Dynamics and Statistics of the Climate System, 2017, 1–16
doi:10.1093/climsys/dzx002
Advance Access Publication Date: 26 June 2017
Research Article

Research Article

A probabilistic Pliocene–Pleistocene stack of benthic $\delta^{18}O$ using a profile hidden Markov model

Seonmin Ahn[a], Deborah Khider[b], Lorraine E Lisiecki[b] and Charles E Lawrence[a,*]

[a]Division of Applied Mathematics, Brown University, Providence, RI and [b]Earth Science Department, University of California, Santa Barbara, CA, USA.

*Correspondence Charles E. Lawrence, Division of Applied Mathematics, Brown University, Providence, RI 02912, USA; E-mail: charles_lawrence@brown.edu

Received 17 October 2016; Revised 1 March 2017; Accepted 24 June 2017

Abstract

We present a 5-Myr probabilistic stack, which we name the Prob-stack. It is constructed from 180 globally distributed benthic $\delta^{18}O$ records using a profile hidden Markov model (HMM). Benthic stacks have been extensively employed to estimate ages in marine sediment cores and lead–lag relationships between paleoclimate proxies. Because this stack is probabilistic, it incorporates the variability among multiple records used in its construction and the uncertainty in alignments of records to the stack that have not been included in previous studies. It represents the uncertainty stemming from these two sources using a normal distribution with mean and variance in benthic $\delta^{18}O$ at each point in the stack. In doing so, Prob-stack captures the time-varying mean and variance of benthic $\delta^{18}O$ in the global ocean, which allows for better assessment of the uncertainty in records' relative ages, such as lead–lag relationships. Algorithms are included.

Key words: Oxygen isotopes, Pliocene–Pleistocene, Stratigraphy, Hidden Markov model.

1. Introduction

A benthic $\delta^{18}O$ stack is a representative time series that describes the global nature of ice volume and deep-water temperature signals by combining benthic $\delta^{18}O$ records from different locations. Due to the usefulness of stacks as measures of global climate change and stratigraphic alignment targets, a progression of stacks (Imbrie *et al.*, 1984; Pisias *et al.*, 1984; Prell *et al.*, 1986; Williams *et al.*, 1988; Raymo *et al.*, 1990; Bassinot *et al.*, 1994; Shackleton, 1995; Karner *et al.*, 2002; Huybers and Wunsch, 2004; Lisiecki and Raymo, 2005) has been created over the past 30 years as more data have become available. The LR04 stack (Lisiecki and Raymo, 2005) is one of the most widely used for age model developments and lead/lag relationship analysis (Clark et al., 2006, 2009; Jouzel *et al.*, 2007; Pollard and DeConto, 2009) because it describes the mean of 57 globally distributed records and continuously spans the Pliocene and Pleistocene (0–5.3 Myr). The LR04 stack was constructed through pairwise alignment of records using segments

of high-resolution records and dynamic programming optimization (Lisiecki and Lisiecki, 2002). This automated algorithm finds the global optimal alignment that minimizes the measures of sequence dissimilarity by comparing stratigraphic features and penalizing unrealistic changes in accumulation rates. After aligning the records, the average of all $\delta^{18}O$ measurements within each time interval was used to represent each stack point. Because the algorithm finds alignments deterministically and returns only the best solution, the LR04 stack does not reflect uncertainties in the alignments. Furthermore, using such a deterministic stack as a stratigraphic reference provides no information about statistical significance. To address this limitation, we present a probabilistic stack, called the Prob-stack, which describes changes in the global mean value of $\delta^{18}O$, including its variability across benthic $\delta^{18}O$ records and uncertainties in the alignments.

The Prob-stack is constructed based on the profile hidden Markov model (HMM). The preceding study (Lin *et al.*, 2014) proposed the HMM-Match algorithm, which is the pairwise alignment method for $\delta^{18}O$ records based on the HMM and this study extends this algorithm for construction of a probabilistic stack. The profile HMM method has been widely applied to in the field of bioinformatics (Eddy, 1998) for multiple sequence alignments (Durbin *et al.*, 1998) and a collection of protein family database (Bateman *et al.*, 2004) to explain the shared information of data sets. The estimated parameters of the profile HMM using the expectation–maximization (EM) algorithm successfully characterize the shared information when there is a sufficient amount of data. Thus, we apply this model to obtain a probabilistic stack with the assumption that each $\delta^{18}O$ record is a sample emitted from the probabilistic stack.

In the following section, we describe the $\delta^{18}O$ data and the construction method used in the Prob-stack. Section 3 presents the Prob-stack and a comparison to the LR04 stack. In addition, we present one more probabilistic stack, the Prob-LR04-stack, which is constructed using only the 57 $\delta^{18}O$ records used in the LR04 stack. In Section 4, we discuss aspects regarding the stack construction method and describe possible applications of the Prob-stack.

2. Methods

2.1. Data collection

The Prob-stack contains all 57 of the records in the LR04 stack plus an additional 123 published benthic $\delta^{18}O$ records. The sites of these records are well distributed in latitude, longitude and depth (Fig. 1), and the records have large variations in terms of resolution and time span (Fig. 2). Each of the collected records covers at least 150 kyr (if a record was obtained by splicing cores, the overall length of the record was at least 150 kyr) and has an average resolution better than or equal to 4 kyr. To reduce computing time and memory use, a few records were sliced into multiple

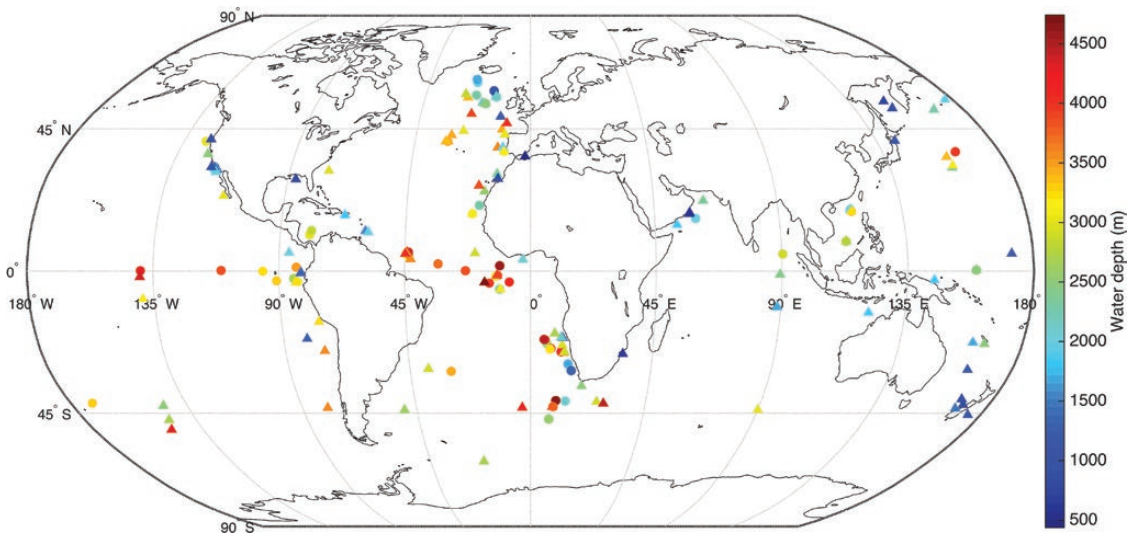

Figure 1. Location of the cores used in this study. The colours represent water depths of the sites. Circle: cores used to construct the LR04 stack. Triangle: cores which are newly added to construct the Prob-stack.

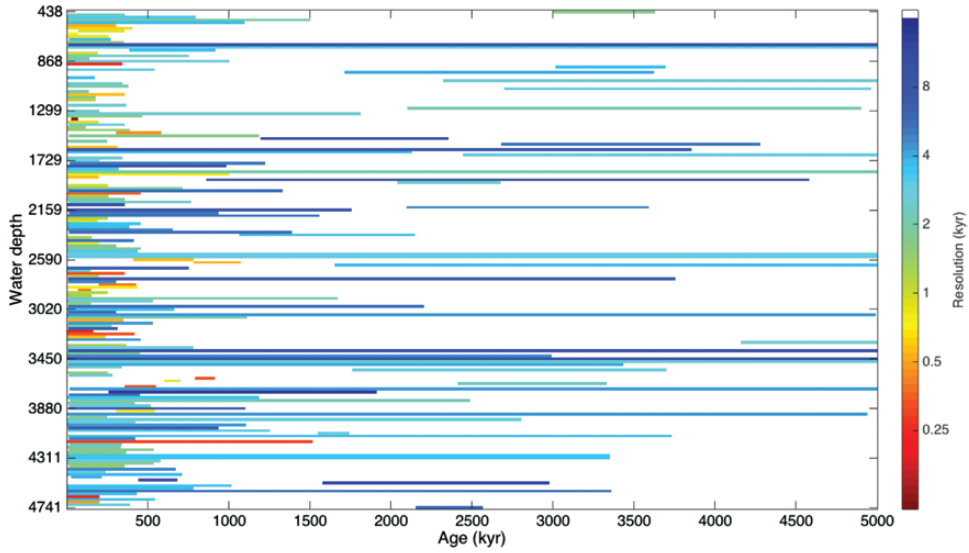

Figure 2. Coverage of the cores. The colours represent the average resolutions of records.

files when a record contained an irregular resolution. Detailed information about the records is given in Metadata Table 1 (Supplementary Information).

Most of the benthic $\delta^{18}O$ measurements comprising the records considered for this study were made on specimens of either *Uvigerina peregrina* or *Cibicidoides wuellerstorfi*, with the appropriate species offset corrections as either defined in the original publications or using the offset from Shackleton and Hall (1984).

2.2. Profile hidden Markov model

The construction of a probabilistic benthic $\delta^{18}O$ stack is done in two steps: determining the probabilistic alignments of individual records to the stack and updating the parameters of the stack based on these alignments. Through iteration of these two steps, we can construct a new stack while probabilistically aligning records to the stack.

To this end, we need a probabilistic model which describes a stack, records and alignments of these records to the stack. This model is based on the HMM-Match algorithm (Lin *et al.*, 2014) which is a probabilistic method to align a single record to a specified target. The profile HMM extends HMM-Match in two ways: it builds a new stack rather than aligning the records to an existing target, and permits the variances of points in the stack to be time-varying. A brief description of the probabilistic model is provided here. The full description of the model can be found in Lin *et al.* (2014). Based on the probabilistic model, we explain how to construct a stack which maximizes the likelihood of records through the Baum–Welch EM algorithm (Rabiner, 1989; Durbin *et al.*, 1998), which is also briefly described below.

2.2.1. Probabilistic model

The probabilistic stack includes the signal of global ice volume and deep-water temperature, and the variability of this signal. Let $\{s_t\}$ be a sequence of random variables, which represents the probabilistic stack, and assume that $\{s_t\}$ follow the normal distribution with mean μ_t and variance σ_t^2.

$$s_t \sim \mathcal{N}\left(\mu_t, \sigma_t^2\right) \tag{1}$$

Then, μ_t and σ_t^2 represent the signal and the variance of the signal, respectively, allowing both to vary with age, t. Note that the variance term represents the variability of $\delta^{18}O$ signals across multiple records and not the error of the mean value. We consider each record to be a randomly drawn sample from the common stack model. In HMM terminology, it is considered as an 'emission' of the model. Thus, we can estimate the distribution of the probabilistic stack from the randomly drawn samples, i.e. benthic $\delta^{18}O$ records.

Systematic differences between the probabilistic stack and individual records arise from two sources: spatial variability between records, and interlaboratory offsets (Ostermann and Curry, 2000). In addition to these systematic differences, there are two random factor: alignment uncertainty and random variations in $\delta^{18}O$. It is known that these two systematic factors may add a mean shift to an individual record compared to the stack. Thus, each record is modelled with a normal distribution allowing for a possible mean shift from the stack. Let $d_{i,j}$ be the jth data point of record i. Then, $d_{i,j}$ follows a normal distribution with mean $\tau_i + \mu_t$ and variance σ_t^2, where τ_i is the mean shift of record i, when the age estimate of the jth data point is t. This is written as follows:

$$d_{i,j} \sim \mathcal{N}\left(\tau_i + \mu_t,\ \sigma_t^2\right) \quad \text{if } A_{i,j} = t \tag{2}$$

where $A_{i,j}$ refers to the synchronous relative age assignment of the jth data point of the record i.

The alignment procedure synchronizes ages from multiple records, i.e. it finds data points assigned to the same age. Because it does not return the age estimate of calendar time, we call it a synchronous relative age estimate. We define an alignment of record i as a vector of synchronous relative age estimate: $A_i = (A_{i,1}, A_{i,2}, \ldots, A_{i,m_i})$, where m_i is the number of data points in record i. If exact age assignments for all data points $\{d_{i,j}\}$ were possible, then we could immediately construct the stack using these assignments. However, age assignments always include uncertainty. The goal of this study is to construct a stack which reflects the uncertainty in synchronous relative age estimates. To this end, we employ a modified version of the HMM-Match algorithm (Lin *et al.*, 2014) to generate samples of age assignments.

Here we briefly describe the probabilistic model of age assignments in the HMM-Match algorithm (Lin *et al.*, 2014). A more detailed description of the probabilistic model and algorithm workflow can be found in Lin *et al.* (2014) and its supplements. This method finds samples of the pointwise age estimate $A_{i,j}$ for each data point $d_{i,j}$. Whereas the stack spans most of the Pliocene–Pleistocene, most records cover only some portions of the total length. Thus, it is required to roughly select an initial target, which is a segment of the stack corresponding to each record.

We employ a procedure that adjusts for initial mismatch of the rough alignment and obtains an accurate alignment. We do this using three possible states for the starting and ending alignment between a target and $d_{i,j}$: matching, insertion or deletion. When a target and a record span the same interval, all data points of a record are considered as emitted data from a target and all alignment states are matching. On the other hand, if their spans are different, the alignment states can be either insertions or deletions at the top or bottom of the records. When a record covers a longer interval than a target, we treat the additional part of the record as inserted data. Alternatively, if a record spans a shorter interval than a target, we assume that the corresponding data in the record are deleted.

In addition to the three alignment states of the HMM-Match algorithm, here we consider one more state: matching-end. The matching-end state indicates that the matched portion is over, and the rest of the states are either insertions or deletions. Figure 3 illustrates the flows of state changes. Begin and End indicate that alignment begins and ends, respectively. Without the matching-end state, a matching state has four possible choices for the next state: matching, insertion, deletion or end. The probability of staying in matching states is almost always nearly one because we have a long series of one-to-one alignments of a record to the stack. Thus, transitions from matching to insertion or deletion rarely occur. Adding the matching-end state prevents such situations, so it enables the algorithm to enter an insertion state or deletion state at the end more frequently, which gives more flexibility in alignments.

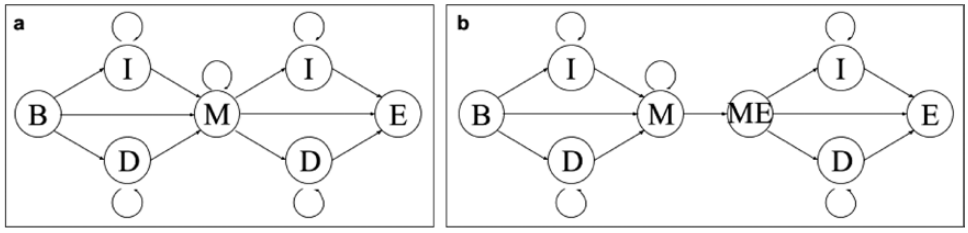

Figure 3. Flows of state changes in (a) the original HMM-Match algorithm and (b) the modified version for this study. The letters are as follows: B stands for begin, E stands for end, I stands for insertion, D stands for deletion, M stands for matching and ME stands for matching-end. Possible state transitions are illustrated with arrows. A rotating arrow represents alignments that can remain in the current state.

Because sedimentation rates vary both spatially and temporally, finding matched portions of a record does not give us a synchronous relative age estimate. Therefore, it is necessary to allow for differences in sedimentation rates between various records and over time within a specific record to determine age estimates. Based on the ratio of two lengths, the length between $d_{i,j}$ and $d_{i,j+1}$ and the length between $A_{i,j}$ and $A_{i,j+1}$, the matching state is categorized into 17 states from a 4:1 (expansion, low sedimentation rate) to a 1:4 (contraction, high sedimentation rate). Because insertions and deletions are now allowed in the middle of records, if there are hiatuses in a record, the alignment would approximate the effect by using a series of low sedimentation rates. If a hiatus was identified in an individual record by the original author, we split the record before finding its age estimates. Conversely, the effect of turbidite deposits would have to be approximated with a series of high sedimentation rates.

This model assumes that a sedimentation rate between points $j+1$ and $j+2$ of a record, given all previous sedimentation rates, depends only on one immediate previous sedimentation rate between points j and $j+1$, which yields the Markov property of this HMM algorithm. That is, the conditional probability of the rate from $d_{i,j+1}$ and $d_{i,j+2}$ depends directly only on the rate from $d_{i,j}$ and $d_{i,j+1}$ and not on any previous rates given the current rate. This property is written with corresponding age estimates as follows:

$$P(A_{i,j+1}|A_{i,j}, A_{i,j-1}, \ldots, A_{i,2}, A_{i,1}) = P(A_{i,j+1}|A_{i,j}, A_{i,j-1}) \tag{3}$$

Thus, while all rates are dependent on one another, the rate at any given point in a record is only indirectly dependent on the historic rates only through the most recent rate. This model is called the HMM (Rabiner, 1989; Durbin *et al.*, 1998) because the unknown states follow the Markov property and these can be inferred through observations, which are the benthic $\delta^{18}O$ records. In HMM terminology, equations (1) and (2) refer to the inhomogeneous emission model, and equation (3) represents the transition model. Because we construct a stack by aligning multiple records simultaneously rather than aligning one single record to a target, in the way the LR04 stack was constructed, the model is called the profile HMM. The advantage of the profile HMM over the deterministic approach employed in the construction of LR04 is that it returns confidence intervals for the stack through the variance term σ_t^2 for each time interval. The variance term can be also used to explain the possible range of an unobserved record because a benthic $\delta^{18}O$ record is a random sample of the probabilistic stack.

2.2.2. Baum–Welch expectation-maximization algorithm

The probabilistic model includes multiple unknown parameters: $\{\mu_t, \sigma_t^2\}$ which describe the mean $\delta^{18}O$ and attached variance for each time step t, $\{\tau_i\}$ which is the set of parameters representing a mean $\delta^{18}O$ shift of each record and $\{\phi_i\}$ which accounts for the set of parameters concerning the alignment of the record i. More precisely, $\{\phi_i\}$ includes state transition probabilities and initial state probabilities.

Let Θ refer to the set of all these unknown parameters. The unknown parameters are learned by the iterative Baum–Welch EM algorithm (Rabiner, 1989; Durbin *et al.*, 1998). This algorithm finds parameters which maximize the likelihood of records. Because we assume that each record is emitted from the stack independently, the total likelihood of records becomes

$$P(\text{record}_1, \text{record}_2, \ldots, \text{record}_n|\Theta) = \prod_{i=1}^{n} P(\text{record}_i|\Theta)\,. \tag{4}$$

The algorithm starts by setting initial values of unknown parameters, say Θ_1, to be their best guesses. Then, it iterates two steps: the expectation-step (E-step) and maximization-step (M-step). In the kth E-step, the algorithm finds the expected sufficient statistics and the likelihood of the data with the current set of parameters Θ_k. Then, in the kth M-step, we find the new set of parameters Θ_{k+1} which maximizes the likelihood of the records.

The E-step is composed of two procedures: the forward algorithm and the stochastic backward sampling algorithm. The forward algorithm finds the likelihood of a record considering all possible alignments starting from the top of each record:

$$P(\text{record}_i|\Theta) = \sum_{A_i} P(\text{record}_i, A_i|\Theta)$$

The total likelihood of all records can then be computed from equation (4). The stochastic backward sampling algorithm draws samples of alignments A_i^* following

$$P(A_i^*|\text{record}_i,\ \Theta_k) = \frac{P(\text{record}_i|A_i^*, \Theta_k)P(A_i^*|\Theta_k)}{\sum_{A_i} P(\text{record}_i|A_i, \Theta_k)P(A_i|\Theta_k)} \tag{5}$$

where every term in the right-hand side, including the summation in the denominator, is evaluated from the forward algorithm. To draw a sample for each data point using the above probability implies choosing an alignment state among the six possible states: begin, end, insertion, deletion, matching, matching-end. We use 1000 stochastic back trace samples for each record. The samples are independent of each other and representative of the space of all alignments. These forward and backward algorithms for each record correspond to the E-step of the HMM-Match method.

The M-step updates the maximum likelihood estimates of unknown parameters Θ, i.e. τ_i, ϕ_i, the means μ_t and the variances σ_t^2 of the stack points using the sample alignments generated from equation (5). Consider that we are updating the first point of the stack, s_1. After collecting all sampled $\delta^{18}O$ values that are aligned to this stack point, we find the maximum likelihood estimators for μ_1 and σ_1^2 which are the mean and the variance of the aligned samples, and similarly for all other points in the stack.

The algorithm used for the construction of a probabilistic stack can be summarized as follows:

1. Set initial values for the unknown parameters.
2. Run the E-step for each record.
    a. Forward algorithm computes the likelihood of a record considering all possible alignments.
    b. Backward sampling draws possible alignments with the stack based on the probability computed from the forward algorithm.
3. Compute the expected sufficient statistics and the total likelihood of all records.
4. Run the M-step for each record and update the stack for each time step.
5. Run the E-step for each record as step 2.
6. Compute the total likelihood of all records. Stop if the total likelihood does not improve by more than 0.1%, otherwise go back to step 4.

We set initial values of the stack to be the $\delta^{18}O$ values from the LR04 stack. This implies that the Prob-stack shares the same time domain as the LR04 stack (Table 1), and that we take the LR04 age model as initial condition. Note that, however, we do not employ the LR04 stack values in updating the probabilistic stack. Only the data points in the individual records are employed according to their age assignment samples. This allows the final solution to converge to a result different from the initial target, as demonstrated below. Also, the profile HMM does not require any other age information except choosing a target for each record. A target can be selected with rough age estimates because the algorithm detects matched portions by allowing insertion and deletion states at the top and the bottom of a record. More discussion on initial values is given in the Technical Validation.

A probabilistic stack reflects the uncertainty by using the samples of age assignments. Analysis of the HMM-Match algorithm demonstrates that the variance of age assignments decreases approximately linearly with the resolution of the records (Lin *et al.*, 2014). Thus, the contribution of data from low-resolution records spreads to a relatively large number of stack points, while the data from high-resolution records tend to be aligned over a smaller number of stack points. In this manner, the algorithm naturally gives more weight to high-resolution records within each stack bin.

Table 1. Interval size

Time, Ma	Interval size, kyr
0–0.6	1.0
0.6–1.5	2.0
1.5–3.0	2.5
3.0–5.0	5.0

2.2.3. Code and data availability

The open source software HMM-Stack is available at https://github.com/seonminahn/HMM-Stack for researchers to construct their own stack using benthic $\delta^{18}O$ records. However, please note that such a construction of the stack will require extensive use of high-performance computing resources.

The Prob-stack can be found in Prob-stack.txt at https://github.com/seonminahn/HMM-Stack. This file contains five columns: age estimate, $\delta^{18}O$ value [‰], SD of $\delta^{18}O$ value, upper bound of 95% interval and lower bound of 95% interval. Detailed information regarding the benthic $\delta^{18}O$ records used to construct the Prob-stack is given in Metadata Table 1 (Supplementary Information). The published data sets used in the compilation are available on the LinkedEarth wiki at http://wiki.linked.earth/Prob-stack.

3. Results

3.1. Prob-stack

The Prob-stack is constructed from 180 individual benthic $\delta^{18}O$ records. For each time interval, we estimate the probability distribution of $\delta^{18}O$ values, i.e. the mean and variance terms of a normal distribution. For purposes of display and comparison with LR04, we describe a time series of the estimated $\delta^{18}O$ mean for each point in the Prob-stack (the 'Prob-mean'), along with the LR04 stack in Figure 4. The Prob-mean is only a partial representation of the Prob-stack because it does not include the essential uncertainty component of the full Prob-stack. The Prob-mean agrees well with the LR04 stack in most time intervals, except for a few small phase shifts between the two stacks. In particular, a ~5-kyr phase shift around 75 ka likely stems from either stretching or lower compression ratio near the tops of most sediment cores. The Prob-stack and LR04 stack show differences in their extremes which are labelled by Marine Isotope Stage (MIS) numbers (Railsback *et al.*, 2015) near G8, K1, KM4 through M1, MG2, MG8 through MG10,

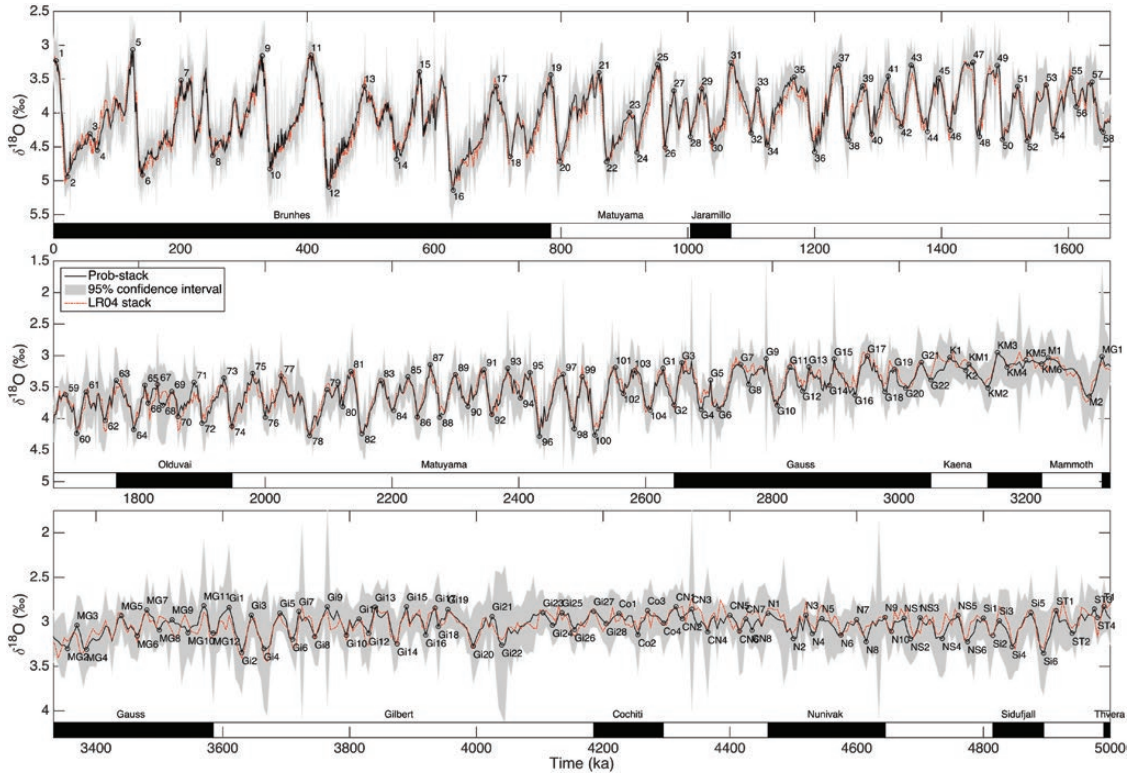

Figure 4. The Prob-stack constructed using 180 globally distributed benthic $\delta^{18}O$ records. The stack is plotted with MIS labels and magnetic polarity reversal stages. Note that the scale of the vertical axis changes across panels.

ST4 and T1. The G8 (2762 ka) glacial max is better resolved in the Prob-mean, but the Prob-stack has a large $\delta^{18}O$ uncertainty at G8. Prob-mean eliminates the double peak at K1 (3080 ka) and has minimal (practically non-existent) change in $\delta^{18}O$ between KM4 (3170 ka) through M1 (3235 ka). MG2 (3355 ka) is somewhat older (>10 kyr) than in the LR04 stack, and intervals between MG8 (3500 ka) and MG10 (3545 ka) have a noticeably different appearance. Additionally, the fluctuation of the Prob-mean near the end [around ST4 (4980 ka) and T1 (4990 ka)] is smaller than in LR04. Overall, the Prob-mean is slightly smoother than the LR04 stack, a reflection of the probabilistic nature of the HMM algorithm. Because HMM incorporates alignment uncertainty, it allocates input points to the sets of stack points corresponding to all possible times for which the input may have been deposited rather than to only the single best time point. The resulting spread tends to smooth the time series corresponding to the mean value at each time step. In addition, the availability of more records may contribute to smoothing.

3.2. Prob-LR04-stack

To evaluate how the results of the profile HMM alignment compare with those from the deterministic pairwise alignment used in LR04, we created the Prob-LR04-stack, which uses the same 57 records that are contained in the LR04 stack. Because the LR04 and Prob-LR04 stacks contain exactly the same data, any differences between these two stacks must derive from the techniques used to create them. The means of Prob-LR04-stack and LR04 are very similar (Fig. 5); the main differences are small age offsets (similar to those observed in the full Prob-stack), and that the Prob-LR04-stack mean is slightly smoother than LR04 due to the probabilistic nature of the alignment, which spreads the effects of uncertain age assignments across multiple time points.

3.3. Comparison between the Prob-stack and the Prob-LR04-stack

It is also of interest to examine the differences that result from the additional 123 records contained in the full Prob-stack compared to the 57 in the Prob-LR04-stack (Fig. 6). The average difference in the mean signal is less than

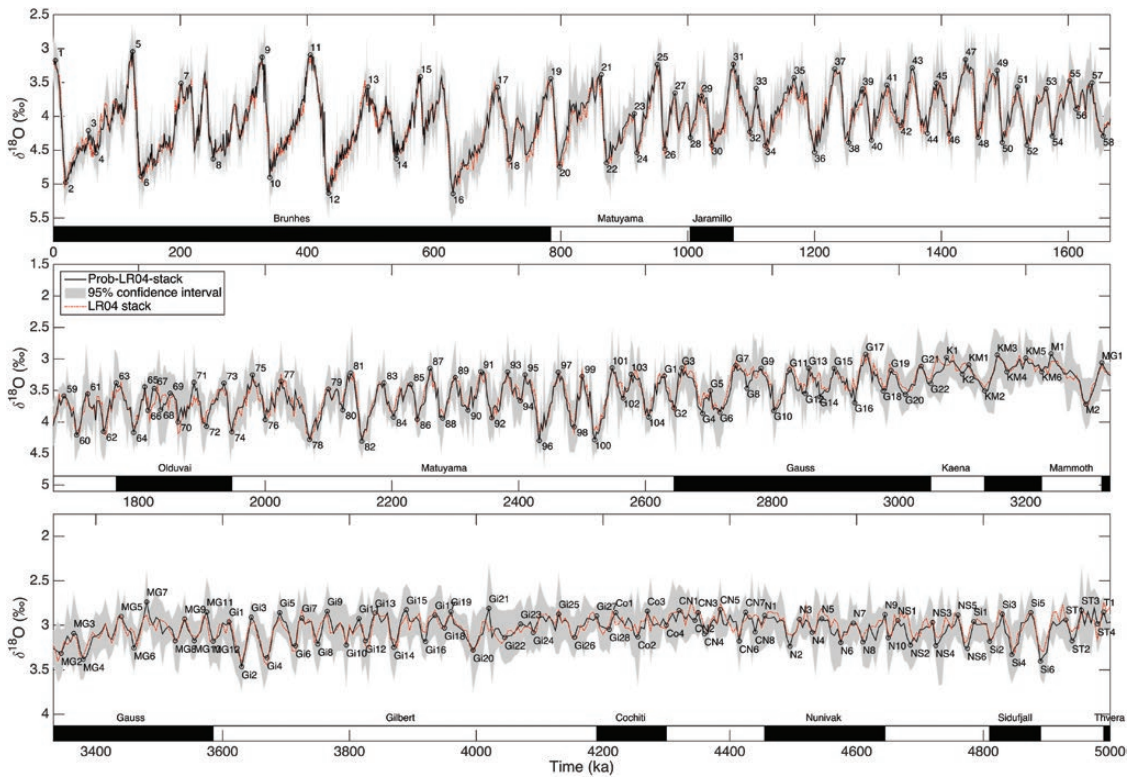

Figure 5. The Prob-LR04-stack constructed using the same 57 records that are contained in the LR04 stack. The stack is plotted with MIS labels and magnetic polarity reversal stages. Note that the scale of the vertical axis changes across panels.

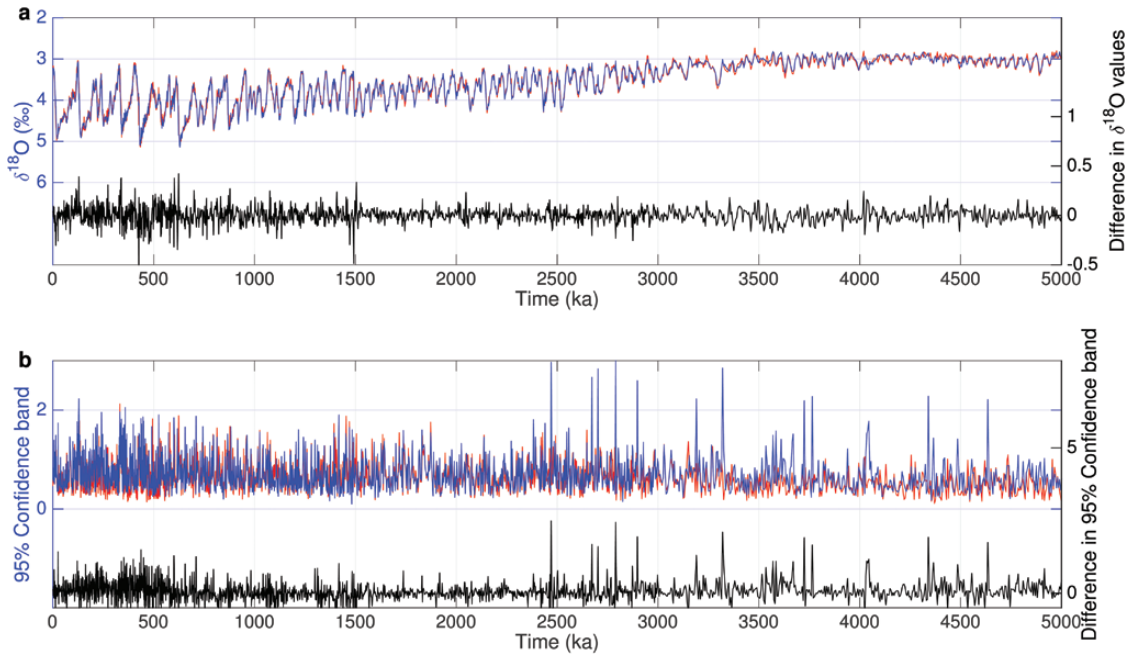

Figure 6. Comparison between the Prob-stack and the Prob-LR04-stack: the red lines and the blue lines represent the results of the Prob-LR04-stack and the Prob-stack, respectively. The black line represents the difference between the two stacks. The upper panel (a) and the lower panel (b) show the mean and the 95% confidence interval of the stacks, respectively.

0.06‰, but the Prob-stack is smoother than the Prob-LR04-stack. This is because the Prob-stack includes a larger number of benthic δ^{18}O records and the mean signal is literally the average of those records rather than a single realization of the Prob-stack. Figure 7 shows the number of records used to construct each point of the two stacks. The confidence limit of the Prob-stack is on average 12% larger than the Prob-LR04-stack overall, but it varies from −5% to 30%, depending on intervals (Table 2). The difference emerges from the diversity of additional records added to the Prob-stack. This indicates that the variability within the 57 records used in the Prob-LR04-stack was less than that captured in the larger data set. Note that the confidence band does not decrease by using a larger number of records because the variance term explains the variability between the various records rather than the error of the mean value. The goal of the study is to construct a stack which describes the pattern of global deep-ocean temperature and salinity variability, so we construct the Prob-stack using all available records instead of choosing high-resolution records. If we construct a stack by using only high-resolution records, we may reduce some smoothing effects because a low-resolution record tends to spread its alignments to a relatively large number of stack points compared to a high-resolution record. However, the resulting stack does not fully characterize the variability of δ^{18}O records over the ocean. Using all available data including low resolution records increases the spatial coverage of the stack, which should improve the estimates of global δ^{18}O distributions.

4. Discussion

4.1. Synchronous relative age estimates

The time axis of the Prob-stack represents synchronous relative age estimates which are the results of the alignment procedure that synchronizes multiple benthic δ^{18}O records. The EM algorithm takes the LR04 stack as initial values. This implies that the constructed stack follows the same age model as the LR04 stack at first, but stack ages are allowed to shift freely afterward according to the alignments of benthic δ^{18}O records.

A sensitivity analysis of initial conditions is conducted to examine the validity of the EM algorithm starting from the LR04 stack. There is a chance for the EM algorithm to converge to a local maximum rather than a global

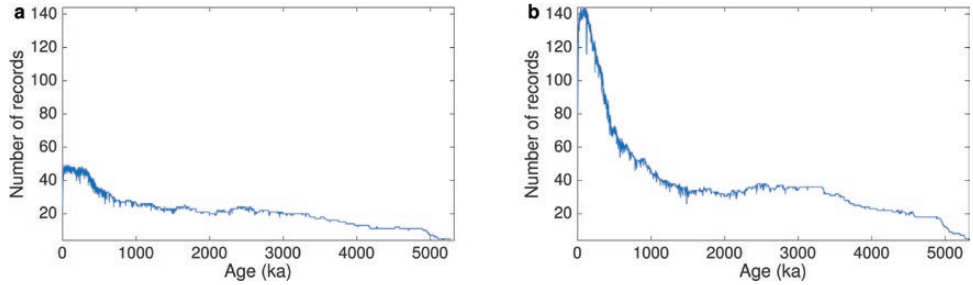

Figure 7. The number of records used to construct each stack point of (a) the Prob-LR04-stack and (b) the Prob-stack.

Table 2. Relative difference in variances between the Prob-stack and the Prob-LR04-stack

Time, Ma	Relative difference in variances between two Prob-stacks, %
0–0.5	25.59
0.5–1.0	5.23
1.0–1.5	–4.81
1.5–2.0	–0.74
2.0–2.5	5.31
2.5–3.0	11.56
3.0–3.5	17.49
3.5–4.0	30.30
4.0–4.5	25.96
4.5–5.0	24.44

A positive value implies that the Prob-stack represents higher variability.

maximum depending on initial conditions (Durbin *et al.*, 1998). To find the global maximum estimator, one may run the EM algorithm multiple times with various initial conditions. However, because running the EM algorithm is computationally expensive, the sensitivity analysis is conducted with different initial conditions for the recent 500-kyr interval using the 57 cores used to construct the LR04 stack. We employ two alternative initial values: a piecewise linear approximation of the LR04 stack and a constant value set to the overall mean of the LR04 stack (Fig. 8a and b). We call the two stacks constructed from these two initial values Prob-LR04-ICL (initial condition linear) and Prob-LR04-ICC (initial condition constant). Note that the Prob-LR04-ICC starts from a flat constant line without taking any information from the LR04 stack.

The stratigraphic features of the resulting three stacks, Prob-LR04, Prob-LR04-ICL, and Prob-LR04-ICC, are similar. The important information that stack construction algorithm takes from the initial conditions is periodic patterns of the $\delta^{18}O$ values rather than the detailed features, so Prob-LR04 and Prob-LR04-ICL are almost identical. However, Prob-LR04-ICC shows differences in its phases compared to Prob-LR04. The total likelihood of Prob-LR04-ICC is smaller than that of Prob-LR04-ICL and Prob-LR04, so Prob-LR04-ICC is a local maximum and not the maximum likelihood solution. Because the alignment procedure synchronizes multiple records without absolute constraints, the resulting stacks involve phase differences; these differences may disappear if we employ other age information as calendar time constraints. The phase differences between Prob-LR04-ICC and Prob-LR04 have a maximum of 11 kyr over the interval 0–250 ka, but the differences are less than 2 kyr for the periods older than 250 ka. We confirmed that the averages of relative accumulation rate are higher in the case of Prob-LR04-ICL than Prob-LR04-ICC where we observe phase differences (Fig. 8c and d). Having a relative accumulation rate at a greater value than one indicates that records are expanded relative to the LR04 stack and the recent records need to be contracted in their alignments to virtually match those of stacks that start with a LR04-like input. Nevertheless, it is remarkable that the detailed stratigraphic features of Prob-LR04 are obtained well, even starting from a constant value. These results indicate that the algorithm does not require any prior

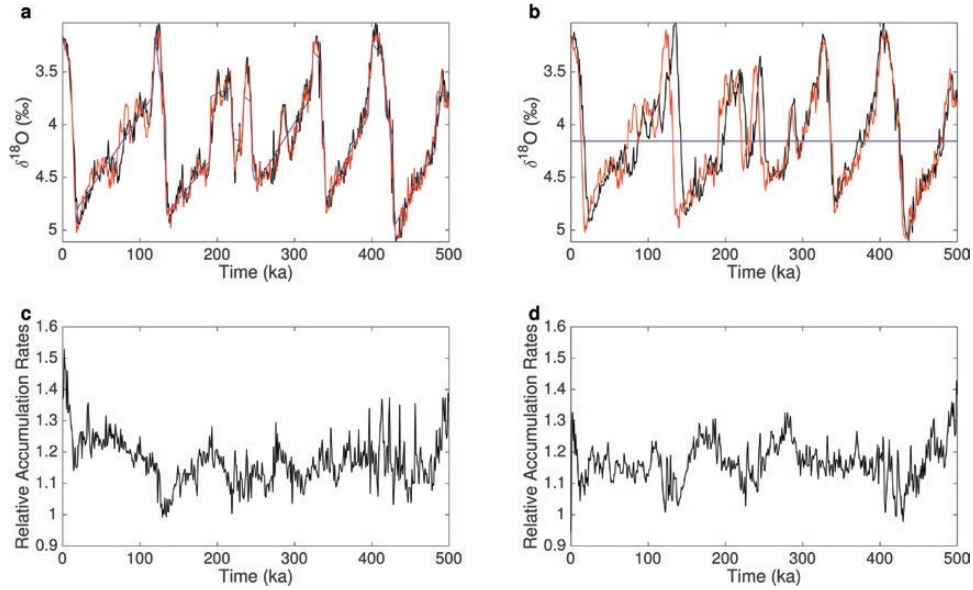

Figure 8. Sensitivity of results to initial values. Top panels: the red lines represent the Prob-LR04-stack when initial values are set to be the LR04 stack. The black lines represent the reconstructed stacks starting from the blue lines where the blue lines represent (a) the linear estimation and (b) the mean value of the LR04 stack. Bottom panels: the overall average of relative accumulation rates over the 57 cores when initial values are set to be (c) the linear estimation and (d) the mean value of the LR04 stack.

information regarding the benthic δ^{18}O signals from initial values or further age model assumptions such as orbital tuning to obtain the detailed stratigraphic features.

In summary, our sensitivity analysis shows that the pattern of ä^{18}O is robust to initial conditions. Therefore, it is not necessary to employ an orbital tuning based starting value to return a solution that is much like Prob-LR04 albeit with relative age estimates that differ from Prob-LR04 and are likely distorted relative to calendar time. Using starting values with LR04 or the piecewise linear version of the LR04 stack return a solution with a higher likelihood and fewer phase differences with the LR04 stack.

Further studies are required to convert synchronous relative age estimates to calendar time estimates. For example, we can incorporate other age proxies such as radiocarbon and magnetic reversals. These age proxies can act as constraints for the alignments, and then the alignment procedures will be able to produce age estimates that are consistent with available calendar time constraints.

4.2. Probability model

The Prob-stack is assumed to follow the normal distribution with mean μ_t and variance σ_t^2. Here we examine the normality assumption of the Prob-stack using a Q–Q plot (Wilk and Gnanadesikan, 1968) comparing the distribution of the aligned samples and the assumed normal distributions of the ä^{18}O proxies at several points in the stack. As shown in Figure 9, most of the points on the Q–Q plot lie on the line $y = x$, which implies that the aligned samples follow the assumed normal distribution of the stack value and the number of potential outliers is small. The largest impact of any outliers would be on the inflation of the variance estimates. Further studies on the identification and adjustment for outliers are expected to tighten confidence limits.

In addition to the normality assumption of the Prob-stack, we assumed that each benthic ä^{18}O record is emitted from the stack. Because we treat each data point in a benthic ä^{18}O record as a random sample drawn from the stack, we could estimate the distribution of the stack from the cores. This also indicates that we assumed δ^{18}O values to be synchronous throughout the ocean. Strictly speaking, generalization to all locations in the oceans without bias would require records to be sampled randomly over the oceans. Although the records comprising the stack seem to follow this assumption, they are not random samples from all locations in the ocean. Thus, we cannot be confident that the

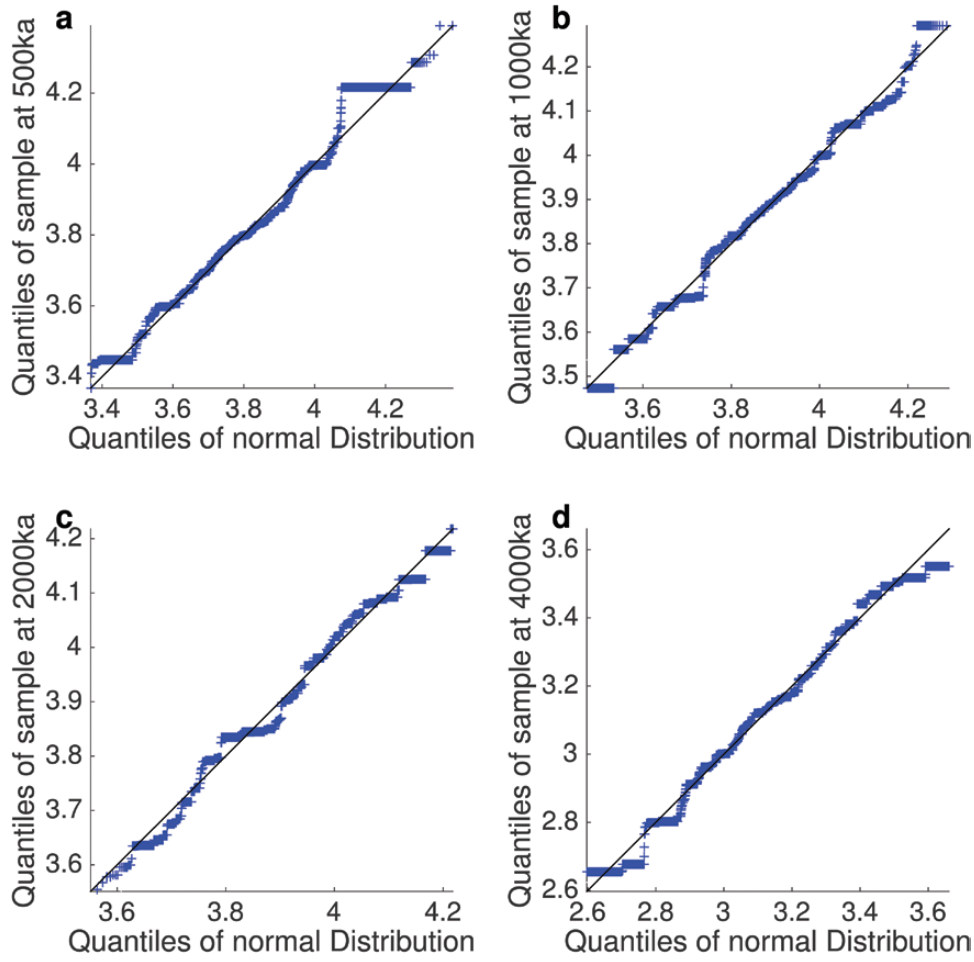

Figure 9. Q–Q plots. The empirical Q–Q plot of the samples aligned at (a) 500 ka, (b) 1000 ka, (c) 1500 ka and (d) 2000 ka versus the distribution of the Prob-stack.

estimates for all ocean sediments are unbiased, but rather that the estimates reflect any biases stemming from the decision made in selecting these locations.

The algorithm constructing the new stacks can be validated through pseudo-proxy experiments. After setting parameter values using Prob-LR04-stack as known values, pseudo-proxies are generated based on the resulting probabilistic model. Two sets of pseudo-proxies are simulated: the first set contains 40 sample records reflecting the resolution of 0–500 ka, and the second set includes 20 sample records reflecting the resolution of 3000–3500 ka. Because we have known values of the parameters, we can evaluate the algorithm by comparing the reconstructed stacks and the exact values. As shown in the two plots in Figure 10, the algorithm can accurately reconstruct the distribution of the stacks. The root mean square errors of the mean values of two stacks are 0.057 and 0.056.

4.3. Suggested applications

The Prob-stack is probabilistic, i.e. it represents the distribution of the stack for each time step. Both the mean and the variance information of the Prob-stack are used simultaneously for the applications of the Prob-stack. Here we introduce two applications: age estimation of a benthic $\delta^{18}O$ core and lead/lag analysis between two events observed from different cores, which can reasonably be completed on a modern laptop or desktop. The open source software HMM-Stack includes the codes for these two applications.

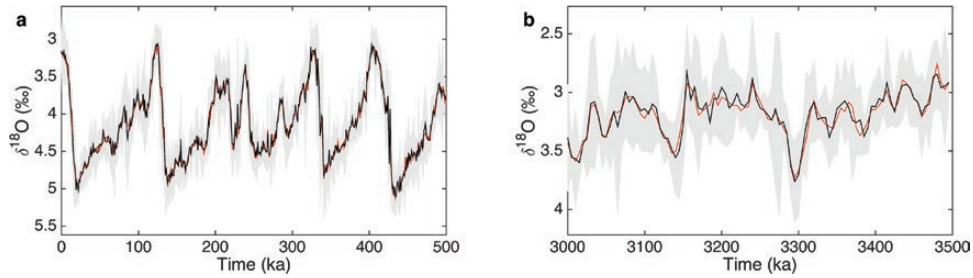

Figure 10. Reconstructed stack from the simulated records reflecting the period of (a) 0–500 ka and (b) 3000–3500 ka: the red lines represent the Prob-LR04-stack while the black lines and the grey areas represent the mean and the 95% confidence interval of the reconstructed stack from simulated records, respectively.

The algorithm of constructing the Prob-stack includes a procedure to find alignments of benthic δ^{18}O records to the Prob-stack. Because the Prob-stack is based on many additional records beyond LR04 and includes uncertainty estimates, it is meant as a replacement to the LR04 stack for age estimations from this point forward. This procedure finds age estimates of a benthic δ^{18}O record using both the means and the variances of the Prob-stack. As an example, Figure 11a shows alignments of the benthic δ^{18}O record from Site GeoB 1041 (Bickert and Wefer, 1996). Among possible alignment samples, the median alignment is chosen as it represents the central tendency of the posterior probability distribution of alignment (Lin *et al.*, 2014). We also display uncertainty information of age estimates (Fig. 11b) and relative accumulation rates with respect to the Prob-stack (Fig. 11c). The length of the confidence band varies from 2.5 to 12.5 kyr. Note that it sometimes increases up to 1/4 of the obliquity cycle. When the signal varies a lot, the confidence band tends to decrease. For example, around 130 ka, the δ^{18}O signal increases significantly, causing the confidence band to decrease.

Lead/lag analysis between two events observed from different cores can be done using age estimates of the two events. For example, let a_1 and a_2 be the age estimates of ODP1143 (Tian *et al.*, 2002) ODP1123 (Hall *et al.*, 2001; Harris, 2008) at 4.7 m below the sediment surface, respectively. The probability that a_1 occurs earlier than a_2 can be estimated by the joint distribution of a_1 and a_2 as follows:

$$\begin{aligned} P(a_1 > a_2 \,|\, \text{record}_{1143}, \text{record}_{1123}) &= \sum_{a_1 > a_2} P(a_1, a_2 \,|\, \text{record}_{1143}, \text{record}_{1123}) \\ &= \sum_{a_1 > a_2} P(a_1 \,|\, \text{record}_{1143}) P(a_2 \,|\, \text{record}_{1123}) \end{aligned} \tag{6}$$

The second equality holds as we assume benthic δ^{18}O records are independent samples emitted from the Prob-stack. After finding the age estimates a_1 and a_2 independently, the distribution of the age difference can be obtained from equation (6) as shown in Figure 12. Although a_1 and a_2 both have relatively large age uncertainties, their age difference is bimodal indicating that the age of a_1 is most likely ~15 kyr older than a_2, with a small chance that the two points are approximately synchronous. Specifically, the results indicate that the probability of a_1 occurring before a_2 is 0.98. The median and 95% confidence interval of age difference ($a_1 - a_2$) are 14.65 kyr and 0.01 to 17.95 kyr, respectively.

Both HMM-Match and HMM-Stack can be used to infer age estimates and lead/lag relationships. These algorithms produce statistical inference that cannot be obtained from alignments or comparisons by hand. However, since HMM-Stack takes into account the substantial time-dependent differences in variances with age while HMM-Match does not, HMM-Stack should be used when using the Prob-stack. HMM-Match remains useful for comparisons with other targets that do not incorporate uncertainty.

5. Conclusions

The Pliocene–Pleistocene Prob-stack presented in this study describes the distribution of global trends in benthic δ^{18}O and associated variance across the individual benthic δ^{18}O records for each time interval. As with all stacks, Prob-stack and stratigraphic alignments to Prob-stack rely on the simplifying assumption that benthic δ^{18}O varies

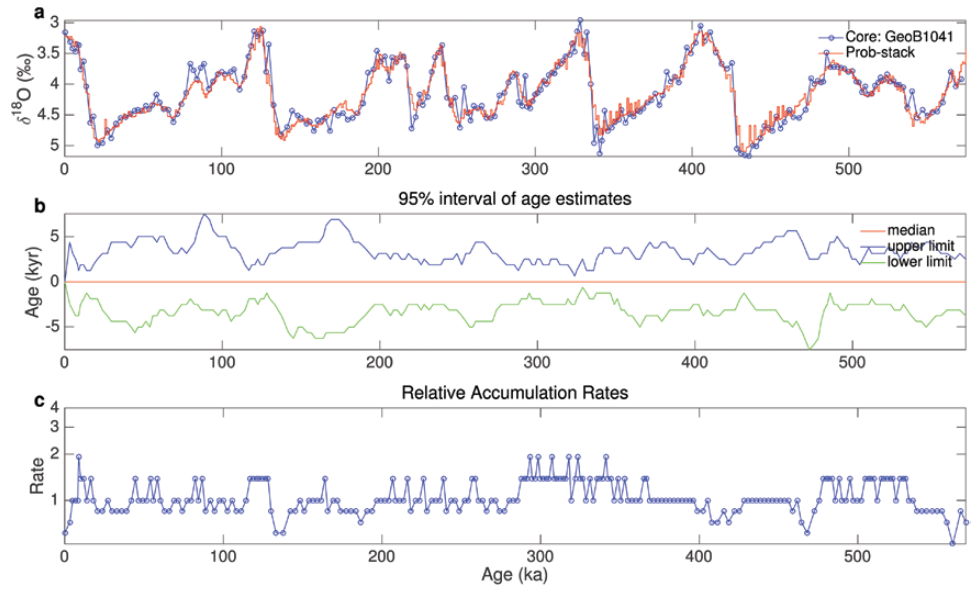

Figure 11. Example of benthic $\delta^{18}O$ record age estimation. (a) represents the median alignment of the GeoB 1041 core (blue line) to the Prob-stack (red line), (b) describes time-varying uncertainty of the age estimates using the 95% interval of the age estimate for each time step and (c) represents the relative accumulation rates of the GeoB 1041 core with respect to the Prob-stack based on the median age estimate.

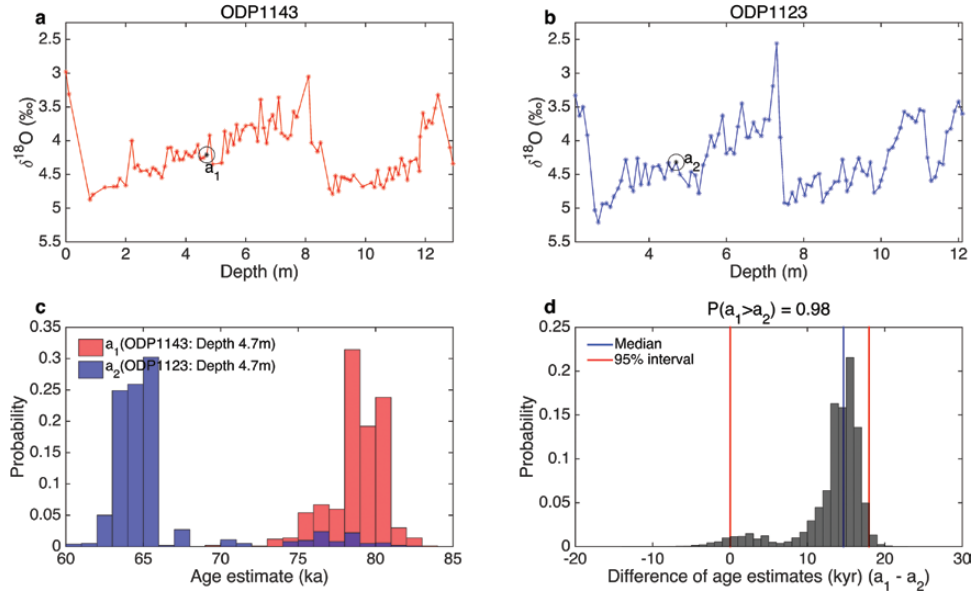

Figure 12. Lead/lag relationships between the two data points marked with circles in (a) ODP1143 (located 4.7 m below the sediment surface) and (b) ODP1123 (located 4.7 m below the sediment surface). The distributions of age estimates are shown in (c), and the distribution of the age difference is plotted in (d).

synchronously throughout the deep ocean. The Prob-mean alone should not be used for stratigraphic alignment because such an approach fails to account for the variability of the stack and is equivalent to reverting to a deterministic stack. Also, the size of 95% confidence band compared to the size of the mean value should be considered when interpreting the Prob-stack. As the Prob-stack contains time-varying variance information, the stratigraphic

Table 3. Summary of abbreviations

Abbreviation	Description
Prob-stack	Probabilistic stack (0–5 Myr) constructed from 180 benthic $\delta^{18}O$ records
Prob-mean	Estimated $\delta^{18}O$ mean of the Prob-stack
Prob-LR04-stack	Probabilistic stack (0–5 Myr) constructed from the 57 benthic $\delta^{18}O$ records used to construct the LR04 stack (Lisiecki and Raymo, 2005)
Prob-LR04-ICC	Probabilistic stack (0–500 kyr) constructed from the 57 benthic $\delta^{18}O$ records with initial conditions being a constant line
Prob-LR04-ICL	Probabilistic stack (0–500 kyr) constructed from the 57 benthic $\delta^{18}O$ records with initial conditions being the linear approximation of the LR04 stack

alignment of a record to the Prob-stack can produce age estimates and confidence limits of these age estimates reflecting the variability of the stack. The Prob-stack serves as a more complete description of global benthic $\delta^{18}O$ variability than the LR04 stack and provides additional functionality as a stratigraphic tool for Pliocene–Pleistocene studies.

Supplementary Data

Supplementary material is available at *Dynamics and Statistics of the Climate System: An Interdisciplinary Journal* online.

Funding

This material is based upon work supported by the National Science Foundation under grant numbers OCE-1025444 and OCE-1025438. This research was conducted using computational resources and services at the Center for Computation and Visualization, Brown University.

Acknowledgements

SA formulated the model, wrote code, ran the model and drafted the manuscript. DK undertook data collection and edited the manuscript. LL conceived and supervised the geoscientific aspect of the study and edited the manuscript. CEL conceived and supervised the mathematical aspect of the study and edited the manuscript. All authors discussed the results and implications as well as commented on the manuscript at all stages.

Conflict of interest

The authors declare no competing financial interests.

References

Bassinot FC, Labeyrie LD, Vincet E et al. The astronomical theory of climate and the age of the Brunhes-Matuyama magnetic reversal. *Earth Planet Sci Lett* 1994; 126: 91–108.

Bateman A, Coin L, Durbin R et al. The pfam protein families database. *Nucleic Acids Res* 2004; 32: D138–41.

Bickert T, Wefer G. Late quaternary deep water circulation in the South Atlantic: reconstruction from carbonate dissolution and benthic stable isotopes. In: Wefer G, Berger WH, Siedler G, Webb D J (eds). The South Atlantic. Berlin, Heidelberg: Springer, 1996, pp. 599–620.

Clark PU, Archer D, Pollard D et al. The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO2. *Quat Sci Rev* 2006; 25: 3150–84.

Clark PU, Arthur SD, Shakun JD et al. The last glacial maximum. *Science* 2009; 325: 710–4.

Durbin R, Eddy SR, Krogh A, Mitchison G. Biological Sequence Analysis. Cambridge, UK: Cambridge University Press, 1998.

Eddy SR. Profile hidden Markov models. *Bioinformatics* 1998; 14: 755–63.

Hall IR, McCave IN, Shackleton NJ, Weedon GP, Harris SE. Intensified deep Pacific inflow and ventilation in Pleistocene glacial times. *Nature* 2001; 412: 809–12.

Harris SE. Data report: late Pliocene-Pleistocene carbon and oxygen stable isotopes from benthic foraminifers at Ocean Drilling Program Site 1123 in the Southwest Pacific. *Proc Ocean Drill Prog Sci Results* 2002; 181: 1–20.

Huybers P, Wunsch C. A depth-derived Pleistocene age model: uncertainty estimates, sedimentation variability, and nonlinear climate change. *Paleoceanography* 2004; 19: PA1028.

Imbrie JZ, Hays DG, Martinson A et al. The orbital theory of Pleistocene climate: support from a revised chronology of the marine δ18 O record. In: Berger A, Imbrie JZ, Hays J, Kukla G, Saltzman B (eds). Netherlands: Milankovitch and Climate, Springer, 1984, pp. 269.

Jouzel J, Masson-Delmotte V, Cattani O et al. Orbital and millennial Antarctic climate variability over the past 800 000 years. *Science* 2007; 317: 793–6.

Karner DB, Levine J, Medeiros BP, Muller RA. Constructing a stacked benthic $\delta^{18}O$ record. *Paleoceanography* 2002; 17: 2-1–2-16.

Lin L, Khider D, Lisiecki LE, Lawrence CE. Probabilistic sequence alignment of stratigraphic records. *Paleoceanography* 2014; 29: 976–89.

Lisiecki LE, Lisiecki PA. Application of dynamic programming to the correlation of paleoclimate records. *Paleoceanography* 2002; 17: 1049.

Lisiecki LE, Raymo ME. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* 2005; 20: PA1003.

Ostermann DR, Curry WB. Calibration of stable isotopic data: an enriched $\delta^{18}O$ standard used for source gas mixing detection and correction. *Paleoceanography* 2000; 15: 353–60.

Pisias NG, Martinson DG, Moore Jr. TC et al. High resolution stratigraphic correlation of benthic oxygen isotopic records spanning the last 300 000 years. *Mar Geol* 1984; 56: 119–36.

Pollard D, DeConto RM. Modelling west Antarctic ice sheet growth and collapse through the past five million years. *Nature* 2009; 458: 329–32.

Prell WL, Imbrie J, Martinson DG et al. Graphic correlation of oxygen isotope stratigraphy application to the Late Quaternary. *Paleoceanography* 1986; 1: 137–62.

Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE* 1989; 77: 257–86.

Railsback LB, Gibbard PL, Head MJ, Voarintsoa NRG, Toucanne S. An optimized scheme of lettered marine isotope substages for the last 1.0 million years, and the climatostratigraphic nature of isotope stages and substages. *Quat Sci Rev* 2015; 111: 94–106.

Raymo ME, Ruddiman WF, Shackleton NJ, Oppo DW. Evolution of Atlantic-Pacific δ13C gradients over the last 2.5 m.y. *Earth Planet Sci Lett* 1990; 97: 353–68.

Shackleton NJ, Hall MA. Oxygen and carbon isotope stratigraphy of deep sea drilling project hole 552A: Plio-Pleistocene glacial history. *DSDP Initial Reports* 1984; 81: 599–609.

Shackleton NJ. New data on the evolution of Pliocene climate variability. In: Vrba ES, Denton GH, Partridge TC, Burckle LH (eds). Paleoclimate and Evolution with Emphasis on Human Origins. New Haven, CT: Yale University Press, 1995, pp. 242–48.

Tian J, Wang P, Cheng X, Li Q. Astronomically tuned Plio–Pleistocene benthic $\delta^{18}O$ record from South China Sea and Atlantic–Pacific comparison. *Earth Planet Sci Lett* 2002; 203: 1015–29.

Wilk MB, Gnanadesikan R. Probability plotting methods for the analysis of data. *Biometrika* 1968; 55: 1–17.

Williams DF, Thunell RC, Tappa E, Rio D, Raffi I. Chronology of the Pleistocene oxygen isotope record: 0–1.88 m.y. B.P. *Palaeogeogr Palaeoclimatol Palaeoecol* 1988; 64: 221–40.