Geosci. Model Dev. Discuss., 7, C2404–C2411, 2014
www.geosci-model-dev-discuss.net/7/C2404/2014/

Geoscientific Model Development Discussions
Open Access

GMDD
7, C2404–C2411, 2014

Interactive comment on "IceChrono v1: a probabilistic model to compute a common and optimal chronology for several ice cores" *by* F. Parrenin

T. Heaton (Referee)

t.heaton@sheffield.ac.uk

Received and published: 24 November 2014

1 Paper Summary:

I should begin my review by stating that I am a statistician and not a geoscientist. As such, my knowledge of the current research into ice core dating is limited. I can therefore only review the paper on the methodology as presented rather than entirely in context with other work. In this regard, I felt that the current paper lacks sufficient explanation for a reader to fully understand the approach being taken and hence judge the appropriateness of the method. This is a shame as it makes the quality of the

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

approach difficult to judge, especially as I think, with careful consideration perhaps including some very simple illustrative examples and figures, it could be much improved and has the potential to be a very useful tool for the community.

From a methodological point of view, the paper would greatly benefit from significantly more explanation and justification (along with a more careful use of technical mathematical language) in order that readers could have confidence in using it themselves. I have tried to provide a possibility for this in my review below.

In addition, as commented by the other reviewers I don't feel that the current code, as available on Github is practically usable for a reader allowing them to reproduce the method or apply it to their own data. Currently it predominantly appears to just contain the code used to run the paper examples rather than acting as a resource for others to enter their own data and examples. A significant user manual with step by step instructions and help files is required if the author wishes others to implement their method.

2 General Comments:

I really found it difficult to understand the method. I describe below what I think the approach intends to do along with what could be a way of semi-formalising the model in a mathematical framework. I apologise if I have misunderstood.

2.1 Idea of paper

The true chronology of an ice core, i.e. age $\chi(z)$ at depth z, is a function of three unknown variables

$$f(z) = f(\mathbf{a}(z), \mathbf{I}(z), \boldsymbol{\tau}(z))$$

GMDD

7, C2404–C2411, 2014

where a is a vector of the accumulation (in m yr^{-1}), I is the lock-in depth, and τ the vertical thinning. To get from these variables to the age at depth z (i.e. the form of this function f) then you use the integral given in Eq. 2 solved by numerical methods. Since we are using a numerical approximation we only consider the value of these variables on a discrete pre-defined grid (quite a dense set of depths z_j) and so they each become a vector e.g. $\mathbf{a} = (a(z_1), a(z_2), \ldots, a(z_n))^T = (a_1, a_2, \ldots, a_n)^T$.

However, these vector valued variables are unknown and, to find our estimated chronology, we would like to estimate them based on:

- Some prior beliefs about their direct values
- Some externally gained information about the chronology: for example the age at a certain depth, the time elapsed between two depths, the synchroneity between two cores, ...

It seems that the approach could be set into a pseduo-Bayesian framework as described below.

2.1.1 Prior for unknown variables

We have some initial beliefs about plausible values of the unknown parameters which we have gained from an expert or previous experience. For example, we might think that a_i is probably close to a_i^b. Using a Bayesian framework we can formalise this belief by placing a initial prior on each of the unknown parameters e.g.

$$\log a_i/a_i^b \sim N(0, \sigma^2).$$

Such a prior suggests that the unknown parameter is centred around a_i^b. Imagine that we have lots of this information together denoted by $\pi(\mathbf{a}, \mathbf{I}(z), \boldsymbol{\tau})$

GMDD
7, C2404–C2411, 2014

2.1.2 External Information

In addition to these prior beliefs about the direct values of $\mathbf{a}(z), \mathbf{I}(z), \boldsymbol{\tau}(z)$ we also have extra information coming from other sources. This extra information can be quite varied, for example an externally found estimate of the age of the core at a specific depth, time elapsed between two depths, ...

Suppose we have one such external piece of information e.g. that the time elapsed between two depths is about T_1. If we knew the true values of $\mathbf{a}, \mathbf{I}, \boldsymbol{\tau}$, then we could work out the true time elapsed as the value $g_1(\mathbf{a}, \mathbf{I}(z), \boldsymbol{\tau})$ for a **known** $g_1(\cdot)$.

If we consider that the estimate T_1 has been observed subject to noise then we might model it as being centred around the true value as

$$T_1 \sim N(g_1(\mathbf{a}, \mathbf{I}(z), \boldsymbol{\tau}), \tau_1^2).$$

We can continue this idea analogously for each additional piece of external information, i.e. the external estimate is centred around a known function of the unknown parameters.

2.1.3 Combining the prior and the external information

Using Bayes theorem, we can combine the prior with the external information to come up with an updated estimate for the unknown parameters:

$$\pi(\mathbf{a}, \mathbf{I}(z), \boldsymbol{\tau} | T_1, T_2, \ldots) \propto \pi(\mathbf{a}, \mathbf{I}(z), \boldsymbol{\tau}) \times L(T_1, T_2, \ldots | \mathbf{a}, \mathbf{I}(z), \boldsymbol{\tau}).$$

The second term on the RHS is the likelihood of the external information.

GMDD
7, C2404–C2411, 2014

2.1.4 A MAP Estimate

If the prior and the likelihood are both normal then this equation simplifies to give e.g.

$$\pi(\mathbf{a}, \mathbf{I}(z), \boldsymbol{\tau} | T_1, T_2, \ldots) \propto \exp\left\{\sum \frac{(\log a_i - \log a_i^b)^2}{2\sigma_i^2)} + \ldots + \sum \frac{(T_i - g_i(\mathbf{a}, \mathbf{I}(z), \boldsymbol{\tau}))^2}{2\tau_i^2}\right\}$$

My interpretation is that the author chooses to find the values of the parameters which maximises this likelihood which just becomes an optimisation of a sum of squares. If this is correct then the formalisation of this is that the author has found the *maximum a posteriori* (MAP) estimate of the unknown parameters. These MAP estimates are then plugged in to the original $f()$ to create the final chronology.

2.1.5 Confidence intervals

The true posterior for the parameters should be a distribution rather than a single value. As such the final chronology created should be a predictive distribution. I am not sure how the variance that you estimate fits in with this - it seems to be a mix of frequentist and Bayesian statistics. Are you considering the posterior distribution to be normally distributed? Is this realistic or is it multimodal?

3 Specific Comments:

- The equations in Section 2 are not sufficiently explained. I think they would benefit from a picture and more justification to explain them. The terms in them are also not sufficiently defined as mentioned by the other reviewers. Specifically I had the following queries

GMDD

7, C2404–C2411, 2014

1. Firstly in Equation 1, what is D_k - the relative density of what compared with what? I also do not understand why this term is in the integral.

   I am not an ice core expert but it would seem to me that if ice at depth z'_k accumulated at the surface at a rate $\alpha(z'_k)\, m\, yr^{-1}$ and is then compressed at depth so that the *post-compression accumulation* rate will be approximately $\alpha(z'_k)\tau(z'_k)\, m\, yr^{-1}$ in a small depth interval from $(z'_k, z'_k + dz'_k)$. Hence the time elapsed in this interval of depth dz'_k will be

   $$\frac{dz'_k}{\alpha(z'_k)\tau(z'_k)} yr$$

   and the total time elapsed from the top of the core will then be

   $$\int \frac{1}{\alpha(z'_k)\tau(z'_k)} dz'_k.$$

   What does the relative density do?

2. Equation 2, and also Equation 4, are very hard to understand. They need to be explained and justified clearly, again I think a picture may help with this. Is z_k^{ie} a dummy variable over which you are integrating or actually a function of z'_k as in Eq. 4? If the latter what do the limits mean in Eq. 2?

- How high dimensional are the vectors $\mathbf{a}, \mathbf{I}(z), \boldsymbol{\tau}$? If high, then how well can one optimise and guarantee the space is searched fully - I would guess the function you maximise could be multimodal. In addition are there not several constraints on the values of the variables, for example the thinning can't be larger than 1. Presumably it's also unrealistic for the values to change rapidly over short depths. How is this accounted for?

- How does a user decide what external information to include? How do you select the covariances and variances for your likelihoods? How could a user decide upon this too?

GMDD

7, C2404–C2411, 2014

- It is not clear what is the difference between IceChrono and DatIce. What do you mean by computation numerically/analytically on pg6822? How much of an advance is IceChrono?
- Care needs to be taken in any conclusions drawn from DatIce and IceChrono giving the same results. Currently it reads as though you are saying that validates the method. Since they seem very similar techniques, it does not say much about the quality of method only that the code seems to do something similar. You should remove this comment since it is open to misinterpretation as being a statement about the quality of the method.

4 Technical Points

- Section 2.2 - what is meant by background information? Requires more formal definition. Also in equations 5, 6 + 7 no probability has been defined by this point and yet this begins talking about transforming densities.
- Section 2.3 is currently unclear to the reader and uses a lot of notation previously undefined e.g. node, correction function, . . .
- Section 2.4. - pdf for what? Again, what the background information actually is is not sufficiently defined. Also when was the change made to multiple cores since this has not been mentioned previously.
- What are the correlation functions on pg 6819? Also incorrect statistical language here — confidence are a range of values and not the standard deviation.
- pg6820 - where have observations come in previously? Not explained sufficiently.
- pg6823 - Why does difficulty of invertability mean that they are a wrong description of knowledge? Also next sentence unclear. How can correlation matrices be

GMDD

7, C2404–C2411, 2014

triangular? Do you mean band-limited or just diagonal? This continues into the example.

- pg6825 - where is z in equation 23? What is ζ in equation 24?
- Appendix A is not very informative and highly repetitive. Space would be better spent explaining the justification behind the method.

Interactive comment on Geosci. Model Dev. Discuss., 7, 6811, 2014.