Supplement of Clim. Past, 19, 1993–2012, 2023
https://doi.org/10.5194/cp-19-1993-2023-supplement

Climate of the Past
Open Access
EGU

Supplement of

Bayesian age models and stacks: combining age inferences from radiocarbon and benthic δ^{18}O stratigraphic alignment

Taehee Lee et al.

Correspondence to: Taehee Lee (taehee_lee@fas.harvard.edu) and Devin Rand (drand@ucsb.edu)

Supplement

S1. Transition Model

At a given core depth n, the transition model returns the probability of an age sample A_n and the sedimentation rate state W_n, given the normalized sedimentation rate and the previous sedimentation rate state W_{n+1}:

$$\pi(A_n, W_n | A_{n+1}, W_{n+1}, \phi, d_n, d_{n+1}, r) = \pi_1(W_n | W_{n+1}, \phi)\pi_2(A_n | A_{n+1}, W_n; d_n, d_{n+1}, r).$$

Here $\pi_1(W_n | W_{n+1}, \phi)$ returns the probability of transitioning from W_{n+1} to W_n. The 3x3 matrix parameter ϕ contains the probabilities of transitioning from each state to all other states. The three sedimentation rate states are expansion, steady, and contraction and have respective normalized sedimentation rate ranges of $(0,0.9220)$, $[0.9220,1.0850)$, and $[1.0850, \infty)$. ϕ can either remain fixed or can be optimized during age model construction.

The second term $\pi_2(A_n | A_{n+1}, W_n; d_n, d_{n+1}, r)$ returns the probability of the required sedimentation rate and is calculated using the mixed log-normal distribution fit to the normalized sedimentation rates from Lin et al., (2014):

$$\pi_2(A_n | A_{n+1}, W_n; d_n, d_{n+1}, r) \propto \left(\sum_{k=1}^{2} \omega_k \cdot Lognormal\left(\frac{A_{n+1} - A_n}{r \cdot (d_{n+1} - d_n)} \middle| \mu_k, \sigma_k \right) \right) \cdot 1_{\left\{ \frac{A_{n+1} - A_n}{r \cdot (d_{n+1} - d_n)} \in \mathbb{I}_{W_n} \right\}}(A_n).$$

Here d_n is the current depth, d_{n+1} is the pervious depth, and A_{n+1} is the previous age. Sedimentation rates are normalized by the depth-dependent average sedimentation rate, or r, which is calculated using the Nadaraya-Watson Kernel (Langrene and Warin, 2019). The variables, ω_k, μ_k, and σ_k are fixed weights, means and standard deviations that describe the log-normal mixture. The last term on the right is equal to 1 when the sedimentation rate is in the range of W_n and 0 otherwise. This effectively truncates the log-normal mixture and only allows sedimentation rates within the range of the given state.

Core	Latitude	Longitude	Depth (m)	Resolution (yrs)
DSDP594	-45.52	-174.95	1204	570
GeoB1711	-23.32	12.38	1967	550
GeoB7920-2	20.75	-18.58	2278	400
GeoB9508-5	15.5	-17.95	2384	170
GeoB9526-5	12.43	-18.05	3233	370
GIK17940-2	20.12	-117.38	1727	270
GIK17961-2	8.51	-112.33	1795	1020
GIK17964-2	6.16	-112.21	1556	760
H214	-36.92	-177.43	2045	340
KF13	37.58	-31.84	2690	1450
KNR159-5-36GGC	-27.51	-46.47	1268	370
KNR31-GPC5	33.69	-57.63	4583	150
M35003-4	12.09	-61.2433	1299	380
MD01-2416	51.27	-167.73	2317	80
MD01-2421	36.02	-141.78	2224	200
MD02-2489	54.39	-148.92	3640	120
MD03-2698	38.24	-10.39	4602	1350
MD07-3076Q	-44.15	-14.22	3770	280
MD84-527	-43.49	51.19	3262	690
MD88-770	-46.02	96.46	3290	590
MD95-2042	37.8	-10.17	3146	100
MD97-2120	-45.53	174.93	1210	120
MD97-2151	8.73	109.87	1598	210
MD98-2181	6.3	125.82	2114	130
MD99-2334K	37.8	-10.17	3146	300
MD99-2339	35.89	-7.53	1177	90
ODP1145	19.58	117.63	3175	1870
PO200-10-6-2	37.82	-9.5	1086	400
RC11-83	-41.6	9.8	4718	340
SO42-74KL	14.32	57.35	3212	360
SO50-31KL	18.76	115.87	3360	300
SU81-18	37.77	-10.18	3135	340
TR163-22	0.52	-92.4	2830	240
V19-30	-3.38	-83.52	3091	360
V35-5	7.2	112.08	1953	640
W8709A-13	42.12	-125.75	2712	970
W8709A-8	42.26	-127.68	3111	1200

Table S1: The 37 cores used to construct the transition model from Lin et al., (2014). The final column "resolution" lists the average number of years between calibrated radiocarbon ages.

S2. Emission Model

The radiocarbon emission model returns the likelihood of an observed radiocarbon measurement $y_{n,1}$ given a proposed calendar age A_n. The likelihood is calculated with a generalized student's t-distribution which depends on the calibration curve $\mu_C(A_n)$, the calibration curve uncertainty $\sigma_C^2(A_n)$, reservoir age ϱ_n, the combination of analytical measurement uncertainty and reservoir age uncertainty ς_n, and the fixed parameters a_1 =10 and b_1=11:

$$g_1(y_{n,1}|A_n) = T\left(y_{n,1}\middle|\mu_C(A_n) + \varrho_n, \sqrt{\frac{b_1}{a_1}(\sigma_C^2(A_n) + \varsigma_n)}; 2a_1\right).$$

The $\delta^{18}O$ emission model returns the likelihood of an observed $\delta^{18}O$ data point $y_{n,2}$ given a proposed age A_n and is also modeled with a generalized student's t-distribution. The $\delta^{18}O$ emission model depends on the target stack's time-dependent mean $\overline{\mu}(A_n)$ and variance $\overline{v}(A_n)$, the core specific shift h and scale σ parameters, and the fixed parameters a_2=3 and b_2=4:

$$g_2(y_{n,2}|A_n) = T\left(y_{n,2}\middle|\sigma \cdot \overline{\mu}(A_n) + h, \sqrt{\frac{b_2}{a_2}(\sigma)^2 \cdot \overline{v}(A_n)}; 2a_2\right).$$

Here we set a_2 and b_2 based on the observed residuals in the DNEA and ITWA stacks (Figure S1). The thicker tails generated by a_2=3 and b_2=4 (compared to the values of 10 and 11 used for ^{14}C) better fit the larger residuals. The degrees of freedom for each generalized student's t-distribution is equal to $2a_i$, thus the radiocarbon distribution has 22 degrees of freedom and the $\delta^{18}O$ emission model has six degrees of freedom.

The emission model for the additional age information can be specified as either a uniform distribution or Gaussian distribution with mean $y_{n,3}$ and uncertainty $\underline{\sigma_n}$ specified by the user:

$$g_3(y_{n,3}|A_n) = N(y_{n,3}|A_n, \underline{\sigma_n^2}) \vee U(y_{n,3}|A_n - \underline{\sigma_n}, A_n + \underline{\sigma_n}).$$

If a Gaussian distribution is specified, $\underline{\sigma_n}$ is the standard deviation; if a uniform distribution is specified, $\underline{\sigma_n}$ is the 50% confidence interval width.

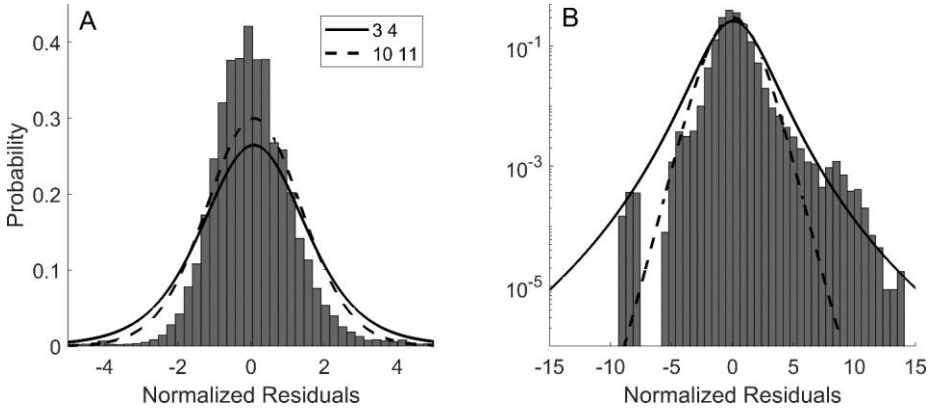

Figure S1: Normalized $\delta^{18}O$ residuals combined from the ITWA and DNEA stacks on a linear y-axis (A) and a log scale (B). Note the expanded x-axis in panel (B). Residuals are normalized by the standard deviation of the stack. The students t-distribution with a_2=3 and b_2=4 better fits the larger residuals.

S3 Additional Figures and Tables

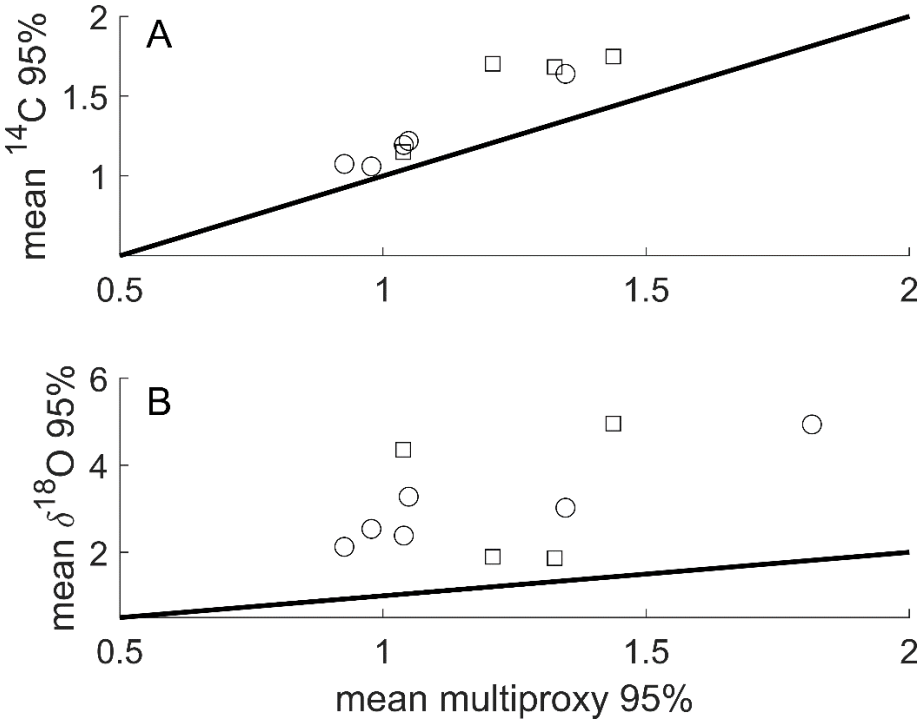

Figure S2: Comparison of the mean 95% confidence interval widths for BIGMACS age models using radiocarbon-only (A) and $\delta^{18}O$-only (B) mode compared to multiproxy age models for cores in the DNEA (squares) and ITWA (circles) stacks. The solid black line marks a 1:1 ratio. The 95% confidence interval widths for radiocarbon-only and $\delta^{18}O$-only age models, respectively, are on average 262 years and 1.92 kyr larger than the 95% confidence interval widths for multiproxy age models

Core	Modern %	LGM %	Shift (‰)	Scale
DNEA				
MD95-2042	26	23	0.25	0.98
MD99-2334	26	24	-0.07	1.04
SU81-18	26	23	0.08	1
GeoB7920-2	23	24	-0.01	0.98
ODP658C	23	24	0.16	0.92
GeoB9508-5	24	26	0.11	0.99
ITWA				
M35003-4	33	24	-0.25	0.99
KNR197-3-53GGC	33	20	0.09	1
KNR197-3-9GGC	48	28	0.3	0.91
GeoB16206-1	31	27	0.06	0.99

Table S2: SCW percentages at core sites based on modern (Gebbie & Huybers, 2012) and LGM (Oppo et al., 2018) water mass reconstructions. Also, the shift and scale parameters applied to each core during alignment and stacking as estimated by BIGMACS.

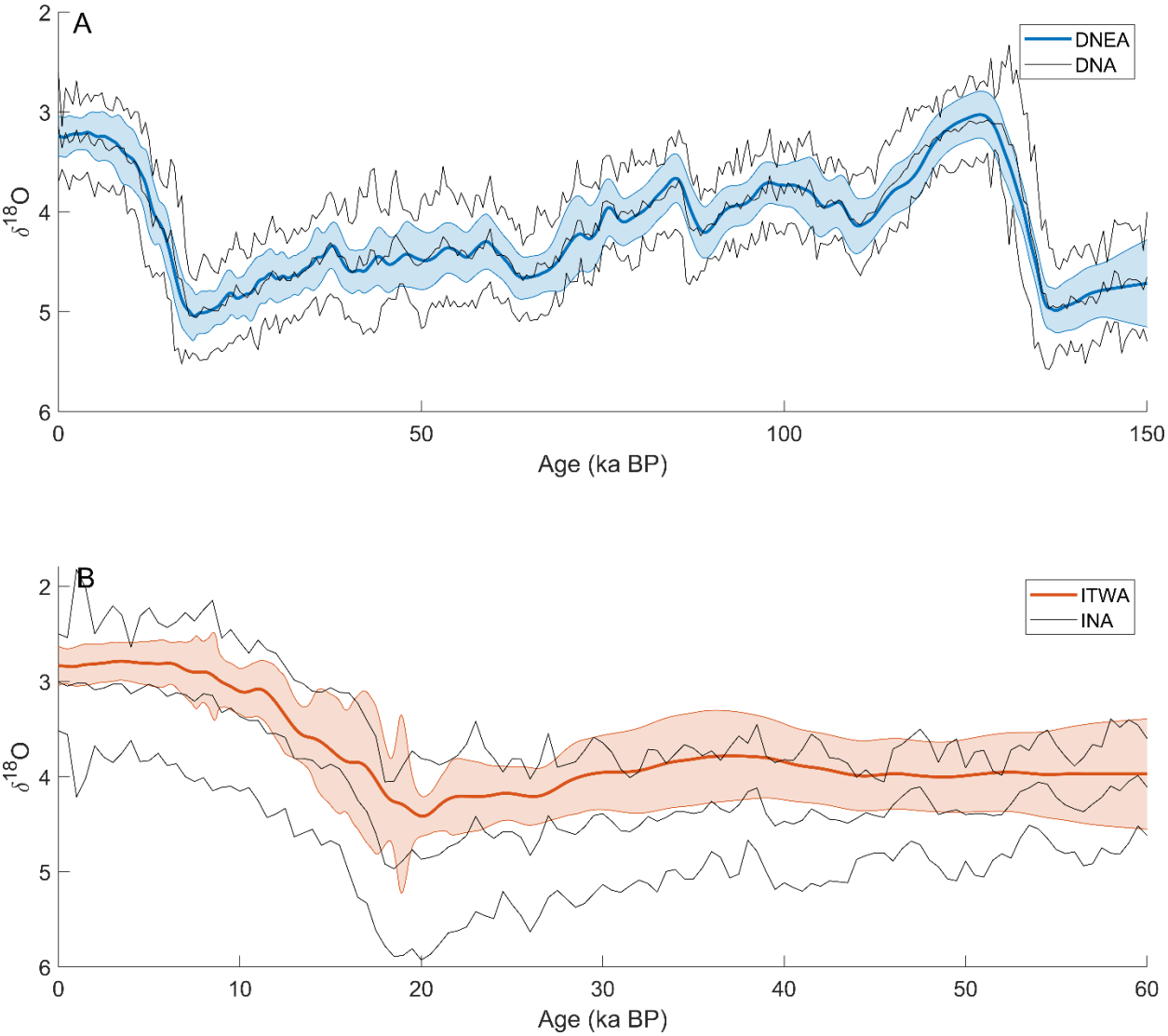

Figure S3: (A) The DNEA stack compared to the DNA stack (Lisiecki & Stern, 2016) used as the initial alignment target. (B) The ITWA stack compared to the INA stack (Lisiecki & Stern, 2016) used as the initial alignment target. Lines mark the mean and 95% confidence interval for each stack. The DNA and INA stacks were constructed using more cores spanning a larger oceanographic region.

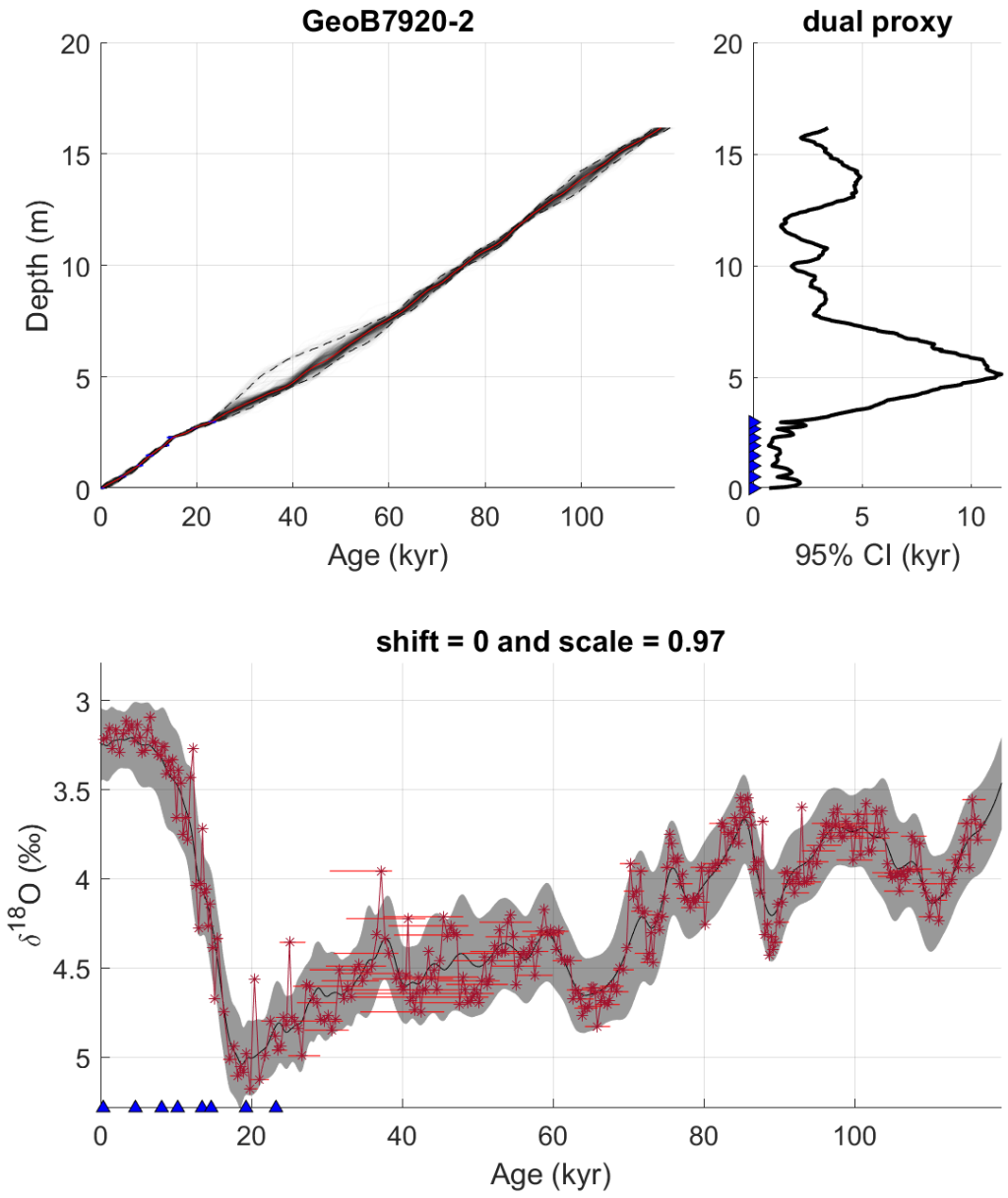

Figure S4: Age model summary for GeoB7920-2. (Top Left) Age vs. Depth plot with the median age model displayed in red and the 95% credible interval displayed as dotted black lines. Shading reveals samples density and radiocarbon ages are shown in blue. (Top Right) Width of the 95% credible interval with the depths of radiocarbon ages displayed as blue triangles on the y-axis. (Bottom) The alignment to the final stack benthic d18O displayed in red, age uncertainty represented by red horizontal lines (95% credible interval), and radiocarbon ages shown as blue triangles.

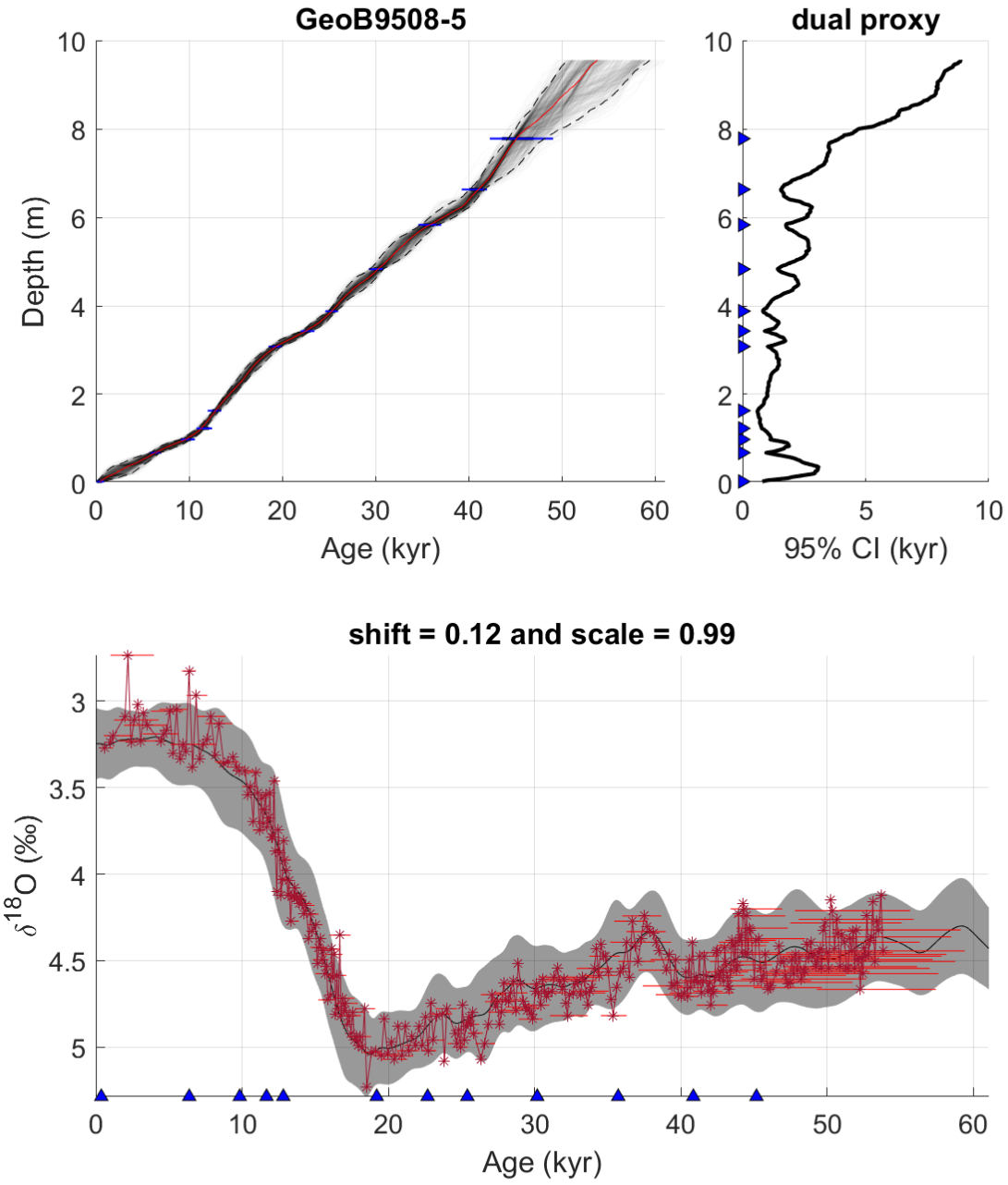

Figure S5: Age model summary for GeoB9508-5

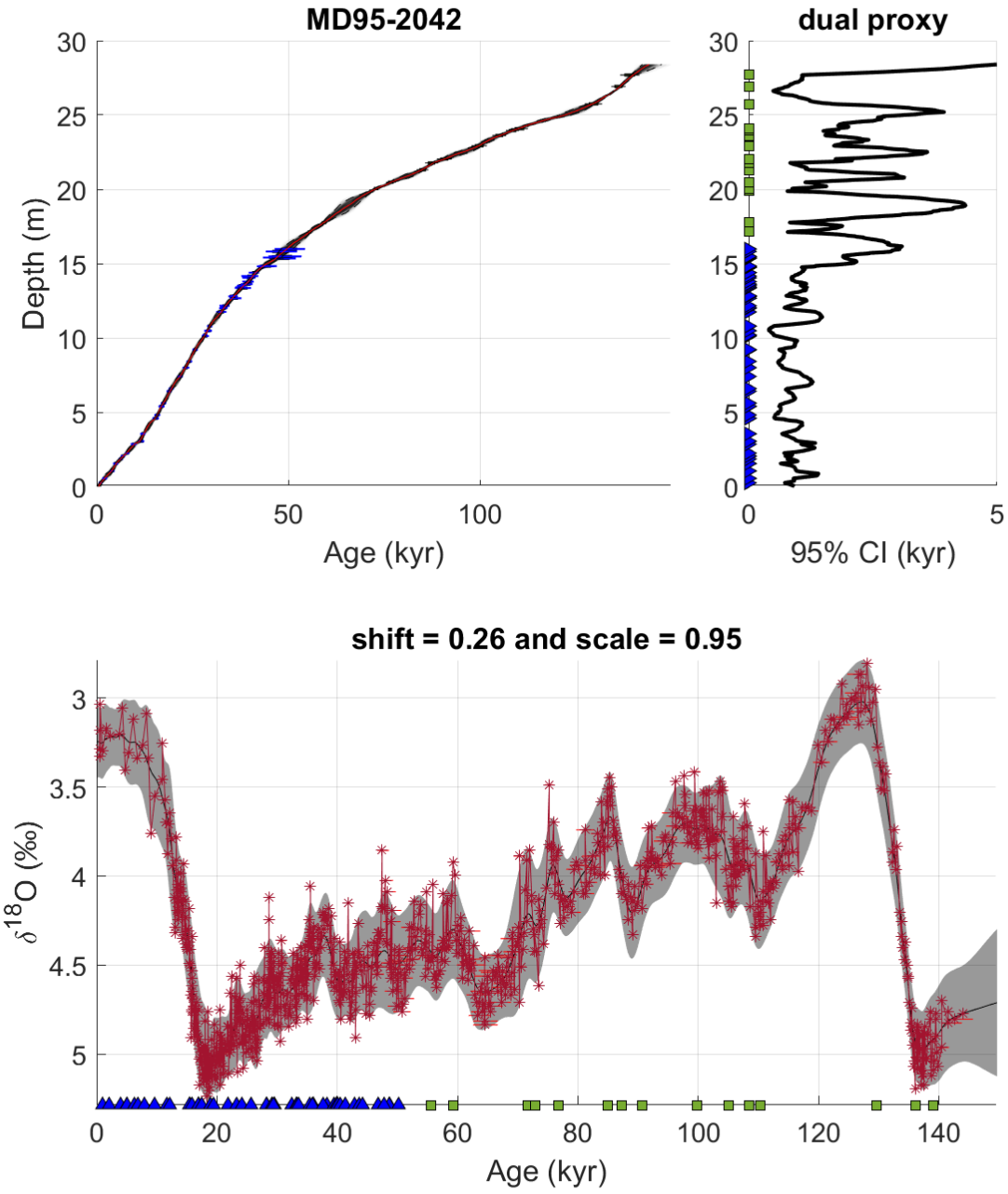

Figure S6: Age model summary for MD95-2042. Additional ages are displayed as green squares.

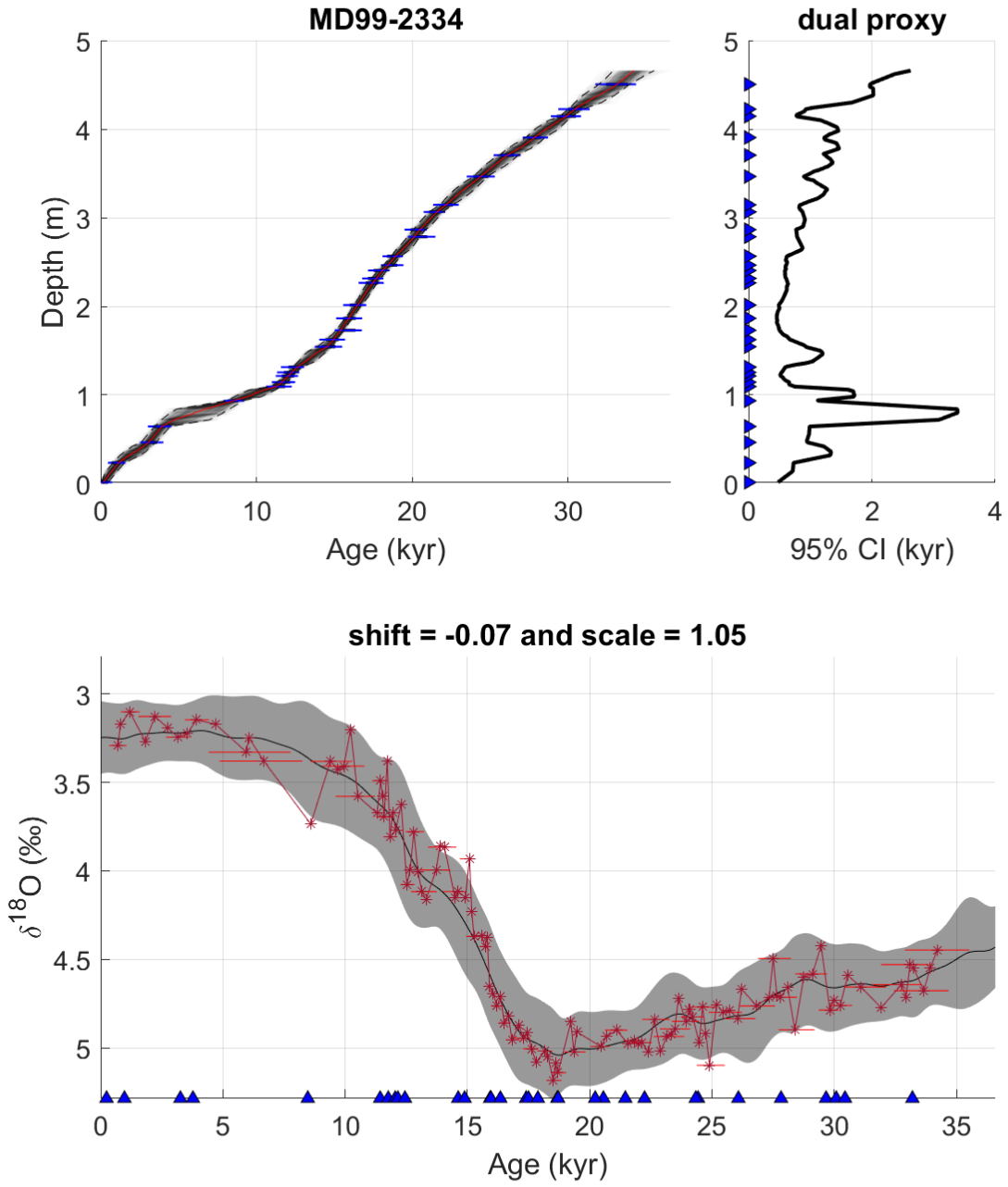

Figure S7: Age model summary for MD99-2334.

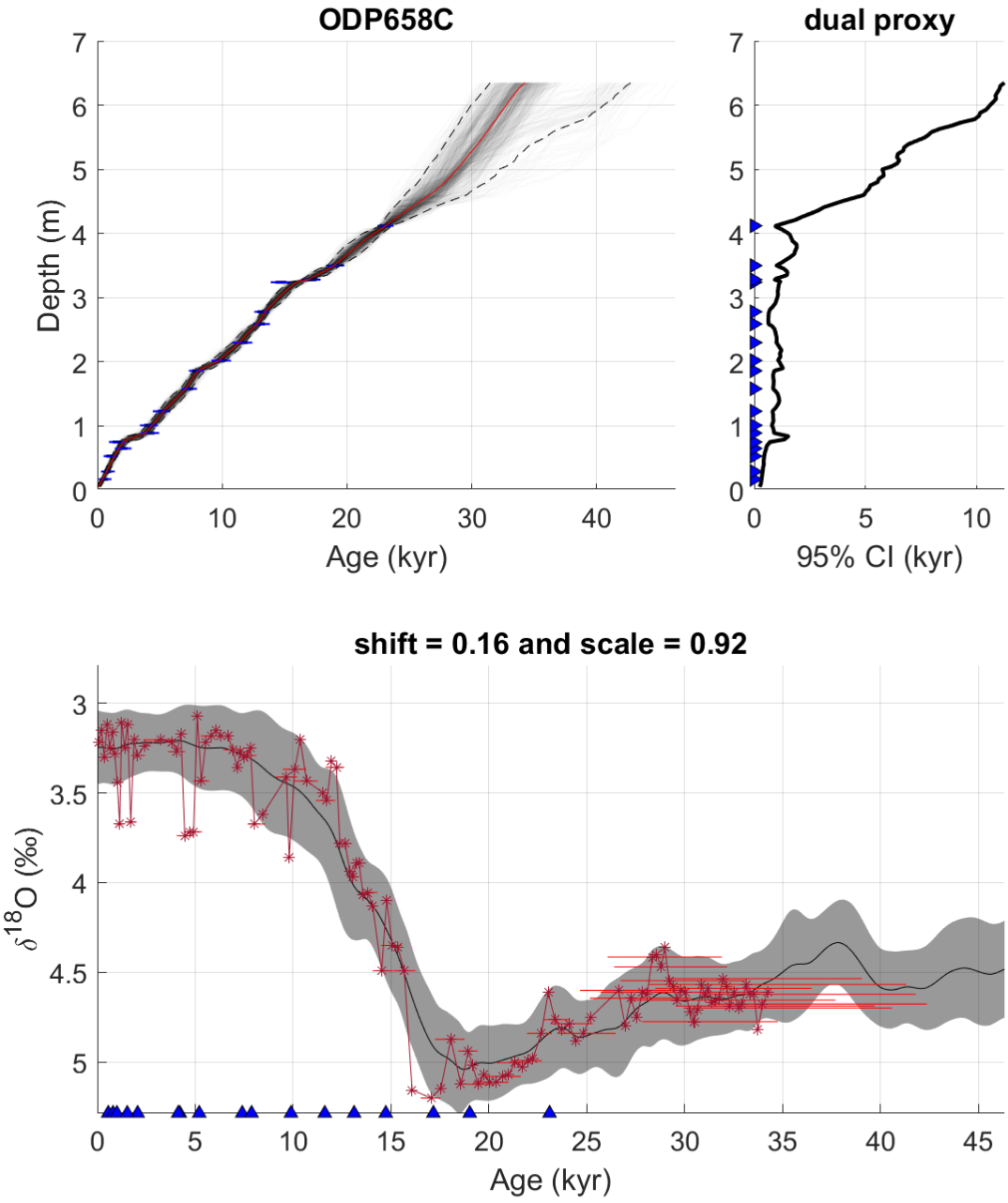

Figure S8: Age model summary for ODP658C.

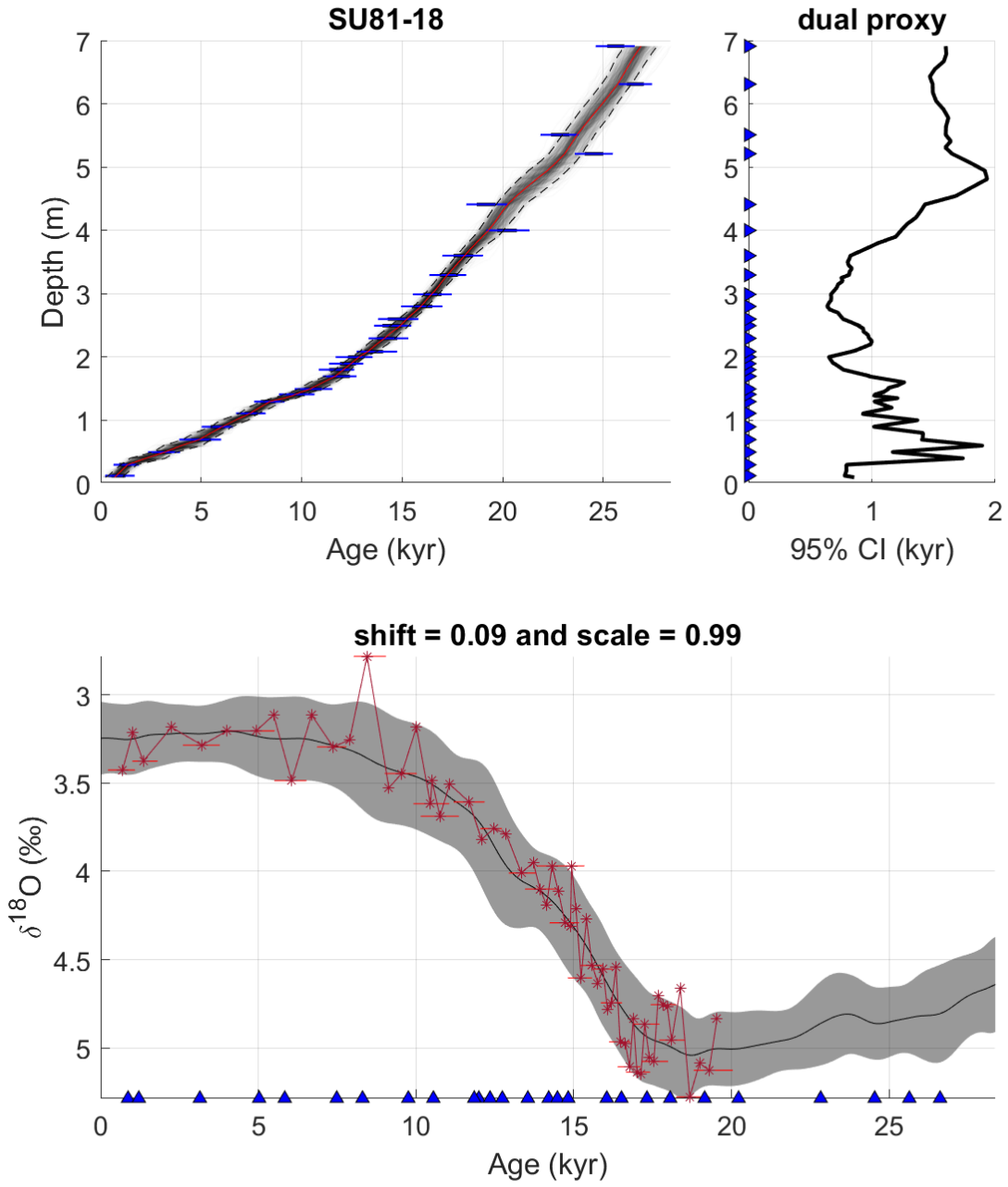

Figure S9: Age model summary for SU81-18.

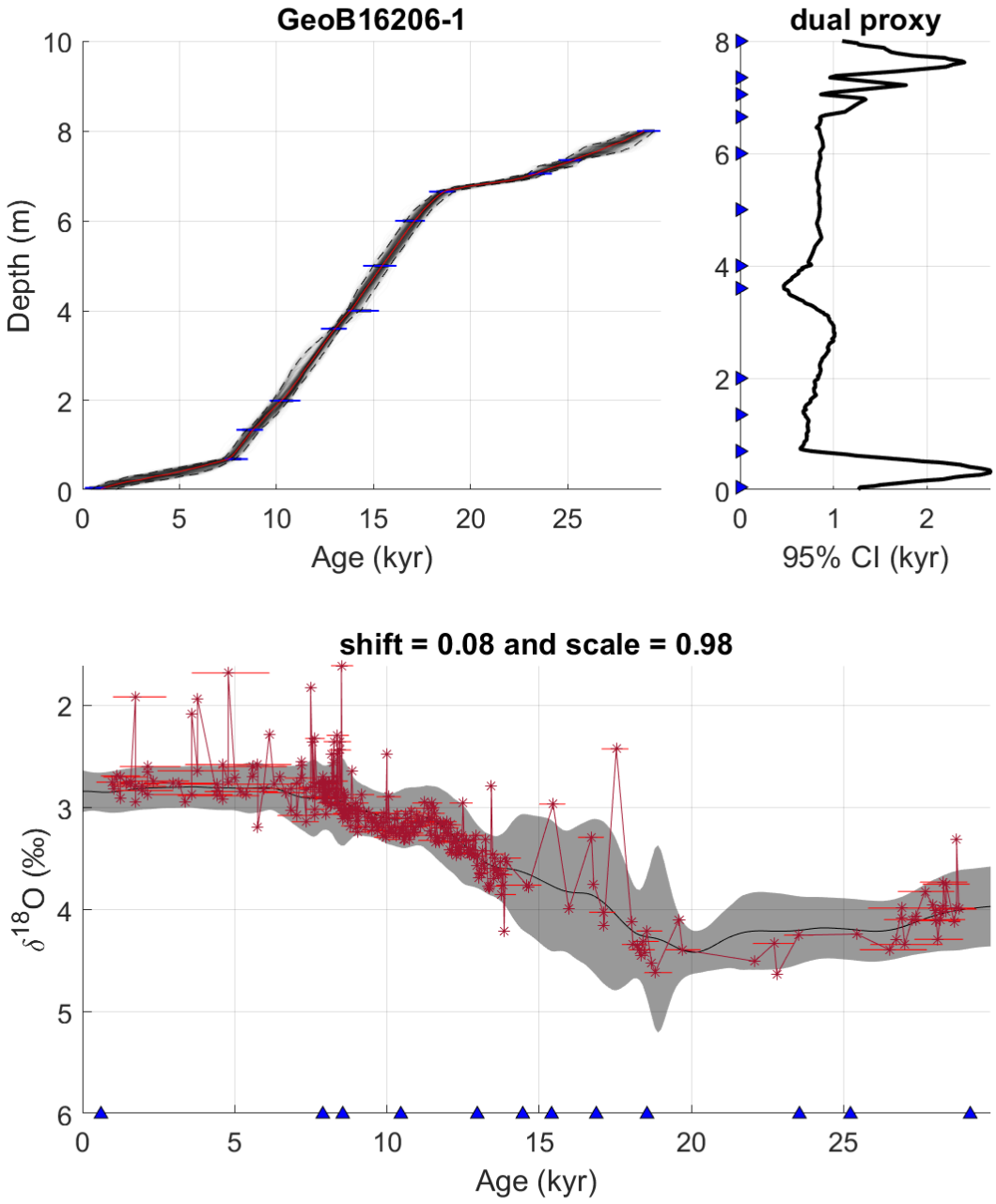

Figure S10: Age model summary for GeoB16206-1

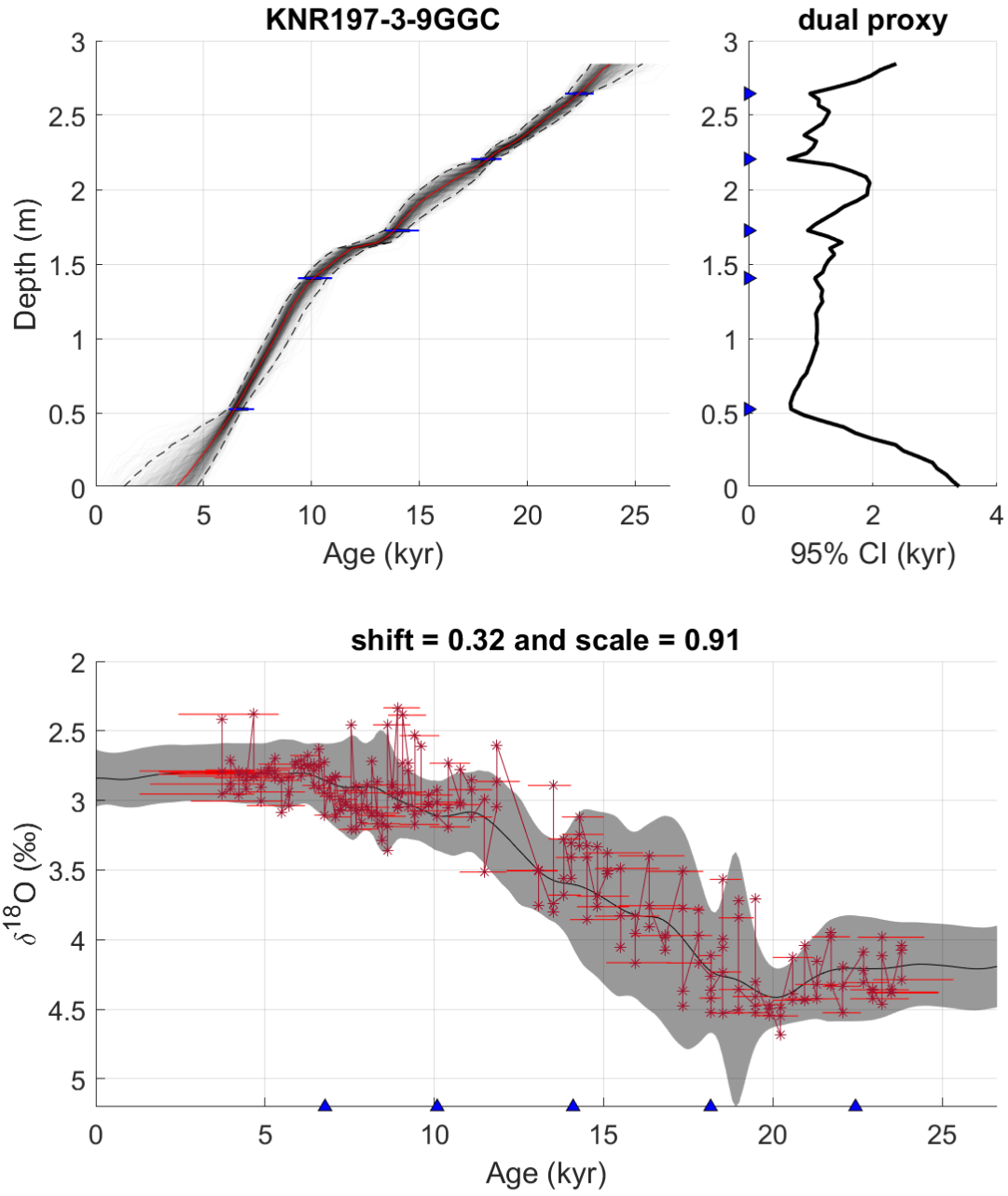

Figure S11: Age model summary for KNR197-3-9GGC.

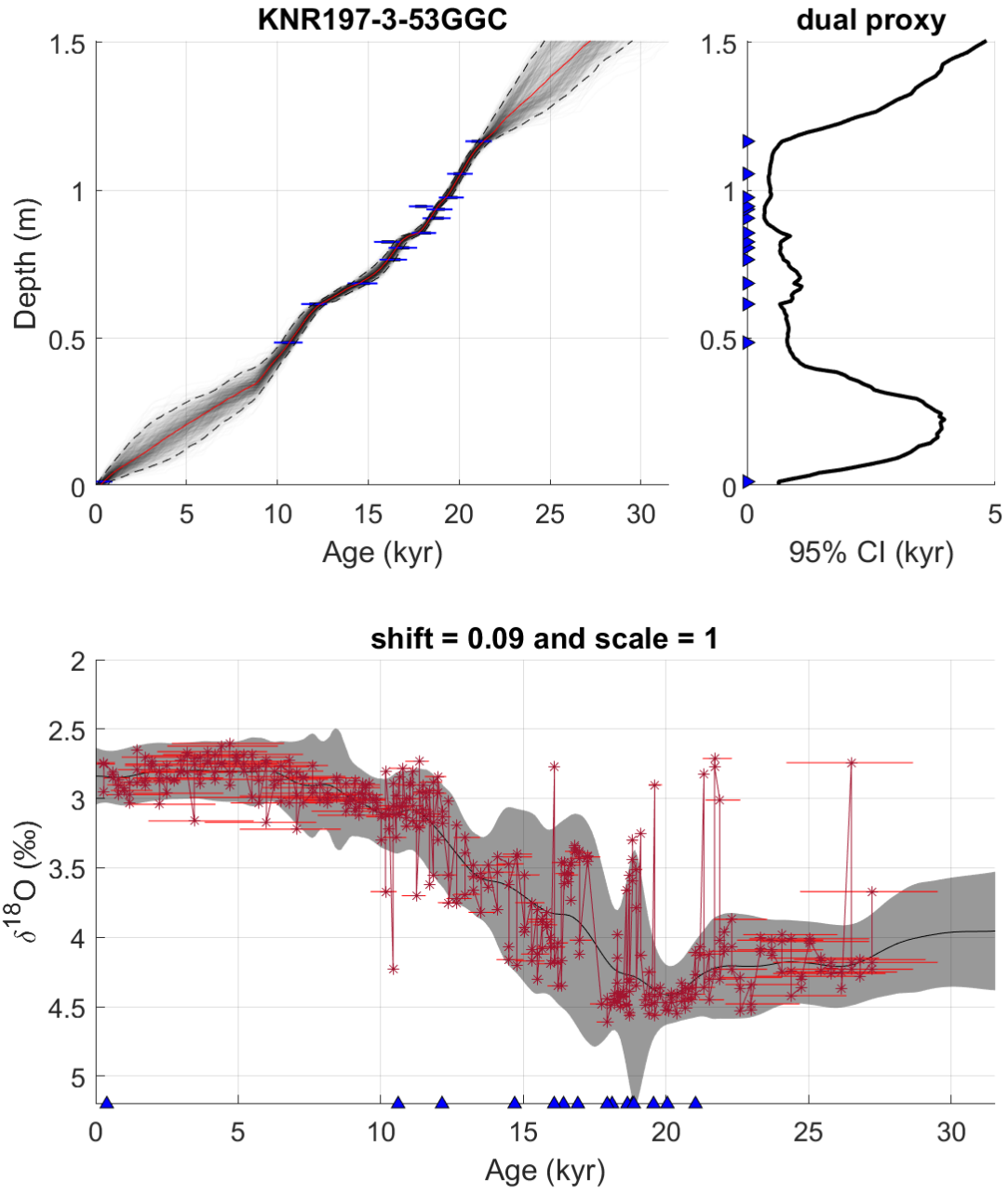

Figure S12: Age model summary for KNR197-3-53GGC.

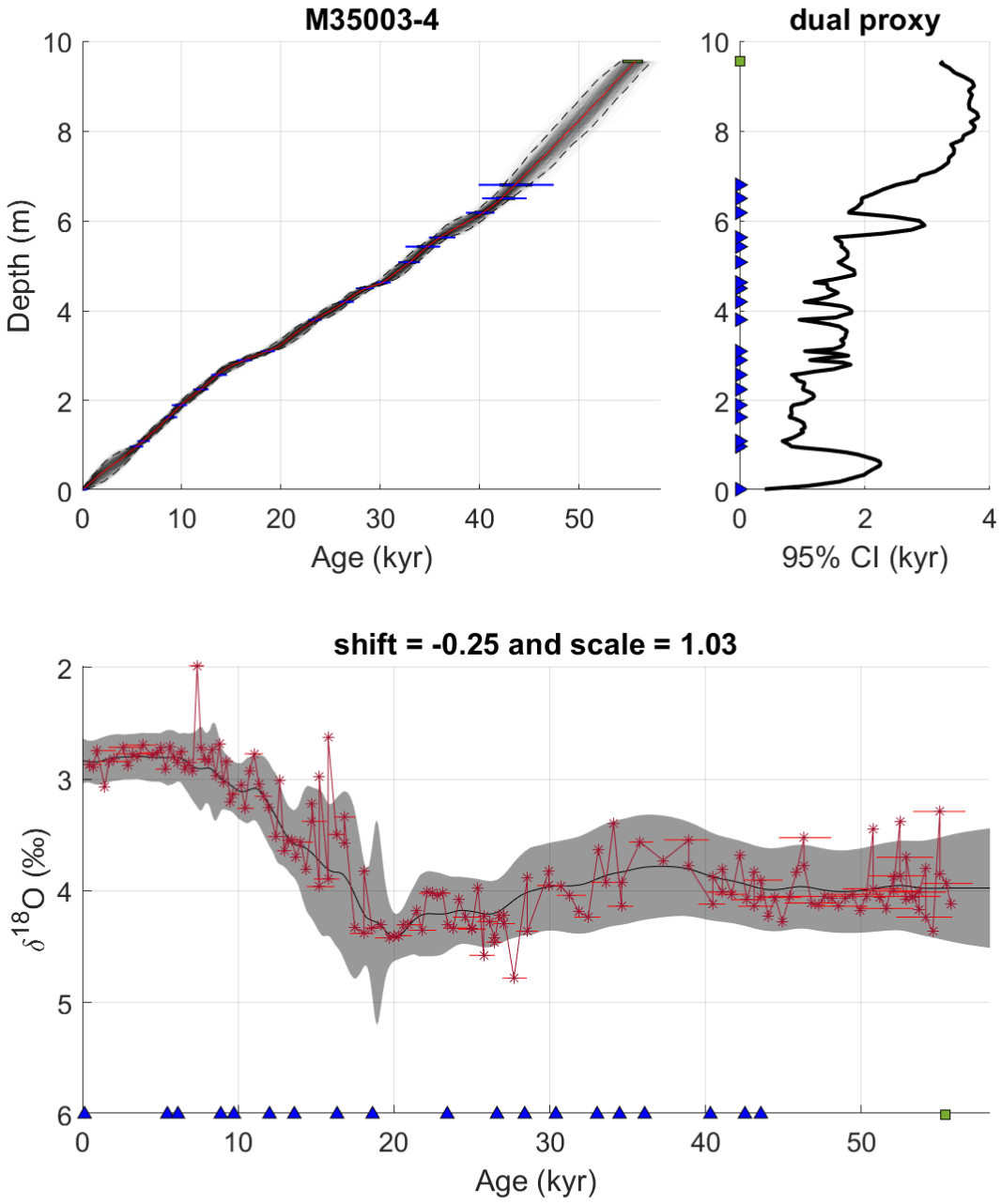

Figure S13: Age model summary for M35003-4.

S4. Alignment Algorithm

The alignment algorithm of BIGMACS consists of two parts: one is to sample age paths from the posterior distribution by the hybrid of particle smoothing [Doucet et al. (2001)] and Markov-chain Monte Carlo (MCMC) algorithm [Martino et al. (2015)], and the other is to estimate the alignment parameters given the sampled age paths. Supplemental sections S1 and S2 introduced quantitative descriptions of the transition and emission models used by BIGMACS. Here, we present detailed formulations of all parts of the algorithm. The following definitions are assumed throughout the supplementary materials. Suppose that there are M sediment cores.

- $\mathrm{D} = \left\{\mathrm{D}^{(m)}\right\}_{m=1}^{\mathrm{M}}$: a set of core depths, where $\mathrm{D}^{(m)} = \left\{d_n^{(m)}\right\}_{n=1}^{\mathrm{L}_m}$ is those of sediment core m.
- $\mathrm{Y} = \left\{\mathrm{Y}^{(m)}\right\}_{m=1}^{\mathrm{M}}$: a set of proxy observations, where $\mathrm{Y}^{(m)} = \left\{y_n^{(m)}\right\}_{n=1}^{\mathrm{L}_m}$ is those of sediment core m.
  - $y_n^{(m)} = \left(y_{n,1}^{(m)}, y_{n,2}^{(m)}, y_{n,3}^{(m)}\right)$: a pair of radiocarbon, $\delta^{18}\mathrm{O}$ observations and other proxies that give age information at depth $d_n^{(m)}$, respectively.
- $\Theta = \left\{\phi^{(m)}, r^{(m)}, \sigma^{(m)}, h^{(m)}\right\}_{m=1}^{\mathrm{M}}$: a set of core-specific parameters that are used in the alignment algorithm.
  - $\phi^{(m)}$ is a transition matrix that maps $\{\mathbb{C}, \mathbb{A}, \mathbb{E}\}$ to itself.
  - $r^{(m)}$ is a depth-scale parameter that rescales $\mathrm{D}^{(m)}$ to adjust the differences in average accumulation rates.
  - $\sigma^{(m)}$ is a core-specific scale parameter for $\delta^{18}\mathrm{O}$.
  - $h^{(m)}$ is a core-specific shift parameter for $\delta^{18}\mathrm{O}$. These scale and shift parameters standardize $\delta^{18}\mathrm{O}$ observations core-specifically to align them to the stack. Details (in formulation) can be found in the definition of likelihood (emission model) in Section S4.1.
- $\mathrm{A} = \left\{\mathrm{A}^{(m)}\right\}_{m=1}^{\mathrm{M}}$: a set of hidden age paths to sample, where $\mathrm{A}^{(m)} = \left\{\mathrm{A}_n^{(m)}\right\}_{n=1}^{\mathrm{L}_m}$ is those of sediment core m.
- $\mathrm{W} = \left\{\mathrm{W}^{(m)}\right\}_{m=1}^{\mathrm{M}}$: a set of medium latent variables, where $\mathrm{W}^{(m)} = \left\{\mathrm{W}_n^{(m)}\right\}_{n=1}^{\mathrm{L}_m}$ is those of sediment core m.
  - $\mathrm{W}_n^{(m)} \in \{\mathbb{C}, \mathbb{A}, \mathbb{E}\}$ stands for contraction, average and expansion, respectively.

S4.1. State-space Modelling

The goal is to sample $\widetilde{\mathrm{A}}^{(m,k)} \sim p\left(\mathrm{A}^{(m)} \middle| \mathrm{D}, \mathrm{Y}, \Theta\right)$ for each sediment core m, and each age sample k, where $p\left(\mathrm{A}^{(m)} \middle| \mathrm{D}, \mathrm{Y}, \Theta\right)$ is the posterior of the hidden age path $\mathrm{A}^{(m)}$ given depths, proxy observations and alignment parameters. To define the posterior, we have the following prior and likelihood, or the transition and emission models in the terminology of the state-space model [Hangos et al. (2001)]:

- Prior (Transition Model)

$$\pi\left(\mathrm{A}^{(m)}, \mathrm{W}^{(m)} \middle| \mathrm{D}^{(m)}, \Theta\right) = \pi\left(\mathrm{A}_{\mathrm{L}_m}^{(m)}, \mathrm{W}_{\mathrm{L}_m}^{(m)}\right) \prod_{n=1}^{\mathrm{L}_m - 1} \pi\left(\mathrm{A}_n^{(m)}, \mathrm{W}_n^{(m)} \middle| \mathrm{A}_{n+1}^{(m)}, \mathrm{W}_{n+1}^{(m)}; d_n^{(m)}, d_{n+1}^{(m)}, \phi^{(m)}, r^{(m)}\right)$$

$$\pi\left(\mathrm{A}_n^{(m)}, \mathrm{W}_n^{(m)} \middle| \mathrm{A}_{n+1}^{(m)}, \mathrm{W}_{n+1}^{(m)}; d_n^{(m)}, d_{n+1}^{(m)}, \phi^{(m)}, r^{(m)}\right) \propto \pi_1\left(\mathrm{W}_n^{(m)} \middle| \mathrm{W}_{n+1}^{(m)}, \phi^{(m)}\right) \pi_2\left(\mathrm{A}_n^{(m)} \middle| \mathrm{A}_{n+1}^{(m)}, \mathrm{W}_n^{(m)}; d_n^{(m)}, d_{n+1}^{(m)}, r^{(m)}\right)$$

, where π_1 and π_2 are defined as follows:

$$\pi_1\left(\mathrm{W}_n^{(m)} \middle| \mathrm{W}_{n+1}^{(m)}, \phi^{(m)}\right) = \phi^{(m)}_{\mathrm{W}_{n+1}^{(m)}, \mathrm{W}_n^{(m)}}$$

$$\pi_2\left(\mathrm{A}_n^{(m)}\middle|\mathrm{A}_{n+1}^{(m)},\mathrm{W}_n^{(m)};d_n^{(m)},d_{n+1}^{(m)},r^{(m)}\right)$$
$$\propto\left(\sum_{k=1}^{2}w_k\cdot LogNormal\left(\frac{\mathrm{A}_{n+1}^{(m)}-\mathrm{A}_n^{(m)}}{r^{(m)}\cdot\left(d_{n+1}^{(m)}-d_n^{(m)}\right)}\middle|\underline{\mu}_k,\underline{\sigma}_k\right)\right)\cdot 1_{\left\{\frac{\mathrm{A}_{n+1}^{(m)}-\mathrm{A}_n^{(m)}}{r^{(m)}\cdot\left(d_{n+1}^{(m)}-d_n^{(m)}\right)}\in\mathbb{I}_{\mathrm{W}_n^{(m)}}\right\}}\left(\mathrm{A}_n^{(m)}\right)$$

, where $\mathbb{I}_{\mathbb{C}}=(0,0.9220)$, $\mathbb{I}_{\mathbb{A}}=[0.9220,1.0850)$, $\mathbb{I}_{\mathbb{E}}=[1.0850,\infty)$ are the intervals that partition $\mathbb{R}_{>0}$ and $\left\{w_k,\underline{\mu}_k,\underline{\sigma}_k\right\}_{k=1}^{2}$ are the fixed weight, mean and standard deviation parameters of the mixture of log-normal distributions, which are trained from [Lin et al., 2014]. To be specific, $\left(w_1,\underline{\mu}_1,\underline{\sigma}_1\right)=\left(0.642432,0.0198,\sqrt{0.0216}\right)$ and $\left(w_2,\underline{\mu}_2,\underline{\sigma}_2\right)=\left(0.357568,0.0297,\sqrt{0.0929}\right)$. In words, each latent variable $\mathrm{W}_n^{(m)}$ confines the transition from an age $\mathrm{A}_{n+1}^{(m)}$ to another $\mathrm{A}_n^{(m)}$ in one of the three regions $\{\mathbb{C},\mathbb{A},\mathbb{E}\}$ that correspond to $\mathbb{I}_{\mathbb{C}}$, $\mathbb{I}_{\mathbb{A}}$ and $\mathbb{I}_{\mathbb{E}}$. In the transition model, the transition matrix $\phi^{(m)}$ and depth-scale parameter $r^{(m)}$ are the parameters to estimate in the training phase, given the sampled age path $\mathrm{A}^{(m)}$, for each core m.

- Likelihood (Emission Model)

$$p\left(\mathrm{Y}^{(m)}\middle|\mathrm{A}^{(m)};\Theta\right)=\prod_{n=1}^{\mathrm{L}_m}p\left(y_{n,1}^{(m)},y_{n,2}^{(m)},y_{n,3}^{(m)}\middle|\mathrm{A}_n^{(m)}\right)=\prod_{n=1}^{\mathrm{L}_m}g_1\left(y_{n,1}^{(m)}\middle|\mathrm{A}_n^{(m)}\right)g_2\left(y_{n,2}^{(m)}\middle|\mathrm{A}_n^{(m)}\right)g_3\left(y_{n,3}^{(m)}\middle|\mathrm{A}_n^{(m)}\right)$$

, where g_1, g_2 and g_3 are defined as follows:

$$g_1\left(y_{n,1}^{(m)}\middle|\mathrm{A}_n^{(m)}\right)=\mathcal{T}_{2a_1}\left(y_{n,1}^{(m)}\middle|\mu_{\mathrm{C}}\left(\mathrm{A}_n^{(m)}\right)+\varrho_n^{(m)},\sqrt{\frac{b_1}{a_1}\left(\sigma_{\mathrm{C}}^2\left(\mathrm{A}_n^{(m)}\right)+\varsigma_n^{(m)}\right)}\right)$$

$$g_2\left(y_{n,2}^{(m)}\middle|\mathrm{A}_n^{(m)}\right)=\mathcal{T}_{2a_2}\left(y_{n,2}^{(m)}\middle|\sigma^{(m)}\cdot\overline{\mu}\left(\mathrm{A}_n^{(m)}\right)+h^{(m)},\sqrt{\frac{b_2}{a_2}(\sigma^{(m)})^2\cdot\overline{\nu}\left(\mathrm{A}_n^{(m)}\right)}\right)$$

$$g_3\left(y_{n,3}^{(m)}\middle|\mathrm{A}_n^{(m)}\right)=\mathcal{N}\left(y_{n,3}^{(m)}\middle|\mathrm{A}_n^{(m)},\underline{\nu}_n^{(m)}\right)\ or\ \mathcal{U}\left(y_{n,3}^{(m)}\middle|\mathrm{A}_n^{(m)}-\sqrt{\underline{\nu}_n^{(m)}},\mathrm{A}_n^{(m)}+\sqrt{\underline{\nu}_n^{(m)}}\right)$$

, where $\varrho_n^{(m)}$ and $\varsigma_n^{(m)}$ are given together with the radiocarbon determination (a measurement of the amount of radiocarbon in a sample) $y_{n,1}^{(m)}$ a priori, μ_{C} and σ_{C}^2 are the mean and variance functions of the radiocarbon calibration curve [Reimer et al. (2020); Hogg et al. (2020); Heaton et al. (2020)], and $\overline{\mu}$ and $\overline{\nu}$ are the mean and variance functions from the target δ^{18}O stack. (a_1,b_1) and (a_2,b_2) are the pairs of fixed hyperparameters for the generalized Student's t-distribution [Christen and Sergio (2009)] to balance observations that follow the given calibration curve (or stack) and potential outliers. For g_3 that reflects our prior knowledge regarding the ages, BIGMACS allows to pick among the Gaussian-based model and the uniform-based model. For the Gaussian-based model, an uncertainty input $\underline{\nu}_n^{(m)}$ works as the variance while $\underline{\nu}_n^{(m)}$ defines the 50% confidence interval for the uniform-based model. In the emission model, the core-specific scale and shift parameters $\sigma^{(m)}$ and $h^{(m)}$ are estimated in the training phase, given the sampled age path $\mathrm{A}^{(m)}$ and the target δ^{18}O stack, for each core m.

The above prior and likelihood, or transition and emission models, define the following joint distribution:

$$p(\mathrm{Y},\mathrm{A},\mathrm{W}|\mathrm{D},\Theta)=\prod_{m=1}^{\mathrm{M}}\pi\left(\mathrm{A}^{(m)},\mathrm{W}^{(m)}\middle|\mathrm{D}^{(m)};\Theta\right)p\left(\mathrm{Y}^{(m)}\middle|\mathrm{A}^{(m)};\Theta\right)$$

Thus, $p\left(\mathrm{A}^{(m)},\mathrm{W}^{(m)}\middle|\mathrm{D},\mathrm{Y},\Theta\right)=p\left(\mathrm{A}^{(m)},\mathrm{W}^{(m)}\middle|\mathrm{D}^{(m)},\mathrm{Y}^{(m)},\Theta\right)\propto\pi\left(\mathrm{A}^{(m)},\mathrm{W}^{(m)}\middle|\mathrm{D}^{(m)},\Theta\right)p\left(\mathrm{Y}^{(m)}\middle|\mathrm{A}^{(m)};\Theta\right)$ for each m, which allows to run the sampling algorithm parallelly over sediment cores. Note that $\mathrm{W}^{(m)}$ is deterministic given $\mathrm{A}^{(m)}$ and the problem is now defined as a state-space model.

S4.2. Sampling

As mentioned earlier, BIGMACS samples the age paths $\mathrm{A}^{(m)}$ (and $\mathrm{W}^{(m)}$) from the posterior by the hybrid of particle smoothing and MCMC algorithms, given all the parameters to be either fixed or estimated. Though the particle smoothing is an efficient method of sampling continuous hidden states from a state-space model, only a small portion of proposed samples, "particles", contribute to the set of sampled paths in practice. Though MCMC can sample the hidden variables in principle, it might require a very long chain before the burn-in phase, especially if a good initialization is not given. Here, we run the particle smoothing for initializing the sampled age paths and then run the Metropolis-Hastings algorithm [Metropolis et al. (1953); Hastings (1970)] to "refine" the previously initialized paths.

The particle smoothing consists of two parts. The forward algorithm samples a set of candidates, or "particles", from a proposal distribution $q_n^{(m)}$ for each step n, and computes weights on those particles to approximate the forward posterior with an empirical distribution. In formulation, it is expressed as follows:

$$p\left(\mathrm{A}_n^{(m)},\mathrm{W}_n^{(m)}\middle|\mathrm{D}_{1:n}^{(m)},\mathrm{Y}_{1:n}^{(m)},\Theta\right)\approx\sum_{k=1}^{\mathrm{K}}\omega_{n,k}^{(m)}\cdot 1_{\left\{\mathrm{A}_n^{(m)}=a_{n,k}^{(m)},\mathrm{W}_n^{(m)}=w_{n,k}^{(m)}\right\}}\left(\mathrm{A}_n^{(m)},\mathrm{W}_n^{(m)}\right)$$

, where $\left\{a_{n,k}^{(m)},w_{n,k}^{(m)}\right\}_{k=1}^{\mathrm{K}}$ are the sampled particles from $q_n^{(m)}$ at step n and $\left\{\omega_{n,k}^{(m)}\right\}_{k=1}^{\mathrm{K}}$ are the associated weights that are sum to 1. Then, the forward posterior of the next step is updated iteratively as follows:

$$p\left(\mathrm{A}_{n-1}^{(m)},\mathrm{W}_{n-1}^{(m)}\middle|\mathrm{D}_{1:n-1}^{(m)},\mathrm{Y}_{1:n-1}^{(m)},\Theta\right)\approx\sum_{k=1}^{\mathrm{K}}\omega_{n-1,k}^{(m)}\cdot 1_{\left\{\mathrm{A}_{n-1}^{(m)}=a_{n-1,k}^{(m)},\mathrm{W}_{n-1}^{(m)}=w_{n-1,k}^{(m)}\right\}}\left(\mathrm{A}_{n-1}^{(m)},\mathrm{W}_{n-1}^{(m)}\right)$$

, where $\left\{a_{n-1,k}^{(m)},w_{n-1,k}^{(m)}\right\}_{k=1}^{\mathrm{K}}\sim_{i.i.d.}q_{n-1}^{(m)}$ and for $\sum_{k=1}^{\mathrm{K}}\omega_{n-1,k}^{(m)}=1$,

$$\omega_{n-1,k}^{(m)}\propto\frac{p\left(y_{n-1,1}^{(m)},y_{n-1,2}^{(m)}\middle|a_{n-1,k}^{(m)};\Theta\right)}{q_{n-1}^{(m)}\left(a_{n-1,k}^{(m)},w_{n-1,k}^{(m)}\right)}\sum_{s=1}^{\mathrm{K}}\omega_{n,s}^{(m)}\pi\left(a_{n-1,k}^{(m)},w_{n-1,k}^{(m)}\middle|a_{n,s}^{(m)},w_{n,s}^{(m)};d_{n-1}^{(m)},d_n^{(m)},\phi^{(m)},r^{(m)}\right)$$

The backward algorithm samples each hidden alignment given the depth above it (from the top of the core down) as well as all inputs and outputs iteratively until a complete path is sampled. In formulation, it is expressed as follows:

$$\begin{aligned}p\left(\mathrm{A}_n^{(m)}=a_{n,k}^{(m)},\mathrm{W}_n^{(m)}=w_{n,k}^{(m)}\middle|\mathrm{A}_{n-1}^{(m)}=\tilde{a}_{n-1}^{(m)},\mathrm{W}_{n-1}^{(m)}=\tilde{w}_{n-1}^{(m)}\right)\\ \propto\omega_{n,k}^{(m)}\cdot\pi\left(\tilde{a}_{n-1}^{(m)},\tilde{w}_{n-1}^{(m)}\middle|a_{n,k}^{(m)},w_{n,k}^{(m)};d_{n-1}^{(m)},d_n^{(m)},\phi^{(m)},r^{(m)}\right)\end{aligned}$$

Note that the particle smoothing is reduced to a hidden Markov model (HMM) [Durbin et al. (1998)] if the proposal distribution is set to have the same finite support and all elements in the support are sampled once as particles. Also, because the particle smoothing does not compute the exact forward posterior, this method has limitations that HMMs do not. First, performance is dependent on the user-specific proposal distributions. Second, a small number of output outliers might ruin the inference, especially if the transition model is too rigid. Third, the weights assigned to the particles are often too small to affect the backward sampling algorithm, which might cause a trouble in learning emission and transition models by the Baum-Welch EM algorithm [Dempster et al. (1977);

Durbin et al. (1998)]. To resolve the first limitation, here, we iterate the sampling part consisting of the particle smoothing and Metropolis-Hastings and the parameter estimation part for Θ until convergence. Suppose that we obtained T sampled age paths $\left\{\widetilde{\mathrm{A}}^{(m,t)}\right\}_{t=1}^{\mathrm{T}}$ in the last round, where each $\widetilde{\mathrm{A}}^{(m,t)} = \left\{\tilde{a}_n^{(m,t)}\right\}_{n=1}^{\mathrm{L}_m}$. Then, each proposal $q_n^{(m)}$ at the current round is designed as follows:

$$q_n^{(m)} = \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} 1_{\left(\tilde{a}_n^{(m,t)}-d,\tilde{a}_n^{(m,t)}+d\right)}$$

, where $d > 0$ is a bandwidth hyperparameter and $\left(\tilde{a}_n^{(m,t)} - d, \tilde{a}_n^{(m,t)} + d\right)$ is an interval. In other words, candidates at step n of the current round are randomly sampled from a randomly chosen interval $\left(\tilde{a}_n^{(m,t)} - d, \tilde{a}_n^{(m,t)} + d\right)$ among $t = 1,2,\cdots,\mathrm{T}$. These reasons prevent us from relying only on the particle smoothing in sampling; we therefore use particles only to initialize the samples.

One advantage of the particle smoothing is that we can quickly sample hidden alignments by the backward algorithm once particles and weights are obtained in the forward algorithm. To guarantee the independence of samples, we first initialize the age paths one-by-one by the particle smoothing and then run the Metropolis-Hastings algorithm on each of them.

The basic framework of the Metropolis-Hastings algorithm starts with computing the following acceptance ratio γ:

$$\gamma = \min\left\{1, \frac{\pi\left(\dot{\mathrm{A}}^{(m)}, \dot{\mathrm{W}}^{(m)} \middle| \mathrm{D}^{(m)}; \Theta\right) p\left(\mathrm{Y}^{(m)} \middle| \dot{\mathrm{A}}^{(m)}; \Theta\right)}{\pi\left(\mathrm{A}^{(m)}, \mathrm{W}^{(m)} \middle| \mathrm{D}^{(m)}; \Theta\right) p\left(\mathrm{Y}^{(m)} \middle| \mathrm{A}^{(m)}; \Theta\right)} \cdot \frac{q\left(\mathrm{A}^{(m)}, \mathrm{W}^{(m)} \middle| \dot{\mathrm{A}}^{(m)}, \dot{\mathrm{W}}^{(m)}\right)}{q\left(\dot{\mathrm{A}}^{(m)}, \dot{\mathrm{W}}^{(m)} \middle| \mathrm{A}^{(m)}, \mathrm{W}^{(m)}\right)}\right\}$$

, where $\left(\mathrm{A}^{(m)}, \mathrm{W}^{(m)}\right)$ is the previously updated age path, $q\left(\cdot \middle| \mathrm{A}^{(m)}, \mathrm{W}^{(m)}\right)$ is the proposal distribution conditioned on $\left(\mathrm{A}^{(m)}, \mathrm{W}^{(m)}\right)$, and $\left(\dot{\mathrm{A}}^{(m)}, \dot{\mathrm{W}}^{(m)}\right)$ is the proposed candidate that is sampled from $q\left(\cdot \middle| \mathrm{A}^{(m)}, \mathrm{W}^{(m)}\right)$. Then, update $\left(\mathrm{A}^{(m)}, \mathrm{W}^{(m)}\right)$ with $\left(\dot{\mathrm{A}}^{(m)}, \dot{\mathrm{W}}^{(m)}\right)$ if γ is bigger than or equal to a uniform random number in $(0,1)$; otherwise, keep $\left(\mathrm{A}^{(m)}, \mathrm{W}^{(m)}\right)$. Once we are in a burn-in phase, stop iteration and return the final $\left(\mathrm{A}^{(m)}, \mathrm{W}^{(m)}\right)$ as the sample.

Note that the Markov structure of the transition model and conditionally independent emission model allow us to efficiently run the algorithm: age samples in a block are simultaneously proposed, evaluated, and potentially updated.

To be more specific, the proposal distribution $q\left(\cdot \middle| \mathrm{A}^{(m)}, \mathrm{W}^{(m)}\right)$ is defined as follows:

$$q\left(\dot{\mathrm{A}}_n^{(m)}, \dot{\mathrm{W}}_n^{(m)} \middle| \mathrm{A}^{(m)}, \mathrm{W}^{(m)}\right) = \begin{cases} \mathcal{N}\left(\dot{\mathrm{A}}_n^{(m)} \middle| \mathrm{A}_n^{(m)}, \frac{1}{8}\left(\mathrm{A}_{n+1}^{(m)} - \mathrm{A}_n^{(m)}\right)\right) \cdot 1_{\left\{\frac{\mathrm{A}_{n+1}^{(m)} - \dot{\mathrm{A}}_n^{(m)}}{r^{(m)} \cdot \left(d_{n+1}^{(m)} - d_n^{(m)}\right)} \in \mathbb{I}_{\dot{\mathrm{W}}_n^{(m)}}\right\}}, & n = 1 \\ \mathcal{N}\left(\dot{\mathrm{A}}_n^{(m)} \middle| \mathrm{A}_n^{(m)}, \frac{1}{8}\left(\mathrm{A}_n^{(m)} - \mathrm{A}_{n-1}^{(m)}\right)\right) \cdot 1_{\left\{\frac{\dot{\mathrm{A}}_n^{(m)} - \mathrm{A}_{n-1}^{(m)}}{r^{(m)} \cdot \left(d_n^{(m)} - d_{n-1}^{(m)}\right)} \in \mathbb{I}_{\dot{\mathrm{W}}_{n-1}^{(m)}}\right\}}, & n = \mathrm{L}_m \\ \mathcal{U}\left(\dot{\mathrm{A}}_n^{(m)} \middle| \mathrm{A}_{n-1}^{(m)}, \mathrm{A}_{n+1}^{(m)}\right) \cdot 1_{\left\{\frac{\mathrm{A}_{n+1}^{(m)} - \dot{\mathrm{A}}_n^{(m)}}{r^{(m)} \cdot \left(d_{n+1}^{(m)} - d_n^{(m)}\right)} \in \mathbb{I}_{\dot{\mathrm{W}}_n^{(m)}}\right\}}, & otherwise \end{cases}$$

, where $\mathcal{U}\left(\dot{\mathrm{A}}_n^{(m)} \middle| \mathrm{A}_{n-1}^{(m)}, \mathrm{A}_{n+1}^{(m)}\right)$ means that $\dot{\mathrm{A}}_n^{(m)}$ follows a uniform distribution on the interval $\left(\mathrm{A}_{n-1}^{(m)}, \mathrm{A}_{n+1}^{(m)}\right)$.

S4.3. Parameter Estimation

To estimate $\Theta = \left\{\Theta^{(m)}\right\}_{m=1}^{\mathrm{M}} = \left\{\phi^{(m)}, r^{(m)}, \sigma^{(m)}, h^{(m)}\right\}_{m=1}^{\mathrm{M}}$, we apply the Baum-Welch EM algorithm by iterating the following steps, with a prior π_0 on $\Theta^{(m)}$, given the sampled age path $\mathrm{A}^{(m)}$, for each m:

- E-step: define the following $\mathcal{Q}\left(\Theta^{(m)}\middle|\Theta^{(m,t)}\right)$:

$$\mathcal{Q}\left(\Theta^{(m)}\middle|\Theta^{(m,t)}\right) = \mathbb{E}_{\mathrm{A}^{(m)},\mathrm{W}^{(m)}|\mathrm{D}^{(m)},\mathrm{Y}^{(m)},\Theta^{(m,t)}}\left[\log\pi\left(\mathrm{A}^{(m)},\mathrm{W}^{(m)}\middle|\mathrm{D}^{(m)};\Theta^{(m)}\right) + \log p\left(\mathrm{Y}^{(m)}\middle|\mathrm{A}^{(m)};\Theta^{(m)}\right) + \log\pi_0\left(\Theta^{(m)}\right)\right]$$

- M-step: find $\Theta^{(m)} = \Theta^{(m,t+1)}$ that maximizes $\mathcal{Q}\left(\Theta^{(m)}\middle|\Theta^{(m,t)}\right)$:

$$\Theta^{(m,t+1)} = \underset{\Theta^{(m)}}{\operatorname{argmax}}\,\mathcal{Q}\left(\Theta^{(m)}\middle|\Theta^{(m,t)}\right)$$

To compute the above $\mathcal{Q}\left(\Theta^{(m)}\middle|\Theta^{(m,t)}\right)$, we approximate it from the samples $\left\{\mathrm{A}^{(m,k,t)},\mathrm{W}^{(m,k,t)}\right\}_{k=1}^{\mathrm{K}}$ drawn from the posterior $p\left(\mathrm{A}^{(m)},\mathrm{W}^{(m)}\middle|\mathrm{D}^{(m)},\mathrm{Y}^{(m)},\Theta^{(m,t)}\right)$ by the hybrid of particle smoothing and MCMC independently, as described in subsection S4.2:

$$\mathcal{Q}\left(\Theta^{(m)}\middle|\Theta^{(m,t)}\right) \approx \frac{1}{\mathrm{K}}\sum_{k=1}^{\mathrm{K}}\left(\log\pi\left(\mathrm{A}^{(m,k,t)},\mathrm{W}^{(m,k,t)}\middle|\mathrm{D}^{(m)};\Theta^{(m)}\right) + \log p\left(\mathrm{Y}^{(m)}\middle|\mathrm{A}^{(m,k,t)};\Theta^{(m)}\right)\right) + \log\pi_0\left(\Theta^{(m)}\right)$$

To optimize $\mathcal{Q}\left(\Theta^{(m)}\middle|\Theta^{(m,t)}\right)$ in the M-step, BIGMACS depends on the gradient ascent algorithm with the above approximation as the objective function to maximize.

Finally, we discuss the depth-scale parameter $r^{(m)}$. Although it may be reasonable to assume $r^{(m)}$is a scalar parameter, BIGMACS considers $r^{(m)}$ to be a continuous function over ages in order to reflect long-term changes of sedimentation rates, by the following definition, based on the Nadaraya-Watson Kernel regression [Langrene and Warin (2019)]:

$$\log r^{(m)}(a) = \frac{1}{\mathrm{K}}\sum_{k=1}^{\mathrm{K}}\left(\sum_{n=1}^{\mathrm{L}_m-1}\mathcal{K}_h\left(a-\tilde{a}_{n+1,k}^{(m)}\right)\log\frac{\tilde{a}_{n+1,k}^{(m)}-\tilde{a}_{n,k}^{(m)}}{d_{n+1}^{(m)}-d_n^{(m)}}\Big/\sum_{n=1}^{\mathrm{L}_m}\mathcal{K}_h\left(a-\tilde{a}_{n+1,k}^{(m)}\right)\right)$$

, where $h > 0$ is a fixed hyperparameter that controls the smoothness of $r^{(m)}$. BIGMACS chose a large $h = 20$ as the default so that the transition model still depends on the transition matrix.

S5. Stack Construction Algorithm

The stack construction algorithm in BIGMACS is designed to construct a set of sample-specific Gaussian process regression models [Rasmussen and Williams (2006)] and average them into a single Gaussian model at each age. However, first we resolve the following three issues: 1) outlier classification from the given δ^{18}O observations, 2) kernel hyperparameter estimation for the Gaussian process regression, and 3) construction of heteroscedastic observational variances of δ^{18}O continuously. In this section, we will discuss these issues together with stack construction. The following definitions are assumed throughout the supplementary materials.

- $\Psi = \{\mathbb{K}, \Lambda\}$: a set of regression hyperparameters.
  - $\mathbb{K}$ is a kernel covariance function controlled by kernel hyperparameters. For example, an Ornstein-Uhlenbeck kernel is defined as follows, for a set of hyperparameters η and η:
$$\mathbb{K}(u,v) = \eta^2 exp(-\xi^2|u-v|)$$
  - Λ is an observational variance function.
- $\vec{\mathrm{Y}}$: a vector that aggregates all δ^{18}O observations of sediment cores, after the standardization.
  - Here, the term 'standardization' means that each observation has been standardized based on the core-specific scale and shift parameters estimated in the alignment part.
- $\widetilde{\mathrm{A}} = \left\{\widetilde{\mathrm{A}}^{(k)}\right\}_{k=1}^{\mathrm{K}}$: a set of vectors that aggregates all sampled age paths of sediment cores.
  - $\widetilde{\mathrm{A}}^{(k)} = \left\{\widetilde{\mathrm{A}}^{(m,k)}\right\}_{m=1}^{\mathrm{M}}$: a vector that concatenates each of the kth path among sediment cores.

- $\underline{\mathrm{A}}$: a set of induced pseudo-inputs. This set is for the variational free energy approximation [Titsias (2009)] and predefined in the same domain of ages.
- $\underline{\mu}$: a constant scalar for the mean value of stack.

The stack construction algorithm first iterates steps in subsections S5.2, S5.3 and S5.4 until convergence and then update the new one by the method in S5.1.

S5.1. Stack Construction

The goal is to construct a generative model of the standardized $\delta^{18}O$ at a query age. In formulation, the stack is in the following form:

$$p(y|a;\mathrm{Y},\mathrm{D},\Theta,\Psi)=\int p\left(y\middle|a;\vec{\mathrm{Y}},\mathrm{A},\Psi\right)p(\mathrm{A}|\mathrm{D},\mathrm{Y};\Theta)\,d\mathrm{A}$$

, where $p(\mathrm{A}|\mathrm{D},\mathrm{Y};\Theta)$ is the posterior distribution of the hidden age paths given depths and proxy observations and $p\left(y\middle|a;\vec{\mathrm{Y}},\mathrm{A},\Psi\right)$ is the regression model given the hidden age paths and a set of regression hyperparameters Ψ. Because it is impossible to represent $p(\mathrm{A}|\mathrm{D},\mathrm{Y};\Theta)$ in a closed distribution, we instead compute the following approximation:

$$p(y|a;\mathrm{Y},\mathrm{D},\Theta,\Psi)=\int p\left(y\middle|a;\vec{\mathrm{Y}},\mathrm{A},\Psi\right)p(\mathrm{A}|\mathrm{D},\mathrm{Y};\Theta)\,d\mathrm{A}\approx\frac{1}{\mathrm{K}}\sum_{k=1}^{\mathrm{K}}p\left(y\middle|a;\vec{\mathrm{Y}},\widetilde{\mathrm{A}}^{(k)},\Psi\right)$$

BIGMACS adopts a Gaussian process regression for modelling each $p\left(y\middle|a;\vec{\mathrm{Y}},\widetilde{\mathrm{A}}^{(k)},\Psi\right)$, after considering outliers and estimating regression parameters Ψ. Suppose that we have already done so, i.e., outliers from $\left(\widetilde{\mathrm{A}}^{(k)},\vec{\mathrm{Y}}\right)$ have been discarded and Ψ is given a priori. Then, we have the following variational free energy approximation [Titsias (2009)] of the Gaussian process regression model for each k:

$$p\left(y\middle|a;\vec{\mathrm{Y}},\widetilde{\mathrm{A}}^{(k)},\Psi\right)=\mathcal{N}\left(y\middle|\overline{\mu}^{(k)}(a),\overline{v}^{(k)}(a)+\Lambda^{(k)}(a)\right)$$

, where:

$$\overline{\mu}^{(k)}(a)=\underline{\mu}+\mathbb{K}_{a\underline{\mathrm{A}}}\left(\mathbb{K}_{\underline{\mathrm{A}}\underline{\mathrm{A}}}+\mathbb{K}_{\underline{\mathrm{A}}\widetilde{\mathrm{A}}^{(k)}}\left(\Lambda^{(k)}_{\widetilde{\mathrm{A}}^{(k)}}\right)^{-1}\mathbb{K}_{\widetilde{\mathrm{A}}^{(k)}\underline{\mathrm{A}}}\right)^{-1}\mathbb{K}_{\underline{\mathrm{A}}\widetilde{\mathrm{A}}^{(k)}}\left(\Lambda^{(k)}_{\widetilde{\mathrm{A}}^{(k)}}\right)^{-1}\left(\vec{\mathrm{Y}}-\underline{\mu}\right)$$

$$\overline{v}^{(k)}(a)=\mathbb{K}_{aa}-\mathbb{K}_{a\underline{\mathrm{A}}}\mathbb{K}^{-1}_{\underline{\mathrm{A}}\underline{\mathrm{A}}}\mathbb{K}_{\underline{\mathrm{A}}a}+\mathbb{K}_{a\underline{\mathrm{A}}}\left(\mathbb{K}_{\underline{\mathrm{A}}\underline{\mathrm{A}}}+\mathbb{K}_{\underline{\mathrm{A}}\widetilde{\mathrm{A}}^{(k)}}\left(\Lambda^{(k)}_{\widetilde{\mathrm{A}}^{(k)}}\right)^{-1}\mathbb{K}_{\widetilde{\mathrm{A}}^{(k)}\underline{\mathrm{A}}}\right)^{-1}\mathbb{K}_{\underline{\mathrm{A}}a}$$

Here, $\mathbb{K}_{\mathrm{AB}}$ is a matrix where each entry is the function value of $\mathbb{K}(a,b)$ for $a\in\mathrm{A}$ and $b\in\mathrm{B}$, and Λ_{A} is a diagonal matrix where each diagonal entry is $\Lambda(a)$ for $a\in\mathrm{A}$, for any sets A and B.

The reason why we consider an approximation instead of the exact Gaussian process regression is to reduce the time complexity stemming from the matrix inversion, especially for the case where the size of $\vec{\mathrm{Y}}$ is large.

To define the stack by a single Gaussian model, BIGMACS again approximates $p(y|a;\mathrm{Y},\mathrm{D},\Theta,\Psi)$ based on the moment-matching [Murphy (2012)], as follows, which results in the stack $\mathcal{N}\left(y\middle|\mu(a),v(a)\right)$:

$$p(y|a;\mathrm{Y},\mathrm{D},\Theta,\Psi)\approx\frac{1}{\mathrm{K}}\sum_{k=1}^{\mathrm{K}}p\left(y\middle|a;\vec{\mathrm{Y}},\widetilde{\mathrm{A}}^{(k)},\Psi\right)\approx\mathcal{N}\left(y\middle|\mu(a),v(a)\right)$$

, where:

$$\mu(a) = \frac{1}{\mathrm{K}}\sum_{k=1}^{\mathrm{K}} \overline{\mu}^{(k)}(a)\,, \qquad \nu(a) = \frac{1}{\mathrm{K}}\sum_{k=1}^{\mathrm{K}} \left(\overline{\nu}^{(k)}(a) + \Lambda^{(k)}(a) + \left(\overline{\mu}^{(k)}(a) - \mu(a) \right)^2 \right)$$

S5.2. Outlier Classification

Because Gaussian process regression is susceptible to outliers, BIGMACS is designed to classify and discard outliers, according to the idea used in [Lee and Lawrence (2019)]. Let $\mathrm{O}^{(k)} = \left\{ \mathrm{O}_n^{(k)} \right\}$ be a set of hidden variables that indicate outliers for $\left(\widetilde{\mathrm{A}}^{(k)}, \vec{\mathrm{Y}} \right)$, where $\mathrm{O}_n^{(k)} = 1$ if the associated $\vec{\mathrm{Y}}_n$ at $\widetilde{\mathrm{A}}_n^{(k)}$ is considered to be an outlier in the stack, 0 otherwise.

We rigorously define outliers as data that do not follow the stack $\mathcal{N}\left(y \middle| \mu(a), \nu(a) + \Lambda(a) \right)$; instead, outliers are assumed to follow an alternative model g. We also assume that outliers are independent from the inputs given core depths D, i.e., we have the following prior and likelihood for a small positive hyperparameter $\delta > 0$:

$$\mathrm{O}_n^{(k)} \sim_{i.i.d.} Bernoulli(\delta)$$

$$p\left(\vec{\mathrm{Y}}_n \middle| \widetilde{\mathrm{A}}_n^{(k)}, \mathrm{O}_n^{(k)} \right) = \begin{cases} \mathcal{N}\left(\vec{\mathrm{Y}}_n \middle| \mu\left(\widetilde{\mathrm{A}}_n^{(k)} \right), \nu\left(\widetilde{\mathrm{A}}_n^{(k)} \right) \right), & \mathrm{O}_n^{(k)} = 0 \\ g\left(\vec{\mathrm{Y}}_n \middle| \widetilde{\mathrm{A}}_n^{(k)} \right), & \mathrm{O}_n^{(k)} = 1 \end{cases}$$

, where g is defined as follows:

$$g(y|a) = \frac{1}{2}\mathcal{N}\left(\vec{\mathrm{Y}}_n \middle| \mu\left(\widetilde{\mathrm{A}}_n^{(k)} \right) + 3\sqrt{\nu\left(\widetilde{\mathrm{A}}_n^{(k)} \right)}, \nu\left(\widetilde{\mathrm{A}}_n^{(k)} \right) \right) + \frac{1}{2}\mathcal{N}\left(\vec{\mathrm{Y}}_n \middle| \mu\left(\widetilde{\mathrm{A}}_n^{(k)} \right) - 3\sqrt{\nu\left(\widetilde{\mathrm{A}}_n^{(k)} \right)}, \nu\left(\widetilde{\mathrm{A}}_n^{(k)} \right) \right)$$

Then, one can easily get the posterior distribution of $\mathrm{O}_n^{(k)}$ as follows:

$$p\left(\mathrm{O}_n^{(k)} = 1 \middle| \widetilde{\mathrm{A}}_n^{(k)}, \vec{\mathrm{Y}} \right) = \frac{\delta \cdot g\left(\vec{\mathrm{Y}}_n \middle| \widetilde{\mathrm{A}}_n^{(k)} \right)}{\delta \cdot g\left(\vec{\mathrm{Y}}_n \middle| \widetilde{\mathrm{A}}_n^{(k)} \right) + (1-\delta)\mathcal{N}\left(\vec{\mathrm{Y}}_n \middle| \mu\left(\widetilde{\mathrm{A}}_n^{(k)} \right), \nu\left(\widetilde{\mathrm{A}}_n^{(k)} \right) \right)}$$

To reflect the ambiguity of outliers, BIGMACS *samples* outlier indicators from the above posterior for each k, instead of classifying them as outliers if $p\left(\mathrm{O}_n^{(k)} = 1 \middle| \widetilde{\mathrm{A}}_n^{(k)}, \vec{\mathrm{Y}} \right) > 0.5$.

S5.3. Kernel Hyperparameter Estimation

From now on, we assume that each $\left(\widetilde{\mathrm{A}}_n^{(k)}, \vec{\mathrm{Y}} \right)$ excludes sampled outliers. To estimate the kernel covariance function $\mathbb{K}$, we first fix the type of function to the OU kernel and just estimate its hyperparameters. The estimated kernel hyperparameters are supposed to be shared throughout the samples and to maximize the following objective function for the variational free energy approximation:

$$\mathcal{L} = \log \mathcal{N}\left(\vec{\mathrm{Y}} \middle| \underline{\mu}, \Lambda_{\mathrm{A}} + \mathbb{K}_{\mathrm{A}\underline{\mathrm{A}}}\mathbb{K}_{\underline{\mathrm{A}}\underline{\mathrm{A}}}^{-1}\mathbb{K}_{\underline{\mathrm{A}}\mathrm{A}} \right) - \frac{1}{2} \cdot trace\left(\Lambda_{\mathrm{A}}^{-1}\left(\mathbb{K}_{\mathrm{AA}} - \mathbb{K}_{\mathrm{A}\underline{\mathrm{A}}}\mathbb{K}_{\underline{\mathrm{A}}\underline{\mathrm{A}}}^{-1}\mathbb{K}_{\underline{\mathrm{A}}\mathrm{A}} \right) \right)$$

Because A is a hidden variable (age paths) and Λ is defined by k, BIGMACS uses a stochastic gradient ascent algorithm that feeds $\widetilde{\mathrm{A}}^{(k)}$ and $\Lambda^{(k)}$ to A and Λ above, respectively, for a randomly chosen k at each iteration. To deal with the matrix inversion of $\Lambda_{\mathrm{A}} + \mathbb{K}_{\mathrm{A}\underline{\mathrm{A}}}\mathbb{K}_{\underline{\mathrm{A}}\underline{\mathrm{A}}}^{-1}\mathbb{K}_{\underline{\mathrm{A}}\mathrm{A}}$, we use the following Woodbury matrix identity [Max (1950)] to convert it more practically:

$$\left(\Lambda_{\mathrm{A}} + \mathbb{K}_{\mathrm{A}\underline{\mathrm{A}}}\mathbb{K}_{\underline{\mathrm{A}}\underline{\mathrm{A}}}^{-1}\mathbb{K}_{\underline{\mathrm{A}}\mathrm{A}} \right)^{-1} = \Lambda_{\mathrm{A}}^{-1} - \Lambda_{\mathrm{A}}^{-1}\mathbb{K}_{\mathrm{A}\underline{\mathrm{A}}}\left(\mathbb{K}_{\underline{\mathrm{A}}\underline{\mathrm{A}}} + \mathbb{K}_{\underline{\mathrm{A}}\mathrm{A}}\Lambda_{\mathrm{A}}^{-1}\mathbb{K}_{\mathrm{A}\underline{\mathrm{A}}} \right)^{-1}\mathbb{K}_{\underline{\mathrm{A}}\mathrm{A}}\Lambda_{\mathrm{A}}^{-1}$$

S5.4. Heteroscedastic Variance Construction

BIGMACS models observational variance as a continuous function over ages (heterscedastic Gaussian process regression). BIGMACS adopts the following close-form update [Lee and Lawrence (2019)]:

$$\Lambda^{(k)}(a) = \sum \left(\left(\vec{\mathrm{Y}}_n - \overline{\mu}^{(k)}\left(\widetilde{\mathrm{A}}_n^{(k)}\right) \right)^2 + \overline{v}^{(k)}\left(\widetilde{\mathrm{A}}_n^{(k)}\right) \right) \mathcal{K}_h\left(a - \widetilde{\mathrm{A}}_n^{(k)}\right) \Big/ \sum \mathcal{K}_h\left(a - \widetilde{\mathrm{A}}_n^{(k)}\right)$$

, where $\mathcal{K}$ and $h > 0$ are a density kernel and a bandwidth hyperparameter that can be tuned as a K-nearest neighborhood bandwidth [Langrene and Warin (2019)], respectively.

6. Time Complexity

All age models and stacks presented here were constructed on a standard desktop machine. However, longer stacks constructed from a large number of high resolution cores may have run times that require a computing cluster. Here we provide time complexity equations to estimate time complexities of future runs.

Age models are constructed in parallel and the time complexity depends on the number of input cores (L) and the number of available CPU processors (C). During age model construction, parameter values are estimated first and then ages are sampled. Parameter estimation requires the particle smoothing algorithm, the Metropolis Hastings algorithm, and the Baum-Welch Expectation Maximization algorithm. Once parameters are estimated, ages are sampled with the particle smoothing algorithm and Metropolis-Hastings algorithms. Particle smoothing requires two steps: a forward step and a backward step. During the forward step the time complexity is quadratic to the number of particles (P, default is 100) and linear to the number of proxy observations (N). The backward step is linear to the number of particles, proxy observations, and age model samples (M_0, default is 100). The Metropolis Hastings algorithm has a time complexity linear to the number of proxy observations, steps until the burn-in phase (B, default is 500) and age model samples (M_0). The total time complexity for a single iteration to learn parameter values is equal to $\mathcal{O}\left(\frac{\mathrm{L}}{\mathrm{C}}(\mathrm{P}^2\mathrm{N} + \mathrm{PNM}_0 + \mathrm{BNM}_0)\right)$. Once the parameters are estimated, ages are sampled. If the number of age model samples is set to M (the default is 1000) and the maximum number of iterations in parameter estimation is equal to R (default is 10), the total time complexity for age model construction is equal to $\mathcal{O}\left(\frac{\mathrm{L}}{\mathrm{C}}R(\mathrm{P}^2\mathrm{N} + \mathrm{PNM}_0 + \mathrm{BNM}_0) + \frac{\mathrm{L}}{\mathrm{C}}(\mathrm{P}^2\mathrm{N} + \mathrm{PNM} + \mathrm{BNM})\right)$. The multiproxy age model for GIK13289-2 (which has 30 δ^{18}O data points and 12 radiocarbon ages) took approximately 86 seconds to run on a standard desktop machine.

Stack construction iterates between an age model construction step and a stack updating step. The latter consists of kernel parameter estimation, δ^{18}O outlier classification, heteroscedastic variance estimation and the Gaussian process regression. Kernel parameter estimation requires a fixed number of iterations (S, default is 3000), with each iteration having a time complexity quadratic to the number of induced pseudo-inputs (fixed to N_0, sampled every 0.5 kyr, see S5 for details) and linear to the number of proxy observations. Time complexity for outlier classification is linear to the number of age model samples and proxy observations. Heteroscedastic variance estimation requires computations proportional to the number of age model samples and quadratic to the number of total proxy observations. Finally, the Gaussian process regression has a time complexity linear to the number of sampled age paths, total proxy observations and the length of the stack (K), and quadratic to the number of induced pseudo-inputs. Therefore, the total time complexity for one stack updating step is $\mathcal{O}\left(\mathrm{T}(\mathrm{SN}_0^2\mathrm{LN} + \mathrm{LNM}_0 + \mathrm{L}^2\mathrm{N}^2\mathrm{M}_0 + \mathrm{N}_0^2\mathrm{LNM}_0\mathrm{K})\right)$.

The stack construction algorithm includes A (default is 5) stack updating steps, and each update includes a new set of age models (i.e., an age model construction step). Thus the total time complexity to construct a stack is equal to $\mathcal{O}\left(\mathrm{A}\left(\frac{\mathrm{L}}{\mathrm{C}}R(\mathrm{P}^2\mathrm{N} + \mathrm{PNM}_0 + \mathrm{BNM}_0) + \mathrm{T}(\mathrm{SN}_0^2\mathrm{LN} + \mathrm{LNM}_0 + \mathrm{L}^2\mathrm{N}^2\mathrm{M}_0 + \mathrm{N}_0^2\mathrm{LNM}_0\mathrm{K})\right)\right)$. The DNEA stack

(which contains 6 cores, 2,112 $\delta^{18}O$ data points, 150 radiocarbon ages, and extends to 150 kyr) has a total run time of 1.8 hours.

References

Christen, J. A.: A NEW ROBUST STATISTICAL MODEL FOR RADIOCARBON DATA, 13, n.d.

Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum Likelihood from Incomplete Data Via the EM Algorithm, Journal of the Royal Statistical Society: Series B (Methodological), 39, 1–22, https://doi.org/10.1111/j.2517-6161.1977.tb01600.x, 1977.

Doucet, A., de Freitas, N., and Gordon, N.: Sequential Monte Carlo Methods in Practice, 12, n.d.

Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G.: Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids, Cambridge University Press, 332 pp., 1998.

Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, 13, n.d.

Heaton, T. J., Köhler, P., Butzin, M., Bard, E., Reimer, R. W., Austin, W. E. N., Bronk Ramsey, C., Grootes, P. M., Hughen, K. A., Kromer, B., Reimer, P. J., Adkins, J., Burke, A., Cook, M. S., Olsen, J., and Skinner, L. C.: Marine20—The Marine Radiocarbon Age Calibration Curve (0–55,000 cal BP), Radiocarbon, 62, 779–820, https://doi.org/10.1017/RDC.2020.68, 2020.

Hogg, A. G., Heaton, T. J., Hua, Q., Palmer, J. G., Turney, C. S., Southon, J., Bayliss, A., Blackwell, P. G., Boswijk, G., Ramsey, C. B., Pearson, C., Petchey, F., Reimer, P., Reimer, R., and Wacker, L.: SHCal20 Southern Hemisphere Calibration, 0–55,000 Years cal BP, Radiocarbon, 62, 759–778, https://doi.org/10.1017/RDC.2020.59, 2020.

Katalin, H., Lakner, R., and Gerzson, M.: Intelligent control systems. An introduction with examples. (Applied optimization 60.), Kluwer, 2001.

Langrené, N. and Warin, X.: Fast and Stable Multivariate Kernel Density Estimation by Fast Sum Updating, Journal of Computational and Graphical Statistics, 28, 596–608, https://doi.org/10.1080/10618600.2018.1549052, 2019.

Lee, T. and Lawrence, C. E.: Heteroscedastic Gaussian Process Regression on the Alkenone over Sea Surface Temperatures, https://doi.org/10.5065/y82j-f154, 2019.

Lin, L., Khider, D., Lisiecki, L. E., and Lawrence, C. E.: Probabilistic sequence alignment of stratigraphic records, Paleoceanography, 29, 976–989, https://doi.org/10.1002/2014PA002713, 2014.

Martino, L., Read, J., and Luengo, D.: Independent Doubly Adaptive Rejection Metropolis Sampling Within Gibbs Sampling, IEEE Transactions on Signal Processing, 63, 3123–3138, https://doi.org/10.1109/TSP.2015.2420537, 2015.

Metropolis, N., Rosenbluth, A. W., and Rosenbluth, M. N.: Equation of State Calculations by Fast Computing Machines, 7, n.d.

Murphy, K. P.: Machine learning: a probabilistic perspective, MIT Press, Cambridge, MA, 1067 pp., 2012.

Reimer, P. J., Austin, W. E. N., Bard, E., Bayliss, A., Blackwell, P. G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P. M., Guilderson, T. P., Hajdas, I., Heaton, T. J., Hogg, A. G., Hughen, K. A., Kromer, B., Manning, S. W., Muscheler, R., Palmer, J. G., Pearson, C., van der Plicht, J., Reimer, R. W., Richards, D. A., Scott, E. M., Southon, J. R., Turney, C. S. M., Wacker, L., Adolphi, F., Büntgen, U., Capano, M., Fahrni, S. M., Fogtmann-Schulz, A., Friedrich, R., Köhler, P., Kudsk, S., Miyake, F., Olsen, J., Reinig, F., Sakamoto, M., Sookdeo, A., and Talamo, S.: The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP), Radiocarbon, 62, 725–757, https://doi.org/10.1017/RDC.2020.41, 2020.

Titsias, M. K.: Variational Learning of Inducing Variables in Sparse Gaussian Processes, 8, n.d.

Williams, C.: Gaussian Processes for Machine Learning, 66, n.d.

Woodbury Max. “Inverting modified matrices”, In *Memorandum Rept. 42, Statistical Research Group*, page 4. Princeton Univ., 1950.