Building better age-depth models by integrating geochronologic nails and springs

David De Vleeschouwer

10/11/2021

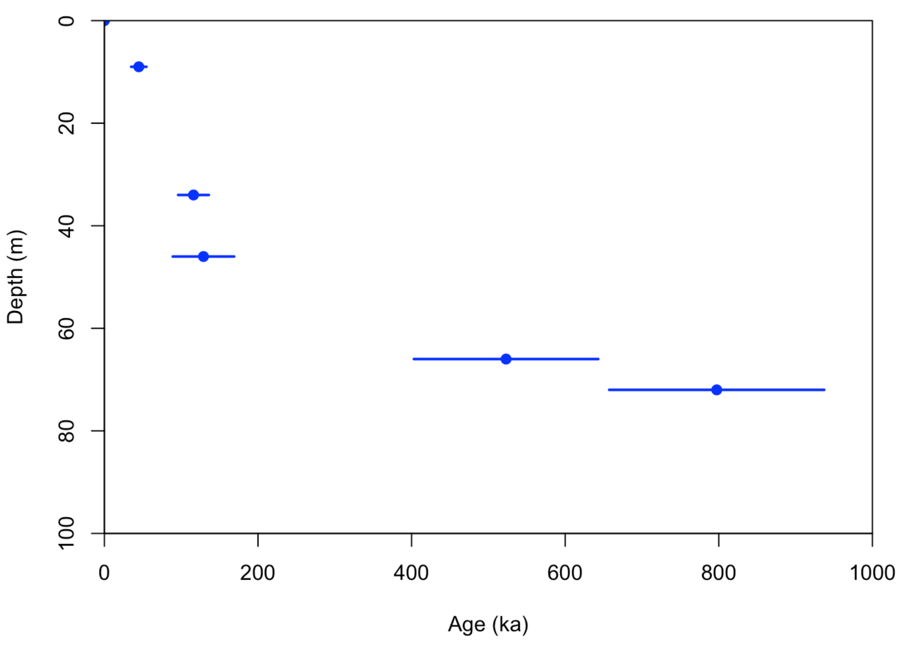
This is a thought-experiment to streamline efforts towards a statistically-sound integration of geochronologic “nails” (e.g. radio-isotopic ages) and “springs” (e.g. cyclostratigraphy). The procedure is strongly inspired by the Bchron methodology for Bayesian age-depth modeling, and it is the scientific continuation of the work that began with the publication of De Vleeschouwer and Parnell (2014, Geology) and continued by Harrigan et al. (2021, GSA Bulletin).

```r
rm(list=ls()) # Empty global environment
library(astrochron) # Load some useful R libraries that must be installed
library(Hmisc)
library(devtools)
```

A dummy dataset

Let’s take a look at the dummy dataset. This is a dataset I often use for teaching purposes. It consists of 5 dated levels with different levels of uncertainty and the information that the top of the core is at t = 0.

```r
par(mar = c(4,4,1,1))
dates = matrix(c(0,9,34,46,66,72,0,44.8,116,129,523,797.3,0.0001,10,20,40,120,140), nrow = 6, ncol = 3)
colnames(dates) <- c("Depth (m)", "Age (ka)", "Uncertainty (2sigma, kyr)")
dates = as.data.frame(dates)
plot(dates$`Age (ka)`, dates$`Depth (m)`, ylim = c(100,0), xlim=c(0,1000),xaxs = "i", yaxs = "i", col = "blue", pch = 19, xlab = "Age (ka)", ylab = "Depth (m)")
for (j in 1:6){segments(dates[j,2]-dates[j,3],dates[j,1],dates[j,2]+dates[j,3],dates[j,1], col = "blue", lwd = 2)
}
```



A compound Poisson process to make an age-depth model

Simple interpolation in-between dated levels ignores the fact that sedimentation rate can change in the intermediate stratigraphy. To alleviate this issue, we use a Poisson process that introduces “break-points” in-between dated levels. At these break-points, sedimentation rate can change. The Poisson distribution reflects the probability of a certain number of events (here: number of sedimentation rate break points) occurring in a fixed interval of time or space (here: stratigraphic thickness in-between dated levels).

To keep things simple, break-points are sampled from a uniform distribution both in the depth- and in the time-domain. Without doubt, this is over-pessimistic because this allows the Monte Carlo procedure to deviate very far away from the straight line between two subsequent dates levels.

Note that the `for (k in 1:5){` loop stipulates that the Poisson process only pertains to the stratigraphic interval in-between dated levels. The age-depth model for the deepest part of the core is constructed by simple extrapolation. This will be different in the next part of this document.

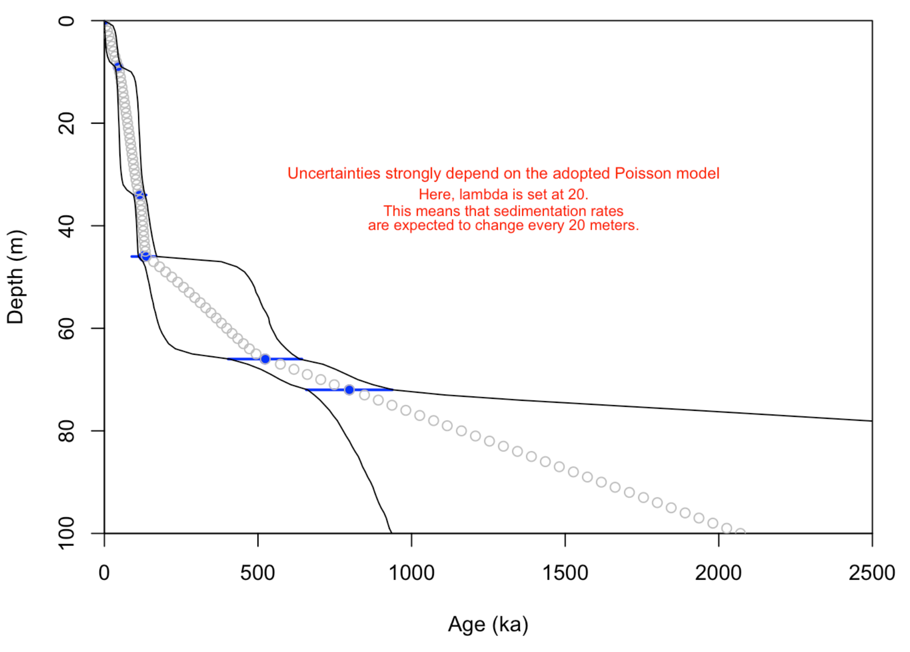
```r
Nsimulations = 10000
ages_lie_MonteCarlo_Poisson=matrix(data = NA, nrow = 101, ncol = Nsimulations ) # This matrix will contain all individual Monte-Carlo-simulated age-depth chronologies
depths_i=seq(0,100,1) # We'll make an age-depth model with a 1-meter resolution

for (i in 1:Nsimulations){
  age_depth_MonteCarlo=c()
  depth_Poisson=c()
  age_Poisson=c()

  for (j in 1:6){age_depth_MonteCarlo[j]=rnorm(1, mean = dates$`Age (ka)`[j], sd = dates$`Uncertainty (2sigma, kyr)`[j]/2)}
    if (length(which(diff(age_depth_MonteCarlo)<0))>0){ages_lie_MonteCarlo_Poisson[,i] = rep(NA, 101) # Prior knowledge: Law of superposition
    }  else{
      for (k in 1:5){
        thickness = dates$`Depth (m)`[k+1]-dates$`Depth (m)`[k]
        lambda = thickness / 20
        N_breaks = rpois(1, lambda = lambda)
        # temp1 represents a depth-level at which a break-point is introduced. temp1 is sampled from a uniform distribution between the two adjacent dated levels
        temp1=runif(N_breaks, min = dates$`Depth (m)`[k], max = dates$`Depth (m)`[k+1])
        # temp2 represents the age of the introduced break-point. temp2 is sampled from a uniform distribution between the two adjacent dates
        temp2=runif(N_breaks, min = age_depth_MonteCarlo[k], max = age_depth_MonteCarlo[k+1])
        depth_Poisson = c(depth_Poisson, temp1)
        age_Poisson = c(age_Poisson, temp2)
      }

      # We concatenate the depths of dated levels and introduced break-points for this specific M-C simulation
      depths_sim = c(dates$`Depth (m)`, depth_Poisson)
      # We concatenate the ages of dated levels and introduced break-points for this specific M-C simulation
      ages_sim = c(age_depth_MonteCarlo, age_Poisson)
      # Temp is the age-depth model (through linear interpolation) of this specific Monte-Carlo simulation
      temp=approxExtrap(depths_sim, ages_sim , xout = depths_i)$y
      # ages_lie_MonteCarlo_Poisson is a large matrix, storing each individual age-depth simulation.
      ages_lie_MonteCarlo_Poisson[,i]=temp
      }
  }
```

Now, we take the 2.5%, 50% and 97.5% quantiles of the 10 000 simulations to calculate the chronology and its confidence levels.

```r
ages_lie_MC_Poisson=matrix(data = NA, nrow = 101, ncol = 3) # This matrix will hold the Median Age and the Confidence Levels
colnames(ages_lie_MC_Poisson) <- c("2.5% CL", "Median Age", "97.5% CL")
for (k in 1:101){
  ages_lie_MC_Poisson[k,]=quantile(ages_lie_MonteCarlo_Poisson[k,], probs = c(0.025,0.5,0.975), na.rm = T)}
```

Let’s plot things up: We observe the typical sausage-shaped uncertainty that we know from Bchron. However, a fundamental difference with Bchron is that Bchron estimates the Poisson-parameter lambda directly from the dataset, whereas here it is provided by the user. In this case, we set lambda = 20. This means that we expect -on average- one sedimentation rate break-point every 20 meters.

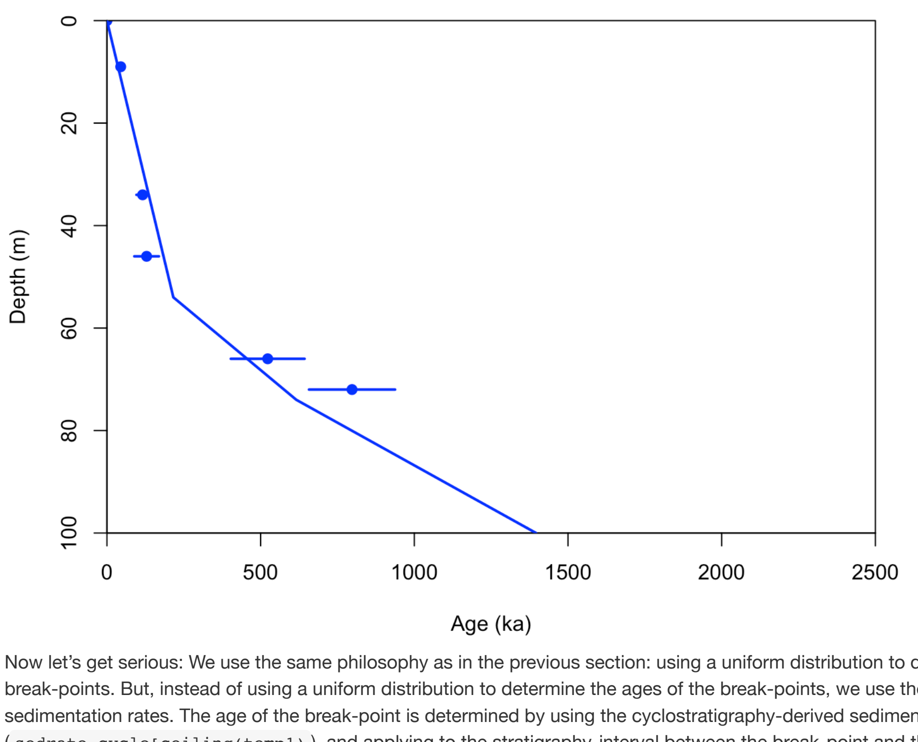
```r
# Let's plot things up
par(mar = c(4,4,1,1))
plot(dates$`Age (ka)`, dates$`Depth (m)`, ylim = c(100,0), xlim = c(0, 2500), xaxs = "i", yaxs = "i", col = "blue", pch = 19, xlab = "Age (ka)", ylab = "Depth (m)")
for (j in 1:6){segments(dates[j,2]-dates[j,3],dates[j,1],dates[j,2]+dates[j,3],dates[j,1], col = "blue", lwd = 2)
}
points(ages_lie_MC_Poisson[,2], depths_i, col = "grey")
lines(ages_lie_MC_Poisson[,1], depths_i) # Upper confidence level
lines(ages_lie_MC_Poisson[,3], depths_i) # Lower confidence level
text(1300,30,"Uncertainties strongly depend on the adopted Poisson model", col = "red", cex = 0.75)
text(1300,34,"Here, lambda is set at 20.", col = "red", cex = 0.7)
text(1300,37,"This means that sedimentation rates", col = "red", cex = 0.7)
text(1300,40,"are expected to change every 20 meters.", col = "red", cex = 0.7)
```



Adding cyclostratigraphy to the mix

We start from the same dataset of geochronologic “nails” as in the previous section. But now, we also consider some cyclostratigraphic information in the form of sedimentation rates (i.e. “springs”). This artificial cyclostratigraphic dataset stipulates that sedimentation rate is 4 cm/kyr between 0 - 55 m, 20 cm/kyr between 55 - 75 m, and 30 cm/kyr between 75 - 100 m. When this information is plotted alongside the geochronologic nails, it becomes clear that both datasets are largely congruent, but several pieces of additional information is provided. For example, the sedimenttion rate change at 55 m becomes clear, as well as the total duration of the dataset which is around 1400 kyr.

```r
Nsimulations = 10000
ages_lie_MonteCarlo_Poisson=matrix(data = NA, nrow = 101, ncol = Nsimulations )
depths_i=seq(0,100,1)

sedrate_cyclo = c(rep(4,55),rep(20,20), rep(30,40)) # These are the "dummy" cyclostratigraphy-derived sedimentation rates.
# We convert sedimentation rates into an age-depth model for visualisation purposes.
cyclo_age_model = c(0)
for (i in 2:101){
  cyclo_age_model[i]=cyclo_age_model[i-1]+sedrate_cyclo[i]
}
par(mar = c(4,4,1,1))
plot(dates$`Age (ka)`, dates$`Depth (m)`, ylim = c(100,0), xlim = c(0, 2500), xaxs = "i", yaxs = "i", col = "blue", pch = 19, xlab = "Age (ka)", ylab = "Depth (m)")
for (j in 1:6){segments(dates[j,2]-dates[j,3],dates[j,1],dates[j,2]+dates[j,3],dates[j,1], col = "blue", lwd = 2)
}
lines(cyclo_age_model, depths_i, col = "blue", lwd = 2)
```



Now let’s get serious: We use the same philosophy as in the previous section: using a uniform distribution to determine the depth-positions of the break-points. But, instead of using a uniform distribution to determine the ages of the break-points, we use the cyclostratigraphy-derived sedimentation rates. The age of the break-point is determined by using the cyclostratigraphy-derived sedimentation rate AT the break-point (`sedrate_cyclo[ceiling(temp1)`), and applying to the stratigraphy-interval between the break-point and the dated level BELOW the break-point (`(temp1-dates$`Depth (m)`[k])`). The “AT” and “BELOW” are user-dependent choices, and other choices could be made here: e.g. using average sedimentation rates rather than the sedimentation rate at the stratigraphic position of the breakpoint. But the “AT” and “BELOW” choice produced the most reasonable results in this artificial case.

```r
for (i in 1:Nsimulations){
  age_depth_MonteCarlo=c()
  depth_Poisson=c()
  age_Poisson=c()

  for (j in 1:6){age_depth_MonteCarlo[j]=rnorm(1, mean = dates$`Age (ka)`[j], sd = dates$`Uncertainty (2sigma, kyr)`[j]/2)}
  if (length(which(diff(age_depth_MonteCarlo)<0))>0){ages_lie_MonteCarlo_Poisson[,i] = rep(NA, 101) # Prior knowledge: Law of superposition
  } else{
    for (k in 1:6){
      if (k==6){thickness = 100-dates$`Depth (m)`[k]
        } else{thickness = dates$`Depth (m)`[k+1]-dates$`Depth (m)`[k]}
      lambda = thickness / 20
      N_breaks = rpois(1, lambda = lambda)
      # temp1 represents a depth-level at which a break-point is introduced. temp1 is sampled from a uniform distribution between the two adjacent dated levels
      if (k==6){temp1=runif(N_breaks, min = dates$`Depth (m)`[k], max = 100)
      } else{temp1=runif(N_breaks, min = dates$`Depth (m)`[k], max = dates$`Depth (m)`[k+1])}
      # temp2 represents the age of the introduced break-point. temp2 is calculated using information on sedimentation rate from cyclostratigraphy
      temp2 = dates$`Age (ka)`[k]+(temp1-dates$`Depth (m)`[k])*sedrate_cyclo[ceiling(temp1)]
      depth_Poisson = c(depth_Poisson, temp1)
      age_Poisson = c(age_Poisson, temp2)
    }
    # We concatenate the depths of dated levels and introduced break-points for this specific M-C simulation
    depths_sim = c(dates$`Depth (m)`, depth_Poisson)
    # We concatenate the ages of dated levels and introduced break-points for this specific M-C simulation
    ages_sim = c(age_depth_MonteCarlo, age_Poisson)
    # Temp is the age-depth model (through linear interpolation) of this specific Monte-Carlo simulation
    temp=approxExtrap(depths_sim, ages_sim , xout = depths_i)$y
    # ages_lie_MonteCarlo_Poisson is a large matrix, storing each individual age-depth simulation.
    ages_lie_MonteCarlo_Poisson[,i]=temp
  }
}
```

Now, we take the 2.5%, 50% and 97.5% quantiles of the 10 000 simulations to calculate the chronology and its confidence levels. Then, we plot things up. The new age-depth model has significantly reduced uncertainty levels, because the sedimentation constraint prevents the Monte Carlo procedure from unrestrainedly deviating from a straight line connecting two dated levels. Also, the new age-depth model clearly incorporates information that comes from cyclostratigraphy: (1) A widening of the confidence levels towards older ages around ~30 meters. (2) A widening of the confidence levels towards youger ages around ~50 meters. (3) A significantly younger age for the bottom of the core, in agreement with cyclostratigraphy. On the other hand, while sedimentation rates are used for constraining the slope of the chronology, it becomes clear that this approach adheres to the “nails” when it comes to determining the absolute age. This is exemplified around 70 m, where cyclostratigraphy and radio-isotopic dating are in disagreement on the age of this level: Here, the cyclo-integrated compound Poisson approach sticks to the radio-isotopic ages. Such a discrepancy between cyclostratigraphy and radioisotopic dating is either the result of an undetected hiatus in the cyclostratigraphy, or it is the result of a systematic error in the radio-isotopic dating. In the former case, the current behavior of the age-depth model is desirable (adhering to the dated levels). But, in case it is suspected that the latter is the case, the age-depth modeling approach should be modified by modifying / extending the uncertainties on those dated levels, so that they overlap with the cyclostratigraphy.

```r
ages_lie_MC_Poisson=matrix(data = NA, nrow = 101, ncol = 3)
colnames(ages_lie_MC_Poisson) <- c("2.5% CL", "Median Age", "97.5% CL")
for (k in 1:101){
  ages_lie_MC_Poisson[k,]=quantile(ages_lie_MonteCarlo_Poisson[k,], probs = c(0.025,0.5,0.975), na.rm = T)}

par(mar = c(4,4,1,1))
plot(dates$`Age (ka)`, dates$`Depth (m)`, ylim = c(100,0), xlim = c(0, 2500), xaxs = "i", yaxs = "i", col = "blue", pch = 19, xlab = "Age (ka)", ylab = "Depth (m)")
for (j in 1:6){segments(dates[j,2]-dates[j,3],dates[j,1],dates[j,2]+dates[j,3],dates[j,1], col = "blue", lwd = 2)
}
points(ages_lie_MC_Poisson[,2], depths_i, col = "grey")
lines(ages_lie_MC_Poisson[,1], depths_i) # Upper confidence level
lines(ages_lie_MC_Poisson[,3], depths_i) # Lower confidence level
text(1300,30,"This age-depth model (black) integrates information", col = "black", cex = 0.75)
text(1300,34,"from geochronologic nails (dates) and springs (cyclostratigraphy) (both in blue)", col = "black", cex = 0.7)
lines(cyclo_age_model, depths_i, col = "blue", lwd = 2)
```