



Quantifying uncertainty in stratigraphic alignment of geological signals using probabilistic dynamic time warping

Aske L. Sørensen^{a,*}, Thomas M. Hansen^a, Frederik A. Falk^a, Jesper Olsen^b, Mads F. Knudsen^a

^a Department of Geoscience, Aarhus University, Aarhus, Denmark

^b Department of Physics, Aarhus University, Aarhus, Denmark

ARTICLE INFO

Handling Editor: Qiuzhen Yin

ABSTRACT

Aligning stratigraphic records is essential for constructing unified and coherent chronological frameworks for Earth's history and advancing our understanding of past environments and geological processes. However, the alignment process is often complicated by noise, hiatuses, and local accumulation dynamics, which introduce significant uncertainty. To address these challenges, we present the Probabilistic Dynamic Time Warping (P-DTW) algorithm, designed to align noisy stratigraphic signals, identify multiple plausible alignment scenarios, and quantify the associated uncertainty. The algorithm incorporates three transition probability factors, which are used to stochastically sample the alignment paths through the DTW cost matrix. The P-DTW algorithm is primarily aimed at aligning stratigraphic signals where peak-to-peak correlations are ambiguous.

Through synthetic test cases, we demonstrate the ability of the P-DTW algorithm to capture a range of plausible alignments, while quantifying the associated uncertainty. This stands in contrast to the traditional deterministic DTW algorithm, which provides a single solution that may overlook geologically plausible alternatives. Furthermore, we show the P-DTW algorithm's capacity to align signals with different amplitudes over the same intervals, and how to incorporate tie points derived from independent constraints.

When applied to real-world $\delta^{18}\text{O}$ data from sediment cores GeOB7920-2 and MD95-2042, the P-DTW algorithm generates an alignment model consistent with established alignments while quantifying the associated uncertainties. Additionally, we demonstrate the algorithm's ability to align the magnetic susceptibility signal from the ENAM93-21 core in the North Atlantic with the $\delta^{18}\text{O}$ record from the NorthGRIP ice core, resulting in an age-depth model that aligns with independent constraints. Finally, we showcase the algorithm's capacity to integrate chronological information across sites within a probabilistic inverse modeling framework, hereby facilitating the construction of a coherent multi-site age-depth model.

1. Introduction

Aligning stratigraphic records is critical for correlating geological events across various regions, as this enables the construction of unified chronological frameworks for Earth's history. Such frameworks underpin reconstructions of past environments, facilitate the exploration of natural resources, and enhance our understanding of geological processes. For instance, our understanding of climate and environmental change during the Quaternary Period hinges largely on correlation between long continuous records from geological archives, such as ice cores, marine sediments, and loess sections. However, some of these records are noisy, embedded with hiatuses, and influenced by local

accumulation dynamics, all of which introduce significant uncertainty into the alignment process.

Traditionally, stratigraphic alignment has been performed visually (e.g. Prell et al., 1986; Nielsen et al., 2018). However, advances in algorithm development have significantly improved the speed and accuracy of stratigraphic alignment, and it has reduced the degree of subjectivity in the alignment process. For instance, Baldwin et al. (1989) and Luthi and Bryant (1997) demonstrated the potential of neural networks for stratigraphic correlation, and more recent work by Brazell et al. (2019) suggests that deep neural networks offer a promising approach for well-log correlation. Sambridge et al. (2022) used optimal transport for aligning seismic waveforms. A widely used algorithm to

* Corresponding author.

E-mail address: als@geo.au.dk (A.L. Sørensen).

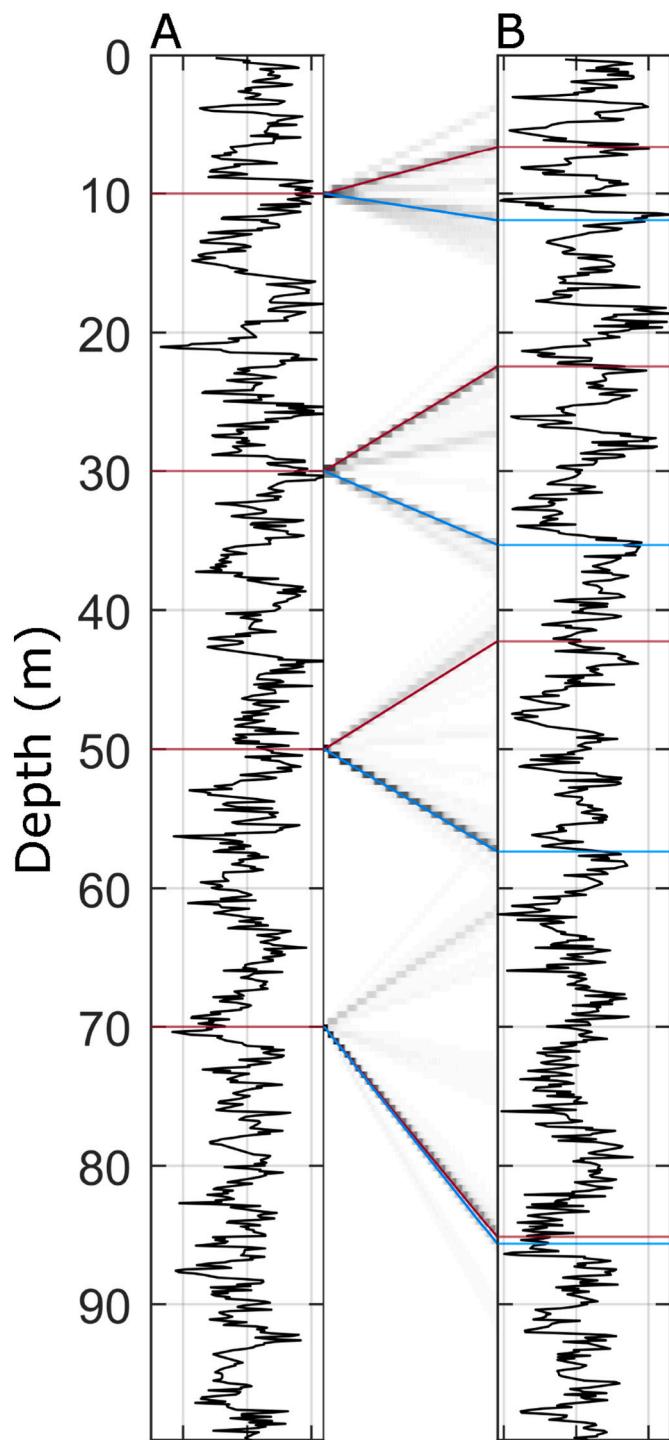


Fig. 1. Multiple distinct solutions appear when aligning two noisy stratigraphic signals. The grayscale depicts the P-DTW alignment model, which reflects multiple distinct alignments with varying probabilities. Notably, the classic deterministic DTW alignment (blue line) diverges from the ‘true’ reference-alignment (red line), whereas the P-DTW model identifies both alignments as plausible. A: signal a, B: signal b.

align stratigraphic signals is the Match-algorithm (Lisicki and Lisicki, 2002), which was instrumental in creating the influential LR04 $\delta^{18}\text{O}$ stack (Lisicki and Raymo, 2005). Despite its widespread adoption, the Match algorithm is deterministic, providing a single optimal alignment without accounting for the uncertainties inherent in the alignment process.

To address this limitation, several algorithms have emerged to

quantify the uncertainty associated with aligning stratigraphic records. For example, the Match algorithm was adapted within a probabilistic hidden Markov model framework to evaluate the uncertainty of $\delta^{18}\text{O}$ alignment with the LR04 stack (Lin et al., 2014; Lee et al., 2023). Additionally, Malinverno (2013) developed a probabilistic alignment method based on inverse Monte Carlo sampling, which was initially applied to bulk density records and further refined in subsequent studies (Muschitiello et al., 2015, 2020; West et al., 2019). These approaches focus on alignment models characterized by unimodal probability distributions, and they work well in scenarios where the signal-to-noise ratio is high, and peak-to-peak correlations are unambiguous.

The Dynamic Time Warping (DTW) algorithm, originally developed for speech recognition (Sakoe and Chiba, 1978), has proven to be an effective tool for aligning stratigraphic records. This is due to its simplicity, computational efficiency, ability to handle non-linear relationships between markers and ease of modification. DTW was initially introduced in stratigraphic applications for facies correlation (Smith and Waterman, 1980; Waterman and Raymond, 1987; Baville et al., 2022), but has also been successfully applied to well-log correlations (Wheeler and Hale, 2014; Wu et al., 2018; Sylvester, 2023), magnetic susceptibility (Hladil et al., 2010), $\delta^{13}\text{C}_{\text{org}}$ (Hay et al., 2019), and for paleomagnetic records (Lallier et al., 2013). Despite its utility, the traditional DTW algorithm is deterministic, yielding a single optimal alignment and neglecting geologically plausible alternatives. Additionally, traditional DTW approaches often lack integration of geological constraints that may help constrain the alignment models.

To overcome these limitations, Lallier et al. (2012, 2013, 2016) developed stochastic adaptations of the DTW algorithm which may integrate geological constraints and hereby enable the generation of multiple plausible correlation scenarios reflecting the uncertainty of the alignment. These algorithms were developed for aligning discrete sequences, such as geomagnetic polarity and stratigraphic well facies, but they can potentially be modified to handle continuous or semi-continuous signals, such as magnetic susceptibility, grain size, or $\delta^{13}\text{C}$ data.

Building on this foundation, we introduce the Probabilistic Dynamic Time Warping (P-DTW) algorithm, which advances beyond the stochastic framework introduced by Lallier et al. (2012, 2013, 2016) through three key methodological innovations specifically suited to align continuous and semi-continuous geological records with ambiguous peak-to-peak correlations. First, we develop a transition probability formulation that uses relative cost differences with a tunable parameter to control the influence of signal similarity on alignment. Second, we introduce a novel path-momentum factor that controls the degree of abrupt changes in the alignment by regulating the local slope direction, addressing a fundamental challenge in noisy data scenarios. Third, we implement a path-steer mechanism that biases alignments toward surface correlation when appropriate, incorporating the geological constraint that surface depths often represent contemporaneous deposition.

We define a continuous signal as a series of data points characterized by gradual, smooth variations; a discrete signal as one marked by abrupt changes or step-like transitions; and a semi-continuous signal as a combination of the two. The P-DTW algorithm generates alignment models that capture a range of plausible alignments while quantifying the associated uncertainties suited to align continuous and semi-continuous geological records characterised by ambiguous peak-to-peak correlations (Fig. 1).

The following sections detail the P-DTW methodology, its implementation, and its application to real-world datasets, showcasing its ability to capture and quantify the complexity associated with aligning noisy signals. Additionally, we demonstrate how the P-DTW algorithm can be integrated into a probabilistic inverse framework to merge age constraints from different sites, hereby facilitating the construction of multi-site age-depth models.

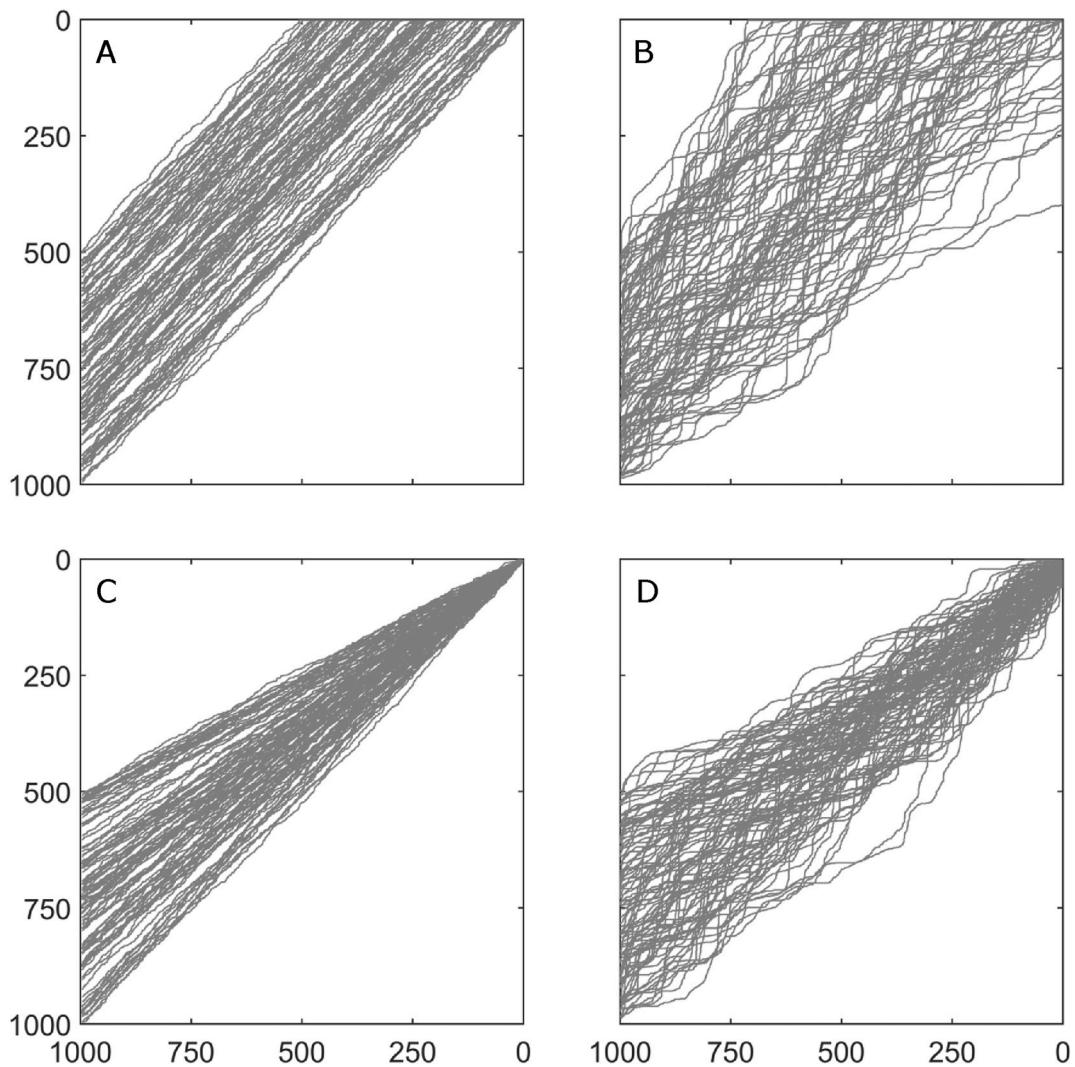


Fig. 2. Illustration of the transition factors *path-steer* and *path-momentum*. **A:** All three transition positions are equally as likely and the paths appears as straight lines in the alignment plot ($ms = 0$ and $ss = 0$), **B:** Adding the *path-momentum* factor to the transition probability ($ms = 1$ and $ss = 0$) causes the paths to follow a more auto-correlated slope, **C:** Adding the *path-steer* factor to the transition probability ($ms = 0$ and $ss = 1$) biases the paths towards the (1,1) position, **D:** Adding both the *path-steer* and *path-momentum* factors to the transition probability ($ms = 1$ and $ss = 0.25$) results in paths that reflect a combination of correlated slope and bias towards the (1,1) position. The *momentum-length* is set to ten ($ml = 10$). The initial path positions are between 500 and 1000 at the at the y-axis. A total of 100 paths are plotted in each panel.

2. Methodology

2.1. Introduction to dynamic time warping

The DTW algorithm operates in two steps. First, it calculates a two-dimensional cost-matrix, \mathbf{C} , where the dimensions correspond to the lengths of the two signals being compared (note that bold symbols represent vectors or matrices, while italicized symbols denote functions or scalars.). Each cell in \mathbf{C} is given by:

$$C_{n,m} = |a_n - b_m|^\Omega + \min\left(\begin{array}{c} C_{n,m-1} \\ C_{n-1,m-1} \\ C_{n-1,j} \end{array}\right), \quad (1)$$

where \mathbf{a} and \mathbf{b} are signals of lengths N and M , respectively. Here, $n = 1, 2, \dots, N$ and $m = 1, 2, \dots, M$, representing positions in the cost-matrix with the size N times M . The term $|a_n - b_m|^\Omega$ represents the distance between \mathbf{a} and \mathbf{b} at the position (n, m) raised to the power of a factor Ω . The minimum function selects the smallest cost from the neighboring cells in \mathbf{C} , at $(n, m-1)$, $(n-1, m-1)$ or $(n-1, m)$.

Once the cost-matrix is computed, the optimal alignment path is determined by tracing back from the endpoint (N, M) to the top point $(1, 1)$ following the path of minimum cost through \mathbf{C} . At each step, the algorithm moves from the current position (n, m) to the neighboring position of $(n-1, m)$, $(n-1, m-1)$, or $(n, m-1)$, that has the minimum cost. This process ensures that the total cost of the alignment path is minimized. The resulting path represents the optimal alignment, allowing for stretching or compressing of the sequences as needed.

2.2. Probabilistic dynamic time warping

We follow the approach of Lallier et al. (2012, 2016) to sample the uncertainty associated with signal alignment. Stochasticity is introduced to the process of defining the alignment-path through the cost-matrix, \mathbf{C} , rather than altering the cost-matrix itself. This allows generation of multiple realizations of the alignment-path which represent the uncertainty of the alignment, analogous to the methods introduced by Pels et al. (1996). While Lallier et al. (2016) and Pels et al. (1996) focus on aligning sequences of sedimentary facies and palynological records, respectively, our approach mainly targets the alignment of continuous,

or semi-continuous stratigraphic signals.

In contrast to the classical DTW, where the alignment-path is determined by transitioning from the current position (n, m) to the position with the lowest cost among $(n-1, m)$, $(n-1, m-1)$, or $(n, m-1)$ in the cost-matrix, our method assigns a probability to each of these transitions. We denote the probability of transitioning to $(n-1, m)$, $(n-1, m-1)$, or $(n, m-1)$ by $P(n-1, m)$, $P(n-1, m-1)$, and $P(n, m-1)$, respectively. The total probability of transitioning to any of these positions must sum to one, such that $P(n-1, m) + P(n-1, m-1) + P(n, m-1) = 1$.

In the following sections, we introduce three factors that influence these transition probabilities, which collectively generate geological founded alignment-paths through the cost-matrix. These paths represent the alignment model, reflecting the uncertainty of the stratigraphic alignment.

2.2.1. Factor one: the cost-matrix

The most critical factor influencing the transition probabilities is the cost-matrix (Eq. (1)) and, by extension, the amplitude of the stratigraphic record or signal. We assume that transitioning to the position $(n-1, m)$, $(n-1, m-1)$, or $(n, m-1)$ with the lowest cost is most likely, meaning that transition probabilities should be inversely related to these costs. For most cases, the costs at the potential transition positions are relatively similar in size. This prevents a direct conversion of costs into probabilities as this would make each transition almost equally probable. To address this, we propose defining transition probabilities based on the relative differences in cost by introducing transition-weights (\mathbf{Mp}) based on the cost-matrix and the parameter p :

$$\mathbf{Mp} = (\mathbf{CT} - \text{minimum}(\mathbf{CT}) + (\text{median}(\mathbf{CT}) - \text{minimum}(\mathbf{CT}))(1-p))^{-1}, \quad (2)$$

where \mathbf{CT} is a vector including the three costs at the transition positions $\mathbf{CT} = [C(n-1, m), C(n, m-1), C(n-1, m-1)]$. The *minimum* function returns the smallest value in the \mathbf{CT} vector, while the median function returns the middle value, or median, of the \mathbf{CT} . The probability of transitioning to either $(n-1, m)$, $(n-1, m-1)$, or $(n, m-1)$ is then calculated by normalizing the transition-weights: $\mathbf{P} = \mathbf{Mp} / \sum_{i=1}^3 \mathbf{Mp}_i$.

The parameter p (typically in the range 0–1) controls the influence of the cost-matrix on the alignment-path. When p is small, the path becomes less dependent on the cost-matrix because the transition probabilities for the three possible positions become more evenly distributed, resulting in a more scattered and random alignment model. In contrast, when p approaches one, the path becomes more strongly influenced by the cost-matrix because the transition probability increasingly favors the position with the lowest cost. This leads to less scattered, more narrow alignment models. In the extreme case where $p = 1$, the transition will always occur at the position with the lowest cost ($P(\text{minimum}(\mathbf{CT})) = 1$), equivalent to the classical deterministic DTW algorithm. In the extreme case where $p = -\infty$, the cost-matrix has no influence on the alignment.

2.2.2. Factor two: path-momentum

During noisy or stationary stratigraphic intervals, the signals display no patterns or trends to guide the alignment. In such cases, the transition-weights, \mathbf{Mp} , become nearly identical on average, leading to alignment-paths that almost appear as noisy straight lines in the depth-depth domain (Fig. 2A). Therefore, the alignment-path must be guided by other factors than the cost-matrix during such intervals to overcome this unrealistic behavior. For this purpose, we assume that abrupt changes in the alignment are unlikely. Therefore, we implement a transition-weight, called *path-momentum*, \mathbf{Mm} , that favors the current slope of the path in the depth-depth domain:

$$\mathbf{Mm} = \left[\frac{n - n_{ml}}{ml}, 1, \frac{m - m_{ml}}{ml} \right]^{ms}, \quad (3)$$

where (n, m) is the current position and (n_{ml}, m_{ml}) is the position ml steps back in the path sequence, and ms controls the strength of the path-

momentum. When ms is zero, the path-momentum has no influence on the transition probability as \mathbf{Mm} becomes $[1, 1, 1]$. ml controls the memory of the path-momentum resulting in smoother paths as ml increases.

This modification offers three advantages. First, it allows the path to maintain a consistent slope during stationary or noisy intervals, avoiding the unlikely consistent 1:1 slope shown in Fig. 2A. Secondly, the path-momentum factor makes the algorithm less sensitive to outliers, hereby reducing the likelihood of abrupt and unrealistic changes in the alignment-path and hence the accumulation rate. Wheeler and Hale (2014) and later Sylvester (2023) suggested using $\Omega < 1$ in equation (1) to reduce the impact of outliers in aligning gamma-log-well curves. However, due to the stabilizing effect of the path-momentum factor in our approach, we maintain $\Omega = 1$, as the momentum factor already reduces the influence of outliers without requiring adjustment to the cost function. Lastly, geological records might vary in amplitude over the same chronology due to local factors. For instance, two magnetic susceptibility records might be influenced by local environmental conditions, leading to non-proportional amplitude variations over the same chronology. This stands in contrast to DTW, which assumes the optimal alignment is the one with the lowest cumulative distance between the signals. By incorporating the path-momentum factor, we move beyond the assumption that alignment is solely driven by minimizing cumulative distance. This provides an alternative and more flexible framework for aligning stratigraphic series.

2.2.3. Factor three: path-steer

We assume the uppermost depths of the sequences correspond to the surface of the sites and should therefore correlate. To account for this, we introduce a third transition-weight called *path-steer* that favors alignment paths to steer toward the $(1, 1)$ position:

$$\mathbf{Ms} = \left[\frac{n}{m}, 1, \frac{m}{n} \right]^{ss}, \quad (4)$$

where (n, m) is the current position and ss controls the strength of \mathbf{Ms} . This ensures that when no other factors significantly influence the transition probabilities (e.g. when $\mathbf{Mp} \sim \mathbf{Mm} \sim [1, 1, 1]$) the path tends toward $(1, 1)$ (Fig. 2C).

Finally, the probability of transitioning to either $(n-1, m)$, $(n-1, m-1)$, or $(n, m-1)$ is calculated by multiplying the three transition weights together and normalizing them: $\mathbf{P} = (\mathbf{Mp} \cdot \mathbf{Mm} \cdot \mathbf{Ms}) / \sum_{i=1}^3 \mathbf{Mp}_i \cdot \mathbf{Mm}_i \cdot \mathbf{Ms}_i$. This combined approach ensures that the alignment path is driven by both the signal characteristics, a tendency to maintain correlation at the surface, and is aligned with the expected amount of abrupt changes in the alignment path.

2.2.4. Initial path position

The initial position of each alignment path is selected based on the following equation:

$$\mathbf{Mi} = (\mathbf{Ci} - \text{minimum}(\mathbf{Ci}) + (\text{median}(\mathbf{Ci}) - \text{minimum}(\mathbf{Ci}))(1-p))^{-1}, \quad (5)$$

where \mathbf{Ci} represents the values in the cost-matrix at all possible initial positions within a chosen range. For example, if the bottom of signal **a** is younger than the bottom of signal **b**, then the bottom of signal **a** should correspond to signal **b** within a given interval x of signal **b**, so that $\mathbf{Ci} = [C(N, x)]$. Alternatively, if the relative age of the signals is unknown, then $\mathbf{Ci} = [C(N, x) \ C(y, M)]$, where y is a given interval of signal **a** of which the bottom of signal **b** might correspond. Equation (5) is similar to Equation (2) but uses \mathbf{Ci} instead of \mathbf{Ct} , ensuring that the initial position with the lowest cost value is the most likely starting point. Similarly, when $p = 1$, only the position with the lowest cost is selected. Conversely, when $p = -\infty$, all possible initial positions, \mathbf{Ci} , are equally probable (Fig. 2A). This approach sets the starting point of the alignment path in a way that reflects the relative costs, with p controlling the degree of preference for the lowest cost position.

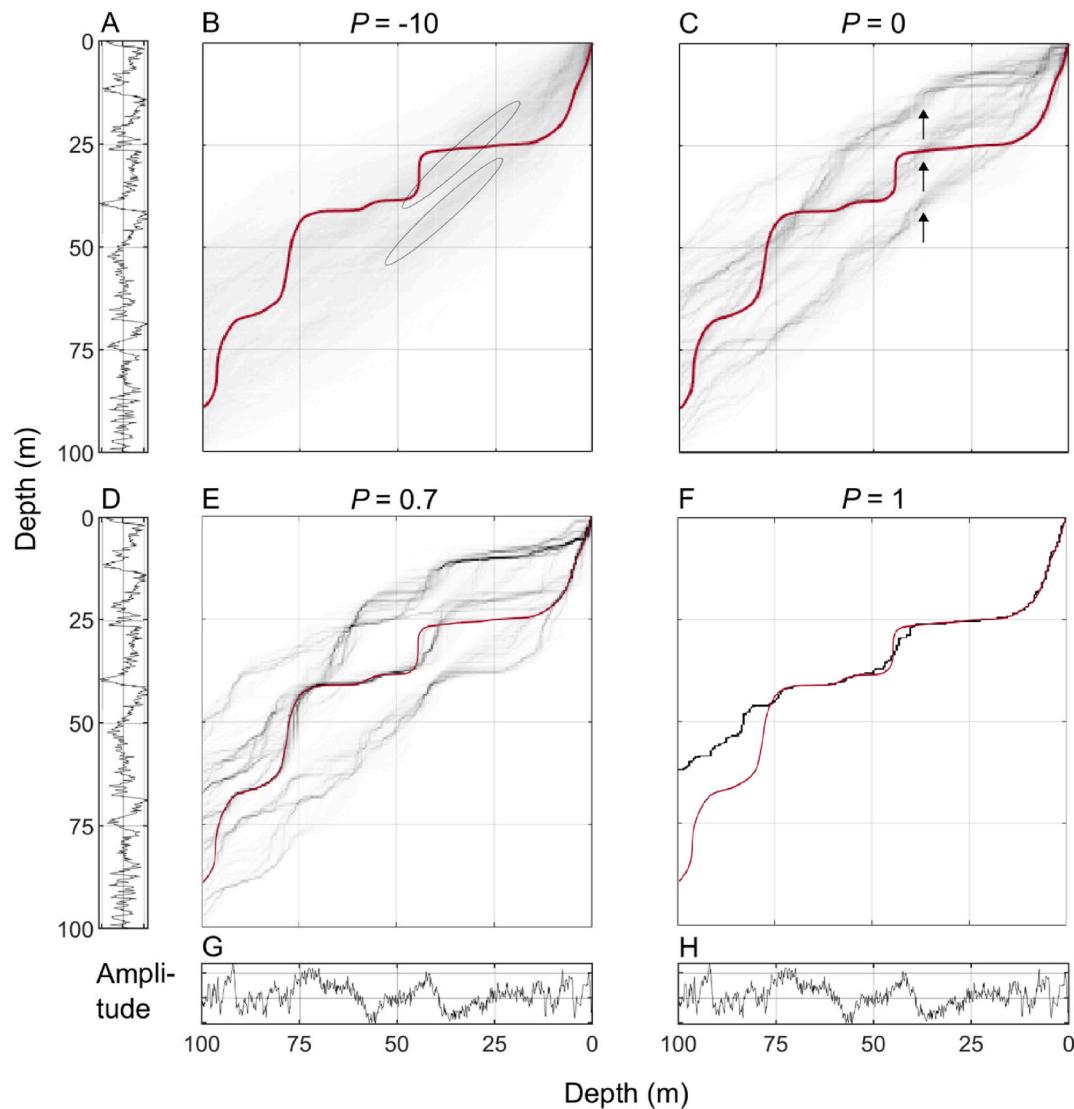


Fig. 3. Test on synthetic data 1: testing the influence of the parameter p (Eq. (2)) on the P-DTW alignment model using the same synthetic test signals and four different values of p . **B, C, E, F:** grayscale represents the P-DTW alignment models with values of p indicated in the title of each panel, and the red curves are the reference-alignments. The black ellipses indicate two denser areas of the alignment model and the black arrows indicate three distinct possible alignments. The parameters ms , ml and ss are fixed at 1, 10, and 0.25, respectively for all models. The initial alignment path positions are between 50 and 100 m on the depth scale of signal a. **A, D:** signal a, **G, H:** signal b.

2.3. Interpolation of the signals

DTW performs point-to-point correlations, requiring each data point in one signal to correspond to a data point in the other signal. This can lead to skewed alignments, particularly around intervals where data are sparse. To address this, we interpolate the datasets to a high resolution with equal depth spacing. This approach ensures that the alignment paths appear smooth and continuous in the depth-depth domain where data is sparse. A sensitivity analysis demonstrates that increasing the interpolation resolution beyond the original sampling length does not necessarily enhance the performance of the algorithm (Fig. S1).

Extended data gaps exceeding the path-momentum length, ml , might counteract the effect of the path-momentum, since noise is effectively correlated across multiple points in those intervals due to the interpolation. However, datasets including such intervals without data are expected to be rare and may not be suitable for alignment. Alternatively, these gaps could be filled using the mean or median value of the signal instead of interpolation, to construct a more natural signal.

2.4. Tie points

Often, independent constraints on the alignment, such as paleomagnetic or biostratigraphic boundaries, are available. These constraints can be incorporated into the algorithm as tie points or overlapping intervals between the signals. This is achieved by modifying the cost-matrix so that areas inconsistent with these constraints are assigned an infinite cost following Lallier et al. (2016). This adjustment ensures that the alignment path will avoid these regions, maintaining consistency with the imposed constraints. While we use infinite costs for hard constraints, soft constraints could potentially be applied by scaling the costs of inconsistent areas to reflect the uncertainty of the constraints.

2.5. Implications for age-depth modelling

The CosmoChron age-depth modelling framework (Sørensen et al., 2024) is a probabilistic inverse modelling approach that constructs age-depth relationships for sedimentary sequences by integrating

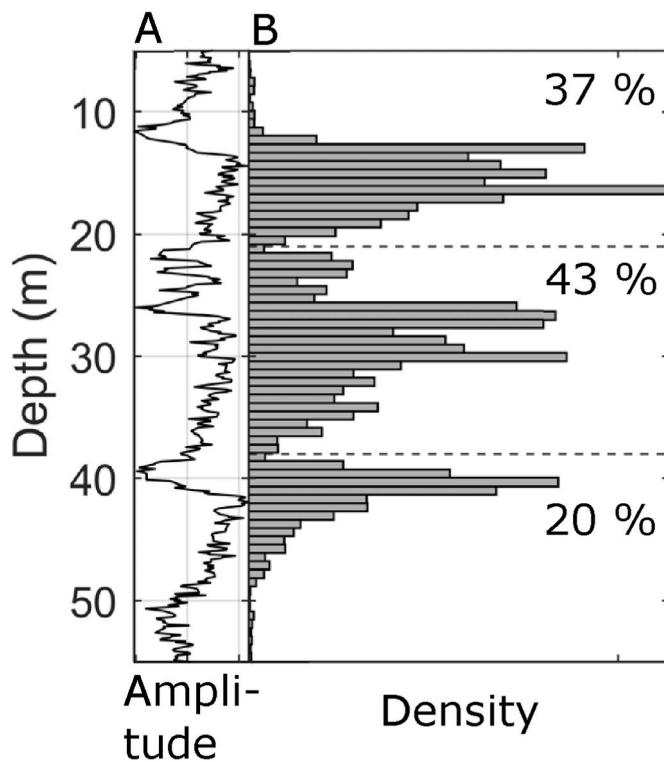


Fig. 4. The correspondence of signal **b** at 40 m depth to signal **a** (panel A), marked with black arrows in Fig. 3C. A: signal **a**, B: a 1D representation of the 2D P-DTW alignment model shown on Fig. 3C. Three distinct alignments emerge when the value of $p = 0$, with their respective probabilities shown in the figure and separated with dashed lines.

cosmogenic nuclide concentrations and direct age constraints. The CosmoChron method follows the probabilistic formulation of Tarantola (2005), where model parameters \mathbf{m} are constrained based on observed data \mathbf{d}_{obs} through a posterior probability distribution, $\sigma(\mathbf{m})$, that combines prior knowledge, $\rho(\mathbf{m})$, with data likelihood, $L(\mathbf{m})$, through $\sigma(\mathbf{m}) = k\rho(\mathbf{m})L(g(\mathbf{m}))$. In its original formulation, the CosmoChron model parameters are $\mathbf{m} = [\Theta, T_0, \mathbf{h}, \mathbf{p}]$, where Θ represents a series of accumulation rates described by correlated gamma distributions, T_0 is the age at the uppermost depth, \mathbf{h} denotes hiatus durations, and \mathbf{p} encompasses pre-burial parameters for cosmogenic nuclide samples. The forward model $\mathbf{d} = g(\mathbf{m})$ computes age-depth relationships by integrating accumulation rates over depth intervals, while accounting for cosmogenic nuclide production and decay processes.

We extend the CosmoChron framework to enable multi-site age-depth modelling by incorporating an additional model parameter $I_{\text{P-DTW}}$, representing the index of alignment paths generated by the P-DTW algorithm. The extended model parameters become $\mathbf{m} = [\Theta, T_0, \mathbf{h}, \mathbf{p}, I_{\text{P-DTW}}]$. The outcome of the P-DTW method presented above directly represents the prior distribution $\rho(I_{\text{P-DTW}})$.

For multi-site scenarios, the forward model computes age-depth relationships where the oldest site derives ages by integrating accumulation rates, while the youngest site's age-depth relationship is calculated by transposing the oldest site's chronology through the P-DTW alignment path specified by $I_{\text{P-DTW}}$. We employ the extended Metropolis algorithm (Mosegaard and Tarantola, 1995) to sample the posterior distribution following Sørensen et al. (2024). This routine is implemented using the SIPPI toolbox for MATLAB (Hansen et al., 2013).

3. Results

3.1. Tests using synthetic data

We apply the P-DTW algorithm to synthetic data to assess its performance across various realistic scenarios and to evaluate the influence of different parameters. Figs. 3 and 5 illustrate how the resulting alignment models are affected by different values of p and ml , respectively. Fig. 6 highlights the fact that the optimal DTW alignment does not necessarily reflect the 'true' reference-alignment. Fig. 7 presents a scenario where the signals exhibit periodic amplitude variations over the same chronology. Lastly, Fig. 8 highlights the impact of tie-points from e.g. biostratigraphic or magnetostratigraphic constraints.

All alignment models are displayed as 2D histograms of 10000 alignment paths generated using the P-DTW algorithm to align signals **a** and **b**. The signals are generated by warping the LR04 $\delta^{18}\text{O}$ stack (Lisiecki and Raymo, 2005) and adding noise to each signal. The warping involves generating an age-depth curve using the CosmoChron forward model (Sørensen et al., 2024) and then transforming the LR04 stack from a linear scale in age to a linear scale in depth based on this age-depth curve. A known reference-alignment curve can be established based on the age-depth relationships of two given signals, including the correspondence between the bottom of the youngest signal and the oldest signal.

The noise is added by $\mathbf{s} + \mathbf{n}$, where \mathbf{s} represents the warped LR04 $\delta^{18}\text{O}$ stack, and \mathbf{n} is Gaussian noise. Uncorrelated Gaussian noise with a standard deviation of $0.2 \delta^{18}\text{O} \text{‰}$ is added to the signals shown in Figs. 3–5 and 8 while uncorrelated Gaussian noise with a standard deviation of $0.3 \delta^{18}\text{O} \text{‰}$ is added to the signals shown in Fig. 6. Correlated noise with a standard deviation of $0.2 \delta^{18}\text{O} \text{‰}$ is added to the signals shown in Fig. 7. The LR04 $\delta^{18}\text{O}$ stack itself has a 1σ of $\sim 0.38 \text{‰}$, implying that the added noise corresponds to $\sim 7 \%$ of the total signal variance. The noise models used for each synthetic dataset are summarized in Table 1.

The correlated noise is defined by a multivariate Gaussian distribution with covariance $\mathbf{C}_R = \exp(-3\mathbf{D}^2/R^2)$, where \mathbf{D} is a distance matrix representing the distance between each depth of the signals in meters, and $R = 20 \text{ m}$ sets the correlation length. We generate realizations of the multivariate distribution using the Cholesky decomposition method (Deutsch and Journel, 1992).

Each signal consists of 1000 data points interpolated from a dataset where the depths of the samples are generated from a uniform distribution along the profile. This introduces a small uncertainty in the signals, reflecting more realistic datasets where perfect depth spacing is unattainable. The original LR04 $\delta^{18}\text{O}$ stack for each signal has a length of about 800 data points depending on the age of the signal.

The individual settings of p , ms , ml and ss and the initial alignment path position can be found in the figure caption related to each synthetic test case and in Table 1. For all synthetic datasets signal **a** is longer (older) than signal **b**, and the initial alignment path position is therefore set at an interval a signal **a**. Note that this does not always have to be the case.

For each synthetic test, we evaluate the resulting alignment models by comparing them to the 'true' alignment, referred to as the reference-alignment. The overall performance of the algorithm is assessed using the coverage probability, defined as the integral of the probability density function (PDF) over the reference-alignment curve. The coverage probability quantifies how much probability mass the P-DTW alignment model assigns to the reference-alignment, with higher values indicating better agreement between the model and the true alignment. It is important to assess the coverage probability in conjunction with a visual representation to fully capture the model's performance. This is because the alignment may match well in some intervals but perform poorly in others which is not captured by a single coverage probability value. We consider a model to be consistent with the reference-alignment if the reference-alignment does not deviate significantly from the

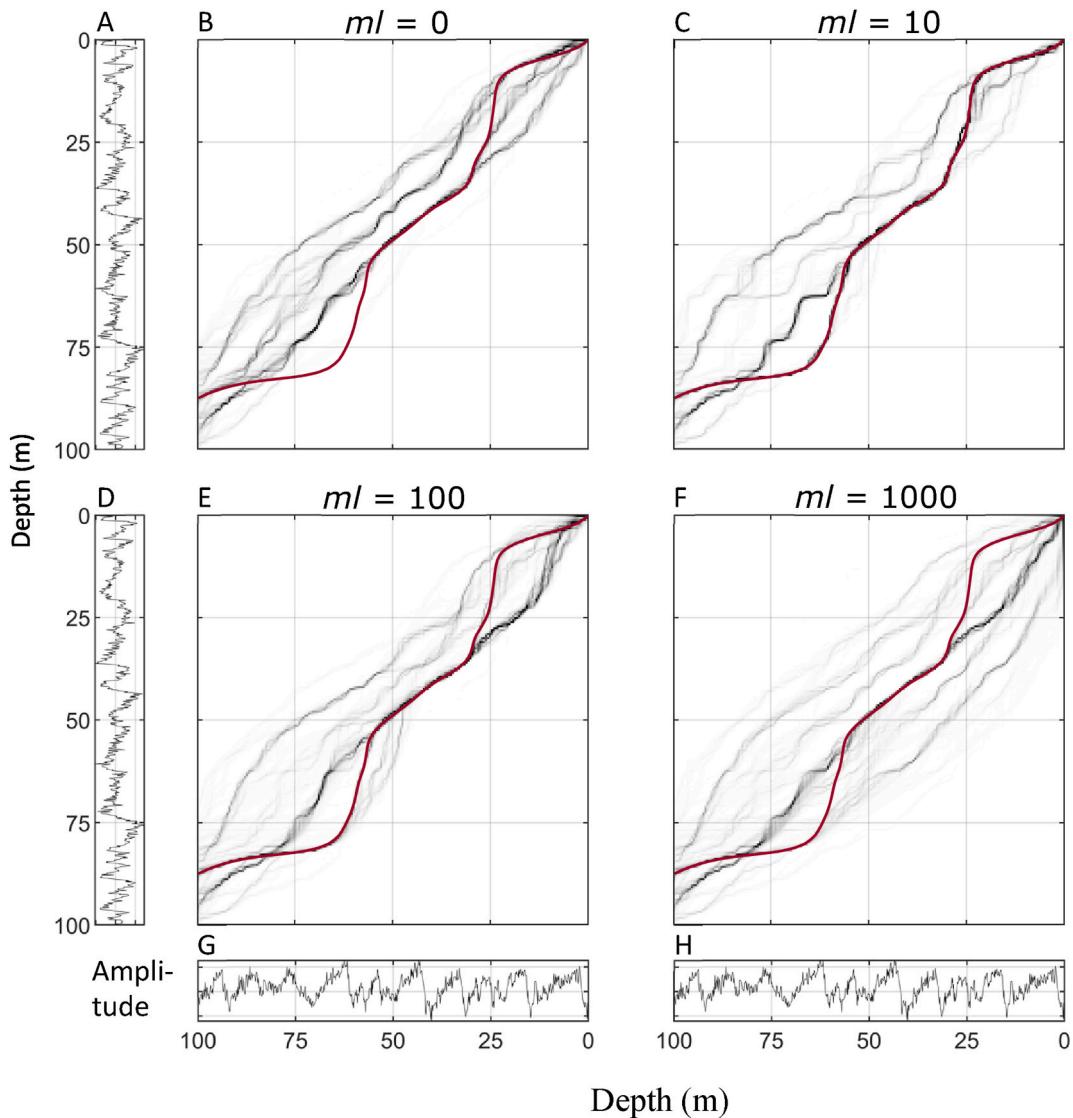


Fig. 5. Test on synthetic data 3: testing the influence of the ml -value (Eq. (3)) on the P-DTW alignment model using the same synthetic test signals and four different values of ml . **B, C, E, F:** grayscale represents the P-DTW alignment models with ml -values indicated in the title of each panel, and the red curves are the reference-alignments. The parameters ms , p and ss are fixed at 1, 0.7, and 0.25, respectively for all models. The initial alignment path positions are between 75 and 100 m on the depth-scale of signal **a**. **A, D:** signal **a**, **G, H:** signal **b**.

alignment model in any interval. In particular, the coverage probability should never be zero.

3.1.1. Test on synthetic data 1: the influence of p

We test the influence of the parameter p (Eq. (2)) on the alignment model by applying the P-DTW algorithm to the same two synthetic test signals while using four different values of p (-10, 0, 0.7, and 1) (Fig. 3).

At $p = -10$ (Fig. 3B), the alignment model appears blurred in the depth-depth domain, indicating high uncertainty in the alignment. Despite this scatter, two denser areas, indicated with black ellipses, suggest potential correlations between the local minima at ~ 38 m in signal **a** (y-axis) and depths of ~ 37 or ~ 57 m in signal **b** (x-axis). Although the alignment is uncertain, the model remains consistent with the reference-alignment, yielding a coverage probability of 0.037.

As the value of p increases to 0, the alignment model shows greater structure, indicating reduced uncertainty. Three distinct parallel maxima emerge, suggesting multiple possible correlations between the signals. For instance, the model suggests that the peak at ~ 40 m in signal **b** could correspond to peaks at either ~ 15 , ~ 30 , or ~ 40 m in signal **a** with a similar probability as indicated with black arrows on Fig. 3C and

shown on Fig. 4. This pattern illustrates how different alignments are possible, generating distinct maxima in the alignment model. The reference-alignment lies within the one of these maxima, and the coverage probability nearly doubles to 0.070 as p increases from -10 to 0.

At $p = 0.7$, the three maxima become narrower, reflecting further reduced uncertainty in the alignment. For example, the model now suggests that the peak at ~ 40 m in signal **a** most likely aligns with the peak at ~ 37 m in signal **b** (Fig. 3E). The model performance improves accordingly, with a coverage probability of 0.11.

Finally, with $p = 1$, the alignment model converges to a single alignment (Fig. 3F), hereby eliminating the uncertainty, which corresponds to the deterministic alignment produced by the standard DTW algorithm. However, the model is not consistent with the reference-alignment, especially in the lower part of signal **b** (75–100 m) as the coverage probability drops to zero in those intervals. In practice, this is problem is encountered whenever $p = 1$, suggesting it is problematic to use a deterministic DTW model.

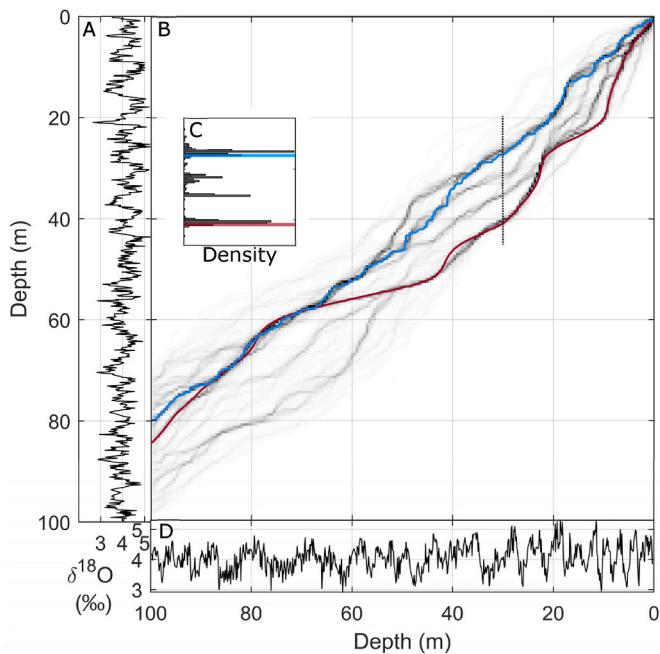


Fig. 6. Test on synthetic data 3: comparing the P-DTW-alignment model and the optimal DTW-alignment to the reference-alignment. A: Signal a, B: grayscale is the P-DTW the alignment model using a p , ms , ml and ss value of 0.5, 1, 10 and 0.25, respectively, while the initial alignment path positions are 70–100 m at the depth-scale of signal a. The blue curve is the optimal DTW-alignment ($p = 1$), and the red line is the reference-alignment, C: P-DTW alignment model of signal b at 30 m (dashed vertical line in panel B) aligned to signal a, blue and red lines are equal to those in panel B, D: signal b. Note that another visualization of this P-DTW alignment model is depicted in Fig. 1.

3.1.2. Test on synthetic data 2: the influence of path-momentum

We test the influence of path-momentum on the alignment model by applying the P-DTW algorithm to the same set of synthetic test signals while using four different ml -values (0, 10, 100 and 1000) in Equation (3)) (Fig. 5).

At $ml = 0$ (Fig. 5B), the alignment model exhibits three distinct maxima in the depth-depth domain, none of which align with the reference-alignment (red curve) for the ~60–80 m interval of signal b. Increasing the ml -value to 10 yields an alignment model closely matching the reference-alignment (Fig. 5C). At $ml = 100$, the model remains largely consistent with the reference-alignment, but diverges more, particularly in aligning the ~60–80 m and ~10–30 m intervals of signal b (Fig. 5E vs. 5C). Finally, at $ml = 1000$, the alignment-model becomes overly smoothed, producing curvy lines in the depth-depth domain similar to those observed with $ml = 0$ (Fig. 5F vs. 5B).

These trends are also reflected in the algorithm's performance, as measured by the coverage probability, which initially increases from 0.11 at $ml = 0$ to a peak of 0.22 at $ml = 10$, before decreasing to 0.14 at $ml = 100$ and further to 0.089 at $ml = 1000$.

These results show the importance of path-momentum as a tuning factor. While excessively high ml -values (>~10 % of the length of the data series, Fig. 5E–F) or none at all (Fig. 5B) result in overly smoothed models that miss alignment trends, while using a ml -value of 10 (1 % of the length of the data series) provides the most reliable alignment model with the largest coverage probability of 0.22 (Fig. 5C).

3.1.3. Test on synthetic data 3: comparison to the deterministic DTW algorithm

In this test, we apply both P-DTW and the classic deterministic DTW algorithm to the same synthetic test signals to compare the resulting alignments with the reference-alignment.

The signals displayed in Fig. 6A and C appear noisy and would likely

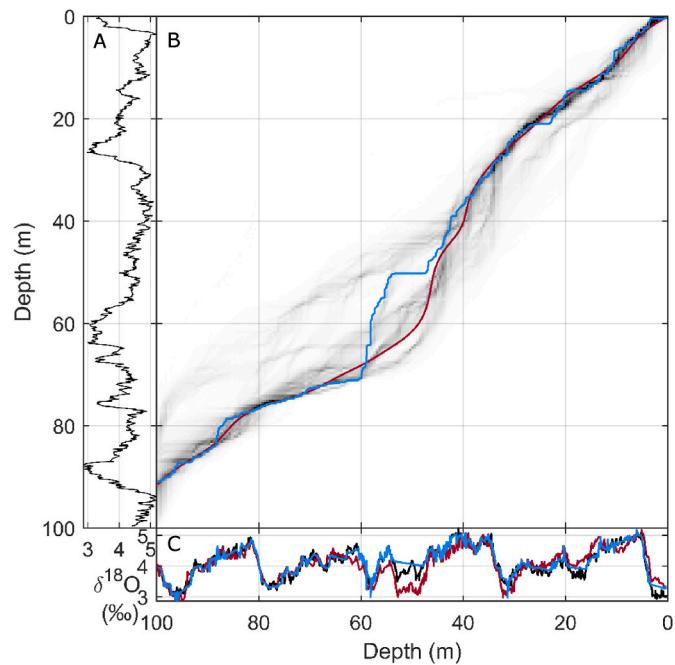


Fig. 7. Test on synthetic data 4: Aligning signals that consistently deviate in amplitude over the same chronology, A: signal a, B: grayscale represent the P-DTW alignment model using a p , ms , ml and ss value of 0, 1, 10, and 0.25, respectively, while the initial alignment path positions are 80–100 m at the depth-scale of signal a. The blue curve is the optimal DTW-alignment ($p = 1$), and the red curve is the reference-alignment curve, C: black curve is signal b, blue and the red curves are signal a transposed onto the depth-scale of signal b using the DTW- (blue curve on panel B) and the correct reference-alignment (red curve on panel B), respectively.

be challenging to align by eye. However, the P-DTW algorithm reveals a range of possible alignments, represented as multiple distinct maxima in the alignment model. Importantly, one of these maxima is consistent with the reference-alignment (red curve on Fig. 6B–C and Fig. 1).

In contrast, the optimal DTW-alignment only produces a single solution that diverges considerably from the reference-alignment, illustrating its limitations in handling noisy data (Figs. 1 and 6).

This test highlights the importance of incorporating uncertainty in the alignment of noisy stratigraphic sequences. Relying solely on the deterministic DTW algorithm can lead to alignments that deviate significantly from the true stratigraphic correlation. The P-DTW approach offers a probabilistic range of alignments that reflect the uncertainty of aligning stratigraphic signals.

3.1.4. Test on synthetic data 4: signals deviate in amplitude over the same stratigraphy

In this fourth test, we evaluate the performance of the P-DTW algorithm in aligning two signals that exhibit persistent amplitude deviations along the same chronology. Traditional DTW struggles with such cases because the cost-matrix (Eq. (1)) relies on the residuals between signals, resulting in alignment errors. To address this, we examine how the P-DTW algorithm handles persistent residuals by introducing correlated noise to the original signals. In Section 3.1, we describe how the correlated noise is included.

As shown in Fig. 7C, the synthetic signals exhibit varying amplitudes, particularly ~50–55 m at the depth-scale of signal b, where the $\delta^{18}\text{O}$ amplitude difference exceeds 0.5 % (red vs. black curves). Despite these amplitude deviations, the P-DTW alignment-model remains consistent with the reference-alignment, although it scatters in the depth-depth domain reflecting a large uncertainty in aligning signals with persistently diverging amplitudes.

In contrast, the DTW alignment displays sharp transitions in the

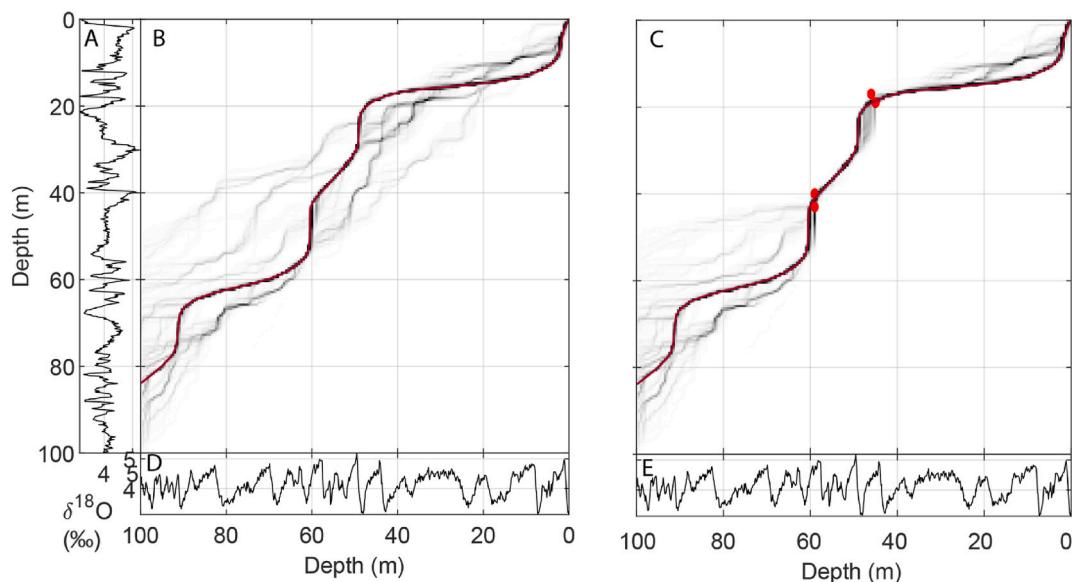


Fig. 8. Test on synthetic data 5: Influence of tie points. A: Signal a, B: no tie points are added, grayscale is the P-DTW the alignment models using a *p*, *ms*, *ml* and *ss* value of 0.5, 1, 10 and 0.25, respectively, while the initial alignment path positions are 50–100 m at the depth-scale of signal a C: Two tied intervals are added (marked with red dots) constraining that a depth within 40–43 m and 17–19 m at the depth-scale of signal a must correspond to 59 m and a depth within 45–46 m, respectively, at the depth-scale of signal b, grayscale and model parameters are equivalent to those in panel B, D-E: signal b.

Table 1

Noise model used for each synthetic dataset and the P-DTW parameter settings. *p*: influence of the cost-matrix, *ms*: path-momentum strength, *ml*: path-momentum length and *ss*: path-steer strength.

Index of synthetic data	Figure	Reference data noise model	Reference data noise level (1σ)	<i>p</i>	<i>ms</i>	<i>ml</i>	<i>ss</i>
1	3	Uncorrelated Gaussian	0.2	-10, 0, 0.7, 1	1	10	0.25
1	4	Uncorrelated Gaussian	0.2	0	1	10	0.25
2	5	Uncorrelated Gaussian	0.2	0.7	1	0, 10, 100, 1000	0.25
3	6	Uncorrelated Gaussian	0.3	0.5	1	10	0.25
4	7	Correlated Gaussian with $R = 20$ m	0.2	0	1	10	0.25
5	8	Uncorrelated Gaussian	0.2	0.5	1	10	0.25

depth-depth domain and is not consistent with the reference-alignment. Particularly at ~50 m at the depth-scale of signal a, where the amplitude difference between the two signals is large, the alignment falsely suggests a significant hiatus in signal a. The DTW algorithm is overfitting the alignment, which is reflected in Fig. 7C (blue curve), where signal a appears to be ‘glued’ to signal b (blue vs the black curve). This overfit results from DTW’s strict reliance on amplitude matching.

The P-DTW algorithm, with its inclusion of the *path-momentum* (Eq. (3)), prevents sharp changes in the accumulation rate. The *path-momentum* factor helps the algorithm maintain a smoother and more realistic alignment, also when amplitude differences are persistent.

3.1.5. Test on synthetic data 5: the influence of tying intervals together

Fig. 8 demonstrates the effect of forcing two intervals to overlap in the alignment-model. This inclusion eliminates multiple optima in the alignment-model, especially near and between the tied intervals, and the coverage probability consequently increases from 0.26 to 0.31 (Fig. 8B vs. 8C).

These tie points, or tied intervals, can be derived from sources such as biostratigraphic or paleomagnetic boundaries, which provide constraints that enhance the alignment model.

3.2. Real world test cases

3.2.1. Real world test case 1: applying the P-DTW algorithm to signals with a known alignment

To evaluate the performance of the P-DTW algorithm, we apply it to real data with published age-depth relationships. This allows us to

construct a reference-alignment curve against which the results of the P-DTW alignment can be compared. We use two benthic (*Cibicidoides wuellerstorfi*) $\delta^{18}\text{O}$ records from sediment cores GeoB7920-2 and MD95 2042 which are located relatively close to each other in the North Atlantic.

The alignment shown in Fig. 9 (red curve) is constructed by plotting the age-depth model of GeoB7920-2 against the one of MD95 2042. The age-depth model of GeoB7920-2 is created by aligning the $\delta^{18}\text{O}$ curve of GeoB7920-2 to the one of MD95 2042 and using the GRIP ss09sea age scale of MD95 2042 (Shackleton et al., 2004; Tjallingii et al., 2008). The alignment was performed by eye and by peak-to-peak correlation using the software AnalySeries (Paillard et al., 1996; Tjallingii et al., 2008). Meanwhile, the age-model of MD95 2042 is constructed by Lisiecki et al. (2021).

It is important to note that the age-depth model of GeoB7920-2 is derived through correlation to MD95-2042, meaning that our comparison between the P-DTW alignment and the reference curve is an indirect validation against this established alignment.

Fig. 9 shows that the P-DTW alignment model is consistent with the established alignment (red curve). Similarly, the deterministic DTW alignment is also consistent with the established alignment for most parts (blue vs red curve in Fig. 9B). Notably, when aligning the peaks spanning ~5–7 m in MD95 2042 with the peak spanning ~2.5–3.5 m in GeoB7920-2 the DTW alignment deviates from the red alignment. This is because the peak in MD95 2042 has a persistently larger amplitude compared to the peak in GeoB7920-2. Nonetheless, the P-DTW alignment model suggests that both the established alignment (red curve) and the deterministic DTW alignment (blue curve) are similarly plausible for

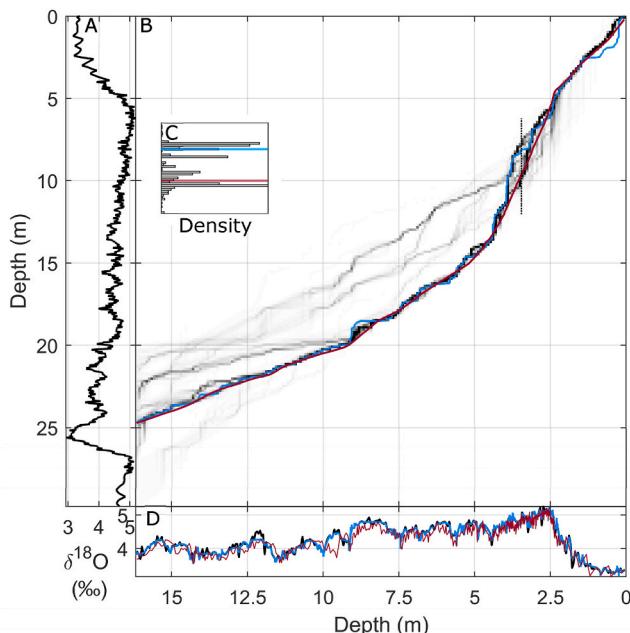


Fig. 9. Test on real data 1: Applying the P-DTW algorithm to signals that have already been aligned, A: $\delta^{18}\text{O}$ curve from sediment core MD95 2042 (Shackleton et al., 2000), B: Greyscale represent the P-DTW alignment model with settings: p , ms , ml , ss equal to 0.8, 1, 500, 0.25, respectively. Red curve is the alignment based on the age-depth models from Lisiecki et al. (2021) and Tjallingii et al. (2008), the blue curve is the classic DTW alignment ($p = 1$), C: P-DTW alignment model of signal in panel B a 2.7 m (dashed vertical line in panel B) alignment to the signal shown in panel A, blue and red lines are equal to those in panel B, D: black curve is the $\delta^{18}\text{O}$ curve from sediment core GeoB7920-2 (Tjallingii et al., 2008) (+0.7 ‰ for scaling), blue and the red curves are the $\delta^{18}\text{O}$ curve from MD95 2042 transposed onto the depth-scale of GeoB7920-2 using the classic DTW (blue curve on panel B) and the alignment based on the age-depth models from Lisiecki et al. (2021) and Tjallingii et al. (2008) (red curve on panel B), respectively.

these peaks.

The P-DTW alignment model mostly displays dense probability around the red alignment with only a few other distinct maxima appearing along the alignment model. Beside the ‘peak’-alignment, these other maxima mostly reflect an uncertainty in the initial alignment path position suggesting that the end of GeoB7920-2 could respond to either ~21 or ~23 m but most likely 24.8 m at MD95 2042.

This test shows that the established alignment by Tjallingii et al. (2008) is the most likely, and the P-DTW alignment model is consistent with it. Moreover, the test shows the presence of alternative plausible alignments, reflecting the inherent uncertainty in the alignment.

3.2.2. Real world test case 2: aligning magnetic susceptibility from the ENAM93-21 core to $\delta^{18}\text{O}$ from the NorthGRIP Greenland ice core

To evaluate the performance of the P-DTW algorithm in aligning different types of stratigraphic signals, we tested it on the magnetic susceptibility record from the ENAM93-21 core (located on the NE Faeroe Margin; Rasmussen et al., 1996) and the $\delta^{18}\text{O}$ record from the NorthGRIP (NGRIP) ice core on the Greenland Ice Core Chronology 2005 (GICC05) (Rasmussen et al., 2022). These datasets capture millennial-scale Dansgaard-Oeschger-cycle climate variability, which are expressed as $\delta^{18}\text{O}$ fluctuations in the NGRIP core that are expected to mirror variations in the magnetic susceptibility signal of the ENAM93-21 core (Rasmussen et al., 1996, 1997).

Since the P-DTW algorithm is sensitive to the relative amplitudes of the signals (Eq. (1)), both datasets were scaled logarithmically to enhance alignment. This transformation reduces the influence of absolute values, allowing the algorithm to focus more on the overall trends rather than specific amplitudes. The magnetic susceptibility signal from ENAM93-21 (denoted as a_1) and the $\delta^{18}\text{O}$ record from NGRIP (denoted as b_1) were transformed using the following equations: $a = \log(a_1/\text{mean}(a_1))+1$ and $b = \log((b_1+50)/\text{mean}(b_1+50))+1$. The constant values 50 and 1 were added to avoid negative values, while logarithmic scaling was applied to reduce the influence of large relative amplitude variations.

Four tephra layers in the ENAM93-21 core have been dated with chemical fingerprinting through correlation to the GICC05 (Rasmussen et al., 1996; Olsen et al., 2014), hereby providing independent

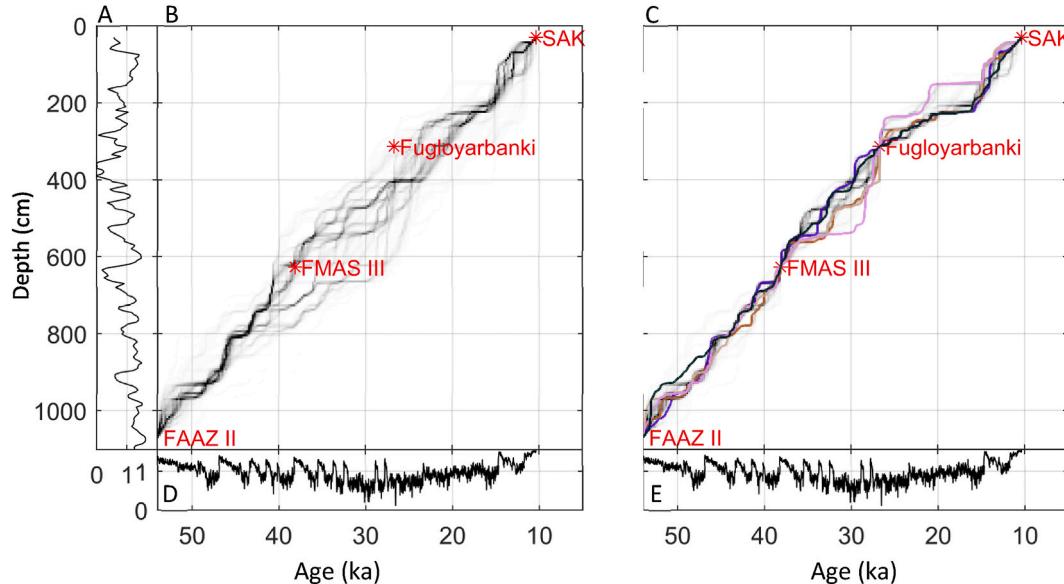


Fig. 10. Test on real data 2: Applying the P-DTW algorithm to two different types of signals. A: Processed (see main text) magnetic susceptibility from the ENAM93-21 core (Rasmussen et al., 1996), B: grayscale represents the P-DTW alignment model without tie points, C: grayscale represents the P-DTW alignment model including the tie points indicated with red stars, the colored curves correspond to the alignment paths represented in Fig. 11. For both panel A and B the P-DTW the alignment models using a p , ms , ml and ss value of 0.5, 1, 10 and 0.25, respectively, D-E: processed (see main text) $\delta^{18}\text{O}$ from the NorthGRIP ice core on the Greenland Ice Core Chronology 2005 (GICC05) (Rasmussen et al., 2022). Abbreviations: SAK: Saksunarvatn.

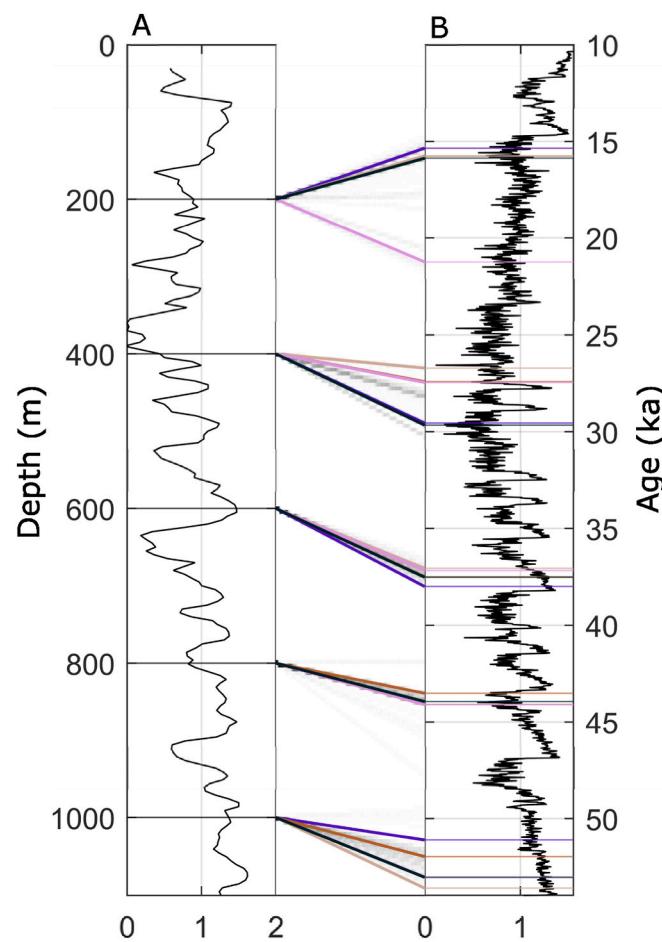


Fig. 11. Test on real data 2: Visualization of different alignments according to the P-DTW model shown in Fig. 10C. The grayscale represents the P-DTW alignment model, while the colored lines correspond to the alignment paths with the same colors shown in Fig. 10C. **A:** Processed (see main text) magnetic susceptibility from the ENAM93-21 core (Rasmussen et al., 1996), **B:** Processed (see main text) $\delta^{18}\text{O}$ from the NorthGRIP ice core on the Greenland Ice Core Chronology 2005 (GICC05) (Rasmussen et al., 2022).

constraints for the alignment/age-model. Here, we use the upper (Sakunarvatn) and lower (NAAZ II) tephra layers to define the time range for the alignment model and hence omit to use the lower ~ 30 cm of the record.

Fig. 10B shows the P-DTW alignment model constructed without incorporating the age constraints from the Fugloyarbanki and FMAS III tephra layers. This model displays a relatively high degree of scatter, representing a range of plausible alignments between the two signals. Importantly, the alignment model is consistent with the independent age constraints derived from the tephra layers despite the different types of the signals. This reflects that it might be feasible to align two different types of stratigraphic signals using the P-DTW algorithm if the signals are normalized correctly.

When the age constraints are directly incorporated into the model, as shown in Fig. 10C and 11, the range of possible alignments is significantly reduced. This shows the refinement of the ENAM93-21 core age model that is achieved by integrating external chronological information.

3.3. Multi-site age-depth modeling

We apply the P-DTW-alignment within an inverse probabilistic framework (Section 2.5) to merge chronological information between

sites. The approach integrates age constraints from two sites through the P-DTW-alignment model, resulting in a combined multi-site age-depth model.

To do this, we use two independent reference age-depth curves generated using the CosmoChron forward model (Sørensen et al., 2024) along with synthetic datasets for the two sites represented with these age-depth curves. Each synthetic dataset consists of four age constraints with a relative uncertainty of 5 %, along with the corresponding LR04 $\delta^{18}\text{O}$ curves that have been warped in depth relative to age according to the reference age-depth curves and added 0.2 $\delta^{18}\text{O}$ ‰ uncorrelated Gaussian noise, as described in Section 3.1 (Fig. 12).

The resulting multi-site age-depth model (Fig. 12G–H) has reduced uncertainties compared to the individual age-depth models (Fig. 12C–D). This is particularly pronounced in regions with sparse age constraints: the bottom of site *a* (~ 80 –100 m; Fig. 12C vs. 12G) and the top of site *b* (~ 20 –40 m; Fig. 12D vs. 12H).

Additionally, the alignment model is also refined. Comparing the prior alignment model (Fig. 12B) with the posterior alignment model (Fig. 12F), several plausible alignment paths in the alignment model have been eliminated. For example, the prior alignment model suggests a correspondence between ~ 60 m at site *a* and ~ 80 m at site *b*, a scenario that is entirely ruled out when merging age constraints from both sites (Fig. 12B vs 12F).

This demonstrates that by merging age constraints across sites using the P-DTW algorithm within a probabilistic inverse framework, it is possible to constrain both the individual site age-depth relationships, and the alignment model between those sites.

4. Discussion

4.1. Limitations

What defines a good alignment? The foundation for aligning two signals lies in understanding the relationship between them. The P-DTW algorithm primarily relies on the amplitude of the signals (Eq. (1)), operating under the assumption that amplitudes remain proportional over the same period. While the algorithm can handle minor deviations from this assumption (Fig. 7), its capacity is limited to cases where the amplitude difference between the signals is relatively small and spans only short intervals. Thus, the algorithm is best suited for aligning signals that are expected to maintain persistent amplitude relationships over the studied period. In general, this limitation applies to all DTW algorithms that rely on the cost-matrix outlined in Equation (1).

Alternatively, preprocessing of signals can facilitate the alignment of signals that share similar trends, but differ in amplitudes over the same period, as in the case where the magnetic susceptibility signal from the ENAM93-21 sediment core is aligned with the $\delta^{18}\text{O}$ signal from the NorthGRIP ice core (Figs. 10 and 11).

Currently, the P-DTW algorithm does not require prior constraints related to specific sedimentation or accumulation processes. However, when such information is available, it may be incorporated to enhance the performance using an approach that resembles the way algorithms like HHM-Match and BIGMACS integrate stratigraphic constraints (Lin et al., 2014; Lee et al., 2023). Nevertheless, it is important to note that the intended use of the P-DTW algorithm primarily is aimed at scenarios where the alignment between two signals is highly uncertain because the peak-to-peak correlations are ambiguous.

The algorithm's performance naturally declines as the signal-to-noise ratio decreases. At higher noise levels, P-DTW may fail to produce reliable alignment models unless the value of p is reduced accordingly (Fig. S2). This necessitates tuning of the parameters so they correspond to the level of noise, although no general function can describe the inverse relationship between noise and the optimal value of p for maximum performance.

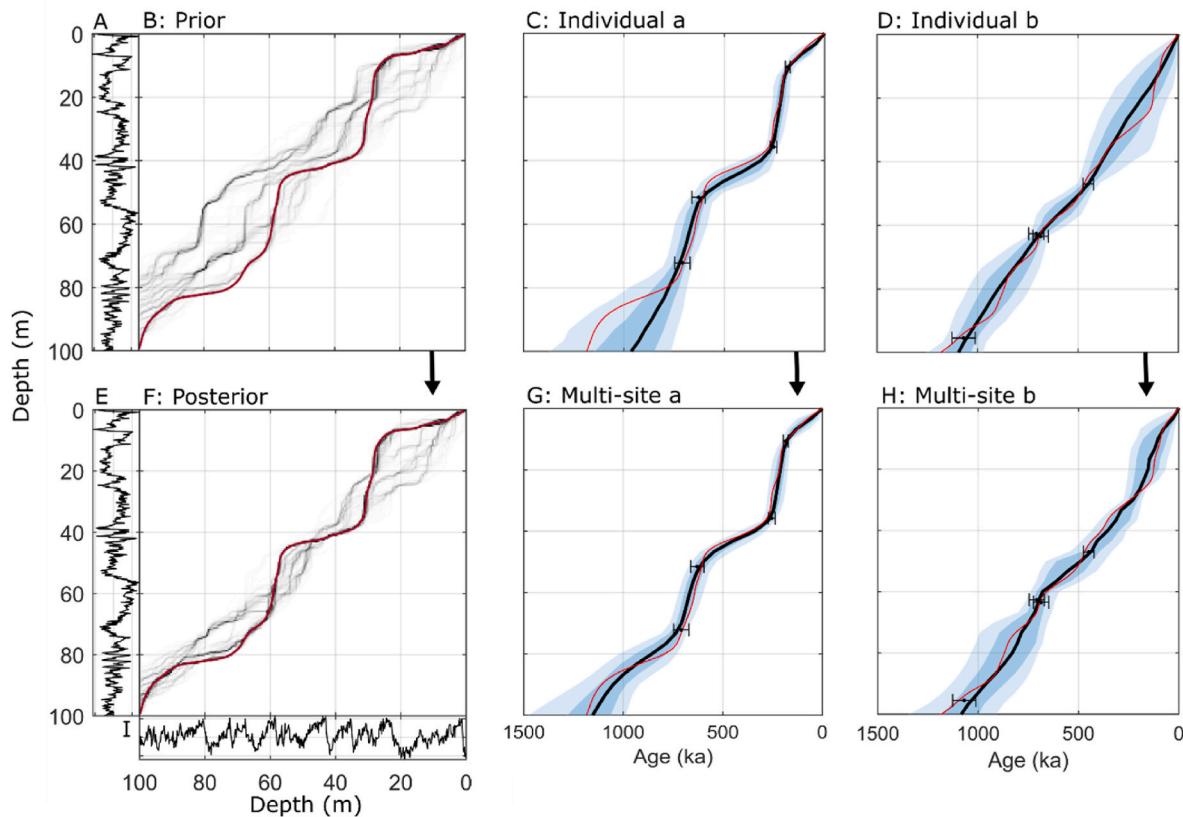


Fig. 12. Synthetic test case 6: Construction multi-site-age-depth models using P-DTW alignment as a prior to merge age constraints from two different sites. A and E: signal a, B: Greyscale represent the P-DTW alignment model with settings: p , ms , ml , ss equal to 0.5, 1, 10, 0.25, respectively (the prior alignment model), the red curve the reference-alignment. C and D: independent CosmoChron (Sørensen et al., 2024) age-depth models of site a and b, respectively F: Greyscale represent the posterior alignment-model, and the red curve is the reference-alignment. G and H: multi-site-age-depth model of site a (G) and site b (H), where data from both sites is merged using the P-DTW alignment-model shown in panel B as the prior alignment-model. I: Signal b. For both panel C, D, G and H: Red lines represent the reference age-depth profiles, black dots with error bars represent the observed age constraints with 5 % uncertainties, black lines, darker and lighter blue shades represent the median, 68 % and 95 % confidence intervals, respectively, of the age-depth models.

4.2. Choice of parameters

The performance and behavior of the P-DTW algorithm are influenced by a set of user-defined parameters that determine how the resulting alignment paths respond to signal variability and noise. Based on empirical experience and sensitivity analyses (Figs. 3 and 5), we offer some practical guidance below for selecting these parameters, acknowledging that the optimal configuration will depend on the specific characteristics of the dataset. Therefore, the choice of parameters should be considered on a case-by-case basis, guided by the expected alignment features and informed by a degree of trial and error.

Using a value of p below zero often produces alignment models that are highly independent of the signals and often contain little useful information about the alignment. Therefore, we recommend using values of p larger than zero but lower than one to achieve the most appropriate estimate of the uncertainty associated with the stratigraphic alignment as illustrated in Fig. 3.

The path-momentum length (ml) is typically set to $\sim 1\%$ of the total dataset length, as this value tends to produce the most reliable alignment models for the synthetic datasets used here (Fig. 5). This length might depend on the resolution of the data series and geological settings. The primary purpose of path momentum is to ensure the alignment path maintains a consistent slope during stationary or noisy intervals, as explained in Section 2.2.2. To achieve this, the ml -value must remain above zero. In certain scenarios, such as when the accumulation rates at the two sites are expected to be very consistent throughout the sections, it may be reasonable to increase the ml -value significantly beyond the 1

%, depending on the expected smoothness of the alignment.

For the steer factor, an ss -value (Eq. (5)) of 0.25 appears effective when the path-momentum length is set to 1 % of the dataset length (e.g. Figs. 2D and 5). When $ml = 0$, an ss -value of 1 steers the alignment path towards (1,1) in the depth-depth domain during noisy and stationary stratigraphic intervals, as illustrated in Fig. 2B. However, for technical reasons, the ss -value should be inversely proportional to ml to maintain this behavior.

4.3. Multiple profiles

In many cases, aligning multiple geological signals is essential. The P-DTW algorithm can be extended to handle multiple signals, where aligning n signals requires an n -dimensional cost-matrix. At each step, the number of possible transitions in tracking the minimum cost path is then $(2^n - 1)$. However, the computational time for multi-signal P-DTW increases exponentially with the number of signals, making it impractical for most datasets (Brown, 1997).

One approach to address this is to start with a single signal and propagate its stratigraphic tops to the others, following a chosen path that avoids geological discontinuities like faults (Wu et al., 2018). Wheeler and Hale (2014) propose an alternative approach by correlating each pair of signals independently and then optimizing the vertical shifts to resolve inconsistencies. While this method is efficient, it is deterministic and, therefore, unsuitable for stochastic versions of the DTW algorithm. Lallier et al. (2016) suggest a different strategy, where signals are first correlated in pairs and then iteratively rebuilt to eliminate the

inconsistencies. This approach could potentially be integrated into the P-DTW algorithm presented here.

5. Conclusion

This study introduces the Probabilistic Dynamic Time Warping (P-DTW) algorithm as a method for aligning noisy continuous stratigraphic signals while quantifying the uncertainty associated with multiple plausible alignments.

Through synthetic test cases, we demonstrate the algorithm's capability to capture a range of alignment scenarios, including the correct reference-alignment. This contrasts with the traditional deterministic DTW algorithm, which is prone to overfitting and may fail to identify the correct alignment. Furthermore, we show that the P-DTW algorithm can also align two signals with different amplitudes over the same period, depending on the magnitude of the difference and the length of the interval.

When applied to real $\delta^{18}\text{O}$ data from sediment cores GeoB7920-2 and MD95 2042, the P-DTW algorithm produces an alignment model consistent with established alignments while quantifying the alignment uncertainty. Furthermore, we demonstrated the algorithm's capability to align the magnetic susceptibility signal from the ENAM93-21 core with the $\delta^{18}\text{O}$ record from the NorthGRIP ice core, creating an age-depth model that is consistent with independent constraints. Lastly, we demonstrated the algorithm's ability to merge age constraints across sites within a probabilistic inverse modeling framework, hereby developing a coherent multi-site age-depth model.

By identifying multiple plausible alignments and quantifying their associated uncertainty, the P-DTW algorithm provides a probabilistic framework for advancing paleoclimate research and geochronological modeling.

Software

The Matlab scripts for running the P-DTW algorithm are available at: <https://github.com/aaaasssskkkeeee/P-DTW>.

Authors' contributions

ALS and **MFK** designed the study. **ALS**: Conceptualization, formal analysis, methodology, software, visualization, writing – original draft. **TMH**: Conceptualization, methodology, supervision. **FAF**: Methodology, software. **JO**: Methodology. **MFK**: Conceptualization, methodology, supervision. All authors reviewed and edited the final draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by NordForsk through the funding to “The timing and ecology of the human occupation of Central Asia”, project number 105204.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.quascirev.2025.109632>.

Data availability

A link to the data and/or code is provided as part of this submission.

References

- Baldwin, J.L., Otte, D.N., Wheatley, C.L., 1989. Computer emulation of human mental processes: application of neural network simulators to problems in well log interpretation. In: SPE Annual Technical Conference and Exhibition. Presented at the SPE Annual Technical Conference and Exhibition, SPE, San Antonio, Texas. <https://doi.org/10.2118/19619-MS>. SPE-19619-MS.
- Baville, P., Apel, M., Hoth, S., Knaust, D., Antoine, C., Carpentier, C., Caumon, G., 2022. Computer-assisted stochastic multi-well correlation: sedimentary facies versus well distality. Mar. Petrol. Geol. 135, 105371. <https://doi.org/10.1016/j.marpetgeo.2021.105371>.
- Brazell, S., Bayeh, A., Petroleum Corporation, Anadarko, Ashby, M., Corporation, Anadarko Petroleum, Burton, D., Geo Southern Energy, 2019. A machine-learning-based approach to assistive well-log correlation. Petro S Journ 60, 469–479. <https://doi.org/10.30632/PJV60N4-2019a1>.
- Brown, I.M., 1997. A new method for correlation of multiple stratigraphic sequences. Comput. Geosci. 23, 697–700. [https://doi.org/10.1016/S0098-3004\(97\)00046-0](https://doi.org/10.1016/S0098-3004(97)00046-0).
- Deutsch, C.V., Journel, A.G., 1992. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York.
- Hansen, T.M., Cordua, K.S., Looms, M.C., Mosegaard, K., 2013. SIPPY: a matlab toolbox for sampling the solution to inverse problems with complex prior information. Comput. Geosci. 52, 470–480. <https://doi.org/10.1016/j.cageo.2012.09.004>.
- Hay, C.C., Creveling, J.R., Hagen, C.J., Maloof, A.C., Huybers, P., 2019. A library of early Cambrian chemostratigraphic correlations from a reproducible algorithm. Geology 47, 457–460. <https://doi.org/10.1130/G46019.1>.
- Hladil, J., Vondra, M., Cejchan, P., Vich, R., Kopitkova, L., Slavik, L., 2010. The dynamic time-warping approach to comparison of magnetic-susceptibility logs and application to Lower Devonian calciturbidites (prague synform, bohemian Massif). Geol. Belg. 13, 385–406.
- Lallier, F., Antoine, C., Charreau, J., Caumon, G., Ruiu, J., 2013. Management of ambiguities in magnetostratigraphic correlation. Earth Planet Sci. Lett. 371–372, 26–36. <https://doi.org/10.1016/j.epsl.2013.04.019>.
- Lallier, F., Caumon, G., Borgomaner, J., Viseur, S., Fournier, F., Antoine, C., Gentilhomme, T., 2012. Relevance of the Stochastic Stratigraphic Well Correlation Approach for the Study of Complex Carbonate Settings: Application to the Malampaya Buildup (Offshore Palawan, Philippines), 370, pp. 265–275. <https://doi.org/10.1144/SP370.12>. SP.
- Lallier, F., Caumon, G., Borgomaner, J., Viseur, S., Royer, J.-J., Antoine, C., 2016. Uncertainty assessment in the stratigraphic well correlation of a carbonate ramp: method and application to the Beauvais Basin, SE France. C. R. Geosci. 348, 499–509. <https://doi.org/10.1016/j.crte.2015.10.002>.
- Lee, T., Rand, D., Lisicki, L.E., Gebbie, G., Lawrence, C., 2023. Bayesian age models and stacks: combining age inferences from radiocarbon and benthic $\delta^{18}\text{O}$ stratigraphic alignment. Clim. Past 19, 1993–2012. <https://doi.org/10.5194/cp-19-1993-2023>.
- Lin, L., Khider, D., Lisicki, L.E., Lawrence, C.E., 2014. Probabilistic sequence alignment of stratigraphic records. Paleoceanography 29, 976–989. <https://doi.org/10.1002/2014PA002713>.
- Lisicki, L.E., Jones, A.M., Rand, D.S., Lawrence, C.E., 2021. Age model of sediment core MD95-2042. <https://doi.org/10.1594/PANGAEA.939691>.
- Lisicki, L.E., Lisicki, P.A., 2002. Application of dynamic programming to the correlation of paleoclimate records. Paleoceanography 17. <https://doi.org/10.1029/2001PA000733>.
- Lisicki, L.E., Raymo, M.E., 2005. A pliocene-pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. Paleoceanography 20, 2004PA001071. <https://doi.org/10.1029/2004PA001071>.
- Luthi, S.M., Bryant, I.D., 1997. Well-log correlation using a back-propagation neural network. Math. Geol. 29, 413–425. <https://doi.org/10.1007/BF02769643>.
- Malinverno, A., 2013. Data report: Monte carlo correlation of sediment records from core and downhole log measurements at sites U1337 and U1338 (IODP Expedition 321). Proceedings of the IODP. Integrated Ocean Drilling Program. <https://doi.org/10.2204/iodp.proc.320321.207.2013>.
- Mosegaard, K., Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems. J. Geophys. Res. 100, 12431–12447. <https://doi.org/10.1029/94JB03097>.
- Muschitiello, F., O'Regan, M., Martens, J., West, G., Gustafsson, Ö., Jakobsson, M., 2020. A new 30 000-year chronology for rapidly deposited sediments on the Lomonosov ridge using bulk radiocarbon dating and probabilistic stratigraphic alignment. Geochronology 2, 81–91. <https://doi.org/10.5194/gchron-2-81-2020>.
- Muschitiello, F., Pausata, F.S.R., Watson, J.E., Smittenberg, R.H., Salih, A.A.M., Brooks, S.J., Whitehouse, N.J., Karlatou-Charalampopoulou, A., Wohlforth, B., 2015. Fennoscandian freshwater control on Greenland hydroclimate shifts at the onset of the younger dryas. Nat. Commun. 6, 8939. <https://doi.org/10.1038/ncomms9399>.
- Nielsen, A.T., Schovsbo, N.H., Klitten, K., Woolhead, D., Rasmussen, C.M.O., 2018. Gamma-ray log correlation and stratigraphic architecture of the cambro-ordovician alum shale formation on bornholm, Denmark: evidence for differential syndepositional isostasy. bgsd 66, 237–273. <https://doi.org/10.37570/bgsd-2018-66-15>.
- Olsen, J., Rasmussen, T.L., Reimer, P.J., 2014. North Atlantic Marine Radiocarbon Reservoir Ages Through Heinrich Event H4: a New Method for Marine Age Model Construction, vol. 398, pp. 95–112. <https://doi.org/10.1144/SP398.2>. SP.
- Paillard, D., Labeyrie, L., Yiou, P., 1996. Macintosh program performs time-series analysis. EoS Transactions 77, 379. <https://doi.org/10.1029/96EO00259>, 379.
- Pels, B., Keizer, J.J., Young, R., 1996. Automated biostratigraphic correlation of palynological records on the basis of shapes of pollen curves and evaluation of next-best solutions. Palaeogeogr. Palaeoclimatol. Palaeoecol. 124, 17–37. [https://doi.org/10.1016/0031-0182\(96\)00017-X](https://doi.org/10.1016/0031-0182(96)00017-X).

- Prell, W.L., Imbrie, J., Martinson, D.G., Morley, J.J., Pisias, N.G., Shackleton, N.J., Streeter, H.F., 1986. Graphic correlation of oxygen isotope stratigraphy application to the late Quaternary. *Paleoceanography* 1, 137–162. <https://doi.org/10.1029/PA001i002p00137>.
- Rasmussen, S.O., Svensson, A.M., Vinther, B.M., 2022. Greenland ice-core chronology 2005 (GICC05) annual layer depths for various Greenland ice cores. <https://doi.org/10.1594/PANGAEA.943195>.
- Rasmussen, T.L., Thomsen, E., Van Weering, T.C.E., Labeyrie, L., 1996. Rapid changes in surface and deep water conditions at the faeroe margin during the last 58,000 years. *Paleoceanography* 11, 757–771. <https://doi.org/10.1029/96PA02618>.
- Rasmussen, T.L., Van Weering, T.C.E., Labeyrie, L., 1997. Climatic instability, ice sheets and ocean dynamics at high northern latitudes during the last glacial period (58–10 KA BP). *Quat. Sci. Rev.* 16, 71–80. [https://doi.org/10.1016/S0277-3791\(96\)00045-5](https://doi.org/10.1016/S0277-3791(96)00045-5).
- Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* 26, 43–49. <https://doi.org/10.1109/TASSP.1978.1163055>.
- Sambridge, M., Jackson, A., Valentine, A.P., 2022. Geophysical inversion and optimal transport. *Geophys. J. Int.* 231, 172–198. <https://doi.org/10.1093/gji/ggac151>.
- Shackleton, N.J., Fairbanks, R.G., Chiu, T., Parrenin, F., 2004. Absolute calibration of the Greenland time scale: implications for antarctic time scales and for $\Delta^{14}\text{C}$. *Quat. Sci. Rev.* 23, 1513–1522. <https://doi.org/10.1016/j.quascirev.2004.03.006>.
- Shackleton, N.J., Hall, M.A., Vincent, E., 2000. Mean stable oxygen isotope ratios of benthic foraminifera from sediment core MD95-2042 on the Iberian margin, North Atlantic. <https://doi.org/10.1594/PANGAEA.58228>.
- Smith, T.F., Waterman, M.S., 1980. New stratigraphic correlation techniques. *J. Geol.* 88, 451–457. <https://doi.org/10.1086/628528>.
- Sørensen, A.L., Hansen, T.M., Nørgaard, J., Buylaert, J.-P., Murray, A.S., Kulakova, E., Kurbanov, R., Knudsen, M.F., 2024. CosmoChron: a versatile age-depth modeling approach using cosmogenic nuclides and direct age constraints. *Quat. Geochronol.* 85, 101618. <https://doi.org/10.1016/j.quageo.2024.101618>.
- Sylvester, Z., 2023. Automated multi-well stratigraphic correlation and model building using relative geologic time. *Basin Res.* 35, 1961–1984. <https://doi.org/10.1111/bre.12787>.
- Tarantola, A., 2005. Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9780898717921>.
- Tjallingii, R., Claussen, M., Stuut, J.-B.W., Fohlmeister, J., Jahn, A., Bickert, T., Lamy, F., Röhl, U., 2008. Coherent high- and low-latitude control of the northwest African hydrological balance. *Nature Geosci* 1, 670–675. <https://doi.org/10.1038/ngeo289>.
- Waterman, M.S., Raymond, R., 1987. The match game: new stratigraphic correlation algorithms. *Math. Geol.* 19 (2), 109–127.
- West, G., Kaufman, D.S., Muschitiello, F., Forwick, M., Matthiessen, J., Wollenburg, J., O'Regan, M., 2019. Amino acid racemization in Quaternary Foraminifera from the yermak Plateau, Arctic Ocean. *Geochronology* 1, 53–67. <https://doi.org/10.5194/gchron-1-53-2019>.
- Wheeler, L., Hale, D., 2014. Simultaneous correlation of multiple well logs. In: SEG Technical Program Expanded Abstracts 2014. Presented at the SEG Technical Program Expanded Abstracts 2014. Society of Exploration Geophysicists, Denver, Colorado, pp. 618–622. <https://doi.org/10.1190/segam2014-0227.1>.
- Wu, X., Shi, Y., Fomel, S., Li, F., 2018. Incremental correlation of multiple well logs following geologically optimal neighbors. *Interpretation* 6, T713–T722. <https://doi.org/10.1190/INT-2018-0020.1>.