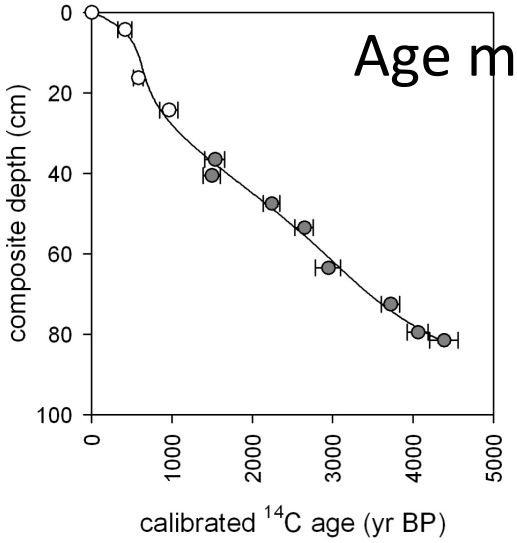
LiPD and `CSciBox`: A case study in why data standards are important for paleoscience

Liz Bradley, Izaak Weiss, Nick McKay, Julien Emile-Geay, Laura de Vesine, Ken Anderson, Jim White, and Tom Marchitto

University of Colorado Boulder

oceanworld.tamu.edu

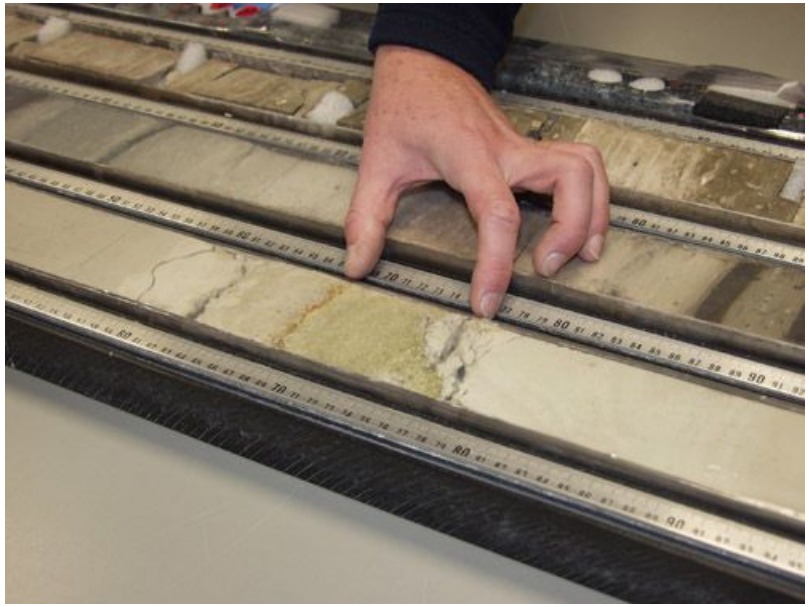
Paleoclimate
proxy data

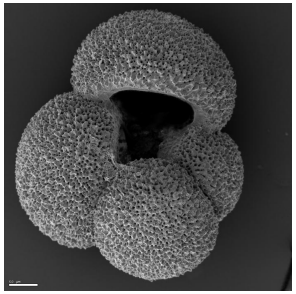


Science!

politicalclimate.wordpress.com

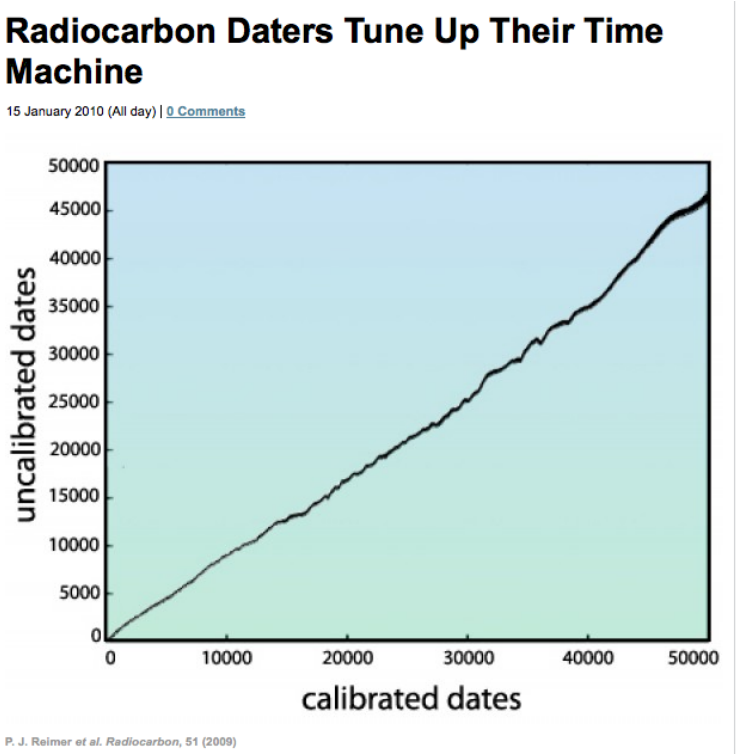
Building age models is hard

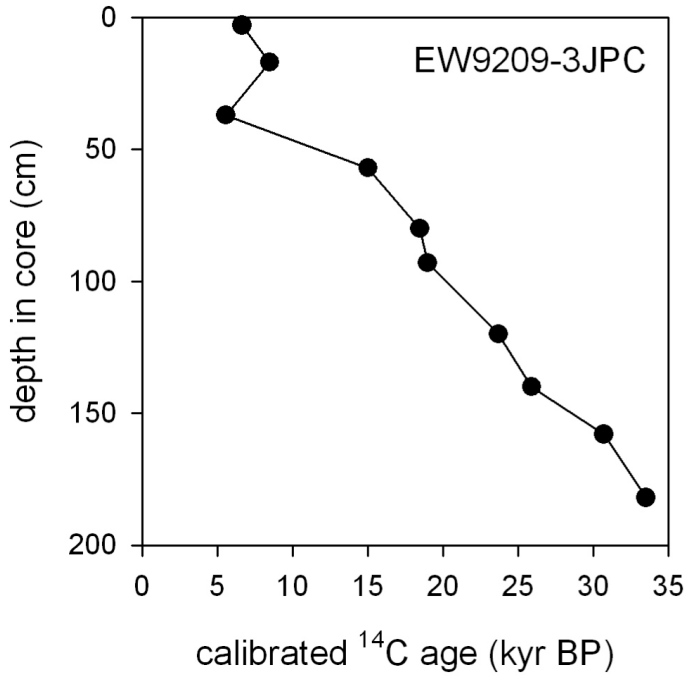




Building age models is hard

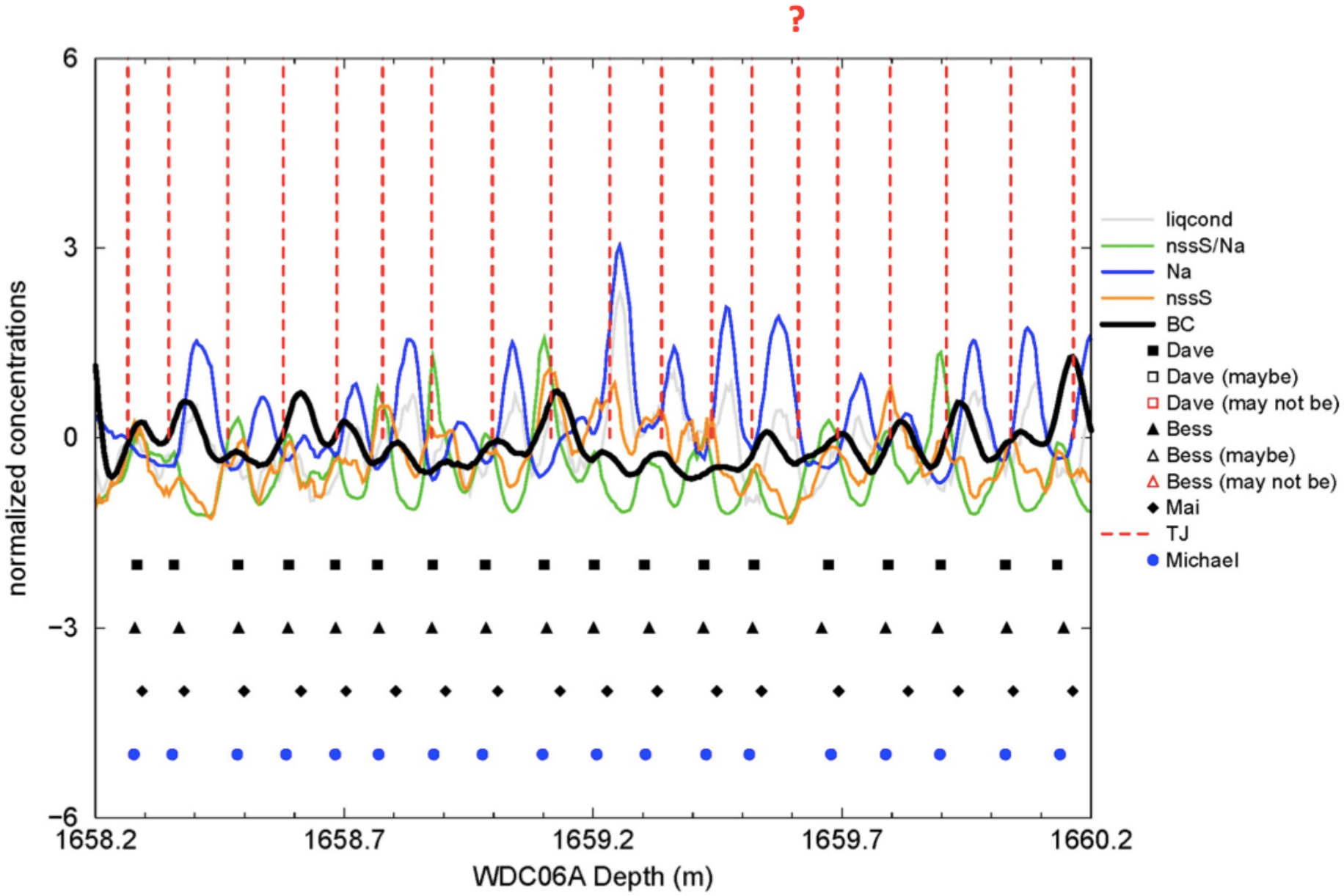
- Requires expert knowledge and forensic reasoning





Building age models is hard

- Requires expert knowledge and forensic reasoning
- Can involve subjective judgements


?
6
3
0
−3
−6
normalized concentrations
1658.2
1658.7
1659.2
1659.7
1660.2
WDC06A Depth (m)
liqcond
nssS/Na
Na
nssS
BC
Dave
Dave (maybe)
Dave (may not be)
Bess
Bess (maybe)
Bess (may not be)
Mai
TJ
Michael

Building age models is hard

- Requires expert knowledge and forensic reasoning
- Can involve subjective judgements
- As well as some fairly complex mathematics

An automated approach for annual layer counting in ice cores

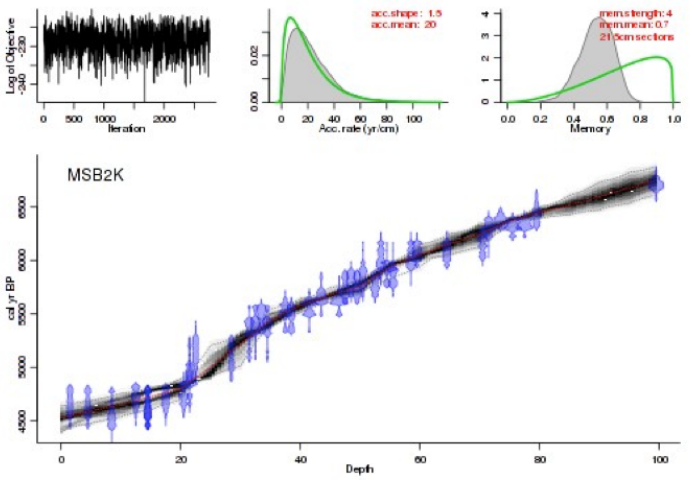
M. Winstrup[1,*], A. M. Svensson[1], S. O. Rasmussen[1], O. Winther[2], E. J. Steig[3], and A. E. Axelrod[4]

Clim. Past **8**:1881 (2012)

Age-modeling software is powerful, but not necessarily user-friendly

Bacon age-modelling software

If you will begin with certainties, you shall end in doubts,
but if you will content to begin with doubts, you shall end in certainties.
After Francis Bacon (AD 1561 - 1626)

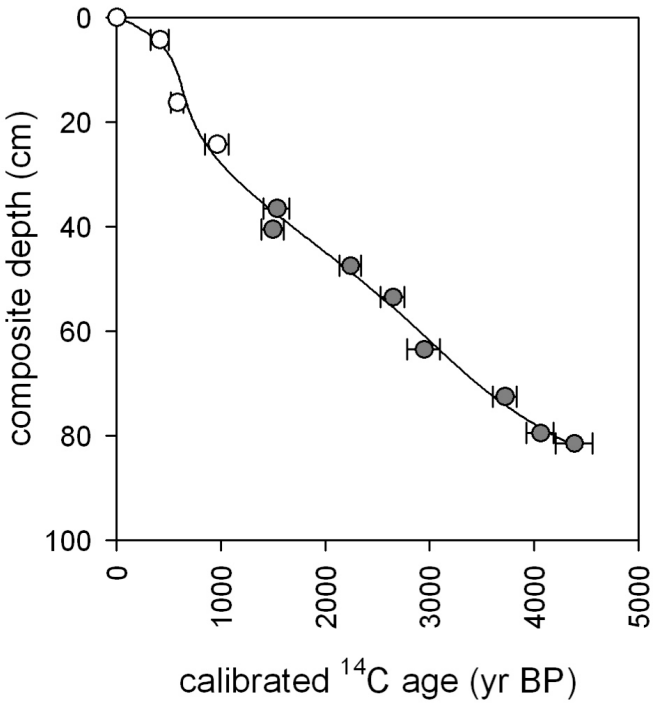



```
Bacon <- function(core="MSB2K", thick=5, prob=0.95, d.min=NA, d.max=NA,
d.by=1, unit="cm", maxcalc=500, depths.file=FALSE, acc.shape=1.5, acc.mean=20,
mem.strength=4, mem.mean=0.7, hiatus.depths=NA, hiatus.shape=1,
hiatus.mean=1000, after=.0001, cc=1, cc1="IntCal13", cc2="Marine13",
cc3="SHCal13", 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=.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,.35), C14.border=rgb(0,0,1,.5), cal.col=rgb(0,.5,.5,.35),
cal.border=rgb(0,.5,.5,.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,.7,.0), bty="l")
```

oceanworld.tamu.edu

Paleoclimate proxy data

CSciBox: automates this

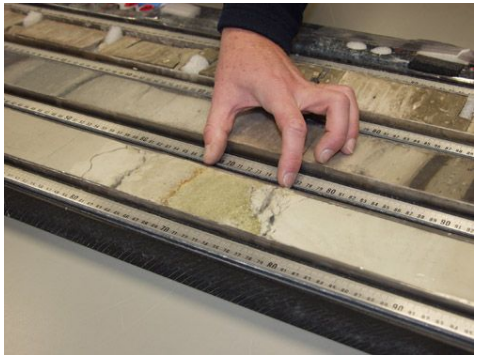


Age model

Challenges in automating age-model building

- Artificial intelligence (another talk)
- The software engineering (yet another talk)
- Outreach, education (best discussed over beers)
- The data! (this talk)

- Data issues
  - Volume
  - Heterogeneity



...

- Data issues
  - Volume
  - Heterogeneity
  - Formats

```
depth_m age_gas_calBP   ch4_ppb
692.165 11067.3 716.8    10.0
692.675 11079.2 672.7    10.0
```

EPICA

```
Bag      Depth_m    GasAge      CH4
3625     1993.48     33678      527
3631     1996.78     33939      438
```

NGRIP

- Data issues
  - Volume
  - Heterogeneity
  - Formats
  - "Flat" structure

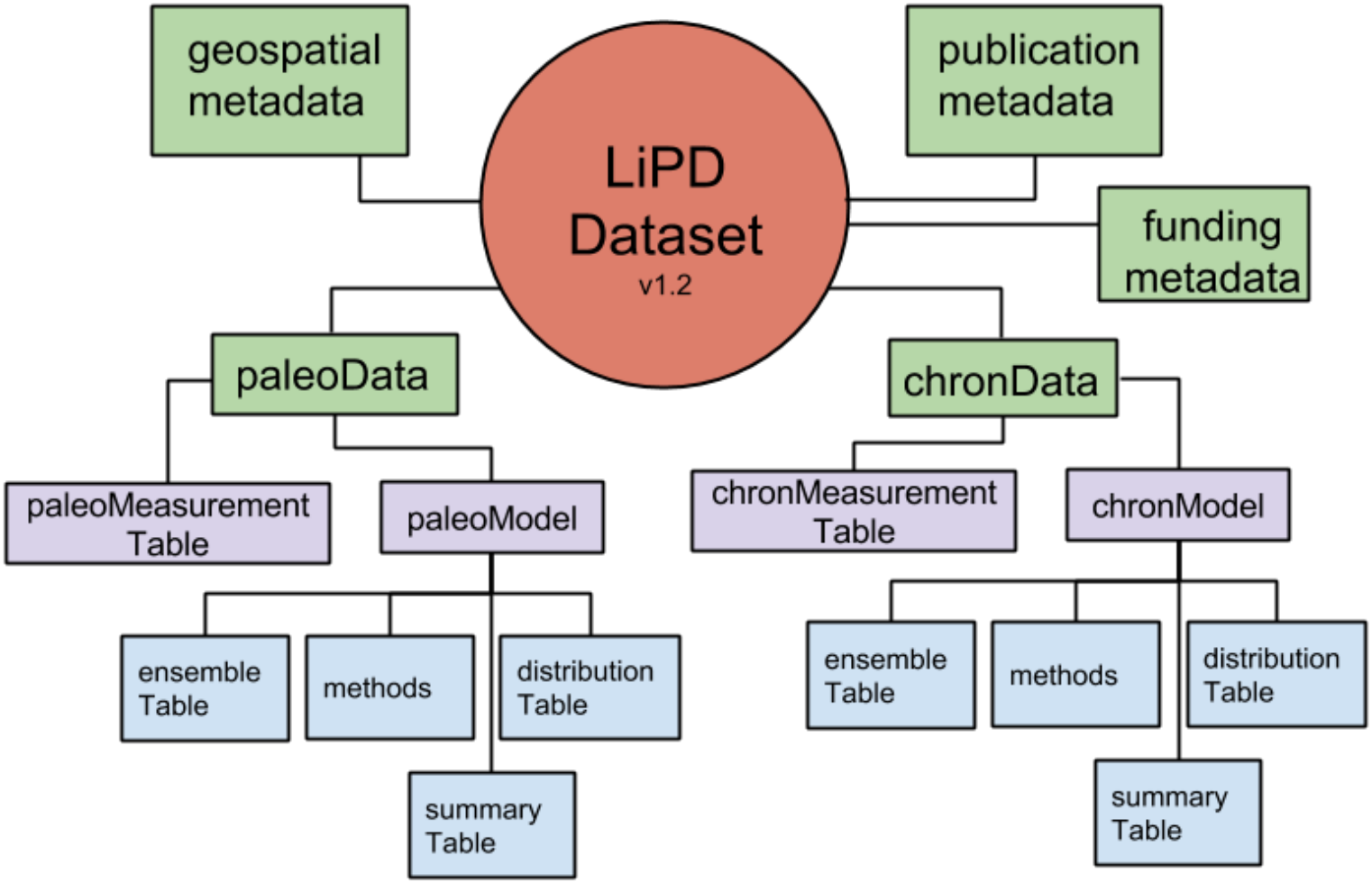
	A	B	C
1	14C Age	Err+	Err-
2	300.2	1.2	0.8
3	310.4	1	1.2

```
#------------------
# Variables
#
## depth_m      depth, , , m, , , EDML depth,N
## age_gas_calBP        Age gas,,,cal yr bp,,,unified EDML gas age scale Lemieux-Dudon et al. 20
## ch4_ppb      CH4 concentration, , , parts per billion, , , ,N
## ch4_1s_ppb   CH4 concentration 1 sigma uncertainty, , , parts per billion, , , , N
## notes        notes - Reference for this sample, , , , , , , C
#--------------------
# Data:
# Missing Value:
depth_m age_gas_calBP   ch4_ppb ch4_1s_ppb      notes
692.165 11067.3 716.8   10.0    EPICA Community Members (2006)
692.675 11079.2 672.7   10.0    EPICA Community Members (2006)
695.165 11138.3 688.2   10.0    EPICA Community Members (2006)
700.015 11253.6 698.0   10.0    EPICA Community Members (2006)
700.045 11254.3 715.8   7.3     "Baumgartner et al., (2012)"
700.165 11257.1 710.7   10.0    EPICA Community Members (2006)
703.165 11325.1 674.6   10.0    EPICA Community Members (2006)
```

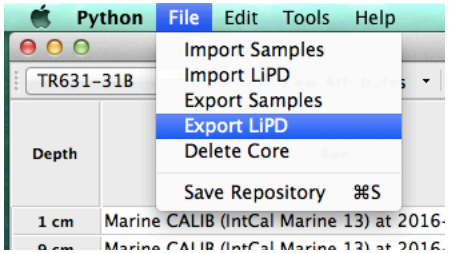
“Metadata”

Linked Paleo Data:

a container for paleoclimate data & metadata



What LiPD gives CSciBox



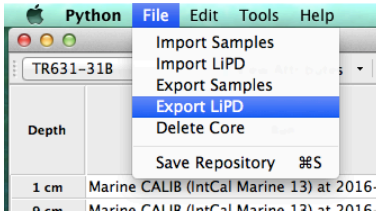
- Uniform file format:

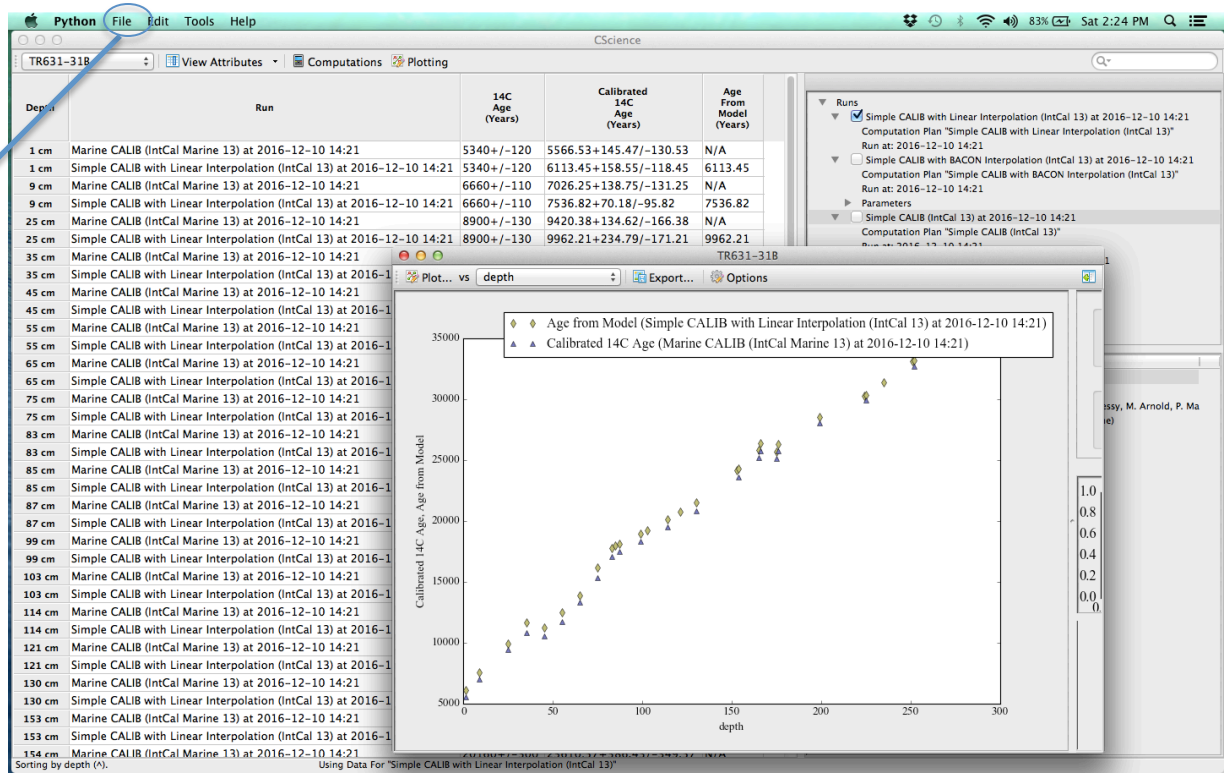
  structured, not flat: explicitly captures semantic relationships between variables
- Specific units with known meanings:

  CSciBox uses this to assure consistency
- Provenance is stored *with the data*:

  location, material, lab procedure, ...
- Citations are also stored in appropriate parts of the LiPD record:

  linked to the data record, the method, etc.
- *Analysis steps* stored with the record, too!
- Interoperable with any other LiPD-aware tool





- Documentable
- Reproducible
- Interoperable

Linked PaleoData

Python File Edit Tools Help

83% Sat 2:24 PM

CScience

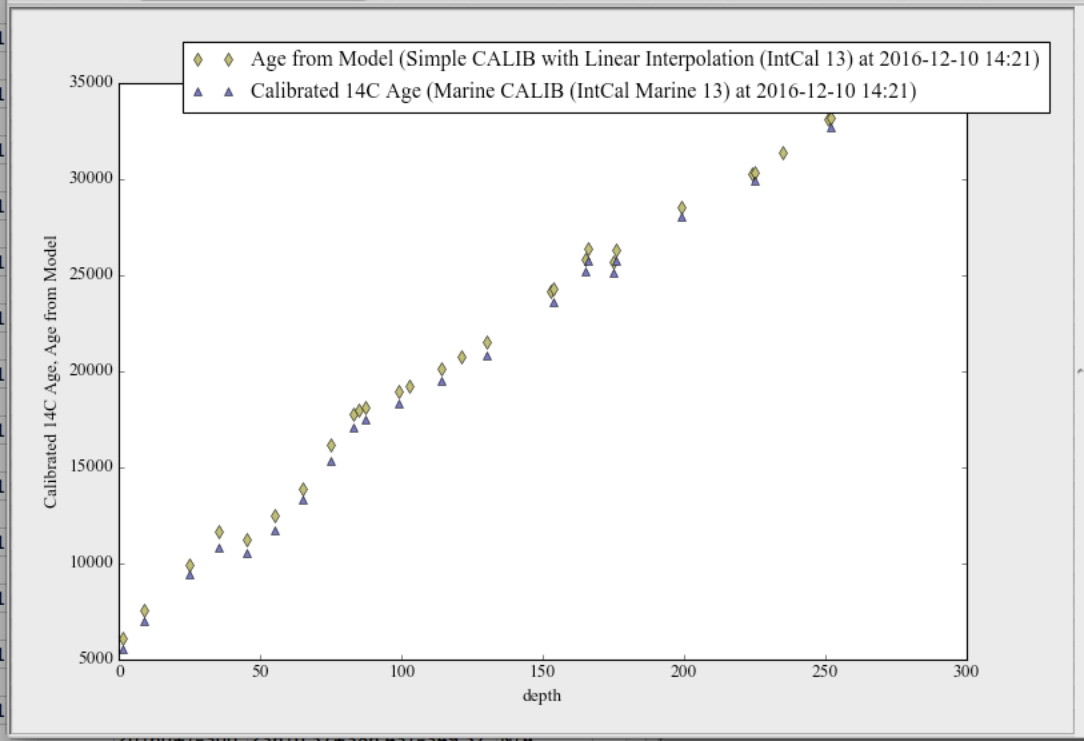
TR631-31B | View Attributes | Computations | Plotting

Depth	Run	14C Age (Years)	Calibrated 14C Age (Years)	Age From Model (Years)
1 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21	5340+/-120	5566.53+145.47/-130.53	N/A
1 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-12-10 14:21	5340+/-120	6113.45+158.55/-118.45	6113.45
9 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21	6660+/-110	7026.25+138.75/-131.25	N/A
9 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-12-10 14:21	6660+/-110	7536.82+70.18/-95.82	7536.82
25 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21	8900+/-130	9420.38+134.62/-166.38	N/A
25 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-12-10 14:21	8900+/-130	9962.21+234.79/-171.21	9962.21
35 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
35 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
45 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
45 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
55 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
55 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
65 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
65 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
75 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
75 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
83 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
83 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
85 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
85 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
87 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
87 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
99 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
99 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
103 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
103 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
114 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
114 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
121 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
121 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
130 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
130 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
153 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			
153 cm	Simple CALIB with Linear Interpolation (IntCal 13) at 2016-1			
154 cm	Marine CALIB (IntCal Marine 13) at 2016-12-10 14:21			

- Runs
  - Simple CALIB with Linear Interpolation (IntCal 13) at 2016-12-10 14:21
    - Computation Plan "Simple CALIB with Linear Interpolation (IntCal 13)"
    - Run at: 2016-12-10 14:21
  - Simple CALIB with BACON Interpolation (IntCal 13) at 2016-12-10 14:21
    - Computation Plan "Simple CALIB with BACON Interpolation (IntCal 13)"
    - Run at: 2016-12-10 14:21
    - Parameters
  - Simple CALIB (IntCal 13) at 2016-12-10 14:21
    - Computation Plan "Simple CALIB (IntCal 13)"

TR631-31B

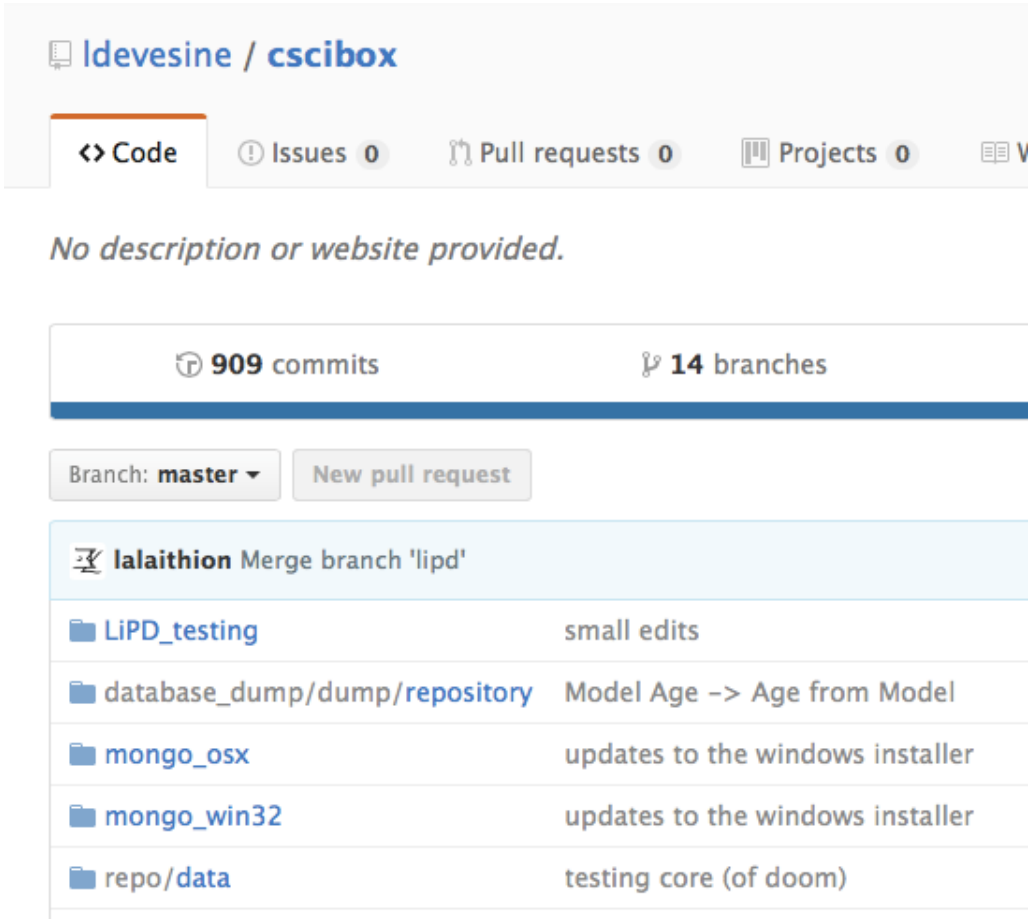
Plot... vs depth | Export... | Options



Sorting by depth (^). Using Data For "Simple CALIB with Linear Interpolation (IntCal 13)"

- Graphical User Interface, powerful plotter, lots of built-in tools, can compose your own analysis workflows, …
- Grab me or Izaak for a demo (and/or help getting it installed on your machine)
- The CSciBox code is open source and freely available on `github`

So you can download/modify/use it as you wish.



One-click installers for Windows, Mac OSX, as well as a collection of tutorial videos, are on our website

Thanks!

This material is based upon work sponsored by the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.