelling of the Lomonosovfonna ice core time scale are displayed in Fig. 2. Note that the core chronology is displayed on the “decompressed”, using Nye relationship, time scale accounting for layer thinning due to compression. The red solid line showing the layer counting-based core timescale lies well within the range of possible chronologies indicating that the proposed accumulation model is adequate along the entire considered ice core section. The theoretical CIs calculated from Eq. (10) lie in a perfect agreement with the MC results. The only time interval of 1925–1935 when the modelling and the actual timescale diverge overlaps with the period where lower than average SNR was found.

Comparison with the timing of spikes in a Lomonosovfonna residual sulfate record (Moore et al., 2006) most likely originating from volcanic eruptions suggests that dates of volcanic events for 4 of 7 sulfate peaks lie within the modeled dating error of max ±5 yr (Fig. 2). Some uncertainty, however, always remains due to an association of a sulphate peak with a specific volcanic eruption. For example, a potential candidate for a spike at ≈1873 ascribed to Krakatau 1883 could well be from a less remote major Icelandic eruption of 1875 Askja (Wastegård and Davies, 2009).

### 3.2 Colle Gnifetti ice core, the Alps

Colle Gnifetti 82-m long ice core was drilled in 2003 on a saddle between Zumstein-spitze and Punta Gnifetti/Signalkuppe of Monte Rosa at (45.93° N, 7.88° E) at an altitude of 4450 m a.s.l. An overview of the core site as well the results of the analyses are presented in (Lüthi and Funk, 2001). The core chronology was established using a combination of glaciological modelling, annual layer counting in chemical and water isotopes stratigraphies, identification of reference horizons and radiocarbon dating of
carbonaceous aerosols (Jenk et al., 2009). Here we apply our approach to the 220-yr long section of the core down to the depth of 38.8 m w.eq. corresponding to the Laki 1783 horizon (Thordarson and Self, 2003). There are in total 6 Saharan dust events (Häberli, 1977), 1963 bomb peak and the Laki sulfate peak during this time interval used as the absolutely dated tie points in the analysis.

The mean annual accumulation rate at the core site was estimated to be 0.44 ± 0.15 m w.eq. since 1977. Figure 3 shows the histogram of the values of annual accumulation for the period of 1783–2003 together with the fitted Gamma pdf, as was estimated from the high resolution core chronology. Despite a substantial wind erosion preserving only a fraction (typically less then 40 %) of the annual accumulation at Colle Gnifetti (Alean et al., 1983), the inferred mean signal to noise ratio is $\gamma = 2.7$ for the considered period. This value actually varies between 2.4 and 3 when estimated in 50 yr long intervals (not shown) and hence comparable with the one found for the Lomonosovfonna core.

To compensate for the layer compression with depth we used a formula of Thompson et al. (1990) with the layer thinning model parameters from (Sigl, 2009). Figure 4 shows the MC modelling results. Note that the sulfate peaks at 38.3 and 76.6 m w.eq. decompressed depth associated with 1912 Novarupta and 1815 Tambora eruptions are not included here in the set of tie points in order to evaluate the method performance. Back to approximately 1850 the actual timescale (red line in Fig. 4) fall well into the range of model results conditioned on the constant expected accumulation rate between the tie points. The Novarupta 1912 horizon is found very close to its most likely position in the ice core; again, given the model applied. The timescale error between the tie points reaches some ±2.5 yr which can be used as a reasonable estimate for the CG timescale uncertainty in the considered core section. For the period before \(\approx 1850\) the constructed timescale and the model-based \(p(t|d)\) diverge, suggesting that the constant expected accumulation between the two tie points of 1863 and 1783 represents a too coarse approximation. The 1815 Tambora sulfate deposition layer does not fall within the model CIs. The latter could be due to an increased accumulation in the first
half of the 19th century – a well-known period of Alpine glaciers expansion (Holzhauser et al., 2005).

### 3.3 Lake Silvaplana sediment core

Lake Silvaplana sediment core is a composite of two freeze and two piston cores recovered from the lake in winters 2004/2005 and 2005/2006 (Blass et al., 2007a; Trachsel et al., 2010). The site is located in the Engadine, Southeastern Swiss Alps, at an altitude of 1800 m a.s.l. The sediments of proglacial Lake Silvaplana are annually laminated (i.e. varved, Blass et al., 2007a) what potentially provides an excellent age control.

Identification and counting of varves was carried out on high-resolution digital core photographs and verified on thin-sections throughout the cores. Varve thickness measurements were taken from the photographs along three scan lines with a digital benchmark and mean values were calculated (Blass et al., 2007a; Trachsel et al., 2010). The annual mass accumulation rate (MAR) was calculated from varve thickness, water and organic matter content according to Berner (1971) and Niessen et al. (1992).

The updated core chronology (Trachsel et al., 2010) for approximately 181 cm-long composite core length spans the period AD 1170–2001. The timescale was constructed from varve counting with the turbidite layers from major historic floods used as additional precise time markers. These well documented floods from AD 1987, 1951, 1834, 1828 (Blass et al., 2007b) and AD 1868, 1793, 1772, 1706, 1566 and 1177 (Trachsel et al., 2010) together with the 1963 $^{137}$Cs peak were used as tie points in our timescale modelling procedure.

Figures 5 and 6 display results of the analysis. The variability in the inferred mass accumulation rate suggests the value of $\gamma = 1.9$. Since it was obtained without taking long term changes in accumulation into account, it represents a conservative estimate of this model parameter. Notable is a good fit of the estimated gamma pdf to the histogram of the varve thickness data, also in support of the proposed accumulation model.
A family of core timescales in Fig. 6 incorporates the actual core chronology throughout most of the core length. Pronounced deviations of the high resolution chronology from the range covered by the “swarm” of modelled timescales are observed in early 1500s, mid 1600s, and early 1900s. This is obviously due to deviations of the accumulation rate from linear between the correspondent tie points during this time periods. The estimated CIs attain the maximum width of about 10 yr for the period before 1500 AD, reaching as high as 20 yr (i.e. some 3 % of the age at most) for the core section between the two oldest dating horizons.

3.4 Marine sediment core MD99-2275

MD99-2275 is a composite of CALYPSO piston 3706 cm long core recovered in 1999 close to Iceland from 66°33.1’ N, 17°41.99’ W, 440 m water depth, and 36 cm long multicore B05-2006-MC04 retrieved at nearly the same location in 2006 (Ran et al., 2011). The core site is located approximately 50 km offshore in the Skjalfandadjup trough. We analyze here the uppermost 358 cm long section of the spliced core down to the settlement tephra layer of 1079 BP (871 AD). This core section was dated by 8 tephra layers and sampled every 1–3 cm to achieve a time resolution of 3–12 yr (950–1999 AD, see Ran et al. (2011) for more details). For the 5 of 8 tephra age markers the dating precision is absolute, whereas the other 3 are dated with an accuracy varying from ±3 to ±10 yr of one standard deviation (Eiriksson et al., 2011).

A precise, with a constant time increment, core chronology in this particular case is unavailable. We, however, hypothesize that the model and the common principles elaborated for so different in their origin ice and lake sediment cores are transferrable to the marine environment. Figure 7 shows the results of the two-steps approach outlined in Sects. 2.1 and 2.2 applied to the MD99-2275 core timescale. As a value of an external γ parameter we used typical γ = 3 and γ = 1.5, and lower than was observed for other series γ = 0.5.

The choice of γ has a profound effect on the spread of the modelled timescales and hence the width of the modelled/theoretical confidence intervals. The corresponding
widths of the CIs can differ up to the order or four between the lowest and highest value of $\gamma$ parameter used. Note that the modelled and theoretical CI agree very well between the absolutely dated tephra layers but substantially diverge around the age markers with dating uncertainties. For comparison red lines in Fig. 7b,c highlight the core age-depth model from (Eiriksson et al., 2011; Ran et al., 2011) together with the respective CIs constructed using Bayesian-based OxCal 4.1 technique (Ramsey, 2008, 2009). As one can see, the results demonstrate that a good agreement between the methods is achieved for $\gamma = 0.5$. It suggests that the Bayesian technique tends to provide in this particular case somewhat more conservative estimate on the timescale accuracy than one could expect from the use of the proposed method.

For this core the estimates of past summer sea surface temperature variability were made using the weighted averaging partial least square regression (WA-PLS) applied on downcore series of fossil diatom assemblages (Ran et al., 2011). Applying interpolation to resample the set of modelled timescales to a common timescale we projected the dating uncertainties to the space of the reconstructed SSST. Figure 7d,h,l demonstrate the effect of timescale errors on uncertainties in the reconstructed environmental variable. The highest inferred $\sigma_{\text{SSST}}$ and hence widest CIs are expectedly found for the lowest used value of $\gamma$. Notable is that even for $\gamma = 3$ for some particular time periods the error due to dating can be of the order of magnitude of the uncertainty of 0.9 °C (root mean square prediction error, RMSEP) associated with the reconstruction method itself.

4 Extension to the case of varying accumulation and general considerations on the choice of $\gamma$ parameter

The presented analysis and results rely on a piecewise linear accumulation model what may pose a substantial limitation on the quality of the inference in case if this condition is severely violated. Although the high-resolution accumulation history at the study site may be unavailable, abrupt shifts in the slopes of depth-age curve at the tie points will
be indicative of the inadequacy of the piecewise linear accumulation model. Bearing in mind that “all models are wrong, but some are useful” (G. Box) and general subjectivity in choosing the most optimal fitting function for \( E(\lambda) \), the proposed approach can be extended to the cases with a time-varying expected accumulation of \( E(\lambda) = f(t) \).

For any core segment of the length \( d_{i+1} - d_i = \Delta d \) between the two dating horizons with \( t_{i+1} - t_i = \Delta t \), \( \Delta t \in \mathbb{Z}_n \) and non-constant expected accumulation rate the conditional density of \((\lambda_1, \ldots, \lambda_m, \ldots, \lambda_{\Delta t})\) given \( \Delta d \) is distributed as (Fearnhead and Meligkotsidou, 2004)

\[
(\lambda_1, \ldots, \lambda_{\Delta t} | \Delta d) \sim \sum_{m_1=0}^{\infty} \ldots \sum_{m_{\Delta t}=0}^{\infty} \left\{ \prod_{m=1}^{\Delta t} \Pr(N_m = m_m) \right\} \\
\times \Delta d \text{Dir}(a + m_1, \ldots, a + m_{\Delta t})
\]

(11)

and

\[
\Pr(N_m = j) = \frac{\Gamma(a + j)}{\Gamma(a)\Gamma(j + 1)} \left( \frac{b_m}{b_{\Delta t}} \right)^a \left( 1 - \frac{b_m}{b_{\Delta t}} \right)^j; \quad j = 0, 1, \ldots
\]

(12)

Here the property of each \( \lambda_m \) to be presented as a mixture of gamma variables is applied, and hence their conditional sum can be written as an infinite mixture of Dirichlet probability density functions. One can see that for a particular case of constant \( \{a, b\} \) parameters Eqs. (11, 12) are simplified to Eq. (3) with \( a_m = a \), since by definition \( \Pr(N_m = j) = 1 \), \( j = 0 \) and \( \Pr(N_m = j) = 0 \) for \( j > 0 \).

The change in accumulation rate is introduced into this scheme via a varying parameter \( b \) inversely proportional to \( E(\lambda) \). The exact sampling procedure conditioned on the length of the core segment is implemented following the exact filtering forward-backward algorithm. For practical computations a truncation of the sums in Eq. (11) to some index \( J \) is applied. On the forward step the \( J \)-by-\( J \) matrix of weights \( \Pr(N_m = j) \) is computed, while on the backward step \((N_1, \ldots, N_{\Delta t})\) are simulated and then \((\lambda_1, \ldots, \lambda_{\Delta t})\) drawn from \( \text{Dir}(a + N_1, \ldots, a + N_{\Delta t}) \) distribution (see Fearnhead and Meligkotsidou, 2004, for more details).
The “signal-to-noise” ratio $\gamma$ is a principal parameter to determine the width of the confidence intervals on the constructed core chronology. For the examples demonstrated above the choice was entirely data-driven which obviously may not be possible if the details on the accumulation process are unknown. As for any free parameter its choice can be quite subjective. However, similar values of $\gamma$ found for Lomonosovfonna and Colle Gnifetti cores suggest that $\gamma \approx 1–2$ could be a reasonable and sufficiently conservative estimate. To corroborate this statement we calculated $\gamma$ for some additional ice core accumulation series from different geographical locations. It includes Antarctic S100 (Kaczmarska et al., 2004, 1737–1999 AD, $\gamma = 3$), B32 (Graf et al., 2002, 167–1997 AD, $\gamma = 3.7$) and Berkner Island cores (Mulvaney et al., 2002, 1000–1994 AD, $\gamma = 3.4$), Greenland GISP2 (Cuffey and Clow, 1999, 1675–1988, $\gamma = 4.4$) and NGT42C95.2 cores (Miller and Schwager, 2000, 1477–1994 AD, $\gamma = 4.8$), Severnaya Zemlya Akademii Nauk ice cap core (Opel et al., 2009; Fritsche et al., 2010; Opel et al., 2011, 1783–1998, $\gamma = 3.0$).

The same procedure applied to some of the high resolution MAR records of varved lake sediments available on the web site of the NOAA World Data Center for Paleoclimatology (http://www.ncdc.noaa.gov/paleo/data.html) yielded the following estimates of $\gamma$ for Lake Pumacocha in Peru (Bird et al., 2011, 1797–2007 AD, $\gamma = 2.1$), lakes Lehmilampi (Haltia-Hovi et al., 2007, 3–2004 AD, $\gamma = 2.0$) and Nautajarvi in Finland (Ojala and Alenius, 2005, 1–1995 AD, $\gamma = 1.6$), Ogac Lake on Baffin Island (Hughen, 2009, 45–1992 AD, $\gamma = 1.1$) and Lago Plomo, Chile (Elbert et al., 2011, 1530–2002 AD, $\gamma = 3.9$). As one can see a typical $\gamma$ value for ice core series exceeds 3 for the considered drilling sites. The lake sedimentation records from various locations show generally lower $\gamma \approx 2$ with the smallest found for the Ogac Lake series. It suggests that $\gamma \approx 1–2$ can well be used as a “rule of thumb” choice in the CIs estimation procedure.
5 Summary

Detailed knowledge of past climate variations is of high importance for gaining a better insight into the possible future climate scenarios. It requires an accurate assessment of timescale errors in proxy-based paleoclimatic reconstructions.

The approach presented here is applicable to paleoproxy archives with age markers of mixed origin and consists in a two step modelling procedure involving

- Step a: MCMC sampling from distributions of age markers to generate a prescribed $K$ sequences of perturbed tie points

- Step b: Modelling the accumulation process between each pair of perturbed fix points $N$ times to yield $N \times K$ age-depths models

The annual accumulation is treated as an independent random Gamma-distributed variable. It allows the exact sampling from the joint distribution of annual accumulations, conditioned on their sum, constituting the considered core segment between the two adjacent tie points. The procedure yields an empirical (simulated) probability density function of the time vs. depth $p(t|d)$. It is then used to estimate the CIs on the constructed paleoproxy archive age-depth model and projection of the dating uncertainties to the space of a reconstructed environmental variable. Note that here we consider an annual time step although the conclusions are transferrable, with some reservations, to any regular time increment.

In a particular case of absolutely dated tie points and stationary in time accumulation rate the method provides an analytical solution. It is demonstrated that the probability density function of $p(t|d)$ is Beta-distributed and the timescale error between the tie points approximately follow the power law with the maximum error attained in the middle of the considered time interval.

Although the assumption of piecewise linear accumulation is primarily used here, the method can potentially be extended to a more general case of slowly varying mean accumulation rate. We note that if the piecewise linear accumulation model is used and...
the high resolution core chronology is available, the approach also provides and indirect test for the periods where accumulation have demonstrated significant deviations from the stationary regime.

The ratio of the expected annual mean accumulation to its standard deviation, $\gamma$, is the only free parameter to be assigned. We however demonstrated using a number of ice and lake sediment core records that a value of $\gamma \approx 1–2$ could be a reasonable and sufficiently conservative estimate to be used as a “rule of thumb” choice in the CIs estimation procedure.

Acknowledgements. This study was financially supported by the Norwegian Research Council via eVita project 176872/v30. The authors also acknowledge the European Commission Project Millennium (Project no. 017008-2). Special thanks are due to T. Opel and D. Fritzche (Alfred Wegener Institute), M. Schwikowski (Paul Scherrer Institute), M. Trachsel and J. Elbert (Univ. of Bern), K. L. Knudsen (Univ. of Aarhus), J. Eiriksson (Univ. of Iceland) and J. Olsen (Queen’s Univ.) for kindly providing the data for analysis.

References

Alean, J., Häberli, W., and Schädler, B.: Snow accumulation, firn temperature and solar radia-

tion in the area of the Colle Gnifetti core drilling site (Monte Rosa, Swiss Alps): distribution patterns and interrelationships, Zeit Glet., 19, 131–147, 1983. 42


Blass, A., Grosjean, M., Troxler, A., and Sturm, M.: How stable are twentieth-century calibra-


Groisman, P. Y. and Easterling, D. R.: Variability and trends of total precipitation and


Kekonen, T., Moore, J., Perämäki, P., and Martma, T.: The Icelandic Laki volcanic tephra layer


Nye, J. F.: Correction factor for accumulation measured by the thickness of the annual layers in an ice sheet, J. Glaciol., 4, 785–788, 1963. 40


Rohling, E. J. and Pälike, H.: Centennial-scale climate cooling with a sudden cold event around 8200 years ago, Nature, 434, 975–979, doi:10.1038/nature03421, 2005. 33


Trachsel, M., Grosjean, M., Larocque-Tobler, I., Schwikowski, M., Blass, A., and Sturm, M.: Quantitative summer temperature reconstruction derived from a combined biogenic Si and chironomid record from varved sediments of Lake Silvaplana (South-Eastern Swiss Alps) back to AD 1177, Quaternary Sci. Rev., 29, 2719–2730,
Timescale modelling for proxy series with chronological errors

D. V. Divine et al.


Fig. 1. Scaled to the length of the series the histogram of the Lomonosovfonna annual accumulation for the period 1783–1999 binned at 0.025 m w.eq. intervals (bars); dashed line shows Gamma pdf for \( \{a,b\} \) parameters estimated from the data using Eq. (2).
Fig. 2. (a) Distribution of 5000 MC generated timescale models for the Lomonosovfonna ice core; solid red line is for the “best” core chronology constructed by summer peaks counting in δ¹⁸O stratigraphy; shades of grey on the vertical bar from the right displays the percentiles of the cumulative probability density function for \( p(t|d) \): the vertical interval between the isolines with equal cdf values \( p \) covers \( p \% \) of all generated chronologies at depth \( d \). (b) Widths of the 95% empirical CIs for the estimated age-depth relationship (solid black line); red triangles indicate depths for the identified tephra layers and 1963 bomb peak; blue dashed line show the theoretical 95% CIs calculated using asymptotic approximation to Eq. (9) by Eq. (10). Vertical dotted lines on both panels indicate the positions of sulphate peaks associated with volcanic eruptions of 1991 Pinatubo, 1956 Bezymianny, 1934 Kharimkotan, 1928 Kuril, 1912 Novarupta, 1883 Krakatau and 1815 Tambora; the time difference between eruption dates and the most likely model dates is shown by red asterisks.
Fig. 3. Same as in Fig. 1 but for the Colle Gnifetti ice core.

\[ a = 7.26 \]
\[ b = 17.1 \]
\[ \gamma = 2.7 \]
Fig. 4. Same as in Fig. 2 but for the Colle Gnifetti ice core.
Fig. 5. Same as in Fig. 1 but for the Lake Silvaplana sediment core.
Fig. 6. Same as in Fig. 2 but for the Lake Silvaplana sediment core.
Fig. 7. (a,e,i) Distribution of 10,000 generated age-depth models for core MD99-2275 obtained by a combination of 100 MCMC fixpoints models and 100 accumulation models for SNR = [0.5, 1.5, 3], solid red line is for the OxCal 4.1 core chronology (Eiriksson et al., 2011; Ran et al., 2011); (b,f,j) Same as in Figs. 2 and 4; blue triangles with the error bars indicate the depths for tephra age markers with their 2σ ranges, red triangles are for the absolutely dated tephra layers; Solid red line in (b) is for the OxCal 4.1 based width of the 95% empirical CI for the estimated age-depth relationship; (c,g,k) Distribution of 10,000 reconstructed SSST profiles derived from a set of generated chronologies, blue and red triangles show the timings of the tephra layers; (d,h,l) Time variations of the width of the 95% empirical CIs on the estimated SSST; blue error bar indicate the RMSEP of 0.9°C for the WA-PLS reconstruction method used.