Eric Wolff

First of all thank you very much for your constructive review.

This paper describes and presents a new model for computing ice core age scales. As the author explains, it has a very similar philosophy and methodology to the already-published DATICE model. The paper claims that it has some (especially computing) advantages compared to DATICE. It presents two examples of the use of the model, one confirming that it obtains similar results to DATICE for the same experiment, and the other producing a first age model for the Berkner Island core.
In general this is a solid piece of work that serves the community by making the code freely available in a (relatively) user-friendly format. Age modelling for ice cores is really important so there is no doubt the work is significant. However there are some issues that need to be dealt with:
a) Clarifying some of the equations and inputs
b) Making sure the code is clearly available on a formal and stable platform
c) The author's statements on the performance relative to DATICE should be discussed by DATICE people
d) There are some issues with the new Berkner age model that need to be explored.

Abstract, lines 2-10 consists of a single very long sentence, and one that seems more suited to the introduction than the abstract. I suggest the abstract needs more thought to ensure it truly explains what is new in this paper.

Abstract has been reworked:

Polar ice cores provide exceptional archives of past environmental conditions. The dating of ice cores and the estimation of the age scale uncertainty are essential to interpret the climate and environmental records that they contain. It is however a complex problem which involves different methods. Here, we present IceChrono1, a new probabilistic model integrating various sources of chronological information to produce a common and optimized chronology for several ice cores, as well as its uncertainty. IceChrono1 is based on the inversion of three quantities: the surface accumulation rate, the Lock-In Depth (LID) of air bubbles and the vertical thinning function. The chronological information integrated into the model are: models of the sedimentation process (accumulation of snow, densification of snow into ice and air trapping, ice flow), ice and air dated horizons, ice and air depth intervals with known durations, Δdepth observations (depth shift between synchronous events recorded in the ice and in the air) and finally air and ice stratigraphic links in between ice cores. The optimization is formulated as a least squares problem, implying that all densities of probabilities are assumed to be Gaussian. It is numerically solved using the Levenberg-Marquardt algorithm and a numerical evaluation of the model's Jacobian. IceChrono follows an approach similar to that of the Datice model which was recently used to produce the AICC2012 chronology for 4 Antarctic ice cores and 1 Greenland ice core. IceChrono1 provides improvements and simplifications with respect to Datice from the mathematical, numerical and programming point of views. The capabilities of IceChrono is demonstrated on a case study similar to the AICC2012 dating experiment. We find results similar to those of Datice, within a few centuries, which is a confirmation of both IceChrono and Datice codes. We also test new functionalities with respect to the original version of Datice: observations as ice intervals with known durations, correlated observations, observations as gas intervals with known durations and observations as mix ice-air stratigraphic links. IceChrono1 is freely available under the GPL v3 open source license.

Abstract, Line 10 “here I propose” seems a slightly awkward wording. Maybe “Here I present”.

Corrected.

Abstract, Line 15 “differences from” better than “differences on”.

Corrected.

Page 6813, line 25-27. It should also be mentioned here that this method requires assumptions, not always fully acknowledged, about synchroneity between changes of similar appearance in different archives.

You are right that such assumptions exists, however, this is not systematically the case. For example, the synchronization of a volcanic ash in an ice core to a dated lava flow can be based on the chemical signatures.
We prefer staying general and not mentioning this assumption in this particular section, but we mention this now in the discussion, in section 4.5. on the “current limitations of IceChrono and possible perspectives” .

Page 6814, line 25. AICC actually is presented as “Chronology” not “Chronologies”.

Corrected.

Section 2.1. There are several aspects I feel could be clearer here.

a) In line 7 “initial surface accumulation” – I don’t really follow the purpose of the word “initial”, surely each layer has only one accumulation rate, and the word “surface” already clarifies that you mean the one it had when it was laid down.

"surface" has been removed but we kept “initial”. “Surface accumulation rate” of a given ice particle is ambiguous since it could mean the present-day accumulation rate at the vertical of the ice particle.

b) It might be worth being precise that in all the integrations, zero is at the surface (since many integrations in glaciology actually treat the bed as zero).

Precision has been made.

c) Line 16: “relative density”: relative to what? If you use water equivalent accumulation rates, then your densities are relative to that of water, and I think that, for eq 4 and 2 to be right this has to be so. However in the files online it looks as if acc rates may be in metres ice equivalent, and in that case for eq 1 to be right the densities would be relative to ice. In any case this needs to be made clear and consistent.

Everything is relative to ice here : *D* is a relative density (1 minus the porosity), *a* is expressed in m ice-equivalent per year. We tried to make this clear in the revised manuscript.

d) And indeed you should be clear what your acc rates are. Many readers would assume they would have units kg mˆ-2 timeˆ-1. But actually they are in m (water or ice) equivalent depth timeˆ-1.

Clarified.

As a general comment, at different times in the text you use "I" (eg p 6817, line 25), "we/us" (6818, line 2), or "one" (6817, line 16). You should standardise.

"I"s have been replaced by "we"s. "one" has a different meaning.

Section 2.8. Probably it was intended but just to avoid doubt, the model code should be a supplement to the paper. I appreciate that the code is available at github. However (a) it looked like a rather confusing set of files; (b) it is not clear which set of files corresponds to the model version presented and tested in the paper; (c) the software archive should be permanent and secure, and with great respect I don't think a personal area on even the most secure server meets that bill. I therefore suggest that the files referring to this version of the paper should be archived at GMD as a supplement.

Agree, the code will also be made available as a supplement.

Section 2.7 (comparison to Datice), also page 6826, line 25 claiming superiority of Ice Chrono. I am unable to judge these sections: one of the developers of DATICE should be strongly urged to comment.

We now give the precision that the comparison is based on the published version of Datice (Lemieux-Dudon et al., QSR, 2010). Also the comparison that we now provide between Datice and IceChrono has been discussed with the developers of DATICE.

Page 6824, line 14: can the author comment what is occurring in the DATICE code that causes it not to respect the confidence interval at the tie point? Knowing this would be very helpful in judging the way the models work.

We now use the raw confidence intervals as produced by Datice and not the AICC2012 confidence intervals. The raw Datice confidence interval does respect the observations confidence interval during the Laschamp event.

Page 6825, Berkner age model. By synchronising on water isotopes the author is assuming synchronous climate changes between Berkner and East Antarctica. This assumption (which precludes testing phase leads and lags) should be made absolutely explicit.

Following the suggestion made by the other reviewers, we have now removed the Berkner experiment in the revised version.

Also re Berkner expt: I wonder why the author has done this in such a way that the Berkner age scale has only the synchro error. It would surely have been straightforward to in addition simply run AICC2012 again but including Berkner, in order to get a realistic uncertainty on the Berkner age model. As there would have been no new absolute age information, I assume that the ages would not have altered for the other cores, but an uncertainty for Berkner would have emerged.

Dito.

Anonymous Referee n°2

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

General comments: Parrenin presents a probabilistic model for computing multiple ice-core chronologies simultaneously. Named IceChrono v1, this model is essentially the same as the Datice model except that the optimization is done numerically rather than analytically. While this slows the computation time, it simplifies the code and makes the program accessible to more people.

IceChrono v1 is publicly available on github, a useful site that provides good version control. There are also a few additional updates of Datice with respect to allowable age constraints. The release of IceChrono v1 is potentially quite useful as this dating method can now be used by other researchers. It could also provide insight into the dating method itself as adoption of the Datice chronologies has been hampered by poor explanation of the methodology. Unfortunately, the paper suffers, as the Datice papers did before it, from mathematical descriptions that are not intuitive and from incomplete diagnostics of the resulting timescale. The IceChrono solver is shown to work, but the improvements to Datice are not tested. Evaluating this paper has been particularly difficult. Should I limit my review to just whether IceChrono v1 functions as described in the paper? Or should the paper be expected to address the limitations that have been identified with Datice and also exist with IceChrono?

We are interested in knowing the limitations of IceChrono. we do not promise to fix everything in this first version but we can at least discuss the current limitations.

The Datice chronologies (first Lemieux-Dudon et al., 2010 and then AICC2012) have not been universally accepted as the best timescales for the various ice cores, and have caused widespread timescale confusion. (A recent paper in Nature by Weber et al. 2014 is a good example of this confusion). I have structured this review in two parts: first, I focus on just the description of IceChrono and second, discuss limitations in the overall Datice/IceCrhono framework. I will let the editor decide if the manuscript should address the more general Datice/IceChrono comments. First, general comments about the manuscript:
1) Remove all discussion of the Berkner Island ice core. The underlying data for the timescale have not been published so there is no ability to evaluate the inputs to the timescale. The Berkner Island core is challenging to date due to the thin, brittle ice and the first timescale for the cores needs to be accompanied by a full description of the methods used.

We agree. This application has been removed.

2) The additional age and uncertainty constraints described in Section 3.1 (particularly the third and fourth points) do not appear to be used in either example. The utility of these additions need to be described and their implementation needs to be evaluated.

Did you mean "section 2.7" ? We now added 4 new test experiments for:

- the use of ice intervals with known durations
- the use of correlated observations
- the use of gas intervals with known durations
- the use of mix ice-gas stratigraphic links

3) Describe the mathematical equations in plain English. For instance, P6817 L2-8 discusses transforming “so-called jeffreys variables” into “Cartesian variables” but does not discuss what the benefit is. This is one example but more description is necessary throughout.

We tried to explain the different equations and their benefit in plain English:

This change of variable allows to transform Jeffreys variables into Cartesian variables (Tarantola, 2005) so as to express our problem into a least-squares problem and will allow us to reduce the number of variables to be inverted (see below).

4) The language in the paper needs significant improvement. I have not tried to edit the writing but there are many instances of basic grammatical errors. For instance, the subject and verb don't agree in the very first sentence of the abstract: Polar ice cores provide, not provides.

The manuscript has been entirely reviewed by several people so we hope that now the language is improved.

5) The references are deficient. From the introduction, one gets the impression that only Europeans (and particularly the French) have done anything noteworthy with ice cores.

The introduction has been entirely rewritten:

Polar ice cores provide continuous records of key past features of the climate and the environment, with local, regional and global relevance (e.g., Ahmed et al., 2013; EPICA community members, 2004; NorthGRIP project members, 2004; WAIS Divide Project Members, 2013). Tracers of polar climate (e.g., Jouzel et al., 2007), ice sheet topography (NEEM community Members, 2013), water cycle (Schoenemann et al., 2014; e.g., Stenni et al., 2010; Winkler et al., 2012), aerosol deposition (e.g., Lambert et al., 2012; Wolff et al., 2006) and global atmospheric composition (e.g., Ahn and Brook, 2014; Loulergue et al., 2008; Marcott et al., 2014) measured in ice cores unveil sequences of events on seasonal to glacial-interglacial time scales.
However, the prior to interpretation of polar ice core records is the complex task of building two robust time-depth relationships, one for the tracers measured in the ice phase (e.g. water isotopes, particulates and chemical impurities) and one for those measured in the air phase (e.g. greenhouse gas concentration, isotopic composition of gases). The firn, where snow is gradually compacted into ice, constitutes the upper 50-120 m part of ice sheets. The firn is permeable and air is only locked-in at its base, at a depth level called the Lock-In Depth (LID). As a result, the entrapped air is always younger than the surrounding ice at any depth level. Through gravitational fractionation processes, LID is closely related to the isotopic composition of δ15N of N2 in air bubbles data (e.g., Buizert et al., 2015; Goujon et al., 2003; Parrenin et al., 2012a; Schwander et al., 1993). The temporal evolution of the age difference between ice and air at a given depth must therefore be estimated using firn densification modeling and air δ15N. This age difference is essential for clarifying the exact timing between changes in atmospheric CO2 concentration and Antarctic surface temperature during deglaciations (Caillon et al., 2003; e.g., Landais et al., 2013; Monnin et al., 2001; Parrenin et al., 2013; Pedro et al., 2011). However, glacial-interglacial Antarctic firn changes remain poorly understood (e.g., Capron et al., 2013).
Several strategies have been developed to build ice and gas chronologies. We briefly describe these methods, their strengths and caveats hereafter:
1) Annual layer counting (e.g., Rasmussen et al., 2006; WAIS Divide Project Members, 2013; Winstrup et al., 2012). Only applicable when accumulation rates are sufficiently high to make this annual layer identification possible, this method provides accurate estimates of event durations and small uncertainties on the absolute age of the upper ice sections. However, the cumulative nature of the errors, associated with the increasing number of counted layers, leads to a decrease of the accuracy of absolute age with depth. For instance, the GICC05 (Greenland Ice Core Chronology 2005) composite timescale for Greenland ice cores (Rasmussen et al., 2006; Seierstad et al., 2014; Svensson et al., 2008), is associated with a maximum counting error of only 45 years at ~8.2 ka B1950 (Before 1950 C.E.). This error increases progressively with depth, reaching more than 2500 years at ~60 ka B1950. Annual layer counting techniques cannot be applied when the annual layer thickness is too small to be resolved visually, e.g. in ice cores from central East Antarctica.

2) Use of absolute age markers in ice cores. Well-dated tephra layers identified in ice cores during the last millennia provide precious constraints (e.g., Sigl et al., 2014). Beyond that period, absolute age markers are very scarce. The links between 10Be peaks and well-dated magnetic events (Raisbeck et al., 2007) have provided an age marker for the Laschamp event (Singer et al., 2009). Promising results have recently been obtained using radiochronologic dating tools (Aciego et al., 2011; Buizert et al., 2014; Dunbar et al., 2008).
3) Orbital dating in ice cores. Because there are few absolute constraints in ice cores beyond 60 ka B1950 (limit for the layer counting in the NGRIP ice core), orbital tuning is the most effective method to provide chronological constraints on ice core deepest sections. In the first orbital dating exercises, tie points were determined from the tuning of water isotopic records on insolation curves (e.g., Parrenin et al., 2004), which limits further investigations of polar climate relationships with orbital forcing. More recent chronologies tried to circumvent this assumption and focused on non-climatic orbital markers. Three complementary tracers are currently used: the δ18O of atmospheric O2 (δ18Oatm) (e.g., Bender et al., 1994; Dreyfus et al., 2007), δO2/N2 (e.g., Bender, 2002; Kawamura et al., 2007; Suwa and Bender, 2008) and the total air content (e.g., Raynaud et al., 2007). While the link between δ18Oatm and precession is explained by variations in the water cycle of the low latitudes, relationships between δO2/N2, air content and local summer insolation are understood to arise from changes in the surface snow energy budget influencing its metamorphism. Without a precise understanding of mechanisms linking these tracers to their respective orbital targets, the associated uncertainties remain large, 6 ka for δ18Oatm and 3 to 4 ka for δO2/N2 and air content (Bazin et al., 2013, 2014; Landais et al., 2012).
4) Ice core record synchronization. Inter-ice core matching exercises are undertaken to transfer absolute or orbital dating information from one ice core to another one. It generally relies on the global synchroneity of changes in atmospheric composition (CO2, CH4 concentration, and δ18Oatm) (Bender et al., 1994; Blunier and Brook, 2001; Monnin et al., 2004), the identification of volcanic sulfate spikes within a given area (Parrenin et al., 2012b; Severi et al., 2007) or the hypothesis of synchronous regional deposition of aerosols recorded as ice impurities (Seierstad et al., 2014). In the first case, limitations are associated with the smooting of atmospheric composition changes through firn air diffusion. In the second case, mismatches may arise through incorrect identification of events in different ice cores.
5) Correlation with other well-dated climatic records. In some cases, high-resolution calcite δ18O records and precise U/Th dates on speleothems have been used to adjust ice core chronologies (Barker et al., 2011; Buizert et al., 2015; Parrenin et al., 2007a). Pinning ice core and speleothem records is attractive to reduce absolute age uncertainties especially during past abrupt climatic events of glacial periods. However, these exercises rely on the assumption of simultaneous abrupt climatic changes recorded in ice core (e.g. water isotopes, CH4) and low latitudes speleothem δ18O records (mostly reflecting changes in regional atmospheric water cycle). A main limitation of this method lies in the validity of this assumption.
6) Modeling of the sedimentation process: snow accumulation, snow densification into ice, air bubbles trapping and ice flow (Goujon et al., 2003; Huybrechts et al., 2007; Johnsen et al., 2001). Glaciological modeling provides a chronology derived from the estimate of the annual layer thickness, and therefore, leads to more realistic event durations when the accumulation history and thinning function are well constrained. A side product of glaciological modelling is the quantification of changes in surface accumulation rates, and the quantification of the initial geographical origin of ice. These additional informations are necessary to convert measurements of concentrations of chemical species in ice cores into deposition fluxes, and to correct ice core records from upstream origin effects (e.g., EPICA community members, 2006; Röthlisberger et al., 2008). Caveats are caused by unknown parameters of such glaciological models, such as amplitude of accumulation change between glacial and interglacial periods, the basal melting or the vertical velocity profile, which have a growing influence at depth.
A common and optimal chronology for several ice cores can be built through the combination of several of these methods in the frame of a probabilistic approach. The first attempts used absolute

and orbitally-tuned age markers along one ice core to constrain the unknown parameters of an ice flow model (e.g., Parrenin et al., 2001, 2004; Petit et al., 1999). This method had however several limitations. First, the uncertainties associated with the ice flow model could not be taken into account, resulting in underestimated uncertainties. Second, the stratigraphic links between ice cores were not exploited, each ice core was dated separately resulting in inconsistent chronologies. A new probabilistic approach based on a Bayesian framework was subsequently introduced. The first tool, Datice, was developed by Lemieux-Dudon et al. (2010a, 2010b). It introduced modeling errors on three canonical glaciological quantities of the dating problem: the accumulation rate, the LID of air bubbles and the vertical thinning function (i.e. ratio between the in-situ annual layer vertical thickness on its initial surface thickness). This method starts from prior (also called "background") scenario for the three glaciological parameters corresponding to an prior chronology for each ice cores. These scenarios, deduced from a modeling of the sedimentation process, are associated with an uncertainty related to the degree of confidence in these prior scenarios. A minimization approach is then applied to find the best compromise between the prior chronological information for each ice core as well as absolute and relative age markers in the ice and in the air phases. This approach has been validated through the Datice tool and applied to build the Antarctic Ice Cores Chronology 2012 (AICC2012), producing coherent ice and air timescales for five different ice cores (Bazin et al., 2013; Veres et al., 2013): EPICA Dome C (EDC), Vostok (VK), Talos Dome (TALDICE), EPICA Dronning Maul Land (EDML) and NorthGRIP (NGRIP). Further developments of Datice were performe to incorporate additional dating constraints such as the depth intervals with known durations and correlation of errors (Bazin et al., 2014). Datice provides an excellent reference for this Bayesian approach. Still, because Datice has been developed over a long term period with a continuous effort in calculation optimization through methodological improvement, the final code is difficult to access for a non-expert and cannot easily be used as a community tool. We thus identified the need for an open and user-friendly program with a performance similar to Datice but that can be more easily used and implemented by different users within the ice core community.
In this paper, we present a new probabilistic model, IceChrono_v1, based exactly on the same approach as Datice but with improvements and simplifications in the mathematical, numerical and programming aspects. We first detail the IceChrono methodology highlighting the differences to Datice (Section 2). We then perform dating experiments described in section 3 using IceChrono1. We first replicate the AICC2012 experiment, and perform 4 additional experiments to test new functionalities of IceChrono1. The results of these experiments are discussed in section 4. We summarize our main findings in the conclusions, and describe perspectives for future developments on IceChrono in section 5.

Line by line comments: Please include a table with all the variables described. There are a lot of very similar symbols and they are difficult to track down in the text.

Done.

P6812,L2 – the second sentence is too long with too many parentheses to understand

This sentence has been simplified.

P6812,L12 – I don't understand the phrase "confidence interval" and how that applies to a chronology. Do you mean the uncertainty in the chronology for each age?

This was a mistake. "confidence interval" has been changed to "standard deviation" throughout the manuscript.

P6813, L5-12 – Please rewrite this paragraph so that the ideas flow more easily.

The introduction and thus this paragraph has been fully rewritten in the revised version.

P6813, L13 – What does "these" refer to? Do you mean that there are 4 broad ways to date ice cores?

The introduction and thus this paragraph has been fully rewritten in the revised version.

P6813, L17 – support your statement that this method is generally accurate for event duration.

Sentence changed to:
"Based on the evaluation of the thickness of annuals layers, this method provides good estimates of event durations and small uncertainties on the absolute age of the shallower ice sections. "

P6815, L12 – I find equation 2 difficult to follow. Please describe in more detail.

Eq. (2) (which is now eq. (3) in the revised manuscript) is described in the text:
"The third equation means that if one virtually un-thins a depth interval between an ice depth and the synchronous air depth, one gets the Unthinned Lock-In Depth in Ice Equivalent (ULIDIE, see Figure 2)."

P6817,L1 – please describe why logarithmic correction functions are needed in plain English. I'm not sure I follow the motivation.

The associated paragraph now reads:
"The variables a_k, l_k and τ_k are always positive in our problem and generally the log-normal density of probability is more adapted for these variables, which Tarantola (2005) call Jeffreys variables. The logarithm is taken so as to transform them into Cartesian variables, with a canonical density of probability being the normal. This change of variable is necessary to describe our optimization problem as a least squares problem (see section 2.4)."

P6817,L13 – how are you solving Equation 2 for deltaDepth?

We added the following description:
"To solve Equation (3), we first integrate D_k/τ_k from the surface down to every depth in the age equation grid, i.e. we have a correspondence table between real depths and unthinned-ice-equivalent (UIE) depths. Then for every air real depth in the age equation grid, we obtain the air UIE depth from the correspondence table. Then we subtract the second member of Equation (3) to this air UIE depth to get the ice UIE depth. Then we use the correspondence table to obtain the ice real depth and the Δdepth."

P6821,L15 – I don't understand "A class exist for the ice core object and does:"
P6821,L14 – This paragraph could use more explanation about object oriented paradigms and object classes

We now introduce the main principle of the object oriented paradigm and we give more details on the object oriented structure of IceChrono v1:

The core of the code is about 1000 lines long (including white lines and comments) and is built using an object oriented paradigm. In such an object oriented language, apart for the classical type of variables (integer, real, characters, etc.), one can define his own classes of objects, containing variables and functions. In IceChrono1, a class exists for the ice core object. It contains the *variables related to this ice core: the age equation grid, the correction function grids, the prior scenarios and their associated standard deviations and correlation matrices, the relative density profile, the correction functions, the observations and their associated standard deviations and correlation matrices and the resulting calculated variables (accumulation, LID and thinning, ice and air ages, Δdepth, ice and air layer thickness, etc.). It also contains functions performing the following tasks: the initialization of the ice core (i.e. reading of the parameters, priors and observations), the calculation of the age model, the calculation of the residuals, the calculation of the forward model Jacobian, the calculation of the standard deviations, the construction of the figures (for ice age, air age, accumulation, LID, thinning, ice layer thickness and Δdepth) and the saving of the results. A class also exists for each ice core pair. It contains all the stratigraphic links and their associated standard deviation and correlation matrices relative to this ice core pair. It also contains functions that perform the following tasks: the initialization of the ice core pair, the calculation of the residuals, the construction of the figures (for ice-ice links, air-air links, ice-air links and air-ice links). The main program is kept as simple and straightforward as possible.*

P6821,L25 – This paragraph needs better explanation. I can't follow what is being done "inside each term of the cost function" and why a "change of variable" is occurring. Please write out the steps and add a plain English explanation.

We wrote the steps using equations so we hope it is now clear:

We used the LM algorithm as implemented in the 'leastsq' function from the 'scipy.optimize' library, which also provides an automatic convergence criteria. It does not try to minimize directly the cost function, but rather a residual vector, the components of this residual vector being supposed independent from each others and with a unit standard deviation. Inside each term of the cost function:

$$\boldsymbol{R}^{\mathrm{T}}\mathbf{P}^{-1}\boldsymbol{R}\,, \qquad (1)$$

*We allow defining a correlation matrix **P** so that the residuals can actually be correlated. We thus used a Cholesky decomposition of **P**:*

$$\mathbf{P}=\mathbf{P}^{1/2}\mathbf{P}^{1/2}\,, \qquad (2)$$

and a change of variable:

$$\boldsymbol{R}'=\left(\mathbf{P}^{1/2}\right)^{-1}\boldsymbol{R}\,, \qquad (3)$$

to transform, the residual vector into a vector composed of independent variables with unit standard deviation. The associated term of the cost function can now be written:

$$\left(\boldsymbol{R}'\right)^{T}\boldsymbol{R}'\,, \qquad (4)$$

that is, the residuals are now independent and with a unit standard deviation.

P6822,L8 – how does this Datice assumption really differ from the IceChrono assumption. In describing equation 2, the author assumes a constant value of 1 for thinning in the firn. So isn't the Datice assumption of a constant thinning function at depth very similar? I'm guessing there is a subtlety here that I'm missing and should be better explained. It also seems like in the AICC2012 comparison later in the paper, the effect of this difference should be specifically diagnosed.

We do not assume that thinning is 1 in the firn. We assume that thinning in the firn at the time of deposition was the same than at present. This is different from the Datice assumption, which assume that thinning between z-Δdepth and z is constant.

P2822,L14 – Please explain why adding mixed ice-gas and gas-ice stratigraphic links is important. You comment that this is new with respect to Datice, but never explain (1) whether the lack of this functionality limits AICC2012, (2) if you used any ice-air or air-ice links in the IceChrono version of AICC-2012, or (3) and concrete example of ice-air and air-ice links.

We added the following sentence to give an example of mix links:
"A concrete example of the use of mix ice-gas stratigraphic links could be the synchronization of Dansgaard-Oechger events recorded in the methane records in Antarctica and in the ice isotope records in Greenland. This would be especially useful if the methane records in Greenland are not yet available."
We also added a test experiment which uses mix ice-air stratigraphic links.

P2822,L17 – The allowance of correlated errors gets considerable attention in this manuscript, yet I get the impression at both the AICC-like and Berkner Island test cases don't make use of this functionality. The writing even seems to emphasize that you don't need to input error correlation for IceChrono to run. If this is an important advance, the impact of it needs to be assessed.

We now present a test experiment on the NGRIP core for dated ice intervals with correlation.
For convenience, the prescription of the correlation matrix in IceChrono is indeed optional and by default, it is assumed equal to the identity matrix.

P2822,L20 – Using a numerical solver of the residuals means that there is the potential to find local minima in the cost function. The paper needs to test whether the solver is robust to different initial conditions.

We tried the solver with 5 different initial conditions taken according to the prior and we checked that they all converge toward the same solution.

P6823,L4 – I was able to find and access IceChrono easily. I did not download and compile it. My quick impression is that it could use more documentation.

The documentation has been reviewed and improved. If you have any other specific request regarding the documentation, please feel free to send it to us.
(Note that IceChrono does not need to be compiled since python is an interpreted language.)

P6823,L11 – I think calling AICC2012 "the last official chronology" is a little strong. AICC2012 is "an official" chronology, though not universally accepted, and I doubt it will be "the last" official chronology.

Corrected to "most recent".

P6823,L14 – I'm confused about correlation matrices. Here you say the Datice background correlation matrices have a Gaussian shape. But on P6822,L18 you write that Datice only allows correlated errors for dated ice intervals. Please explain clearly in the text.

The most recent version of Datice actually allows defining correlation matrices for all types of observations. This sentence has therefore been corrected.

P6823,L17 – Please diagnose and describe the effect of changing from Gaussian to triangular matrices. What is the resulting difference in the timescale and uncertainty from making this change alone? What is the change in computation time?

The problem of gaussian-shape correlation matrices is not a computation time problem. The usual solver to perform a Cholesky decomposition simply do not work for gaussian-shape matrices, because they are too close to singular matrices. Instead of spending a lot of time to solve this numerical problem just for the sake of comparison with Datice/AICC2012, we decided to use triangular-shape matrices.

P6823,L11 – More description is needed of the AICC2012-like inputs. Please describe at least all the different types of dating information used and provide specific references to the AICC2012 work (i.e. if you are using the same gas tie points, reference the appropriate tables in Veres et al. 2013 and Bazin et al. 2013). There is new functionality in IceChrono, is any of this employed? The reader should not be expected to have read all the Datice papers in detail.

We now describe in more details the AICC2012 dating information used and make a reference to the supplementary material of Bazin et al. (2013) and Veres et al. (2013) where all the data are available:

IceChrono1 is similar in scope to the Datice model (Lemieux-Dudon et al., 2010a, 2010b). Datice has been used to build AICC2012, the most recent official chronology for the 4 Antarctic ice cores EPICA Dome C (EDC), Vostok (VK), Talos Dome (TALDICE), EPICA Dronning Maud Land (EDML) and the Greenland ice core NorthGRIP (NGRIP) (Bazin et al., 2013; Veres et al., 2013). The AICC2012 experiment was based on a previous experiment (Lemieux-Dudon et al., 2010a) on 3 Antarctic ice cores (EDC, VK, EDML) and one Greenland ice core (NGRIP), but with updated chronological information. All chronological informations are available in the supplementary material of Bazin et al. (2013) and Veres et al. (2013). This experiment integrates orbital tuning constraints based on the $\delta^{18}O_{atm}$, O_2/N_2 *and air content records (Bazin et al., 2013, p.201; Dreyfus et al., 2007, 2008; Landais et al., 2012; Lipenkov et al., 2011; Raynaud et al., 2007; Suwa and Bender, 2008), layer counting on NGRIP back to 60 kyr B1950 (Svensson et al., 2008 and references theirein), a tephra layer (Narcisi et al., 2006) dated independently at 93.2±4.4 kyr B1950 (Dunbar et al., 2008), the Laschamp geomagnetic excursion at 40.65±0.95 kyr B1950 (Singer et al., 2009) and the Brunhes-Matuyama geomagnetic reversal at ~780.3±10 kyr B1950 and its precursor at 798.3±10 kyr B1950 (Dreyfus et al., 2008) identified in the* ^{10}Be *records (Raisbeck et al., 2006, 2007; Yiou et al., 1997), a Holocene* ^{10}Be*-dendrochronology tie point on Vostok at 7.18±0.1 kyr B1950 (Bard et al., 1997; Raisbeck et al., 1998),* Δ*depth observations at NGRIP obtained by comparing the* $\delta^{18}O_{ice}$ *and* $\delta^{15}N$ *records (Capron et al., 2010; Huber et al., 2006; Landais et al., 2004, 2005), air synchronization tie points using the* CH_4 *records* (Buiron et al., 2011; Capron et al., 2010; Landais et al., 2006; Lemieux-Dudon et al., 2010a; Loulergue et al., 2007; Schilt et al., 2010; Schüpbach et al., 2011) *and the* $\delta^{18}O$ *records (Bazin et al., 2013) and ice synchronization tie points using the volcanic records (Parrenin et al., 2012b; Severi et al., 2007, 2012; Svensson et al., 2013; Udisti et al., 2004; Vinther et al., 2013) and the* ^{10}Be *records (Raisbeck et al., 2007).*

P6824 – as discussed above, this section is lacking sufficient detail in the comparison with AICC2012.

See above.

P6824, L9 – Please provide more description of the different uncertainties compared to AICC2012. Why does Datice have such a larger uncertainty at the Laschamp event if both methods are using the same inputs?

For the AICC2012-like experiment, we now compare directly with the output of Datice and not with the AICC2012 stated uncertainty. The abnormally large uncertainty of AICC2012 at the Laschamp event is actually not present in the output of Datice.

P6824,L22 – remove this entire section. The first dating of the deep Berkner Island core needs its own paper with the data published. The records are not straightforward and will require significant explanation.

The application on the Berkner Island is now removed in the revised manuscript.

P6827,L10 – adding in glaciological models is a great next step. I think this paper would benefit from an extended discussion of why this is a limitation of IceChrono.

We added the sentence:
"Indeed, sedimentation models also have their own poorly-known parameters and it is why they need to be integrated into the optimization process."

P6827,L18 – The appendix seems both repetitive and under-explained. All of the odd-numbered equations appear to be the same except that they are for different classes of age markers. I think the appendix would be improved by condensing the odd equations and providing explanation (not just a description of the variables) for the even equations.

We agree this appendix is a bit repetitive. The basic principle of all the terms is explained in equations (18) and (19). Now during the submission process, the editor reckoned that a detailed description of all the terms individually would help to better understand, hence this appendix.

Figure 2 – This figure would benefit from a second panel with the most recent 60 ka (when the chronologies are tied to the GICC05 annual chronology). The ice and gas chronologies as well as the ice and gas uncertainties should be shown.

We now show both the ice and gas chronologies and we added figure 5 which is a zoom on the last 60 kyr.

Figures – The removal of the Berkner Island figures will allow many more figures diagnosing the differences of the Datice/IceChrono methologies and the AICC example.

The Berkner example has been removed. The AICC2012-VHR experiment and its results are now provided as a supplementary material. This will allow the readers to analyze and modify this experiment. We did not include all the figures of this experiment in the main text since there are so many of them (7 figures per drilling and 4 figures per drilling couple!).

This second part of the review describes some of the limitations of the Datice/IceChrono methodology. While I will let the editor decide if this manuscript should address these issues, I want to emphasize that the manuscript would be much improved by taking a more comprehensive view. If the goal is to get IceChrono accepted as the best way to date ice cores, then a full description and real evaluation of the resulting chronologies will go a long way to achieving this goal.
The Datice and IceChrono methodologies have recognized limitations, which is part of why the chronologies have not been universally adopted. But there are two other reasons as well:
1) The methods have never been well explained. The Lemieux-Dudon papers are nearly impenetrable and the AICC2012 papers add very little in methodological description.
2) The methodologies have never been shown to actually yield improved timescales. Yes, Datice and IceChrono can produce timescales for all the various ice cores, but there has never been a test case that truly evaluates the resulting timescales. A synthetic test case has never been performed; why not develop "known" timescales, add noise, employ Datice/IceChrono, and then compare the inferred timescales to the "known" timescales? This would allow a much improved understanding of the methodology and greatly improve the confidence in the inferred timescales.
This work by uses the agreement with AICC2012 as validation for IceChrono. I admit that I bring a

bias against the Datice chronologies because there are too many oddities: -the small glacial delta-ages of EDML -the reversals in the thinning function of EDML and Talos Dome in glacial-transition ice -the same uncertainties for the ice and gas timescales that have not been explained.

We think we should separate two different kinds of critics regarding AICC2012/DatIce: the critics toward the DatIce method (which aims at bringing different kind of chronological information together) and the critics toward the AICC2012 experiment, which uses a particular set of chronological information.
Here we developed a probabilistic dating method, comparable to DatIce in its scope, but which fixes some of its issues. So we are ready to discuss the methodological aspects (the model), but not the AICC2012 experiment in particular (the simulation).

I worry IceChrono may perpetuate many of the Datice methodological problems and create more, not less, confusion about ice-core timescales.
A few of the issues identified with Datice:
A) Ice-ice stratigraphic links are predominantly matches in sulfur peaks between cores. These matches are either correct, or wrong. However, the Datice/IceChrono methology assigns an uncertainty (I think Gaussian) to them. The final timescale results in the links no longer being exact, invalidating the premise of the links in the first place.
This limitation needs to described with guidance about what types of analysis the Datice/IceChrono chronologies are not appropriate for.

This is clearly a limitation of the IceChrono (and DatIce) model in its current state.
All uncertainties are assumed gaussian and we agree that the gaussian model does not represent correctly our state of knowledge regarding the volcanic peaks matching. So for now, we assume that the volcanic matching is perfectly done by the operator within the width of the volcanic peak.
We added a section in the discussion regarding the limitations of IceChrono v1.
We also mention in the perspective the coupling with the volcanic matching problem:

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.

B) The only glaciological constraints are in background scenarios. This leads to chronologies that are not glaciologically consistent. For the inferred thinning functions, this is revealed through reversals where deeper, older ice has thinned less than shallower, younger ice. For most ice core sites, this is not physically realistic (compressive flow can allow these reversals, but this is not appropriate for most dome sites). Impurity concentrations have been suggested as the cause, but these ideas ignore continuity and are implausible.

We disagree that a reversal in thinning is not glaciologically consistent. For example, Parrenin et al. (JGR, 2004) show that this kind of reversal occurs for the Vostok ice core, due to a thinner ice thickness upstream of the drilling site. We agree this is more difficult to explain for a dome site, but one is never sure that a dome has stay stable during the past and some anomalous flow can also occur (Dreyfus et al., CP, 2007).
The idea of the IceChrono (or DatIce) model was that ice flow models are imperfect, and that we wanted to use their information, but in a weak way, not in a strong way.

C) The gas-age and ice-age uncertainties are the same. This is odd because there should be different uncertainties based on whether the age markers are in the ice or gas. In tracking down this oddity, I found this was instituted after a reviewer of the Bazin et al. (2013) AICC paper found that the ice age uncertainties were smaller even when the age markers were in the gas phase and the gas age uncertainty should have been smaller. The authors decided to make the uncertainty for both the ice and gas equal to the larger of the two uncertainties. This approach concerns me because (1) the error in the Datice methodology was never diagnosed and (2) the uncertainties are incorrect. While this may seem a minor point, it shows clearly that the output of Datice is not fully understood. If Datice cannot do a simple uncertainty correctly, why should anyone have faith that the more complicated implementations find an "optimal" chronology. This work has not provided enough detail to diagnose whether IceChrono suffers from the same problem.

There is a known limitation in the way the confidence intervals were calculated in AICC2012 using Datice. For example, the air age uncertainty at EDC is too large during the last glacial part with respect to the observations which have been used.
As far as we have tested, this problem does not exist in IceChrono:

During the last glacial period, there are many CH_4 Antarctica-NGRIP stratigraphic links with uncertainties of a few centuries. The NGRIP chronology being tightly constrained to GICC05 within 50 yr, it is expected that the gas age uncertainty at EDC sometimes decreases below 1000 yr during this time period. The posterior uncertainty calculated by IceChrono is therefore consistent with the chronological information used, in contradiction to that calculated by Datice.

While possibly outside the scope of this manuscript, addressing these (and other concerns) of the Datice/IceChrono methodology would provide confidence in the resulting chronologies.

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 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. a = (a(z 1), a(z 2), . . . , a(z n) T = (a 1 , a 2 , . . . , 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.

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 bi . Using a Bayesian framework we can formalise this belief by placing a initial prior on each of the unknown parameters e.g.

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

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

Your equation in 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 **a**(z), **I**(z), **τ**(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 **a**, **I**, **τ** , then we could work out the true time elapsed as the value $g_1(\mathbf{a}, \mathbf{I}(z), \tau)$ 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:

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

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.

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

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.

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 $\alpha(z'_k)$ $m\ yr^{-1}$ 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?

D_k is the relative density of the material with respect to pure ice. It is equal to 1-ρ where ρ 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 τ_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 τ_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.

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?

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'^{ie}_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), τ ? 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 ~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 τ 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^τ_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 τ 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 $\mathbf{C}^X$. When the optimal solution is found, it also allows to evaluate the uncertainty of the model $\mathbf{C}^G$ 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 ζ 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.

Anonymous Referee 4

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

The manuscript describes a new implementation of the Datice model, which most notably was used to create the AICC2012 ice-core time chronological framework. The new implementation, IceChrono, is only marginally conceptually different from the Datice model, but is based on a Python platform that makes the model more accessible to the wider community. I believe IceChrono is appropriate for release/publication in GMDD. However, a number of open questions about the function of the model remain, and the presentation is not very clear. Without being an expert on the mathematical formulation of Bayesian models, I agree on the comments provided by Tim Heaton and the other anonymous reviewer:

- The model description needs to be more detailed, especially because none of the previous papers on Datice have clearly presented the modelling framework in an accessible format discussing in depth how to address problems with meaningful uncertainty assignment etc.. Some concrete examples are given below.

We tried to clarify the description of the model.

- The glaciological validity of the results needs to explored further,

The glaciological validity is taken into account using the prior information.

- The section on the Berkner Island dating needs massive improvements or could be removed to free space for a more thorough discussion of the model. I provide comments about the Berkner Island section separately below.

The section on Berkner Island has been removed.

- The manuscript is in serious need of significant improvements in language and clarity. Just to add to the list provided by the other reviewers: use dash in “ice-core records” and similar expressions. The manuscript should pass a thorough grammar check before resubmission.

The language of the manuscript has been checked independently by several people.

With regard to the first point above, I find the best approach to be to describe the workings of the model conceptually first and then, in a separate section that can be skipped by readers with limited technical interest and/or skill in Bayesian modelling, go into the level of detail asked for by reviewer Tim Heaton.

We now have a initial “method summary” sub-section to give the reader a rough idea on how the model works. We did not adopt entirely the Bayesian framework as described by T. Heaton since we could not come up with a clear description using this framework. So we kept our framework based on the books by A. Tarantola:

The true chronology of an ice core, i.e. the ice and air ages at any depth, is a function of three variables (also functions of the depth): the initial accumulation rate (the accumulation rate when and where the particle was at surface), the lock-in depth of the air and the vertical thinning function (the ratio between the thickness of a layer in the ice core to the initial accumulation rate). This is what we call the forward model. These variables are unknown, and to find our optimal chronology we estimate them based on:

- *Prior information about their values on each ice core;*
- *Chronological observations, such as (see Figure Erreur : source de la référence non trouvée): the ice or air age at a certain depth, the time elapsed between two depths, the synchroneity between two ice or air depths within two different ice cores or the depth shift between synchronous ice and air depths within the same ice core.*

All these different types of information, mathematically described as probability density functions (PDF), are assumed to be independent and are combined using a Bayesian framework to obtain posterior estimates of the three input variables (accumulation, LID and thinning) and of the resulting chronologies. Uncertainties on the prior estimates and on the observations are further assumed to be Gaussian and the forward model is linearized, which allow to use the Levenberg-Marquardt (hereafter LM) algorithm (Levenberg, 1944; Marquardt, 1963) to solve this least-squares optimization problem. The philosophy of the method is similar to that of the Datice method (Lemieux-Dudon et al., 2010a, 2010b).

In addition, I would like to raise three other central points, first one regarding the presentation of the IceChrono (and Datice) results, and two of more technical nature. Firstly, I find it unjustified to claim that the Datice or IceChrono models produce “optimal” chronologies. The models use optimization techniques, and therefore, the use of the word “optimization” is acceptable. However, optimization only produces optimal results if the model underlying assumptions are justified, the simplifications insignificant, and the data basis is correct and with correct representation of uncertainty (which is particularly problematic in this context, see comment below on volcanic matching). I know that “optimal” is used in the title of Lemieux-Dudon et al 2010b, but I still think that this use of “optimal” should be discontinued. I think that the word “consistent” in the title of the original Datice paper of Lemieux-Dudon et al. 2010a is the most appropriate description of the approach.

Consistent was used to mean that the chronologies for the different ice cores are consistent (following the stratigraphic information). “Optimal” means that the chronology is optimal with respect to the different chronological informations which are used. This term is very often used in the field of data assimilation and we see no reason not to use it here. Of course we do not claim we produce “perfect” chronologies, but optimal chronologies in a certain sense.

Secondly, a more technical point is the balance between data and model inputs. The Datice and IceChrono models essentially make trade-offs between background scenarios (that are known to be wrong as there would otherwise not be a reason to apply the model) and data-based constraints. The trade-off is made in the models’ costs functions, which gets contributions from the misfit between model and data constraints and the deviation of the model from the background scenarios. From eq. 9-14, it appears to me that all contributions are added weighted only by their uncertainties. There may be no “optional” way to do this, but at least the question of how to obtain a good balance between the data constraints and the backgrounds should be thoroughly discussed. In judging whether a fair balance has been obtained, I encourage the author to consider if a measure and/or figures showing the contribution to the cost function from the different terms in the cost function could be useful.

Following Tarantola's framework, we combine different information (prior or observations). Each prior and observation has a weight which is related to its variance-covariance matrix.

Some specific questions to address:

- It seems clear that the resolution of the background scenarios influences the balance between data constraints and backgrounds. The experiment on page 6823 explores the effect of doubling the resolution of the correction functions of the background scenarios, which is a good test to make. Does the resolution of the background scenarios themselves (and not only their correction functions) influence the results? In other words: Is the relative contribution of the deviation of the model from the background scenario dependent on the grid resolution for the background, the grid resolution of the correction function, or both?

The prior scenario are given on the age equation grid. This age equation grid is used to solve Eqs (1-3). Eq (2) is simply solved using a linear interpolation, so it is certainly robust to a change of resolution. Eq (1) and Eq (3) are integrals that clearly converge when the depth step decreases to zero. We made this point clear in the revised manuscript:

It is important to assess whether the formulation of IceChrono1 is robust to a change of resolution:

when the resolution increases, the simulations should converge toward a meaningful result. IceChrono1 uses different two different types of grids to optimize the ice cores age scales: the age equation grids and the correction function grids.
The age equation grids are used to solve Eqs (5), (6) and (7). Eq (6) is the value of the ice age function at a given depth, so it is clearly robust to a change of resolution. Eqs (5) and (7) are integrals and are therefore also robust to a change of resolution.
Concerning the correction functions grids, we made two test experiments with different resolutions: AICC2012-VHR and AICC2012-V2HR. The fact that the AICC2012-VHR and AICC2012-V2HR experiments agree well indicates that the formulation of the optimization problem in IceChrono1 is robust to a change of resolution of the correction functions.

- Especially if the only weighting factor of the background scenarios is determined by the width of their confidence intervals, the assignment of confidence intervals/uncertainty is of central importance. What measures have been taken to ensure consistent assignment of background scenario uncertainty between different cores and through time?

The uncertainties on the prior and observations should be determined by an in-depth analysis of the methods which were used to determine them. There is no common solution to this problem and it is up to the user of IceChrono to ensure consistent error assignment to the different chronological information.

- Given that the models include background scenarios for cores that cover from 1 to 8 glacial cycles, and given that the flow regime and accumulation reconstructions are likely much better constrained for some cores and some time intervals than others, is it a reasonable assumption to use the same resolution of background scenarios for different cores and times? For example, by using a temporal resolution of 1 kyr for the accumulation background correction functions, I guess there will be 6-8 times more points that relate to the EDC accumulation reconstruction than the NGRIP reconstruction? If so, could/should the gridding/background scenario resolution be made variable to adapt to this?

It is normal that a 800 kyr long ice core has more weight than the first 100 kyr of the same ice core, since it brings more information to the optimization problem. But the weight also depends on the uncertainties prescribed for the prior: the smaller the uncertainty, the larger the weight. Concerning the resolution of the correction function, we generally assign a correlation length to the priors. This way, if one increases the resolution below this correlation length, we indeed increase the number of points of the prior but these points are not independent anymore, so there is no real new information.

- A specific question along the same lines: Assume that two cores have been linked stratigraphically using 1000 volcanic fix points and 50 methane horizons with the same uncertainty. Does each horizon enter the cost function with the same weight? If so: Is this reasonable? If not: How is the weighting determined?

As explained above, we use the Tarantola framework and we assume that each independent observation is treated the same way. Of course, the weight of an observation is related to its uncertainty. This is even more complicated when there is a correlation between the various observations.

Thirdly, the author is encouraged to discuss how to better represent the uncertainty of volcanic matches in the model. I understand that no new such data are introduced here, but are adopted from the AICC2012 data basis. However, this is an obvious place to improve Datice to increase the confidence of the results. The issue is that volcanic matches come in different categories. In rare cases, tephras have been found in several cores and geochemistry has confirmed that the tephra are indeed coming from the same volcano. However, more commonly, a series of acidity/sulphate peaks representing a characteristic pattern are matched. Assuming that the pattern match is correct, the uncertainties of the individual horizons are on the order of centimeters (i.e. years to decades) and largely uncorrelated between horizons. However, there is a small risk that the pattern match is incorrect, in which case the error can be several meters or more (centuries or millenia), while being highly correlated between the horizons that belong to the same matched pattern. In contrast to this, as far as I know, the uncertainty of volcanic horizons is represented in Datice and IceChrono as Gaussian errors of typically 20-200 years. It would be great to hear if the author has ideas about how to implement a more realistic uncertainty estimate, and to respect the different types of volcanic ties.

The volcanic matching is, as you described, an ambiguous and complex problem. IceChrono v1 assumes gaussian uncertainties, so in this case we assume the volcanic matching has been done correctly. We added this point in the discussion. Note that IceChrono v1 could be used to test different plausible volcanic matching scenarios, but it is not able to synthesize the different results at the moment. The automatic synchronization of records is outside the scope of the present study and will be done later one:

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.

Concrete comments to the manuscript:

6812

7: “use of dated depth intervals” is slightly misleading as the depth intervals are not dated, but represent a certain duration in the record. Please use another word, e.g. “use of intervals with known duration” or similar. This change should be applied consistently throughout the manuscript.

Changed to “intervals with known durations”.

6813

3: . . . field STRENGTH.

The introduction has been entirely rewritten:

Polar ice cores provide continuous records of key past features of the climate and the environment, with local, regional and global relevance (e.g., Ahmed et al., 2013; EPICA community members, 2004; NorthGRIP project members, 2004; WAIS Divide Project Members, 2013). Tracers of polar climate (e.g., Jouzel et al., 2007), ice sheet topography (NEEM community Members, 2013), water

cycle (Schoenemann et al., 2014; e.g., Stenni et al., 2010; Winkler et al., 2012), aerosol deposition (e.g., Lambert et al., 2012; Wolff et al., 2006) and global atmospheric composition (e.g., Ahn and Brook, 2014; Loulergue et al., 2008; Marcott et al., 2014) measured in ice cores unveil sequences of events on seasonal to glacial-interglacial time scales.
However, the prior to interpretation of polar ice core records is the complex task of building two robust time-depth relationships, one for the tracers measured in the ice phase (e.g. water isotopes, particulates and chemical impurities) and one for those measured in the air phase (e.g. greenhouse gas concentration, isotopic composition of gases). The firn, where snow is gradually compacted into ice, constitutes the upper 50-120 m part of ice sheets. The firn is permeable and air is only locked-in at its base, at a depth level called the Lock-In Depth (LID). As a result, the entrapped air is always younger than the surrounding ice at any depth level. Through gravitational fractionation processes, LID is closely related to the isotopic composition of δ15N of N2 in air bubbles data (e.g., Buizert et al., 2015; Goujon et al., 2003; Parrenin et al., 2012a; Schwander et al., 1993). The temporal evolution of the age difference between ice and air at a given depth must therefore be estimated using firn densification modeling and air δ15N. This age difference is essential for clarifying the exact timing between changes in atmospheric CO2 concentration and Antarctic surface temperature during deglaciations (Caillon et al., 2003; e.g., Landais et al., 2013; Monnin et al., 2001; Parrenin et al., 2013; Pedro et al., 2011). However, glacial-interglacial Antarctic firn changes remain poorly understood (e.g., Capron et al., 2013).
Several strategies have been developed to build ice and gas chronologies. We briefly describe these methods, their strengths and caveats hereafter:
1) Annual layer counting (e.g., Rasmussen et al., 2006; WAIS Divide Project Members, 2013; Winstrup et al., 2012). Only applicable when accumulation rates are sufficiently high to make this annual layer identification possible, this method provides accurate estimates of event durations and small uncertainties on the absolute age of the upper ice sections. However, the cumulative nature of the errors, associated with the increasing number of counted layers, leads to a decrease of the accuracy of absolute age with depth. For instance, the GICC05 (Greenland Ice Core Chronology 2005) composite timescale for Greenland ice cores (Rasmussen et al., 2006; Seierstad et al., 2014; Svensson et al., 2008), is associated with a maximum counting error of only 45 years at ~8.2 ka B1950 (Before 1950 C.E.). This error increases progressively with depth, reaching more than 2500 years at ~60 ka B1950. Annual layer counting techniques cannot be applied when the annual layer thickness is too small to be resolved visually, e.g. in ice cores from central East Antarctica.
2) Use of absolute age markers in ice cores. Well-dated tephra layers identified in ice cores during the last millennia provide precious constraints (e.g., Sigl et al., 2014). Beyond that period, absolute age markers are very scarce. The links between 10Be peaks and well-dated magnetic events (Raisbeck et al., 2007) have provided an age marker for the Laschamp event (Singer et al., 2009). Promising results have recently been obtained using radiochronologic dating tools (Aciego et al., 2011; Buizert et al., 2014; Dunbar et al., 2008).
3) Orbital dating in ice cores. Because there are few absolute constraints in ice cores beyond 60 ka B1950 (limit for the layer counting in the NGRIP ice core), orbital tuning is the most effective method to provide chronological constraints on ice core deepest sections. In the first orbital dating exercises, tie points were determined from the tuning of water isotopic records on insolation curves (e.g., Parrenin et al., 2004), which limits further investigations of polar climate relationships with orbital forcing. More recent chronologies tried to circumvent this assumption and focused on non-climatic orbital markers. Three complementary tracers are currently used: the δ18O of atmospheric O2 (δ18Oatm) (e.g., Bender et al., 1994; Dreyfus et al., 2007), δO2/N2 (e.g., Bender, 2002; Kawamura et al., 2007; Suwa and Bender, 2008) and the total air content (e.g., Raynaud et al., 2007). While the link between δ18Oatm and precession is explained by variations in the water cycle of the low latitudes, relationships between δO2/N2, air content and local summer insolation are understood to arise from changes in the surface snow energy budget influencing its metamorphism. Without a precise understanding of mechanisms linking these tracers to their respective orbital targets, the associated uncertainties remain large, 6 ka for δ18Oatm and 3 to 4 ka for δO2/N2 and

air content (Bazin et al., 2013, 2014; Landais et al., 2012).
4) Ice core record synchronization. Inter-ice core matching exercises are undertaken to transfer absolute or orbital dating information from one ice core to another one. It generally relies on the global synchroneity of changes in atmospheric composition (CO_2, CH_4 concentration, and $\delta^{18}O_{atm}$) (Bender et al., 1994; Blunier and Brook, 2001; Monnin et al., 2004), the identification of volcanic sulfate spikes within a given area (Parrenin et al., 2012b; Severi et al., 2007) or the hypothesis of synchronous regional deposition of aerosols recorded as ice impurities (Seierstad et al., 2014). In the first case, limitations are associated with the smooting of atmospheric composition changes through firn air diffusion. In the second case, mismatches may arise through incorrect identification of events in different ice cores.
5) Correlation with other well-dated climatic records. In some cases, high-resolution calcite $\delta^{18}O$ records and precise U/Th dates on speleothems have been used to adjust ice core chronologies (Barker et al., 2011; Buizert et al., 2015; Parrenin et al., 2007a). Pinning ice core and speleothem records is attractive to reduce absolute age uncertainties especially during past abrupt climatic events of glacial periods. However, these exercises rely on the assumption of simultaneous abrupt climatic changes recorded in ice core (e.g. water isotopes, CH_4) and low latitudes speleothem $\delta^{18}O$ records (mostly reflecting changes in regional atmospheric water cycle). A main limitation of this method lies in the validity of this assumption.
6) Modeling of the sedimentation process: snow accumulation, snow densification into ice, air bubbles trapping and ice flow (Goujon et al., 2003; Huybrechts et al., 2007; Johnsen et al., 2001). Glaciological modeling provides a chronology derived from the estimate of the annual layer thickness, and therefore, leads to more realistic event durations when the accumulation history and thinning function are well constrained. A side product of glaciological modelling is the quantification of changes in surface accumulation rates, and the quantification of the initial geographical origin of ice. These additional informations are necessary to convert measurements of concentrations of chemical species in ice cores into deposition fluxes, and to correct ice core records from upstream origin effects (e.g., EPICA community members, 2006; Röthlisberger et al., 2008). Caveats are caused by unknown parameters of such glaciological models, such as amplitude of accumulation change between glacial and interglacial periods, the basal melting or the vertical velocity profile, which have a growing influence at depth.
A common and optimal chronology for several ice cores can be built through the combination of several of these methods in the frame of a probabilistic approach. The first attempts used absolute and orbitally-tuned age markers along one ice core to constrain the unknown parameters of an ice flow model (e.g., Parrenin et al., 2001, 2004; Petit et al., 1999). This method had however several limitations. First, the uncertainties associated with the ice flow model could not be taken into account, resulting in underestimated uncertainties. Second, the stratigraphic links between ice cores were not exploited, each ice core was dated separately resulting in inconsistent chronologies. A new probabilistic approach based on a Bayesian framework was subsequently introduced. The first tool, Datice, was developed by Lemieux-Dudon et al. (2010a, 2010b). It introduced modeling errors on three canonical glaciological quantities of the dating problem: the accumulation rate, the LID of air bubbles and the vertical thinning function (i.e. ratio between the in-situ annual layer vertical thickness on its initial surface thickness). This method starts from prior (also called "background") scenario for the three glaciological parameters corresponding to a prior chronology for each ice cores. These scenarios, deduced from a modeling of the sedimentation process, are associated with an uncertainty related to the degree of confidence in these prior scenarios. A minimization approach is then applied to find the best compromise between the prior chronological information for each ice core as well as absolute and relative age markers in the ice and in the air phases. This approach has been validated through the Datice tool and applied to build the Antarctic Ice Cores Chronology 2012 (AICC2012), producing coherent ice and air timescales for five different ice cores (Bazin et al., 2013; Veres et al., 2013): EPICA Dome C (EDC), Vostok (VK), Talos Dome (TALDICE), EPICA Dronning Maul Land (EDML) and NorthGRIP (NGRIP). Further developments of Datice were performed to incorporate additional

dating constraints such as the depth intervals with known durations and correlation of errors (Bazin et al., 2014). Datice provides an excellent reference for this Bayesian approach. Still, because Datice has been developed over a long term period with a continuous effort in calculation optimization through methodological improvement, the final code is difficult to access for a non-expert and cannot easily be used as a community tool. We thus identified the need for an open and user-friendly program with a performance similar to Datice but that can be more easily used and implemented by different users within the ice core community.
In this paper, we present a new probabilistic model, IceChrono_v1, based exactly on the same approach as Datice but with improvements and simplifications in the mathematical, numerical and programming aspects. We first detail the IceChrono methodology highlighting the differences to Datice (Section 2). We then perform dating experiments described in section 3 using IceChrono1. We first replicate the AICC2012 experiment, and perform 4 additional experiments to test new functionalities of IceChrono1. The results of these experiments are discussed in section 4. We summarize our main findings in the conclusions, and describe perspectives for future developments of IceChrono in section 5.

11: "pros and cons" - > strengths and weaknesses

Ditto.

17: "but is generally accurate for event duration". Would this not only be true when the accumulation history is well known also far back in time?

Ditto.

18: It seems like a circular argument here: surface accumulation is modelled in line 13 and suddenly it is a result that can be used for interpretation in line 18.

It is not circular. Modeling of snow accumulation is useful in itself (e.g. for determining chemical fluxes) but it is also used for dating.

24: low-accumulation sites.

The introduction has been entirely rewritten.

27: Add that these methods rely on the existence of climate-independent horizons or the assumption that the synchronized records indeed show the same changes synchronously.

The introduction has been entirely rewritten.

28: (4) The synchronization of ice-core records can be done.

Ditto.

6814:

8: "Optimal". See above.

See answer above.

8: "therefore": There is really no argument presented to support this statement.

The introduction has been entirely rewritten.

10: “calibrating 14C ages” is better than using the word “chronology”.

This sentence has been removed.

17: Please specify what these errors are.

This sentence has rewritten.

23-24: Please replace 1 and 4 with one and four.

This sentence has been rewritten.

25: AICC ChronoloGY not IES

Corrected.

6816:

5: Remove “just”.

Removed.

5: “Un-thins” is modelling slang. Please revise.

We don't see which term could better represent what we mean here. We added the “virtually” word to clarify.

7: “Second member” . . . do you mean “term”? Or right-hand side?

Changed to “right-hand side”.

9-10: Very unclear sentence.

The thinning function is very close to 1 in the firn in all circumstances. So we can't make a big error on it.

6818:

26: J k ii is linked to ICE, right? If not, J k ii and J k ai seems to essentially be the same.

As it is explained in the text:
J k ii is the term linked to ice intervals with known durations.
J k ai is the term linked to air intervals with known durations.
So these are two different terms.

6820:

10: Annex -> Appendix

Changed.

6821:

24: The section starting here is unclear to me. In particular, is it possible to evaluate to which degree the residual vectors are indeed independent and whether their standard deviations are unity (which is what the word “unit” in line 27 means, right)?

We now give more details about this point:

We used the LM algorithm as implemented in the 'leastsq' function from the 'scipy.optimize' library, which also provides an automatic convergence criteria. It does not try to minimize directly the cost function, but rather a residual vector, the components of this residual vector being supposed independent from each others and with a unit standard deviation. Inside each term of the cost function:

$$\boldsymbol{R}^{\mathrm{T}}\mathbf{P}^{-1}\boldsymbol{R}\,, \qquad (5)$$

We allow defining a correlation matrix $\boldsymbol{P}$ *so that the residuals can actually be correlated. We thus used a Cholesky decomposition of* $\boldsymbol{P}$*:*

$$\mathbf{P}=\mathbf{P}^{1/2}\mathbf{P}^{1/2}\,, \qquad (6)$$

and a change of variable:

$$\boldsymbol{R}'=\left(\mathbf{P}^{1/2}\right)^{-1}\boldsymbol{R}\,, \qquad (7)$$

to transform, the residual vector into a vector composed of independent variables with unit standard deviation. The associated term of the cost function can now be written:

$$\left(\boldsymbol{R}'\right)^{T}\boldsymbol{R}'\,, \qquad (8)$$

that is, the residuals are now independent and with a unit standard deviation.

6822:

15: The use of annual-layer-counted intervals in Datice is described in a manuscript in revision for Climate of the Past by Bazin et al., so this is only partially true.

We now acknowledge that a new version of Datice to be published allows for observations as intervals with known durations.

22: What is meant by “development”? The rest of the sentence is very convoluted.

This paragraph has been clarified:

7) In IceChrono1, the Jacobian of the model is computed numerically by a finite difference approach while it is computed analytically in Datice. This Jacobian is needed by the minimizer to find the optimal solution X^{opt} *and its uncertainty* $\boldsymbol{C}^{X}$*. When the optimal solution is found, it also allows to evaluate the uncertainty of the model* $\boldsymbol{C}^{G}$ *through equation (20). In Datice, analytical expressions of the Jacobian with respect to X have been derived and these expressions are used to numerically compute the Jacobian 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, which is a complex task.*

6823:

15: Can you really conclude that the matrices do not describe the physical reality well because they are hard to invert?

When we asked mathematicians about how to invert this kind of matrices, they told us: “If you try to invert such matrices, it is very likely that the formulation of your problem is wrong.”
But we changed a bit the argument, saying that the across-diagonal Gaussian shape leads to a too high correlation for neighboring points.
The definition of realistic correlation matrices is a complex problem that will be dealt in future studies.

6824:

2: It means that IceChrono is robust to a change of the resolution of the correction functions by a factor of 2. It may INDICATE robustness on a more general level.

Changed to “indicate”.

13: A more thorough analysis of WHY IceChrono and Datice differs at the Laschamp event would be useful.

We now use the a posteriori uncertainty estimates obtained directly from Datice and not the official AICC2012 uncertainties. The differences at the Laschamp event disappeared so we modified this paragraph.

19: The consistency of the results confirms that the codings of Datice and IceChrono are performing similarly, which can be taken as an indication that they are correct. It shows nothing about the validity of the assumptions or the method itself. The formulation (and especially the similar statement in the conclusion) should clearly reflect this.

We changed this paragraph into:
“The fact that IceChrono v1 and Datice codes, which have been developed independently using different programming languages and different numerical schemes, agree fairly well in this experiment indicates that both codes perform correctly. Note however that the main principles of these codes are the same, so this agreement is not a confirmation of these main principles.”

6825:

1 and 7: repetition

Corrected.

Comments about the Berkner Island dating section:

As requested by other reviewers, the section on the Berkner ice core dating has been removed. It requires an in-depth study and it is not appropriate to include it in this methodological paper.

I’m not opposed to including the section on Berkner Island dating as an example of IceChrono, however if this is to be the first official timescale for the core, then more detail and figures are certainly required.

A discussion of possible reasons for the large accumulation correction around 80 kyr and (in particular) the physical realism of the reversal of the thinning function in fig. 7 below 830 m musty be included.

Moreover all assumptions and parameters that would be needed to replicate this dating by another user of IceChrono should be provided/tabulated or included as supplementary information. Examples of information that should be provided in a revised version as follows:

- p. 6826, line 22-23: 'values of alpha, beta, gamma etc have been chosen to obtain a good fit with independent age markers along the core'. These values should be listed. Please detail which independent age markers have been used and if they are different from the constraints applied later and listed in the tables (and if not, if they are independent from these constraints).
- Figure 3: On the scale of this figure it is difficult to see the differences, if any, between the background scenario and the corrected age. Two things would help: Use colours with more contrast, and add a subplot showing (1) the age difference versus age and (2) the age difference versus depth between the background scenario and the corrected age. It is hardly relevant to state for each figure whether it has been produced by IceChrono. Mentioning it once the text is sufficient.
- Figure 4: As for Figure 3.
- Figure 5: As above use colours with greater contrast. Also please comment on the substantial differences between the background and corrected scenarios around 30-40 ka and 80-90 ka.
- Figure 6: As above for colours. Also please comment on the deviation from the background at 30-40ka is this due to a particular constraint?
- Figure 7, as above for colours.
- The isotope and gas records on the corrected timescale should also be shown in figures (see also below).

It is difficult to compare figure 7 (plotted versus depth) with figure 3-6 (plotted versus age). Please add a secondary age axis to the right axis of fig. 7.

Table 1 shows age ties based on 'comparing the deuterium records' of EDC and Berkner Island. The uncertainties attributed to these ties range between 150 and 300 years. Some more explanation and details should be provided here:

- Why is EDC used for the comparison, instead of EDML, which is closer to Berkner Island and better resolved?
- Also, how are these ties made? Is it by visual matching of Antarctic Isotope Maximum events, or perhaps by some statistical method? In any case, a figure illustrating this would be appropriate: i.e. a comparison of the deuterium records from EDC (or EDML etc) and Berkner Island and same for the CH_4 records.
- It seems optimistic to allocate centennial-scale time-scale uncertainties for deuterium ties. Noise and internal variability between Antarctic ice core sites is at least this large. The onset of deglacial warming in the deuterium record is a good example of this: at WAIS the onset

of warming begins 2000 years earlier than at EDC. Yet the Table 1 is making the assumption that the onset of deglacial warming Berkner Island occur within 170 years of the onset of deglacial warming at EDC.

- Along the same lines: Please explain how the synchronization uncertainties in Table 2 are derived and why EDC was preferred over EDML or even a Greenland CH4 record. Uncertainties in these methane ties as low as 50 years also seem highly optimistic. As above, figures illustrating these ties would be appropriate.
- It is not appropriate to give ages in Table 1 and 2 to two decimal points.

If this is to be the first official/recommended Berkner Island timescale then the ice and gas phase age-depth profiles must be provided as supplementary data and/or deposited in a well-established data repository.