TurbIFAv3.1 User Guide

1. What is TurbIFA?

Tracking uncertainty of reworking & bioturbation in IFA (TurbIFA) is a software designed for paleoceanographers, who use the individual foraminiferal analyses (IFA) as a tool to characterize ocean and climate variability in the past. TurbIFA consists of a suite of MATLAB routines that explore and quantify the effects of postdepositional sediment particle mixing on IFA data and assesses statistical significance of the attendant ocean and climate variability reconstruction. TurbIFA combines existing algorithms, such as the SEAMUS bioturbation routine (SEdiment AccuMUlation Simulator; Lougheed, 2020) with two IFA uncertainty routines, namely QUANTIFA (Quantile Analysis of Temperature using Individual Foraminiferal Analyses; Glaubke et al., 2021) and INFAUNAL (Individual Foraminiferal Approach Uncertainty Analysis, Thirumalai et al., 2013). By doing so, TurbIFA quantifies the uncertainties arising from methodological protocols (sample size, that is, number of individuals analyzes per sample, and instrumental precision) in IFA-$\delta^{18}O$ and Mg/Ca-based sea surface temperature (SST) datasets collected from a bioturbated sediment archive.

TurbIFA is aimed at IFA users that want interpret their IFA datasets in a quantitative manner, by determining the impacts of bioturbation along with other sources of uncertainty to fully exploit the paleoclimate value of their datasets. Accordingly, TurbIFA evaluates:

i) The probability that the collected IFA dataset contains individuals that grew during a different (colder, warmer) mean climate state;

ii) To what extent bioturbation distorts the paleoclimate interpretation of downcore changes in IFA distributions;

iii) Whether a downcore change on the IFA distribution is statistically once the impact s of bioturbation and other IFA uncertainties are taken into account.

To provide an overview to users of the of the TurbIFA's utilities, two example scripts scripts *example_TurbIFA_Tv3_1.m* and *example_TurbIFA_d18Ov3_1.m* are written and made available, so that they can be run by clicking ➡. This allows exploring TurbIFA's outputs without the need of input any data.

➔ Note that the main point of the example scripts (*i.e., example_TurbIFA_Tv3_1.m, example_TurbIFA_d18Ov3_1.m)* is that the user just has to open and run them. This means that we have removed the lines that calls to the opening user-input data windows. In both examples, the sedimentation rate, bioturbation depth and number of IFA individuals remains constant through time. As such, any change in the code (e.g., switch to a scenario with dynamic variables) may result in an error.

2. TurbIFA routines

2.1 *TurbIFA_Tv3_1.m* & *TurbIFA_d18Ov3_1.m*

TurbIFA is branched into two main models: *TurbIFA_Tv3_1.m* and *TurbIFA_d18Ov3_1.m*. *TurbIFA_Tv3_1.m* is written for IFA datasets made of Mg/Ca measurements, while *TurbIFA_d18Ov3_1.m* is written for IFA datasets made of $\delta^{18}O$ IFA measurements.

By default, input data for both *TurbIFA_Tv3_1.m* and *TurbIFA_d18Ov3_1.m* are the monthly resolved air temperature output and the decadally resolved ocean surface salinity output of the Simulation of the Transient Climate of the Last 21,000 Years (TraCE-21ka; He, 2011). Both datasets are available in this repository (*Trace_SAT.mat*, *SSS_Trace.mat*). The TurbIFA user has also the option of using as an input other long-term monthly resolved temperature or predicted-$\delta^{18}O$ time series. Since TurbIFA is based on the comparison of "modern" (core-top) IFA distributions *versus* the targeted time slice in the past (downcore), the time series is expected to reach modern ages e.g., 1990 CE as TraCE-21ka.

➔ Note that before running the *TurbIFA_T.m* or *TurbIFA_d18O.m* scripts input data should be defined: *i.e.,* coretop ("X") and downcore IFA datasets ("Y"); climate time series to be bioturbated ("Trace=1" by default, the **program** loads T° or predicted-$\delta^{18}O$ time series from TraCE-21ka; "Trace=0", the **user** loads other monthly resolved T° or predicted-$\delta^{18}O$ time series "clim_full"), start year of the clime time series to be bioturbated ("clim start", 1990 CE by default), IFA specie abundance ("abu", constant by default), number of bioturbated synthetic cores ("run", 50 by default) . After clicking "run", windows open for the user to add the input data to the software.

The bioturbation simulation is based on a continuous process of homogeneous and random mixing of the "foraminiferal signal carriers" within the sediment mixing layer. The bioturbation model implemented is based on Berger and Heath (1968). To calculate the uncertainty, individual foraminiferal carrier signals are randomly picked 1,000 times in the discrete depth with the closest mean age to the IFA discrete sample depth.

The sedimentation rate, bioturbation depth, and the abundance of IFA individuals within the sample or time slice can be simulated as constant or dynamical. For dynamical abundances, data can be derived from foraminiferal census counts ("abu=1", downcore changes in IFA abundance) or e.g., temperature-transfer functions ("abu=2", temporal changes in IFA abundance). Two different routines are used to avoid biases at the time of quantifying the heterogeneity in ages due to the finite length of the input climate time series:

- <u>Constant parameters:</u> Foraminifera are "picked" from an exponential age distribution calculated using SMLD/SAR as a scaling parameter.

- <u>Dynamic parameter(s):</u> A step-by-step random bioturbation simulation is performed where the SAR, SMLD, and/or abundance of IFA individuals may vary downcore/over time. The age from each individual foraminifera is "picked" at the discrete depth of sediment where after bioturbation, their age is closer to the age of the input IFA discrete depth.

➔ Note that in dynamical scenarios the user should define the length of the age time series to be bioturbated ("le"). The age time series is then bioturbated to obtain the probability of pick 1 to 20 individuals older than 1 to 10 kyr than the mean age of the IFA discrete depth (TurbIFA graphical output nº1, see section 4.2). The value assigned to "le" is especially relevant for situations where the ratio between the BD and SAR is high, *i.e.,* the length of the TraCE-21ka time series is no longer enough to, after bioturbation, sample it and obtain the expected long tail age distribution (age heterogeneity subestimation) of the individuals archived in a given discrete depth sample.

2.2 *bioprobv3_1.m*

Bioprobv3_1.m is a MATLAB script that calculates the probability of picking 1 to 20 individuals that precipitated their shells during time slices 1-to-10 kyr older than the targeted IFA time slice for a given discrete sediment sample depth. *Bioprobv3_1.m* is designed for IFA time slices that are currently outside the range of transient simulations (e.g., TraCE-21ka; He, 2011), so that uncertainties cannot be assessed using these input data.

As in the *TurbIFA_Tv3_1.m* and *TurbIFA_d18Ov3_1.m* models, sedimentation rate, bioturbation depth, and abundance of individuals selected for IFA can be constant or dynamic.

3. TurbIFA input data

3.1 *TurbIFA_Tv3_1.m* and *TurbIFA_d18Ov3_1.m* inputs

Variable	code name	Description	Model
Latitude	lat	In decimal degrees. A value from –90 to 90ºN (e.g., 2.13).	Both
Longitude	lon	In decimal degrees. A value from –180 to 180ºE or 0–360ºE (e.g., 90.36).	Both
Trace	Trace	If Trace = 1, the program automatically loads TraCE-21ka model data By setting Trace = 0, the user will load climate time series ("clim_full") other than TraCE-21ka.	Both
Start year simulation	clim_start	Youngest year of the input climate time series	Both
run	run	Number of times the climate time series is bioturbated. That is, number of bioturbated synthetic sediment archives. By default, run = 50.	Both
Analytical (1σ)	anerr	Analytical protocol, i.e., measurement (instrumental) uncertainty.	Both
Extra gaussian noise (1σ)	exerr	Gaussian noise related to e.g., Mg-T calibration or $\delta^{18}O_{sw}$–S relationship.	
Mean age IFA core top	IFAtop_mage	<u>NOTE</u>: If more than 1 discrete depth sample is picked, the mean age of each (e.g., 1,000, 1,300 a BP) should be used as an input.	Both
Start year IFA core top	ystart_top	Oldest age within core-top IFA discrete depth sample (e.g., 2,000 a BP)	Both

End year IFA core top	`yend_top`	Youngest age within core-top IFA discrete depth sample (e.g., 100 a BP)	Both
Mean age downcore-IFA depth	`IFApast_mage`	<u>NOTE:</u> If more than 1 discrete depth has been picked, the mean age of each (e.g., 5,000, 5,300 a BP) should be used as an input.	Both
Start year downcore-IFA	`ystart_past`	Oldest age within paleo-IFA discrete depth sample (e.g., 14,000 a BP)	Both
End year downcore-IFA	`yend_past`	Youngest age within paleo-IFA discrete depth sample (e.g., 13,000 a BP)	Both
Number quantiles	`num_q`	Number of quantiles for quantile-quantile (q-q) plot (e.g., 50)	Both
Sedimentation rate (cm kyr^{-1})	`sed_r`	Can be set as constant or dynamical: ➔ <u>If constant</u>, the mean sedimentation rate during ~ last 22,000 years should be used as an input; ➔ <u>If dynamical</u>, age and depths values should be used as inputs. Data can be spaced e.g., depth = [0, 15, 33, 55] age = [1, 5, 17, 20] or continuous, e.g., depth = [0:1:77]; age = [1:×:20]; ➔ If SAR is constant but the SMLD and or planktic foraminiferal abundance are dynamical, depth and age values according to a constant SAR should be used as inputs.	Both
Bioturbation depth	`BD`	Can be constant or dynamical: ➔ <u>If constant</u>, the mean BD during ~ last 22,000 years should be used as an input; ➔ <u>If dynamical</u>, age, depth, and SMLD values should be used as inputs. In both dynamical cases the depths where the age and SMLD are known should be inserted as integers (i.e., 0 3 5 15). ➔ If BD is constant but SAR and/or the planktic foraminiferal abundance are dynamical, just insert the same BD value as data for the other variables are inserted;	Both
Abundance of IFA species	`abu`	➔ = 0, constant abundance of the specie selected for IFA. No abundance input is needed. ➔ = 1, IFA data is interpolated downcore. Input can be obtained from foraminiferal census counts, *i.e.,* number of individuals of the selected species in a 1 cm-thick sample. ➔ = 2, IFA abundance is interpolated along time: temporal changes in IFA abundance e.g., abundance data from a temperature-transfer function. <u>NOTE:</u> If planktic foraminiferal abundance is constant but SAR and/or BD are dynamical,	Both

Variable	code name	Description	
		just insert an arbitrary value as times as needed to match the length of the SAR and BD data.	
time series length	`le`	The oldest age in the T (ºC) or δ^{18}O input time series determines the oldest age of the foraminifera that can be picked in the TurbIFA routines. For dynamic scenarios, the user should indicate the length of the time series to be bioturbated. This input parameter will be used to calculate graphical output nº1 (see section 4.2).	Both
δ^{18}Osw-S relationship	-	Select local δ^{18}O$_{sw}$–S relationship from LeGrande & Schmidt (2006) or indicate slope and intercept from new δ^{18}O$_{sw}$–S datasets.	δ^{18}O
scale	`scale`	The dynamical abundance values are normalized according to the scale parameter, e.g., 100 prior to the simulation of bioturbation. The higher the value of the scale parameter, the smaller the abundance changes that may be simulated.	both

3.2 *bioprobv3_1.m* inputs

Variable	code name	Description
Start year simulation	`sstart_y`	Oldest year of the age time series to be bioturbated.
End year simulation	`send_y`	Youngest year of the age time series to be bioturbated.
run	`run`	Same as in TurbIFA_T and TurbIFA_d18O.
Mean age IFA discrete depth(s)	`IFApast_mage`	Mean age of the IFA discrete sample depth.
Sedimentation rate (cm kyr^{-1})	`sed_r`	Same as in TurbIFA_T and TurbIFA_d18O.
Bioturbation depth	`BD`	Same as in TurbIFA_T and TurbIFA_d18O.
Abundance of IFA species	`abu`	Same as in TurbIFA_T and TurbIFA_d18O.
Number of IFA individuals	`num_p`	Size of the IFA dataset.
scale	`scale`	Same as in TurbIFA_T and TurbIFA_d18O.

➔ Note that for constant scenarios, the user has to input the SAR, BD and the size (*i.e.,* numbers of individual analyses) of the IFA dataset.

➔ Note that the point of define sstart_y and send_y is to avoid biases in the "picked" IFA age distribution related to finite length of the age time series to be bioturbated. Also, the idea is that IFApast_mage is ~close to send_y

4. TurbIFA outputs

4.1 *TurbIFA_Tv3_1.m* and *TurbIFA_d18Ov3_1.m* outputs (vars)

Variable (code name)	Short description	Dimensions
qx	Coretop IFA q-q space (mean removed)	$1 \times$ n° quantiles
qy	Downcore IFA q-q space (mean removed)	$1 \times$ n° quantiles
errTnbio	Coretop IFA q-q uncertainty (P2.5 & P97.5) before bioturbation	$2 \times$ n° quantiles
errPnbio	Downcore IFA q-q uncertainty (P2.5 & P97.5) before bioturbation	$2 \times$ n° quantiles
errTbio	Coretop IFA q-q uncertainty (P2.5 & P97.5) after bioturbation	$2 \times$ n° quantiles x n° runs
errPbio	Downcore IFA q-q uncertainty (P2.5 & P97.5) after bioturbation	$2 \times$ n° quantiles x n° runs
errsigmaX	Uncertainty (P2.5 & P97.5) in coretop IFA distribution 1σ value (std) after bioturbation	$2 \times$ n° runs
errsigmaY	Uncertainty (P2.5 & P97.5) in downcore IFA distribution 1σ value (std) after bioturbation	$2 \times$ n° runs
std_i	Probability of a significant downcore increase in the IFA std	n° runs
std_d	Probability of a significant downcore decrease in the IFA std	n° runs
qTnbio	"picked" pseudo-IFA coretop datasets before bioturbation (qq space)	n° quantiles $\times$ mc*
qPnbio	"picked" pseudo-IFA downcore datasets before bioturbation (qq space)	n° quantiles $\times$ mc
qTbio	"picked" pseudo-IFA coretop datasets after bioturbation (qq space)	n° quantiles $\times$ mc x n° runs

qPbio	"picked" pseudo-IFA downcore datasets after bioturbation (qq space)	nº quantiles × mc × nº runs
qqsignbio_in	Probability of a significant downcore increase in the shape of IFA distribution before bioturbation**	nº quantiles × 1
qqsignbio_de	Probability of a significant downcore decrease in the shape of IFA distribution before bioturbation**	nº quantiles × 1
qqsigbio_in	Probability of a significant downcore increase in the shape of IFA distribution after bioturbation**	nº quantiles × nº runs
qqsigbio_de	Probability of a significant downcore decrease in the shape of IFA distribution after bioturbation**	nº quantiles × nº runs
pbt	Probability of picking 1 to 20 individuals that precipitated their shells during time slices 1-10 kyr older than the mean age of IFA time slices (if sed_r, BD and abu are set to be constant and the size of X and Y IFA datasets is the same)	age offset × nº individuals
tpbt	probability of picking 1 to 20 individuals that precipitated their shells during time slices 1-10 kyr older than the mean age of coretop IFA time slice	age offset × nº individuals
ppbt	probability of picking 1 to 20 individuals that precipitated their shells during time slices 1-10 kyr older than the mean age of Past IFA time slice	age offset × nº individuals

*mc=number of "picked" pseudo-IFA datasets; ** compares "picked" pseudo-IFA data on the input climate time series e.g., trace data. Hence, it depends on the amplitude of climate variability simulated by e.g., TraCE-21ka along the IFA time slices (before bio) & at the IFA discrete depth (after bio).

4.2 *TurbIFA_Tv3_1.m* and *TurbIFA_d18Ov3_1.m* graphical outputs

1. A heatmap chart, which displays the probability of picking in a given discrete depth sample 1 to 20 individuals, which precipitated their shells during periods 1-10 kyr older than the targeted IFA time slice;

➔ Note that if SAR, BD, and the number of IFA individuals stays constant, the shape of the long tail age distribution does not change along IFA discrete depths after bioturbation. Hence, the probability heatmap chart is the same for both core-top and downcore IFA discrete depth(s).

2. Comparisons of core-top versus downcore IFA distributions in the quantile-quantile space with 95% confidence intervals (based on QUANTIFA software MATLAB code for q-q graphical output);

3. Summary table of the uncertainties of the IFA Mg/Ca distribution, and of the probability that the change in the standard deviation (IFA 1σ) of IFA distribution is significant;

4. Probability significant increase/decrease variability for pseudo-IFA datasets resampled in a non-bioturbated vs bioturbated sediment archive (qq space).

4.2 *bioprobv3_1.m* outputs

A probability matrix (pp), that stores the probability of picking in the IFA discrete depth(s) sample 1 to 20 individuals that precipitated their shells during time slices 1-10 kyr older than the IFA sample time slice.

The graphical output is the same heatmap chart concept that in *TurbIFA_T.m* and *TurbIFA_d18O.m.*

5. Step by Step *TurbIFA_Tv3_1.m*

Let's say that you have core-top and downcore Mg/Ca IFA datasets and so you have to run *TurbIFA_Tv3_1.m*

Before run the script, you need to input some basic information. Let's see how it looks:

```
138    %% 1. User Input data
139
140    %% 1.1 Before run TurbIFA model !
141
142    Trace= 1;          % bioturbation simulation using SAT Trace data.
143                      % Set Trace= 0 to load other temperature time series
144                      % "clim_full"
145
146    % clim_full= ;     % should be monthly resolved and ordener
147                      % from older (January) to younger ages (December)
148
149    clim_start = -40;  % 1990 CE for Trace, edit as needed
150    abu= 0;            % constant foraminiferal abundance
151    run=50;            % number of bioturbated synthethic cores (recomendable)
152
153    % X=               % CoreTop  IFA dataset (ºC)
154    % Y=               % Downcore IFA dataset (ºC)
155
156    % if SAR,BD or abu varies over time:
157    le = 40;           % only for dynamical inputs: length of age time series (kyr) to be bioturbated
158                      % for probability heatmap (Graphical output 1)
159                      % should be longer enought to avoid a substimation of
160                      % the long tail age distribution in cases of
161                      % hight bioturbation depth respect sed. rate
162    % -------------------------------------------------------------------
```
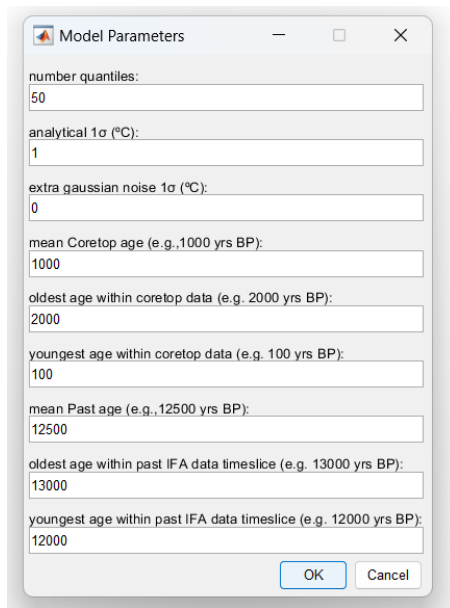
➜   Note that Mg/Ca IFA distributions should be input in ºC.

Now you are ready to just **click ➡ and run** the script. Then, a series of opening windows will require from your input. Let's see them one by one:

1. **Core location:** You need to input the latitude and longitude of your core site. Then click ok.
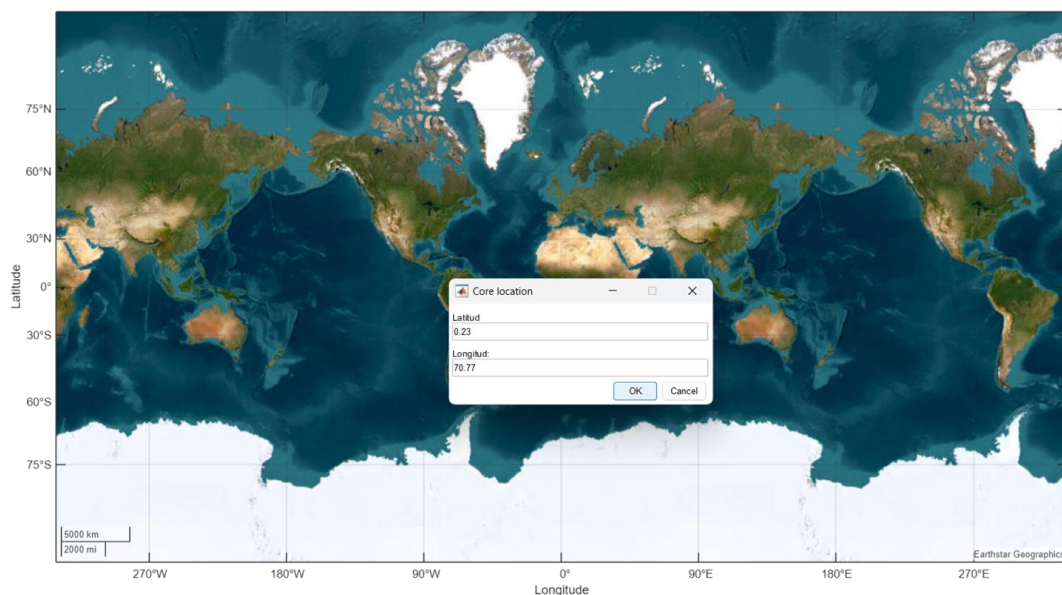
2. **General model inputs:** Now, you have to define some general inputs. After that, click ok.
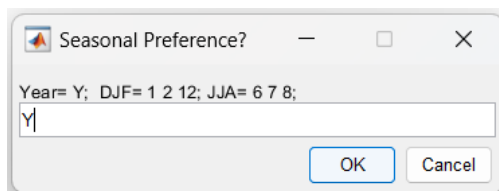
3. **Seasonal preference?** Here you have to indicate if the IFA species grew during the year or instead its seasonal locked.
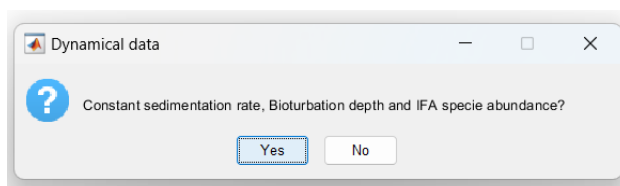

For a yearly species, just write Y

For seasonal locked species, write the months from 1 to 12 where the species grew *e.g.,* for DJF write 1 2 12

Then, click ok.

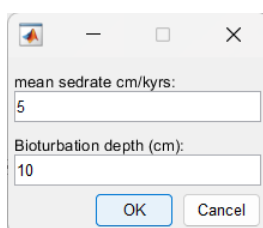4. **Do you want to simulated a constant SAR, BD, and IFA abundance?**

➔ Note that if abu is set to be 1 or 2, this window will not appear.

If **Yes**, you have to input the mean SAR and BD. Then, click OK.
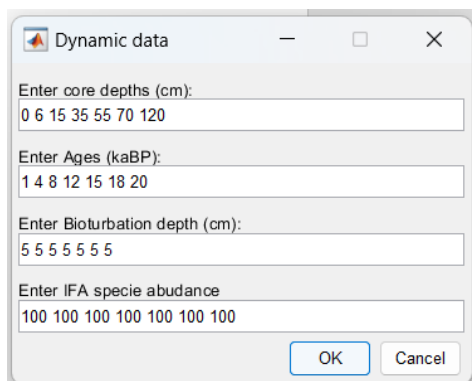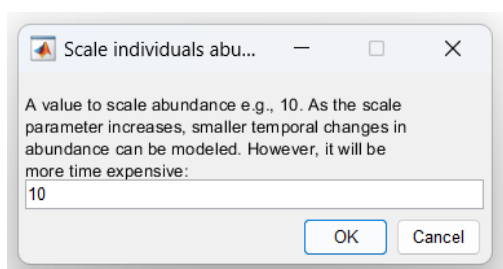
If **No**, you have to input the core depths (in cm) and its mean age (kaBP), the bioturbation depth for each depth and the abundance of IFA individuals.

Some notes:
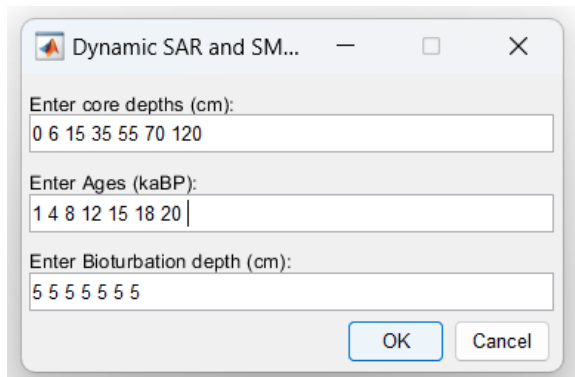
➔ Let's say that you're just interested in simulated a dynamical SAR. Then you just have to repeat the same BD and abundance value in its corresponding window row. You should input the same number of data points for each variable, if not, and error will occur.

➔ If abu is set to be 0, that is, you want to simulated constant foraminiferal abundance, then the input values for IFA specie abundance will not be used. You can basically insert random data. Just be sure that number of values inserted do not change between the rows of the window.

➔ If abu is set to be 1, that is, you want to simulated dynamical foraminiferal abundance and the data is derived from foraminiferal census counts, you have to input abundance counts in the same window that the showed above. After that, you will have to indicate and scale parameter in another opening window:
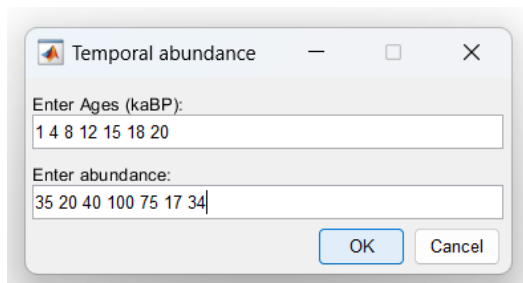

The abundance data will be re-scaled to this parameter. This scale parameter will be used as the maximum number of times that a monthly T° or $\delta^{18}O$ value is repeated before bioturbated the time series. Note that it's important to remove outliers (if exist) before input the abundance data. Then, click OK.

➔ If abu is set to be 2, that is, you have abundance data for a given time slices, three windows will appear:

First, you have to input the depths of the core where its mean age its known i.e., to simulate changes/no changes in the sedimentation rate. You have also to indicate the bioturbation depth at each depth/time slice. If the bioturbation depth don't change, just write the same value. The number of values for each of the three variables should be the same.
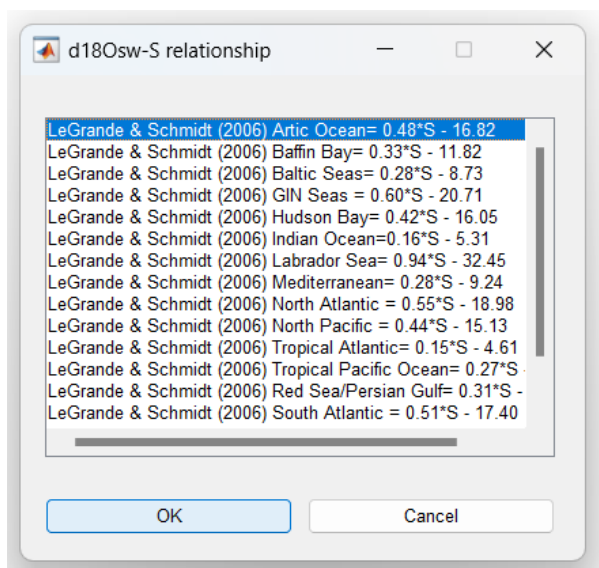

Second, you have to input the number of IFA individuals per time slice. Then, click OK.

Third, as in the case of abu=1, you have to input a value for the scale parameter.

6. Step by Step *TurbIFA_d18O.m*

In essence and to not be repetitive, the main difference between *TurbIFA_d18Ov3_1.m and TurbIFA_Tv3_1.m* is that in the first you have also to indicate a $\delta^{18}O_{sw}$–S relationship:

TurbIFA offers by default the local relationships from Legrande & Schmidt (2006). Then, click OK.

If you don't want to use the relationships by Legrande & Schmidt (2006), you have another option. If you scroll down the bar, you'll see an option called "other". If you click on it, you will be able to define the slope and intercept of the $\delta^{18}O_{sw}$-S relationship.
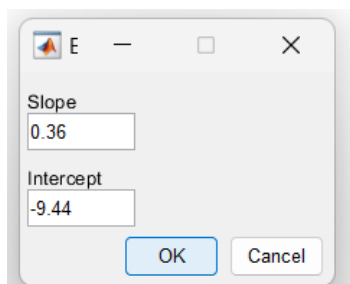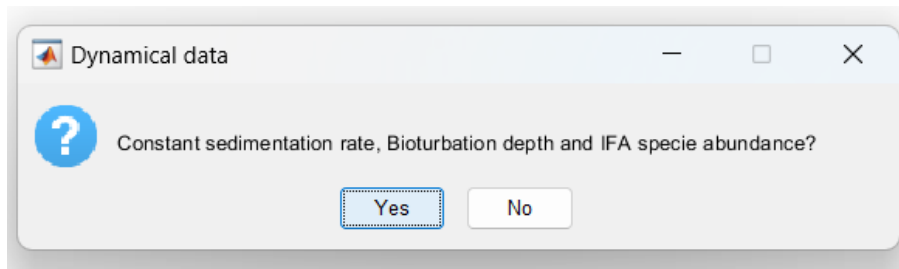

Then, click OK. That's it! Now you just have be patient and wait until de script stops.

7. Step by Step *bioprobv3_1.m*

In bioprobv3_1.m you don't need to define input data before run it. So, open *bioprobv3_1.m* and **click ➡ and run** the script:

1. **Do you want to simulate a constant SAR, BD, and IFA abundance?**

Then click **yes**. User will be asked to input constant SAR and BD, and the size of the IFA dataset:

That`s all, in a few seconds the analysis will end.

If **no** click it.

2. **General inputs for dynamical scenarios.** The user will have to add as input data: (*a*) the number of IFA individuals; (*b*) the start and end year of the simulation (in ka); and (*c*) the mean age of the IFA discrete depth (in ka); *(d)* number of synthetic cores to be bioturbated (var run). As to the abundance of the selected IFA species, there are three options: (*i*) constant (abu = 0); (*ii*) dynamical abundance: abundance data that come from foraminifera census counts (abu=1); and (*iii*) dynamical abundance: abundance data that come from e.g., temperature transfer functions, abundance vs time (abu=2). Then, click OK.

➔ If abu=0: You have to indicate if the IFA specie grew along the year or not. Then, click OK.

After that, you will need to input: the core depths (in cm) where its mean age its known (kaBP); the bioturbation depth (cm) at each depth/age. As the abundance do not changes, just repeat the same value in the last row as many times as you have data for depth, age, and BD. Then, click OK. That's all!

➔ If abu=1: You have to indicate if the IFA specie grew along the year or not. Then, click OK. After that, an opening window (same than just above) will ask you to input depth, age, BD and IFA abundance at each depth. Then, click OK.

➔ Finally, you have to input a value to scale the abundance data e.g., 10 (from 0 to 10).

Then, click OK. That's all!