

# T. Heaton

First of all thank you very much for your constructive review, which greatly helped improve the clarity of the manuscript.

## 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 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.

Based on your review, we tried to improve the clarity of the manuscript as well as of the code documentation. The code documentation has been independently reviewed and corrected.

## 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.

You have almost perfectly understood the method! However, we have been working on improving the clarity of our manuscript in the revised version.

### 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

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

where  $\mathbf{a}$  is a vector of the accumulation (in  $\text{m yr}^{-1}$ ),  $\mathbf{I}$  is the lock-in depth, and  $\boldsymbol{\tau}$  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.  $\mathbf{a} = (\mathbf{a}(z_1), \mathbf{a}(z_2), \dots, \mathbf{a}(z_n))^T = (a_1, a_2, \dots, a_n)^T$ . 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 synchronicity between two cores, ...

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

We added a “method summary” sub-section at the beginning of the “method” section, partially inspired by your review, to give the reader a rough general idea of the method before going into the detail.

### 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  $\pi(\mathbf{a}, \mathbf{I}(z), \boldsymbol{\tau})$

Your equation is only partially correct since we allow the prior estimates on, e.g. the  $a_i$ , to be correlated. Other than that your reformulation is correct.

### 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.

You are correct except that, as for the prior estimates, we allow the observations to have correlated errors (within the same type of observations).

### 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:

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

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

It is perfectly right that we use the Bayes theorem. However, we tried to express the problem using

a Bayesian framework as you did in your review but we came up with something which was in our opinion less clear than our initial formulation, which is more in-line with the framework presented by A. Tarantola. So we decided to keep our initial framework, although we tried to improve its clarity.

#### 2.1.4 A MAP Estimate

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

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

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.

Again your equation is correct except that we used a more general approach by allowing the prior estimates and observations to have correlated errors.

#### 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?

Yes, we do consider the posterior distribution to be normally distributed. Because the prior estimates and observations have normal distributions, this is also true for the posterior if the model is linear. So our approach assumes that the model is 'not too far' from a linear model, as we acknowledged in the discussion.

### 3 Specific Comments:

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  $\alpha(z'_k) \text{ m yr}^{-1}$  and is then compressed at depth so that the post-compression accumulation rate will be approximately  $\alpha(z'_k)\tau(z'_k) \text{ m yr}^{-1}$  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

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

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?

$D_k$  is the relative density of the material with respect to pure ice. It is equal to  $1-\rho$  where  $\rho$  is the porosity.  $D_k$  enters in the ice age equation because  $a_k$ , the accumulation rate, is expressed in meters ice-equivalent per year and because  $\tau_k$  is the thinning function with respect to this ice-equivalent accumulation rate. This is a convention in glaciology, used because accumulation rates are

expressed in water or ice equivalent per year and because ice sheet models usually consider only pure ice (they do not represent the densification process). We tried to make this clearer in the text:

*Note that in our convention inherited from glaciology, accumulation  $a_k$  is expressed in ice-equivalent per year and that the thinning function  $\tau_k$  is expressed with respect to this ice-equivalent accumulation rate, hence the appearance of  $D_k$  in equation (1). We used this convention because most of the ice flow models, which give prior estimates on the thinning function (see below), only consider pure ice, i.e. they assume the snow is instantaneously densified into ice when it falls on the surface of the ice sheet.*

- 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?

We have now added Figure 2 which explains Eq. (2) (now renamed Eq. (3) in the revised version). Again,  $z_k^{ie}$  is a convention inherited from ice flow models which consider that ice sheets contain only pure ice.  $z'_k$  is indeed a dummy variable over which we integrate (we added a 'prime' to make this clearer).

- How high dimensional are the vectors  $a$ ,  $I(z)$ ,  $\tau$ ? 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?

There are several questions here. Yes, the  $X$  vector can be high dimensional: from  $\sim 100$  to several thousands. It is difficult to guarantee the space is fully searched. We would need some in-depth mathematical analysis of the problem which is clearly outside the scope of the current manuscript. The only thing we could do here is to start the optimization problem with different values of  $X_0$ , all in agreement with the prior information, and check that the optimal solutions reached are the same. This is something we did in the revised version of the manuscript.

The 'glaciological' constraints for the variables are comprised in the prior information. For example, the uncertainty of the prior thinning decrease to 0 when the depth tends to zero. Also, the covariance matrix on the priors ensures that the values of the correction functions on  $a$ ,  $l$  and  $\tau$  cannot vary abruptly on short depth or time intervals. We added the following sentence to make this clear: *These three terms  $J^a_k$ ,  $J^l_k$  and  $J^\tau_k$  bring the 'glaciological constraints' of the problem given by the sedimentation models. For example, they ensure that the optimum values for  $a$ ,  $l$  and  $\tau$  will be close to the modeled values and also that their rates of change with respect to depth will be close to the rates of change of the modeled values. That means the model giving the prior scenario for the accumulation, the LID and the vertical thinning function should have a quantified error.*

- 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?

Only an in-depth analysis of the methods used to determine the external information can help the user decide which variance vectors and correlation matrices should be associated to each type of external information. This is a common problem in the field of data assimilation in geophysics. We added the following sentence:

*The determination of the standard deviation vectors and of the correlation matrices for the prior and for the observations can be a difficult problem which requires an in-depth analysis of the methods used to determine the prior and the observations.*

- 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?

The comparison with Datice is done in details in sections 4.2 and 4.4. A numerical computation of a Jacobian just means that you perturb alternatively each component of the variable vector  $X$  and you see how your forward model is modified. This requires to run  $\dim(X)$  times the forward model. Alternatively, you can derive analytical expressions of the Jacobian as a function of  $X$  and use these expressions for the numerical computation of the Jacobian. The relevant paragraph has been modified as:

*Sixth, in IceChrono v1, the Jacobian of the model is computed numerically while it is computed analytically in Datice. This Jacobian is needed by the minimizer to find the optimal solution  $X^{opt}$  and its uncertainty  $C^X$ . When the optimal solution is found, it also allows to evaluate the uncertainty of the model  $C^G$  through equation (20). In Datice, analytical expressions of the Jakobian with respect to  $X$  have been derived and these expressions are used to numerically compute the Jakobian for a particular  $X$ . In IceChrono, each component of the  $X$  vector are alternatively perturbed and the forward model  $G$  is run to evaluate how the model  $G(X)$  is modified. In other words, the Jacobian is evaluated by a finite difference approach. While a numerical computation of the Jacobian leads to a slower computation time, it leads to a more flexible use of the model since if one modifies the formulation of the cost function, one does not need to derive again analytical expressions for the Jacobian.*

- 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.

We did not write it is a confirmation of the *\*methods\**, but we wrote it is a confirmation of the *\*codes\**, which uses different numerical schemes. We think it is important that these codes have been confirmed, which mean they probably don't contain any obvious bug. Of course the methods have the same spirit so they might be biased the same way. We now talk about the current limitations of the method in a dedicated sub-section:

*IceChrono1 is already a useful tool to define a common and optimized chronology for several ice cores all together. It however has several limitations that we will discuss below.*

*1) All uncertainties are assumed Gaussian. While the Gaussian probability density functions are the most often encountered in Science, it is not always appropriate. For example, the volcanic synchronization of ice cores (e.g., Parrenin et al., 2012b) is ambiguous: a volcanic peak in one ice core can sometimes be synchronized to several other volcanic peaks in the other ice core. IceChrono1 therefore assumes that the features recognition has been performed without ambiguity and that the only uncertainty remaining is the exact match of the two features within their durations.*

*2) The forward dating model is assumed to be “almost linear” in the plausible vicinity of the solution. Further developments would be necessary to diagnose if this assumption is justified.*

*3) IceChrono1 is not appropriate for optimizing the chronology of many ice cores at a very high resolution: the computation time would be too high, due to the numerical finite differences approach to evaluate the Jacobian matrices. In practice this is not a problem as a high resolution is only necessary for recent periods. If, in the future, the need for a more efficient dating model appears, we could develop an analytical gradient for the forward model, as it is done in the Datice software.*

*4) The age observations on different cores or on different ice cores pairs are not always*

*independent (because the dating methods are the same). We might allow in the future to take into account correlation of observations in-between ice cores and ice cores pairs.*

*5) IceChrono1 requires the need for external sedimentation models to infer prior scenarios and their uncertainties. In practice, these sedimentation models also need to be optimized using age observations. In a next step, we will incorporate sedimentation models directly into IceChrono. The uncertainties of these sedimentation models could be inferred automatically by comparing them to the age observations.*

*6) The stratigraphic observations in-between the different ice cores need to be derived externally and imported as stratigraphic observations into IceChrono1. This step also requires some prior knowledge about the sedimentation process. Therefore, it would be best to incorporate it directly into the IceChrono software. Automatic methods for ice core synchronization would eliminate this step which is most of the time done visually, in a subjective, fastidious and undocumented way.*

*7) The layer counting dating of ice core needs to be done externally and imported as intervals with known durations observations into IceChrono1. Again, this step also requires prior knowledge about the sedimentation process (e.g., the typical annual layer thickness). Therefore, it would be best to incorporate it directly into the IceChrono software. An automatic method for layer counting has already been proposed (Winstrup et al., 2012).*

*8) The definition of realistic prior correlation matrices is a difficult issue which will be dealt with in details in future studies.*

#### **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.

*“Background” has now been renamed into “prior”. Some scientists use “background”, some scientists use “prior”. All this is the same. We now introduce the term “prior” in the “method summary” sub-section.*

- Section 2.3 is currently unclear to the reader and uses a lot of notation previously undefined e.g. node, correction function, . . .

*We think the “nodes” of a numerical mesh is a very classical term and does not need to be defined. The correction functions are defined in the previous paragraph. Other than that, we tried to clarify as much as possible this sub-section.*

- 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.

*Again, the prior information is now defined in the method summary sub-section.*

- What are the correlation functions on pg 6819? Also incorrect statistical language here —

confidence are a range of values and not the standard deviation.

These are the correlation matrices of the prior, we clarified this point. “Confidence interval” was a mistake and has been changed to “standard deviation”.

- pg6820 - where have observations come in previously? Not explained sufficiently.

Again, observations are now introduced into the method summary sub-section.

- pg6823 - Why does difficulty of invertability mean that they are a wrong description of knowledge? Also next sentence unclear. How can correlation matrices be triangular? Do you mean band-limited or just diagonal? This continues into the example.

This paragraph has been modified as follow:

“Only one aspect is modified: in AICC2012, the prior correlation matrices are supposed to have an across-diagonal Gaussian decrease in amplitude. We believe that this Gaussian shape leads to a too high correlation for neighboring points. As a consequence, these matrices are numerically very difficult to invert (with a poor conditional state). We therefore opt for correlation matrices with an across-diagonal linear decrease in amplitude. The definition of realistic prior correlation matrices is a difficult task which will be dealt with in details in future studies.”

- pg6825 - where is  $z$  in equation 23? What is  $\zeta$  in equation 24?

The Berkner example has been removed.

- Appendix A is not very informative and highly repetitive. Space would be better spent explaining the justification behind the method.

We kept Appendix A for reference, since it was asked by the editor at the time of submission. We agree it is a repetitive section but its reading is optional.