BIGMACS v.1.0 Manual

Taehee Lee[1], Devin Rand[2], Lorraine Lisiecki[2], Geoffrey Gebbie[3], Charles Lawerence[4]

[1]Statistics Department, Harvard University, Cambridge, USA
[2]Department of Earth Science, University of California Santa Barbara, Santa Barbara 93106, USA
[3]Physical Oceanography Department, Woods Hole Oceanographic Institute, Falmouth, 02543, USA
[4]Applied Mathematics, Brown University, Providence, 02906, USA

Table of Contents

1. Introduction

BIGMACS can construct radiocarbon age models, benthic $\delta^{18}O$-aligned age models, multiproxy age models (age models which leverage age information from both dating techniques), and benthic $\delta^{18}O$ stacks. Modeled sedimentation rates are realistically constrained using an empirically derived prior distribution based on the observed sedimentation rates of 37 radiocarbon dated cores (Lin et al., 2014). The stacking capability allows users to construct improved benthic $\delta^{18}O$ alignment targets that share similar water mass histories with their input cores.

Here we provide instructions to download BIGMACS, prepare the required input files, construct age models or a stack, and analyze the results. While we give a brief explanation of each parameter, most settings should not be changed from the default values. It is our hope that BIGMACS provides its users with a less subjective and more realistic method to construct ocean sediment core age models and benthic $\delta^{18}O$ stacks.

Happy chronology building!

2. Installation

The open-sourced software package can be installed from github using the following link: https://github.com/eilion/BIGMACS. BIGMACS was developed and tested on MATLAB R2021b and requires the statistics and machine learning toolbox and the parallel computing toolbox. An example run is included with the installation. To confirm that BIGMACS is running properly, open MATLAB in the BIGMACS directory and enter "BIGMACS('GIK13289_2','age_model_construction','show')" into the command line. This will begin a multiproxy alignment for the example core GIK13289-2 to the DNEA stack (Lee & Rand et al., in review). Also included are the input and output folders for the DNEA stack and the ITWA stack.

3 Preparing input files

To construct age models or a stack, create a new folder in BIGMACS\Inputs with the desired run name. BIGMACS will query all the necessary settings and proxy data from text files saved within this folder and the default folder ('BIGMACS\Default'). All proxy data text files should be tab delimited and include column headers. See one of the example input folders for structure and formatting. Settings that remain unchanged from their default values do not need to be specified in the run folder. See BIGMACS\Defaults for the default settings text files.

3.1 Proxy Data

Create a new folder titled 'Records' in the run folder. Then create a folder for each input core within the records folder (e.g., BIGMACS/Inputs/[*run name*]/Records/[*core name*]. These folders will hold the proxy data for each core and any core specific settings specified in 'setting_core.txt'. Each core folder can store the following text files depending on the available proxy data and desired settings.

1. **C14_data.txt**: stores the radiocarbon and calibration data for each core and has the following six columns: (1) depth m, (2) ^{14}C age kyr, (3) ^{14}C standard deviation kyr, (4) reservoir age kyr, (5) reservoir age standard deviation kyr, and (6) the desired calibration curve ('1' for Intcal20, '2' for Marine20, '3' for SHCal20, and '4' for a custom curve).

2. **δ^{18}O_data.txt**: stores the δ^{18}O data for each core and has two columns: (1) depth m, and (2) δ^{18}O ‰.

3. **additional_ages.txt**: This text file stores any additional age information (e.g., tephra layers, tie points, magnetic reversals, etc.). The user has the option to model each additional age as a Gaussian distribution or a uniform distribution. This text file has 4 columns: (1) depth m, (2) age kyr, (3) error kyr, and (4) mode. To model a given age as a Gaussian distribution, enter a '0' in column four. A '1' will indicate a uniform distribution. The error specified in column 3 will be the standard deviation for the Gaussian distribution and the half-width for the uniform distribution, such that the uniform distribution is centered on the value in column 2 and has the range +/- column4.

4. **setting_core.txt**: This file does not need to be included if all settings are the same as the default settings. For a description of each setting, see the appendix below. For an example of this text file with the default settings see BIGMACS\Defaults\setting_core.txt.

If the user is constructing a stack, all cores must have benthic δ^{18}O data. However, during age model construction, input cores with different combinations of proxy data can be included in the same run. If a core does not have a certain proxy type, that text file does not need to be included in the core folder.

3.2 Run Settings

The following are the text files that set the run-specific settings. For an example of these text files with their default settings, see BIGMACS\Defaults. If you are not constructing a stack, you do not need to include setting_stacking.txt. If you are not changing the settings of a text file from the default values, you do not need to include that text file in the run folder.

1. **setting_alignment.txt**
    - data_type ('both', 'd18O', 'C14'): Indicates the type of proxies that will be included in the run. The additional age data will be used independent of the data type selected. If 'both' is selected, all available proxy data saved in every core folder will be used. If constructing a stack, the user must select either 'both' or 'd18O'.
    - Islearn_transition ('yes or 'no'): Indicates whether the state change probabilities saved in 'transition_parameter.txt' should remain fixed or optimized throughout the run. To maintain the state change probabilities, set to 'no'.
    - Stack_min (number): Defines the minimum age of the alignment target. This setting along with Stack_max truncates the alignment target to reduce computational complexity. Default is –inf.
    - Stack_max (number): Defines the maximum age of the alignment target. Default is inf.

2. **setting_stacking.txt**
    - start_age (number): Defines the start age of the new constructed stack. Default is –inf, meaning the start age of the new stack will be automatically truncated based on the 2.5% age quantiles of the first depths of the input cores.
    - end_age (number): Defines the end age of the new constructed stack. Default is inf, meaning the end age of the new stack will be truncated based on the 97.5% age quantiles of the last depths of the input cores.
    - interval (number): Defines the output interval length between ages of the new constructed stack. Default is 0.1 kyr.

3. **stack.txt**: The initial stack for stack construction or the alignment target for age model construction. Users can enter their own stack using a text file with three columns: (1) age kyr, (2) $\delta^{18}O$ ‰, (3) standard deviation ‰. The custom stack/target file should be named stack.txt and saved in the run folder. To use one of the default stacks provided, which are the regional 150 kyr stacks from Lisiecki & Stern (2016) and the LR04 stack, write the abbreviated name of the stack in 'stack.txt'. For example, to use the deep North Atlantic stack, just write DNA in the stack.txt input file. The abbreviated names for each stack are the names of the text files saved in 'BIGMACS\Defaults\Regional_Stacks. For a description of each stack, see 'BIGMACS-master\Defaults\Regional_Stacks\README.txt'.

4. **calibration_curve_14C.txt**: The custom calibration curve ('4' in 'C14_data.txt') for utilizing the radiocarbon proxy. The format is the same as 'Defaults\Calibration_Curves'.

5. **hyperparameters.txt**: We discourage users from changing the settings saved in this text file. For a description of each setting, see the appendix.

6. **transition_parameter.txt**: We discourage users from changing the settings saved in this text file. For a description of each setting, see the appendix.

4 Running BIGMACS

BIGMACS can be run directly from the command line using the function 'BIGMACS' or by executing the script 'main.m'. To run BIGMACS from the command line, enter type 'BIGMACS(inputfile, inputMode, printfig)' and press enter. The variable 'inputfile' should be set to the run name (e.g., 'GIK13289_2' for the example run). The varible 'inputMode' should be set to 'stack_construction' if the user wants to construct a stack or 'age_model_construction' if the user wants to only create age models. The optional variable 'printfig' indicates whether the output figures are created and saved. To save and show the output figures set 'printfig' to 'show', to save but close the output figures after storing them set 'printfig' to 'hide'. To skip figure generation, just enter two inputs into the function (i.e., 'BIGMACS(inputfile, inputMode)'). Before executing a run, make sure MATLAB's

directory is set to the BIGMACS-master folder. The script 'main.m' defines each of these variables and executes the function BIGMACS(). After setting each variable name, press the run button on the editor ribbon or type 'main' in the command line.

5 Output Files

All output files will be saved in a folder titled 'BIGMACS/Outputs/[*run name_proxy type*]. The output folder will include the following text files.

1. Age models for each core will be saved in the folder BIGMACS/Outputs/[*run name_proxy type*]/ages. Each text file will contain depth, 95% and 68.2% confidence bands, the mean and the median.
2. If radiocarbon ages were used, the calibrated radiocarbon ages will be saved in BIGMACS/Outputs/[*run name_proxy type*]/C14_ages.
3. The transition parameters (whether optimized or held constant) will be saved in a text file titled 'transition_parameter.txt'. Rows and columns correspond to {'Initial','Contraction','Steady','Expansion'} and {'Contraction','Steady','Expansion'}, respectively.
4. All figures will be saved in BIGMAS/Outputs/[*run name_proxy type*]/figures. There will be one summary figure for each core and a stack summary figure if a stack was constructed.

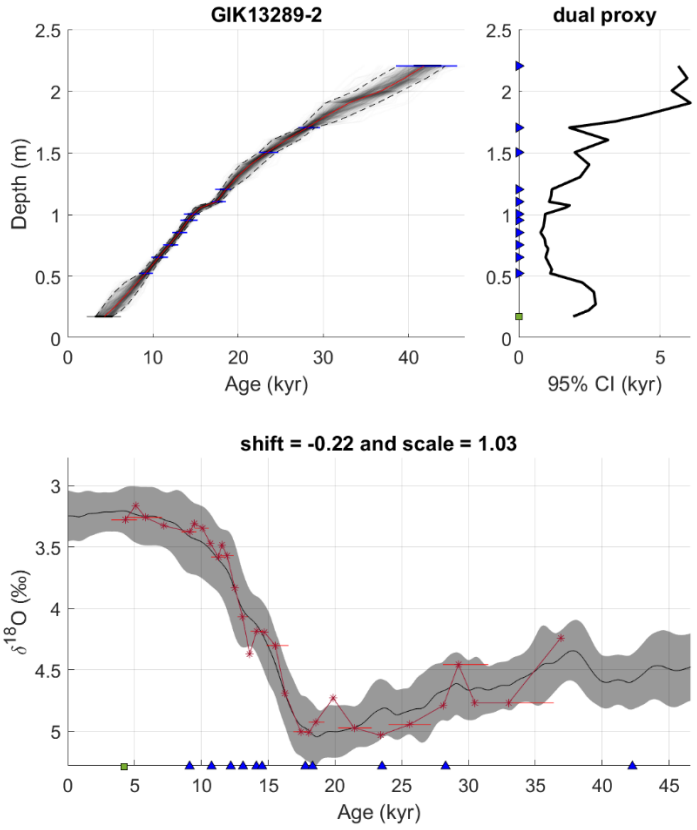

Figure 1: A core summary figure for GIK13289-2 aligned to the DNEA stack using both $\delta^{18}O$, radiocarbon, and an additional age (Lee & Rand, in Review). (A) Age vs. depth plot shaded

according to sample density. The 95% confidence interval is indicated with dotted black lines and the median age model is displayed as a solid red line. Calibrated radiocarbon and additional ages are shown as blue and green horizontal lines respectively. (B) The 95% confidence interval width vs. depth with radiocarbon and additional ages displayed as blue triangles and green squares respectively. (C) Shifted and scaled benthic $\delta^{18}O$ (red *) with 95% error bars (red lines). The shift and scale parameters ('summary.d18O_shift' and 'summary.d18O_scale', respectively in 'results.mat') are displayed as the figure title. The shaded region displays the upper and lower 2 sigma of the alignment target with the solid black line indicating the median. Median radiocarbon and additional ages are displayed as blue triangles and green squares respectively.

Figure 2: The stack summary figure for the DNEA stack (Lee & Rand et al., in review). The shifted and scaled $\delta^{18}O$ data plotted for each core with the radiocarbon data (triangles) and additional age data (squares). The shaded region indicates the upper and lower 2-sigma for the final stack and the solid black line displays the median. The dotted black lines show the 2-sigma for the initial alignment target (in this case the deep North Atlantic stack from Lisiecki & Stern, 2016).

5. If a stack was constructed, the final stack will be saved in a text file titled 'stack.txt'. In addition, two types of stack samples will be saved in the text files stack_samples_mean.txt and stack_samples_noisy.txt The file stack_samples_mean.txt includes averages of the Gaussian process regressions while stack_samples_noisy.txt saves sample stack paths that include the appropriate variance.
6. All input settings, proxy data, and finalized parameters are saved in 'results.mat'. This file includes seven structures.
    a. CI_C14: contains the calibrated radiocarbon ages, including the upper and lower 1 and 2 sigma, the median and mean, and the calendar age samples.
    b. Summary: includes all input proxy data, the $\delta^{18}O$ shift and scale parameters, the final MCMC age model samples, an outlier flag for each $\delta^{18}O$ data point on each age model sample (with a '1' indicating an outlier), the age model 95% confidence bands, 68% confidence bands, the median, the mean, and the average learned sedimentation rate.
    c. target: contains the final stack if constructed ('stack'), the calibration curve ('cal_curve'), the initial target stack ('init_stack'), and both types of stack samples ('stack_samples_mean' and 'stack_samples_noisy'). Each cell of 'target.cal_curve' has

three columns: CAL_BP, 14C_age and Sigma in the calibration curves, respectively. We depend on the linear extrapolation above the range of the given calibration curves.

d. 'hyperparameter' contains all settings in the 'hyperparameter.txt' file.
e. 'setting_alignment' contains all settings in the 'setting_alignment.txt' file.
f. 'setting_core' contains all settings in the 'setting_core.txt' file
g. 'setting_stacking' will only be saved if a stack was constructed. This structure saves all settings saved in the 'setting_stacking.txt' file.

Appendix

1. setting_core.txt

This text file stores core specific settings and should be saved in the core folders if default settings are not used.

- start_depth (number, m): Sets the first depth of the age model. The default is NaN and the first depth is determined by the proxy data.
- end_depth (number, m): Indicate the last depth to infer ages. The default is NaN and the last depth is determined by the proxy data.
- initial_shift (number, ‰): Specify the initial shift for $\delta^{18}O$ alignment. The default is NaN.
- initial_scale (number): Specify the initial scale for $\delta^{18}O$ alignment. The default is 1.
- initial_average_sed_rate: Suggest an initial average sedimentation rate. The default is NaN.
- islearn_shift ('yes' or 'no'): Specify whether the core specific shift and scale parameters are learned during alignment using the Baum-Welch Expectation Maximization algorithm. The default is 'yes'.
- islearn_scale ('yes' or 'no'): Specify whether the core specific shift and scale parameters are learned during alignment using the Baum-Welch Expectation Maximization algorithm. The default is 'yes'.
- islearn_average_sed_rate ('adaptive', 'no', 'constant'): Indicates how the average sedimentation rate will be learned. If 'adaptive', BIGMACS will learn a smoothed depth-dependent average sedimentation rate, 'constant' will learn the average sedimentation rate as a time-independent single scalar parameter, 'no' will keep the average sedimentation rate set to the initial_average_sed_rate.
- smoothness_bandwith (number, kyr): the smoothness of the regression of the core's average sedimentation rate. The default value is 20 kyr. Lower values will cause less smoothing of the core's average sedimentation rate (i.e., allowing more absolute variability in sedimentation rate).
- lower_bound (number, kyr): The lower limit of age samples. Default is NaN.
- upper_boud (number, kyr): The upper limit of age samples. Default is NaN.
- min_resolution (number): Sets the maximum length between depths for which a sedimentation rate is calculated. This setting creates additional depths in the core's age model if the $\delta^{18}O$ and/or radiocarbon data have gaps larger than min_resolution. The default is 0.145679 m, which was derived from the radiocarbon compilation from Lin et al., (2014).
- lower_sed_rate (number): The lowest allowable normalized sedimentation rate. Default is 1/8.
- upper_sed_rate (number): The highest allowable normalized sedimentation rate. Default is 8.
- particle_bandwidth (number, kyr): the bandwidth of the proposal distribution for the particle smoothing algorithm. The default value is 7 kyr. We recommend users to increase this value if the age models look to be too narrow and not to count all possible paths.

2. hyperparameters.txt

We discourage users from changing these settings unless they have good reason to do so.

- q: the prior outlier probability for $\delta^{18}O$. The default value is 0.05, or 5%.
- d: Number of standard deviations that $\delta^{18}O$ outliers are deviated from the alignment target. The default value is 3.
- nParticles: the number of particles in the initialization step of particle smoothing. Increasing this number might be useful for alignments or stacks longer than ~500 kyr but will increase runtime. Default is 500.

- v: The efficiency of the sampling in particle smoothing. Default value is 7
- max_iters: the number of iterations in the Metropolis Hastings algorithm for sampling ages. The default is 500.
- a_d18O: Parameter that defines the generalized student's t-distribution for the $\delta^{18}O$ emission model. The default value is 3.
- b_d18O: Parameter that defines the generalized student's t-distribution for the $\delta^{18}O$ emission model. The default value is 4.
- a_C14: Parameter that defines the generalized student's t-distribution for the C14 emission model. The default value is 10.
- b_C14: Parameter that defines the generalized student's t-distribution for the C14 emission model. The default value is 11.
- nSamples_learning (number): Indicates the number of sample age models that are used to learn parameters. Default value is 100.
- nSamples_drawing (number): Defines the number of sampled age paths used to draw the final age model. The default is 1000.

6. Defaults/transition_parameter.txt

- The default state change probabilities during age model construction. Rows and columns correspond to {'Initial','Contraction','Steady','Expansion'} and {'Contraction','Steady','Expansion'}, respectively.