Shared Chronologies: developing tools to improve age–depth modeling by incorporating common event layers among several sedimentary records

Rodrigo Vega[1] and Daniel Melnick[1]

[1]Instituto de Ciencias de la Tierra, Universidad Austral de Chile

November 21, 2022

Abstract

Sedimentary records provide an invaluable background for understanding of complex phenomena that vary within multiple spatio–temporal scales, such as climate and the seismic cycle. Understanding the latter in southern Chile has yielded motivation to develop new tools to deal with such records, in order to build a comprehensive peleoseismic catalog from them. The region inherited an extensive chain of lakes from the pleistocene glaciations, and a strong tephrochronological framework has been developed during the last two decades. Lake deposits have been extensively studied and shown to contain an incredibly sensitive paleoseismic record in the form of lacustrine turbidites. The task is thus to build the best possible chronology making use of all available data. Age–depth modeling is now routinely done by means of bayesian techniques, by using a sedimentation model as prior information and a set of age determinations as data. This approach provides the best results for any single record, but not necessarily for a set of records taken together. This is the goal of the shared chronologies approach, to build the tools for estimating the best chronologies for a set of sedimentary records given some chronological data for each and a set of shared events or stratigraphic markers. We use for this purpose the fact that two or more of such layers should yield age differences close to zero, within the general age uncertainty. This fact is incorporated to the model as prior information, along with the sedimentation model. The idea is clearly usable in a wide range of contexts, and for this reason we would like to share the implementation in a very early stage of development in order to incorporate feedback into design decisions that could affect extensibility and modularity, and to forge collaboration. This contribution shares an early experiment against a simulated data set, as well as the current R implementation and future plans.

CYCLO
MILLENNIUM NUCLEUS
THE SEISMIC CYCLE ALONG SUBDUCTION ZONES

Instituto de Ciencias de la Tierra
Universidad Austral de Chile

Ministerio de Economía, Fomento y Turismo
Gobierno de Chile

milenio
INICIATIVA CIENTÍFICA

Shared Chronologies:

Developing tools to improve age–depth modeling by incorporating common event layers among several sedimentary records

Rodrigo M. Vega[a b], Daniel Melnick[a b],

[a] Núcleo Milenio CYCLO

[b] Instituto de Ciencias de la Tierra, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile

Abstract

Sedimentary records provide an invaluable background for understanding of complex phenomena that vary within multiple spatio–temporal scales, such as climate and the seismic cycle. Understanding the latter in southern Chile has yielded motivation to develop new tools to deal with such records, in order to build a comprehensive peleoseismic catalog from them. The region inherited an extensive chain of lakes from the pleistocene glaciations, and a strong tephrochronological framework has been developed during the last two decades. Lake deposits have been extensively studied and shown to contain an incredibly sensitive paleoseismic record in the form of lacustrine turbidites. The task is thus to build the best possible chronology making use of all available data.

Age–depth modeling is now routinely done by means of bayesian techniques, by using a sedimentation model as *prior* information and a set of age determinations as *data*. This approach provides the best results for any single record, but not necessarily for a *set of records* taken together. This is the goal of the *shared chronologies* approach, to build the tools for estimating the best chronologies for a set of sedimentary records given some chronological data for each and a set of *shared events* or stratigraphic markers. We use for this purpose the fact that two or more of such layers should yield age differences close to zero, within the general age uncertainty. This fact is incorporated to the model as *prior* information, along with the sedimentation model.

Improving precision

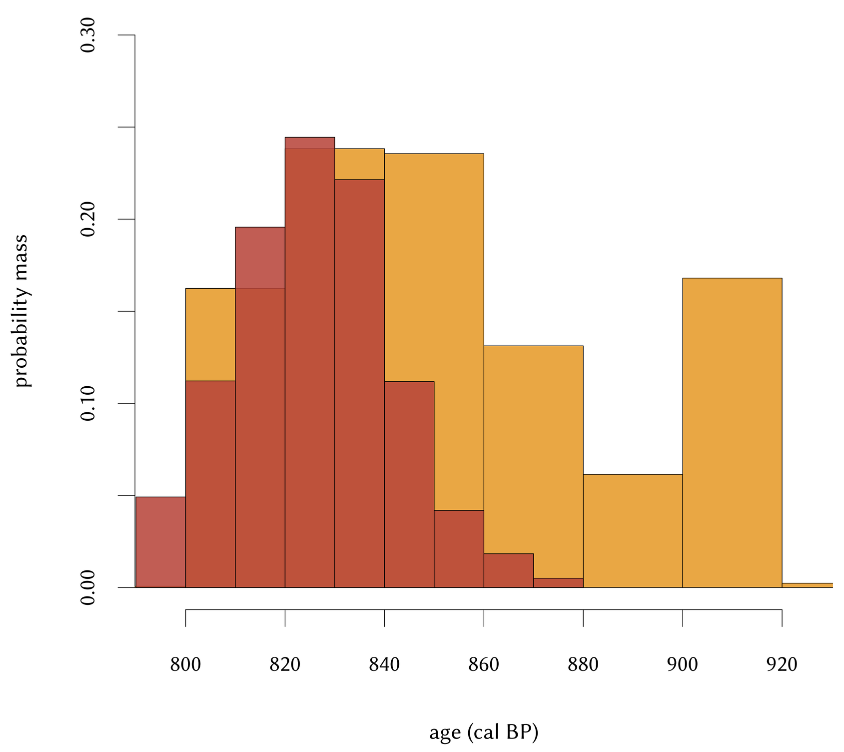

Fig. 1: Age distributions for a ^{14}C–dated layer within a set of related records (see below) as derived from: (•) the calibrated ^{14}C sample; and (•) the modeled *shared chronologies* for the whole set. The modeled age takes into account other dates in the same sedimentary sequence and their stratigraphic relations, thus it will only sample sets of ages in which the overlying ones are younger. By using the stratigraphic markers ammong sequences, all ages within the set of records are considered. This renders the best estimate considering all available data.

Outlook

These preliminary results demonstrate that the proposed procedure is practicable (the ~ 100.000 iterations presented here took less than 2 minutes to run in my laptop), yet the sampler's working is suboptimal and therefor it is not covering the whole parameter space. Prospectively, we intend to design the means to integrate with existing software for MCMC sampling, gaining the ability to choose from different strategies in a case-by-case basis. Specifically, we intend to develop a set of functions for model description within R, which can be passed to R–based MCMC algorithms but also to external software such as JAGS. Further extension will aim to include other kind of geological sequences and dating methods, such as raised beaches or coastal marsh tsunami deposits, and used to integrate such diverse records into a coherent *shared chronology*.

Modeling sets

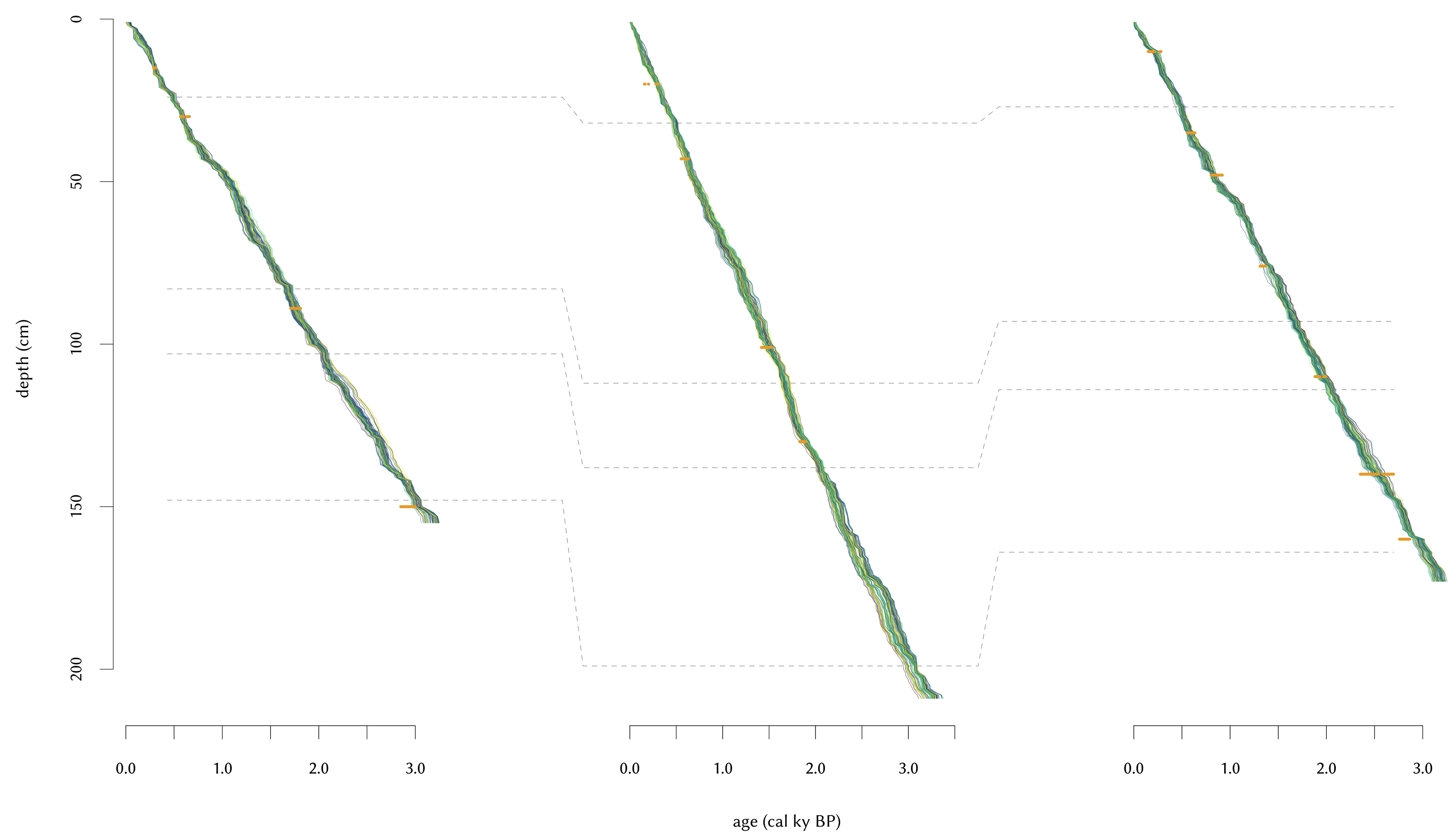

Fig. 2: *Shared chronologies* for three sintetic sediment cores. Lines are a sample of 100 possible age–depth trajectories from the posterior distribution. (•) bars represent higher (posterior) density intervals for ^{14}C dates. (•) dashed lines mark *event layers* shared ammong records. This sintetic data set was produced by simulating 3000 years of sediment accumulation for the three records. Four events produced instantaneous deposits within each record. ^{14}C ages were simulated from the SHcal13 calibartion curve [3] for the known time of deposition, and these dates where then treated as laboratory estimates. Posterior estimates are based on ~ 25000 samples drawn by Markov Chain Monte Carlo (MCMC) sampling. All the code is written in plain R [4], which leaves much space for optimization by porting the core routines to a low level language, suggesting it would be practicable to approach larger data sets.

Shared events

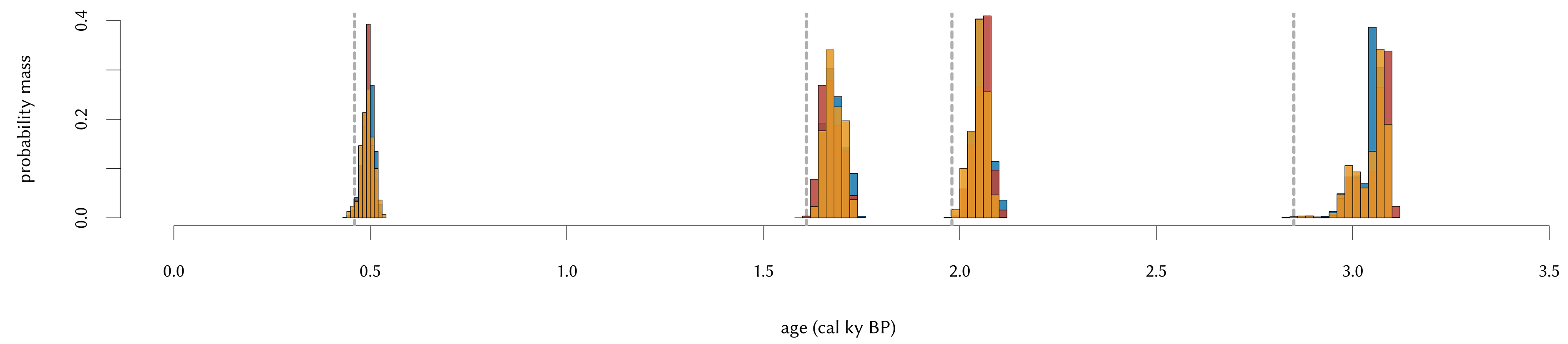

Fig. 3: Modeled age distributions for common event layers. (•) dashed lines mark the true age for each event. The fact that they generally fall outside the highest probability region is related to the poor density of dating data. Nevertheless, the close resemblance of modeled ages ammong records demonstrates that the knowledge about these layers being derived from a single event across sedimentary records has been successfully incorporated. This also means that dates from one record are being used to inform the chronology of every other record within the set.

A first approximation

We use a bare–bones version of the *rbacon* model[1], based solely on a regular increment in depth associated to a random accumulation rate, from which we can derive an age. The formulation is as follows:

Let $j = 1, 2, \ldots, m$ be a core index within a set of m cores, and let $i = 1, 2, \ldots, n_j$ be a depth index denoting the depths for which we will compute an age for core j. Depths d_{ij} will be equaly spaced and conditioned by a desired resolution c (in cm) such that $\Delta d = c$ is a constant for all cores. This is excluding any *event layer*. We model a set of ages x_{ij} as a function of depth and an accumulation rate $\theta_{ij} \sim \text{Gamma}(a_j, b_j)$ (in years cm^{-1}), as:

$$x_{ij} = x_{0j} + \sum_{i=1}^{i} \theta_{ij} c$$

Where x_{0j} is the jth core's extraction year. If we denote the probability of each particular accumulation rate within a random realization of the model as $p(\theta_{ij}|a_j, b_j)$, then the prior probability for the given set of sedimentation histories will be:

$$\mathbf{p}(\theta) \propto \prod_{j=1}^{m} \prod_{i=1}^{n_j} p(\theta_{ij}|a_j, b_j)$$

Let δ_k, $k = 1, 2, ..., l$ be a set of age differences between age estimates x_{ij} for known event layers ammong the m cores. We can represent our prior knowledge by stating that $\delta \sim N(\mu_\delta, \sigma_\delta)$, where $\mu_\delta = 0$ and σ_δ resembles the uncertainty related to the dating method. For this run we used $\mu_\delta = 0$ and $\sigma_\delta = 1$. If we denote the probability of any particular δ_k as $p(\delta_k|\mu_\delta, \sigma_\delta)$, then the prior for any iteration will be:

$$\mathbf{p}(\delta) \propto \prod_{k=1}^{l} p(\delta_k|\mu_\delta, \sigma_\delta)$$

The overall prior, considering our knowledge of sedimentation rates and shared event layers, would be given by:

$$\mathbf{p}(\theta, \delta) \propto \mathbf{p}(\theta)\mathbf{p}(\delta)$$

For ^{14}C calibration of independent samples, we follow Bronk Ramsey [2] and borrow the following method from OxCal: let $y_{rj} \sim N(\mu_y, \sigma_y)$ be the rth ^{14}C determination from core j, where $r = 1, \ldots, s_j$ is an age index, and each age is associated with a depth d_{rj}. Let ψ be the calibration curve, a function of age t providing an estimate for the ^{14}C age such that $f(t) \sim N(\mu_\psi, \sigma_\psi)$. The probability for any year t can be calculated by:

$$p(t|y, \psi) \propto \frac{\exp\left(-(\mu_y - \mu_\psi)^2 / 2(\sigma_y^2 + \sigma_\psi^2)\right)}{\sqrt{\sigma_y^2 + \sigma_\psi^2}}$$

If we denote the likelihood of any age estimate related to depths within d_{rj} as $p(x_{rj}|y_{rj}, \psi)$, the likelihood for any given set of chronologies will be given by:

$$\mathbf{p}(\mathbf{x}|\mathbf{y}, \psi) \propto \prod_{j=1}^{m} \prod_{r=1}^{s_j} p(x_{rj}|y_{rj}, \psi)$$

The posterior probability can be estimated from the product of the overall prior $\mathbf{p}(\theta, \delta)$ and the likelihood $\mathbf{p}(\mathbf{x}|\mathbf{y}, \psi)$, as:

$$\mathbf{p}(\theta, \delta|\mathbf{y}, \psi) \propto \mathbf{p}(\theta, \delta)\mathbf{p}(\mathbf{x}|\mathbf{y}, \psi)$$

Posterior densities are estimated by sampling with a Metropolis–Hastings random walk algorithm, with steps coming from a jump distribution $J \sim N(\mu = 0, \sigma = 0.8)$

References

[1] Blaauw, M., Christen, J. A., 09 2011. Flexible paleoclimate age-depth models using an autoregressive gamma process. Bayesian Analysis 6 (3), 457–474.

[2] Bronk Ramsey, C., 2001. Development of the radiocarbon calibration program. Radiocarbon 43 (2A), 355–363.

[3] Hogg, A., Hua, Q., Blackwell, P., Niu, M., Buck, C., Guilderson, T., Heaton, T., Palmer, J., Reimer, P., Reimer, R., Turney, C., Zimmerman, S., 2013. Shcal13 southern hemisphere calibration, 0–50,000 years cal bp. Radiocarbon 55 (4), 1889–1903.

[4] R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.