**Part 2**

**Chronological modelling using the ‘rbacon’ package**

**"R" you ready?**

**Dr Larissa Schneider**

Larissa.Schneider@anu.edu.au



**Table of Contents**

[**1 – R and R Studio** 3](#_Toc520666004)

[**2- Bacon: an R package** 4](#_Toc520666005)

[**3- Setting up Bacon** 4](#_Toc520666006)

[3.1. Install rbacon package: 4](#_Toc520666007)

[3.3. Set RStudio to read your files in the correct directory (to the folder Bacon\_runs). 5](#_Toc520666008)

[**4- Running Bacon using the example core** **MSB2K** 6](#_Toc520666009)

[4.1- What do you need to know to make the magic happen? 6](#_Toc520666010)

[4.2- Now, what is all that info in the RStudio console? 11](#_Toc520666011)

[4.3- What is the result? 11](#_Toc520666012)

[**5 - Download files for this workshop** 12](#_Toc520666013)

[**6- Lynch’s Crater** 12](#_Toc520666014)

[**Supplementary Exercises** 14](#_Toc520666015)

[**1- Anteojos: a dream core!** 14](#_Toc520666016)

[**2– Chumbrumba: a more complicated core** 16](#_Toc520666017)

[**3- Lake Kutubu: when limestone is not helping us** 18](#_Toc520666018)

[**4– Palau sinkhole: the spray of outliers** 19](#_Toc520666019)

[**5- Type five error fixes** 22](#_Toc520666020)

[**6- Bacon functions** 23](#_Toc520666021)

[**7- References** 23](#_Toc520666022)

# **1 – R and R Studio**

**R** is a language and environment for statistical computing and graphics.

**R-Studio** uses the graphical environment of the computer to facilitate interactions with R (Figure 1):

* Console window for typing code
* Data window for reviewing data frame and other data structures
* Work space for viewing all of the data loaded into R
* Plot area for showing visualizations



Figure 1 – A view of RStudio showing its different environments.

**Why bother using R-Studio?**

It is more flexible than R, and provides direct access to R code

Just trust me….you will do more data analysis faster and with fewer headaches if you use **Rstudio**!

# **2- Bacon: an R package**

Bacon is an approach to age-depth modelling that uses Bayesian statistics to reconstruct **B**ayesian **ac**cumulati**on** histories for deposits, through combining radiocarbon and other dates with prior information.

Any age-depth model produces estimates of accumulation rates, but **Bacon gets its fame because it has the ability to divide a core into many vertical sections** (by default of thick=5 cm thickness), and through **millions of Markov Chain Monte Carlo (MCMC) iterations estimates the accumulation rate** (in years/cm; so more correctly, sedimentation times) for each of these sections. Combined with estimated starting dates for the first section, these accumulation rates then **form a precise age-depth model.**

Bacon works with any kind of absolute dates (e.g., 210Pb, OSL, tephra or other dates on a calendar scale; all dates are assumed to come with an error, or estimated uncertainty).

For help and explanations of functions and options within functions, type **?Bacon**

There is a list of functions and options to use in Bacon at the end of this handout.

# **3- Setting up Bacon**

## 3.1. Install rbacon package:

You will first have to install Bacon in R. It is now available as an R (CRAN) package, so you just need to type and run the functions:

install.packages (“rbacon”) 🡪 note the name and the **lower case**. Choose a mirror near to you.

require (rbacon)



* **install.packages** is the function used to install packages. You only need to install a package once per computer (note that new versions of packages in R are released now and then, make sure you are using the latest one).
* **Require** is the function used to load packages. You will need to run it every time you open RStudio.

3.2. Create a new folder called Bacon\_runs:

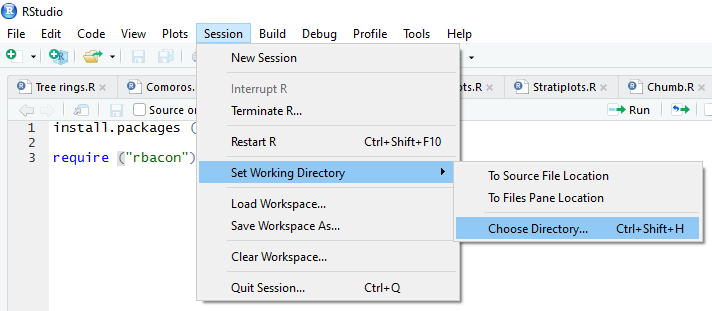
You can create a **Bacon\_runs** folder wherever you want it to be. For this workshop thought, please create a **Bacon\_runs** folder in a USB stick or on your desktop.

**Bacon\_runs** folder to read your core data from and store the output files created. Bacon will read your core data from the Bacon-runs folder and output the results of your model to the same folder. To facilitate this, create sub-folders with the name of your cores. Example

**Desktop\Bacon\_runs\Anteojos**

* There is a core within the rbacon folder (in your R library) called MSB2K. This is an example (peat core from the Netherlands) that comes with the rbacon package. If you get stuck in your data entry, you can always refer to this core for reference.

## 3.3. Set RStudio to read your files in the correct directory (to the folder Bacon\_runs).



# **4- Running Bacon using the example core** **MSB2K**

## 4.1- What do you need to know to make the magic happen?

* You need to provide Bacon with a file containing dates for your core (this is the sub-folder we were talking about earlier).
* Your datafile containing dates should be in a **.csv** format (.csv means your data values are separated by comma).
* Have four columns: labID | age | error | depth

1. LabID should be unique and meaningful (but Bacon is not picky).
2. Age is radiocarbon age (more about using non-radiocarbon dates later).
3. Error is lab error in this case. Error cannot equal zero!
4. Depth is in cm and it refers to the middle of the sediment layer.

* If you chose to have a file with core depths, it should:

1. Have a single column without header and stored within the core's folder.
2. Should be in a **.txt** format.
3. The file name should start with the core's name and end with '\_depths.txt'.
4. You need to specify in your code that **depths.file=TRUE** (default is FALSE).

**Bacon()** is the function you use to create an age-depth model. Type this function in your editor and run it, Bacon will run the model using the default core (MSB2K).

**This core has lots and lots of dates!** (this is not always the case as 14C analyses are expensive).

Pause here. Observe the following in Figure 2:

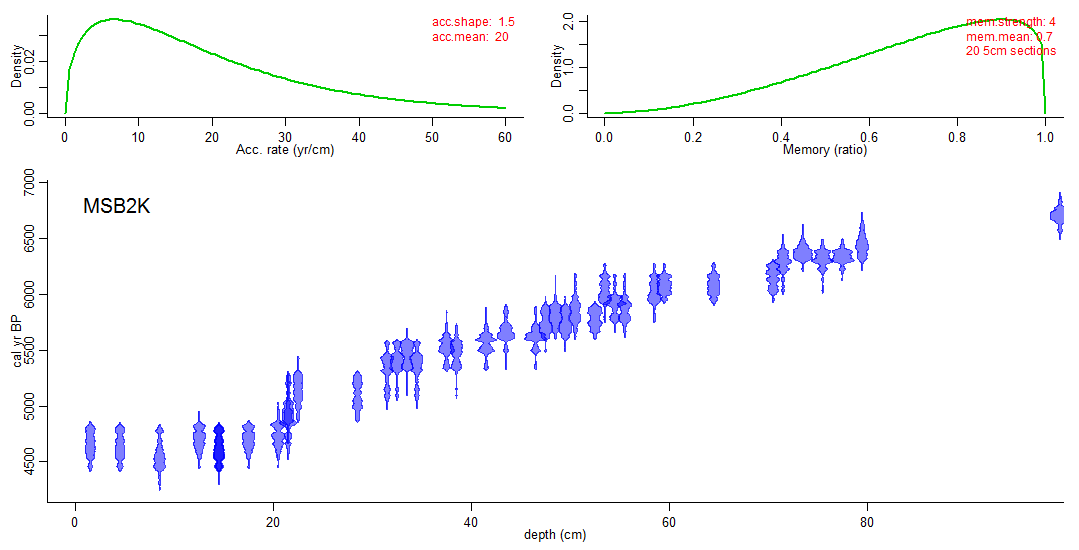


Figure 2 – Bacon output of dates and prior information for the example core MSB2K.

Observe the blue symbols in the graph panel. These are probability distribution functions (pdfs) of the radiocarbon dates in this model. This blue colour denotes that they’re all 14C dates.

In the R window, there is a message “run Bacon with 20 sections?” This is because Bacon wants to divide the cores in 20 sections of 5 cm **(thick = 5)**. Bacon usually does a good job in setting the number of sections based on the dates that you provided it with, so let it run with the number of sections it has suggested, that means, **type y.**

Now Bacon does its thing! You may feel your fan working and your computer getting a little warm.

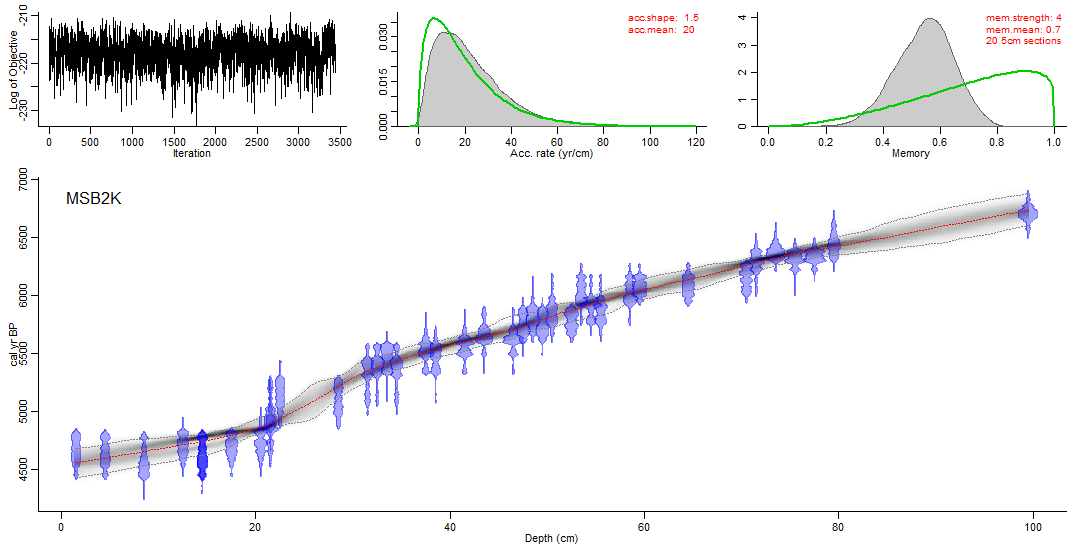


Figure 3 –Bacon output after running core MSB2K.

Observe the following in Figure 3.

* Red line is weighted mean maximum probability date for that depth.
* The intensity of black is related to the probability around that line.
* Full width is 95% confidence interval.
* Notice that the confidence interval is tighter near where there are more dates.
* This is only a 1m core! With all those dates. No wonder it’s the example.

4.2- How to interpret Bacon plots

|  |  |
| --- | --- |
| Top left plot | **What does this plot mean?** This is a plot that shows how the millions of Markov Chain Monte Carlo (MCMC) iterations performed in the model run.  **What should it look like?** It should like an overall straight noise. You don’t want any systematic pattern happening. This black noise should be oscillating linearly.  **What it should NOT look like?** It should not look uniform as well as you don’t want the noise to wonder from the linear plot. |
| Top middle plot | **What does this plot mean?** The top middle pane is a representation of the accumulation rate distribution (deposition time).  The green line is the prior distribution Bacon sets based on your dates. The accumulation rate prior consists of a gamma distribution which is much like a normal/Gaussian  distribution but is often asymmetric and always positive. (since we can often safely assume that deposits did not  accumulate backwards in time).  The grey curve is the accumulation rate that Bacon found in your ore after running millions of iterations in your model. It has two parameters:  **acc.shape:** this is the setting for the prior distribution. Default is 1.5, higher values result in more peaked  shapes.  **acc.mean:** default 20, which can be changed to, e.g., 5, 10 or 50 for different kinds of deposits. In the case of the MSB2K example, acc.mean is 20 years per cm.  20 5 cm = means this core was run using 20 sections of 5 cm.  Bacon is pretty good in setting these parameters automatically. In case initial rough age estimates indicate an accumulation rate much different from the default, Bacon will suggest an alternative acc.mean based on round values such as 1, 2, 5, 10, 20, 50, …, 500 yr/cm. Ensure that you agree with the prior distributions applied to your core.  **What should it look like?** The green and grey curves should be matching. You don’t want the grey curve to go too much out of the green curve. If that happens, you need to reset your prior distribution.  **What it should NOT look like?** You don’t want the grey curve to be offset from the green curve. |
| Top right plot | **What does this plot mean?** This shows the memory used to run your core. The memory defines how much the accumulation rate of a particular depth in a core depends on the depth above it.  A low memory assumes that the accumulation rate has changed significantly over time.  A high memory will imply a smooth, more constant accumulation history.  Memory values are between 0 (no assumed memory) and 1 (100% memory).  **mem.strength** will determine how much the memory will be taken into account in the model (higher values result in more peaked shapes).  mem.mean=0.7 allow for a large range of posterior memory values.  **What should it look like?** The prior information should be a beta distribution with values between 0 and 1. The posterior information looks like a normal distribution.  **What it should NOT look like?** They should not be entirely separate (i.e. the green peak and shoulders being entirely separate of the green one). |
| Bottom plot | **What does this plot mean?** This is the output of your model.  **What should it look like?** Really depends on the nature of your core.  **What it should NOT look like?** It should not be elbowy and the model should fit your dates, if you don’t have outliers or other issues. |



**Avoid changing** prior information. Bacon usually does a good job on that, there is no need to change it in most of the cases.

If you know your corse site has a fast or slow deposition rate, you could change the **acc.shape.** Note that the default 1.5 is a quite permissive accumulation, so don’t worry too much about it.

If you have time at the end of the workshop, try to play with these functions:

**acc.shape** = (default 1.5, higher values result in more peaked shapes).

Try chaging it to: **Bacon (acc.shape=50)**

**acc.mean =** (default 20, which can be changed to, e.g., 5, 10 or 50 for different kinds of deposits).

Try changing it to: **Bacon (acc.mean = 50)**

**mem.strength =** 4 (higher values result in more peaked shapes).

Try changing it to: **Bacon (acc.mean = 50)**

**mem.mean =** 0.7 (allow for a large range of posterior memory values).

Try changing it to: **Bacon (mem. mean = 1)**

## 4.2- Now, what is all that info in the RStudio console?

It is a sum of information on what was happening in the model while R was running is. There is a lot of technical information, but no need for concern… basically what you have to check is:

* Is R running the right model?
* Has R run the model using the calibration curve you wanted?

**Just for your information:**

twalk: tells you how many iterations Bacon has tried to fit in your model. In the case of core MSB2K it has performed 4,840,000 iterations.

twalk thinning: tells how many iterations were accepted and stored for your core.

suggested burn in= 22,000 means that Bacon removed the first 22,000 iterations from the model as they were not fitting well.

## 4.3- What is the result?

* R calculated the age for each of the depths you have requested it to do. Default ages are calculated for each cm and stored as files ending in **'\_ages.txt'** within the core's folder.
* Look at MSB2K\_20\_ages.txt. This is the output file. It gives the mean and weighted mean probability ages for each centimetre, as well as the minimum and maximum possible ages.

A pdf of the graph image is automatically saved in the folder for that core. Save/export graph image commands are in the RStudio window. Different file formats are available.

Note that the new files (the ones that Bacon has generated) have the number of sections (in the case of MSB2K, 20) as part of the file name. If you run the same core again with the same number of sections, Bacon will overwrite the files without confirming.

# **5 - Download files for this workshop**

The Lynch’s Crater data file required for the following exercise can be found at <https://github.com/felihopf/CABAH>. Right click on the file “LC.csv” and select “Save link as” to download the file. Create anew folder inside your “Bacon\_runs” folder and name it: LC. Copy the LC.csv file into this folder.

# **6- Lynch’s Crater**

Lynch’s Crater is located in northeastern Australia and 44 14C ages were derived from swamp peat sediments by Turney et al. (2017) which we will now run in Bacon.

Open LC.csv and have a look at the file:

|  |  |  |  |
| --- | --- | --- | --- |
| **labID** | **age** | **error** | **depth** |
| Wk-36775 | 29645 | 181 | 697.9 |
| Wk-36774 | 29601 | 187 | 683.4 |
| Wk-36773 | 28673 | 162 | 667.9 |
| Wk-36772 | 27885 | 149 | 662.7 |
| Wk-36771 | 27693 | 146 | 657.5 |
| Wk-36770 | 27560 | 151 | 650.3 |

Observe the following:

* Reported errors should be at 1 standard deviation and always be > 0.
* Avoid using spaces within the individual names.

The default calibration curve for 14C dates is IntCal13 (cc=1). You can change this to cc=2 (Marine13), cc=3 (SHCal13), or cc=4 (an alternative curve). We want **cc=3** as Lynch’s Crater is in the Southern Hemisphere. You can either add a 5th columm to the spreadsheet with the heading “CC” and a value of “3” for each date, but we will simply specify cc=3 in the code when we run the age model as illustrated below:

Type in RStudio editor:

Bacon("LC", cc=3)

Then Bacon will ask you:

* Ballpark estimates suggest changing the prior for acc.mean to 50 yr/cm. OK? (y/n)

Type y

* Run LC with 59 sections? (y/n)

Type y

Then we get something like Figure 4:

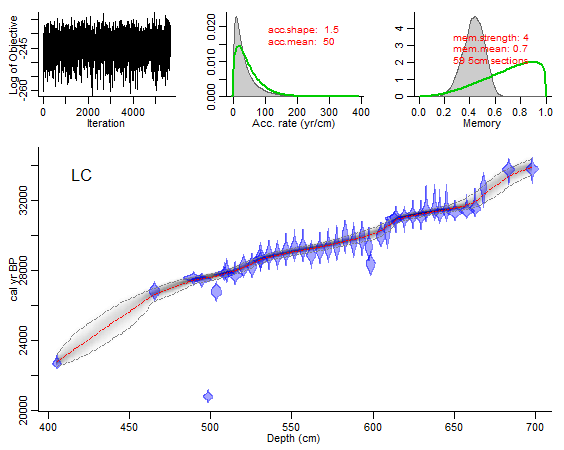


Figure 4 – Output for the Lynch’s Crater core.

Now, try this:

**agedepth(rotate.axes=T)**

# **Supplementary Exercises**

If you would like to explore Bacon further, you can work through the following examples from a variety of cores.

The files required for these examples can be downloaded from: <https://github.com/laribio/Palaeo>

Please create **four** new folders inside “rbacon\extdata\Cores” and name them:

* Anteojos
* Chumbrumba
* Kutubu
* Palau

Each of these folders represents a core for which we will build an age-depth model. You need at least a file containing labID | age | error | depth.

# **1- Anteojos: a dream core!**

Laguna de los Anteojos is a small headwater alpine lake in Venezuela. It is located in the Northern Hemisphere. All radiocarbon dates for this core are behaving very well and it should be straight forward to run it in Bacon.

This core was dated using bulk sediment.

Open Anteojos.csv and have a look at the file:

|  |  |  |  |
| --- | --- | --- | --- |
| **labID** | **age** | **error** | **depth** |
| UCI-37537 | 6420 | 20 | 262 |
| UCI-37511 | 8850 | 20 | 326 |
| UCI-37538 | 10180 | 25 | 372 |
| UCI-37539 | 11060 | 30 | 406 |
| UCI-37540 | 11880 | 35 | 425 |
| UCI-37623 | 12430 | 80 | 446 |

Observe the following:

* Reported errors should be at 1 standard deviation and always be > 0.
* Avoid using spaces within the individual names.

The default calibration curve for 14C dates is IntCal13 (cc=1). You can change this to cc=2 (Marine13), cc=3 (SHCal13), or cc=4 (an alternative curve). We want **cc=1** as Laguna de Los Anteojos is in the Northern Hemisphere.

Type in RStudio editor:

Bacon("Anteojos", depths=TRUE)

Then Bacon will ask you:

* Ballpark estimates suggest changing the prior for acc.mean to 50 yr/cm. OK? (y/n)

Type y

* Run Anteojos with 37 sections? (y/n)

Type y

Then we get something like Figure 5:

A close up of a map

Description generated with high confidence

Figure 5 – Output for the Anteojos core.

Now, try this:

**agedepth(rotate.axes=T)**

# **2– Chumbrumba: a more complicated core**

Chumbrumba Swamp lies within a mosaic ecosystem complex that was presumed to be maintained through environmental factors (e.g. drainage) and disturbance regimes (e.g. fire, grazing, clearing). Accumulation rate is expected to be variable in this core.

This site is in Curtain Fig National Park, Queensland (Southern Hemisphere). It was dated using bulk sediment.

* **What to note about this core:**
* We are working with 14C and 210Pb.
* As Chumbrumba is in the Southern Hemisphere, we need to add a fifth column called cc (remember: cc=3 for the Southern Hemisphere, cc=0 to cal BP ages).
* We are working with samples after 1950, (negative 14C ages, we live in the future!) so you need to specify a postbomb curve.
* See the global map at the end of this handout. Postbomb = 1 for the NH1

2 for NH2

3 for NH3

4 for SH1-2

5 for SH3

* For calendar dates ( i.e. 210Pb dates that are already on the calendar scale and thus should not be calibrated), set cc=0.
* Bacon will divide the core into numerous sections of equal thickness specified by **thick= (default 5)**, from the uppermost to the lowermost dated depth. Bacon calculates accumulation rates for each of these sections.

Type:

**Bacon (“Chumbrumba”, cc=3,**  **postbomb = 4)**

You will get this:

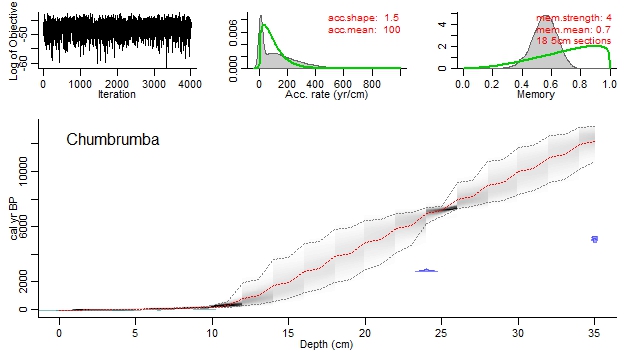


Figure 5 – A not desired age-depth model for Chumbrumba.

This is not what we are after really, note that the model is not following the dates. This is because the logarithm of 210Pb age-model is different than the 14C. Marco Antonio Aquino Lopez ([maquinolopez01@qub.ac.uk](mailto:maquinolopez01@qub.ac.uk)) is developing a package to be able to harmonise the two dating systems. The package is called Plum and is already available for Mac users.

But while we do not have Plum ready for PC, lets sort this out via other means:

Try this: **Bacon("Chumbrumba", thick=2, acc.mean=100**)

A close up of a map

Description generated with high confidence

Figure 7 – A good age-depth model for Chumbrumba.

Que RICO!!!!!!!!!!!!!!!!!!!!

What we have just done was to set our model to 18 sections. It then considered 210Pb in one session and 14C dates in the other. With more sections, Bacon is more perceptive of changes in accumulation rates and other things in your core.

Observe the following:

* cc=3
* See dates in different colours
* More 14C dates would help here.

# **3- Lake Kutubu: when limestone is not helping us**

Lake Kutubu is a freshwater lake in a limestone karst in Papua New Guinea. This core was dated using bulk sediment.

* **What to note about this core:**
* We are working with 210Pb and 14C dates.
* Although Lake Kutubu is in the SH, you should use NH calibration curve (The Queen of Radiocarbon Rachel can tell you why).
* We are working with another core after 1950, so we need to specify a postbomb curve.
* We are dealing with a core with one of the highest reservoir effected reported in the literature.
* We have tephras in this core (they are our saviors!!!!)

The reservoir effect on this core was calculated using the R package ResAge (<https://darchive.mblwhoilibrary.org/bitstream/handle/1912/7508/ResAge_package.zip?sequence=2>).

Open the Kutubu excel file and it will be all there for you. We have two new columns:

* dR
* dRSTD

They are the reservoir effect to this core in 14C years.

Type: **Bacon(“Kutubu”, postbomb = 4)**

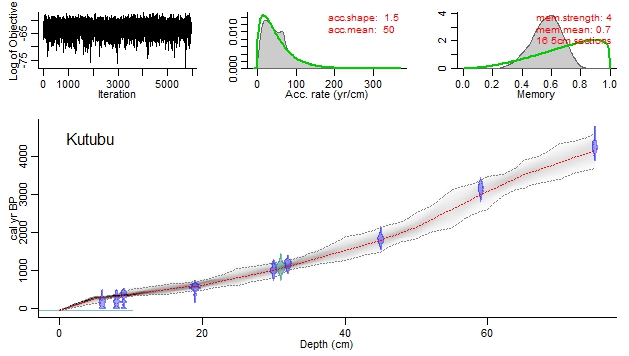


Figure 8 – Lake Kutubu age-depth model with reservoir effect corrected.

# **4– Palau sinkhole: the spray of outliers**

As the other cores for this workshop, Palau has also been dated using bulk sediment.

This core was collected from a sinkhole that had marine influxes up to 3000 cal BP (corresponding to the depth of 220 cm). After this period, the sediment of this sinkhole has received more *in situ* material.

This means that the part below 200 cm should not be really taken into consideration. Let’s see why not:

Type: **Bacon(“Palau”, postbomb= 3)**

You will get this:

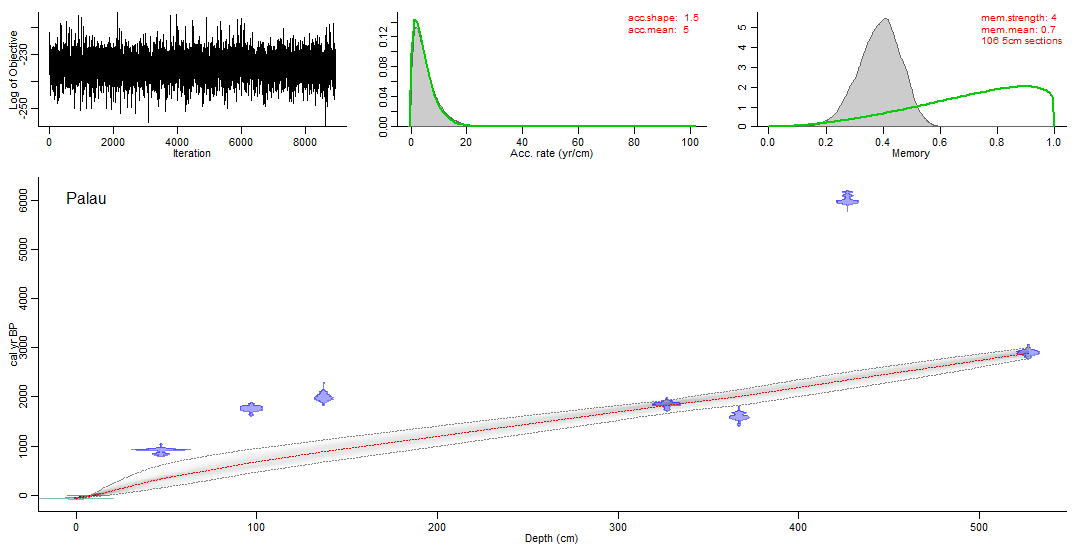


Figure 9: A “undesired” age-depth model for the Palau core.

It is really not what we are after, it seems there are many outliers and the model is lost. As we have this information about the part below 200cm, let’s exclude the bottom part of the core and only run the model for the top 200 cm.

Type:

**Bacon("Palau", postbomb=3, d.max=200)**

A close up of a map

Description generated with high confidence

Figure 10 – A good age-depth model of Palau, but not reflecting the entire core.

It looks ok, but it is not reflecting the entire core. We need to show to the reader that the core is longer and mixed. Our solution here is to add the rest of the core in the plot without running the model. You can do this by typing:

**agedepth(d.max=550)**

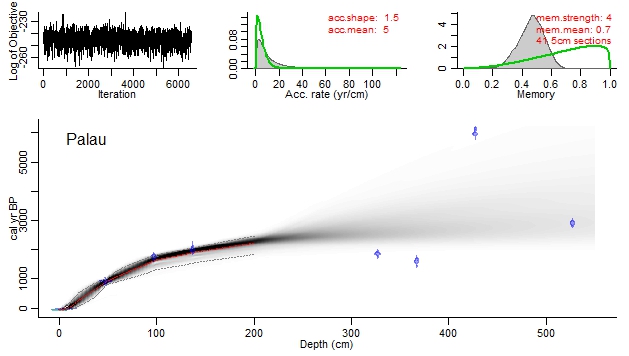


Figure 11- An ideal age-depth model for Palau core.

# **5- Type five error fixes**

1. Make sure you have set the directory in R to read the folder Bacon\_runs.
2. Make sure the names of your cores do not have spaces.
3. Make sure you have the parameters right (ages in 14C or cal BP, correct calibration curves for both 14C and postbomb, etc.)
4. If your core is too long, sometimes R will lose itself (the line will not follow the ages). Maarten is trying to fix that. If this happens to you someday, just email him to sort out the problem.
5. If the plot you get after running your model is too elbowy, try to play with the functions:

**thick=**

**acc.shape=**

**mem.strength=**

For cores longer than a few metres it could make sense to increase the section thickness for an initial quick run (e.g., setting thick=50), followed by decreasing thick until the model appears sufficiently smooth.

When playing with these functions, you will have to find a sweet spot for those parameters to be able to have a nice output that does not look too artificial.

# **6- Bacon functions**

By typing args(Bacon) you get the following list of possible arguments and their default values:

function (core = "MSB2K", thick = 5, prob = 0.95, d.min = NA, d.max = NA, d.by = 1, unit = "cm", maxcalc = 500, depths.file = FALSE, acc.shape = 1.5, acc.mean = 20, mem.strength = 4, mem.mean = 0.7, hiatus.depths = NA, hiatus.shape = 1, hiatus.mean = 1000, after = 1e-04, cc = 1, cc1 = "IntCal13", cc2 = "Marine13", cc3 = "SHCal13", cc4 = "ConstCal", postbomb = 0, d.R = 0, d.STD = 0, t.a = 3, t.b = 4, normal = FALSE, suggest = TRUE, reswarn = c(10, 200), remember = TRUE, cleanup = TRUE, ask = TRUE, run = TRUE, defaults = "default\_settings.txt", sep = ",", dec = ".", runname = "", slump = NA, BCAD = FALSE, ssize = 2000, rounded = 1, th0 = c(), burnin = min(200, ssize), MinYr = c(), MaxYr = c(), find.round = 4, bins = c(), cutoff = 0.001, plot.pdf = TRUE, rotate.axes = FALSE, rev.yr = FALSE, rev.d = FALSE, yr.min = c(), yr.max = c(), normalise.dists = TRUE, plot.title = core, title.location = "topleft", d.lab = "Depth", yr.lab = c(), d.res = 200, yr.res = 200, date.res = 100, grey.res = 100, width = 15, dark = 1, greyscale = function(x) grey(1 - x), C14.col = rgb(0, 0, 1, 0.35), C14.border = rgb(0, 0, 1, 0.5), cal.col = rgb(0, 0.5, 0.5, 0.35), cal.border = rgb(0, 0.5, 0.5, 0.5), range.col = grey(0.5), range.lty = "12", hiatus.col = grey(0.5), hiatus.lty = "12", wm.col = "red", wm.lty = "12", med.col = NA, med.lty = "12", mar = c(3, 3, 1, 1), mgp = c(1.5, 0.7, 0), bty = "l")

# **7- References**

Turney, C.S.M., Jones, R. T., Phipps, S. J., Thomas, Z., Hogg, A., Kershaw, A. P., Fogwill, C. J., Palmer, J., Bronk Ramsey, C., Adolphi, F., Muscheler, R., Hughen, K.A., Staff, R. A., Grosvenor, M., Golledge, N., R., Rasmussen, S. O., Hutchinson, D. K., Haberle, S., Lorrey, S., Boswijk, G. and Cooper, A., 2017. Rapid global ocean-atmosphere response to Southern Ocean freshening during the last glacial. *Nature Communications***volume 8**, Article number: 520 (2017).

World map showing zonal atmospheric bomb 14C. From: Hua, Q; Barbetti, M.; Rakowski, A.Z. 2013. Atmospheric Radiocarbon for the period 1950-2010. Radiocarbon 55: 2059-2072.

