Dear editor,

The three reviews received for our paper were sound and very useful. They made us realised that more work on the method validation and implementation of errors were needed.

We have thus proposed new methodological developments to address these comments. The paper has been strongly reworked to present these new developments and include proper answers to the questions and comments of the reviewers. We have also chosen, following the suggestions of the reviewers, to give less importance to the production of a new chronology (which is actually just an example) and devote more text and explanations to the method validation and tests to best take into account the layer counting errors.

We have also removed the part dedicated to glaciological parameters. This should go into a next paper more focused on these particular aspects with other important associated methodological developments (like implementation of glaciological thinning models as discussed in answer to the technical comment of reviewer 2).

We hope that these modifications are helpful to better present the methodological improvement described in this manuscript.

Reviewer 1 (PG Blackwell)

It took quite some effort to understand just what this paper was doing. The title suggests something methodological; but the explanation is then tightly bound up with discussion of the specifics of GICC05.

→ This is now fully modified and we now strongly focus on the method.

Similarly, the comment about "the implementation in Datice of a new type of markers" I found a bit misleading. It wasn't clear what these markers really were until much later. In fact, they don't really seem to be NEW markers; it is simply that the dependence structure of layer-counting measurements is being appropriately taken into account, t give sensible uncertainties on age differences, "instead of only absolute ages with artificially small uncertainties" and, presumably, unrealistic assumptions of independence. The implementation is new; but that needs to be made clear.

→ We have fully rewritten the text and this is hopefully now clear.

Such proper allowance for dependence is important. But a lot of details of this are glossed over, even though it is the central methodological idea of the paper. The authors say that "[a] last parameter permits to correlate or not this uncertainty with other duration uncertainties at various depth levels." How is that correlation implemented? If there is no correlation, is that equivalent to R being diagonal for these observations? Given the topic, and the comment that the R "matrices play a key role", it is surprising that no detail is given on their construction.

→ We have taken into account this comment and much more details on the correlation of uncertainty is now given in the manuscript. The construction of the matrix R is also explained in the text and with a specific figure.

Independence between these "new" markers and other measurements on the cores is a very strong assumption, and it is not clear what the justification is. Is there a clear scientific basis for it, or is it a simplifying assumptions in the analysis (or does that depend on the particular application?). Either MIGHT be acceptable, but much more clarity is needed.

→ For NorthGRIP, only counting layers are constraining the chronology. This should now be clearer from the new text with simulations performed only on this specific ice core (for validation) and hence using only one type of markers.

The MCE approach to assigning errors to layer counting will interact in a complex way with the frequency of the measurements used; again, this is essentially about dependence, but the choice of spacing will affect the actual MCE substantially. I fear that the issues are more complex than are fully explored by the currently reported experiments.

→ This is indeed true and we have now devoted a full section (2.4) to address how sampling is important and how it relates to the assumptions of the counting errors. As explained in the new text, there is not a simple solution to address sampling but we propose some clues after some sensitivity tests that could be used for a future dating exercise.

Given that there is quite a bit of discussion of the MCE approach, at least it would be good to acknowledge that other approaches to assessing this kind of error are possible e.g. Wheatley et al (2012) [Wheatley, J. J., Blackwell, P. G., Abram, N. J., McConnell, J. R., Thomas, E. R., and Wolff, E. W.: Automated ice-core layer-counting with strong univariate signals, Clim. Past, 8, 1869-1879, doi:10.5194/cp-8-1869-2012], which lead to the prospect of a more coherent representation of their dependence structure.
I think that this kind of improvement is important, not least because of the artificiality o the tricks that were needed in some of the earlier analysis described in the Introduction. However, I think the nature of the measurements and the modelling of the size and dependence of the errors need to be defined and explained much more clearly, and separated from the details - however important - of GICC05 (and of the coding). Only then can the merits of the methodological change be properly judged.

→ A full section (2.3) is now devoted to the treatment of the MCE and it should address this comment.

Specific comments
The terminology relating to estimation in Datice is unclear for those not already familiar with it. Is the "first guess" essentially a prior mean? If so, the terminology is rather misleading, to me; it may be established, for this particular case, but should nevertheless be explained a little. Since the method is specifically described as Bayesian, it seems important to relate the description of the method to standard terminology.

→ We do not fully understand this comment since, to our knowledge, we have used the standard terminology. However, we have kept this comment in mind and section 2.1 also includes now some more precisions on terminology.

The bayesian problem is dealt through variational optimisation with a cost function and the adapted language is indeed to use « background » and « first guess ». "First guesses" Ab, Tb and Cb are the median of "prior probability distributions" that are given as lognormal distributions.

As presented in section 2.1, the bayesian problem is formulated and solved through variational optimisation with a cost function J :

J(x) = - log(P(x|y)) = - log (P(y|x) Pb(x))

where P(x|y) is the posterior probability distribution, product of the prior distribution Pb(x) with the likelihood function P(y|x). x is the model that we want to find (i.e., A, T and C) and y are observations.

The prior probability distributions (Pb) on A, T and C are assumed to be lognormal distributions and what is described in the text as "first guess" or "background" are the medians of these lognormal distributions:

Ab = median of the prior distribution Pb(A)
Tb = median of the prior distribution Pb(T)
Cb = median of the prior distribution Pb(C)

In the text, we also used the denominations "first guess" or "background" for the ice and gas chronologies (Psi_b and Xi_b) that we can calculate from Ab, Tb and Cb (see equations A1 and A3, Appendix A).

These multivariate lognormal distributions contain error covariance matrices (Balpha, Btau, Bgamma) that define the asymmetric error for the medians Ab, Tb and Cb. Still, to work with normal probability distributions and thus a cost function made of a sum of quadratic terms, i.e., || x - xb ||², we change the variables to tilde_alpha, tilde_tau, tilde_gamma so that:

tilde_alpha = log (A/Ab)
tilde_tau = log (T/Tb)
tilde_gamma = log (C/Cb)

Similarly, we call tilde_alphab, tilde_taub, tilde_gammab "background" or "first guess". We can use such a language since all these quantities are linked by formulas:

tilde_alphab = log (Ab/Ab) = 0
tilde_taub = log (Tb/Tb) = 0
tilde_gammab = log (Cb/Cb) = 0

We do not feel that the above discussion should be included in the manuscript but it can easily be implemented in the first annex if needed.

The same applies to the discussion of the cost function in 2.1. In a way this is even more confusing - or at least distracting - because the cost function suggests something different in a Bayesian context.

→ We are sorry but do not understand what language problem the reviewer refers to. We hope that the precisions added along section 2.1 removed the confusion.

Equation 4 seems unnecessarily off-putting - or redundant for those readers less likely to be put off. Relating the function J(.) to a set of independent multivariate normals would be clearer and would justify its form. In addition, the details of the R matrices are needed (as mentioned above) as well as a description in more standard statistical Some of the discussion of the fitting talks about curves meeting the constraints of the error bars. Again, that seems to be at odds with the Bayesian approach. Are the error bars part of the process, or is this description purely figurative?

→ We think that the new manuscript, now strongly devoted to methodological developments and validation / treatment of the error answer the comments raised here. The J(.) function is indeed the logarithm of a product of « independent multivariate normals », and it does fully account for errors by the mean of the error covariance matrices.

In general, it would be much better to talk about the model structures and parameters rather than focus on features of the code.
Technical corrections
There are some technical corrections, mostly minor, to follow. For now, I would just like to say that, in the title, "implementing counted layers" doesn't sound right, and doesn't quite reflect the emphasis on age differences within the body of the paper.

→ The title of the manuscript has also been modified.

Reviewer 2

Bazin et al. present a new way of incorporating depth-age constraints from an annual timescale into the Datice method. The previous AICC2012 chronology for 5 ice cores used the GICC05 timescale of the NorthGRIP ice core as the base chronology; however, using GICC05 as the reference chronology was not straightforward with the Datice methology and violated the assumption that the initial (background) chronology was independent from the age markers. This modified formulation allows the GICC05 age constraints to be included without violating the independence of the initial chronology. The result is timescales for the 5 cores that are similar to AICC2012 although th authors note that these new chronologies should not be used in analyses of ice cores. Bazin et al., using the inferences of firn thickness from previous authors, infer lower glacial accumulation rates at North GRIP than were inferred from GICC05 and ice-flow modeling.
The approach of Bazin et al. is logical and fits in well with the overall Datice methodology. The authors demonstrate that the new age constraints are functional within the Datice framework. However, I have two major issues with the work, one general and one technical.
My general issue with the presented work is that I'm not convinced this improves the NGRIP timescale or the Antarctica timescales. The authors seem to agree that the new timescale for NGRIP is not an improvement as they note that it should not be used. That the final inferred timescale remains similar to GICC05 is likely because the age-interval markers used in the new formulation are still the primary age constraints - what data from the Antarctic cores they choose (EDC, EDML, Taldice, Vostok) are going to improve the dating of NorthGRIP? In the absence of an analysis of whether the NorthGRIP timescale was improved, I at the least expected a discussion of how these age markers would in the future lead to better timescales. The lack of discussion was especially disappointing as an annually resolved timescale for the past 30 ka was published for the WAIS Divide ice core a year ago. EDML also has an annual timescale for the past 10 ka (although the in preparation paper cited in the AICC2012 papers appears to still be in preparation). Is the methodology developed in this paper going to be able to make use of these annual timescale for Antarctica

→ Indeed, we realize that the previous manuscript was misleading. Our aim was not to provide a new timescale but to test a different (and more correct) way to implement absolute dating constraints for NGRIP. This is more a methodological development that could be used later for many dating exercises and indeed for layer counting in many other ice cores (including the WAIS ice core of course). This is now clearly mentioned in the conclusion with a reference to the WAIS paper using layer counting for the chronology of the younger part.

I would have preferred to see an analysis of whether this new implementation can improve the timescale. Why not use synthetic timescales and test the approach? The authors have already defined the uncertainties for the all the different ice cores, so creating synthetic timescales should be relatively straightforward. Then applying Datice to these timescale where the “true” timescale is already known could yield significant insight into not just the new formulation for Datice but for Datice itself

→ This is an excellent suggestion and we have followed it by implementing section 2.2 with twin experiments to validate the new methodological development.

My technical issue relates to the thinning function. The authors were kind enough to supply the Datice output. As the authors briefly allude to, the thinning function is no longer smooth. It also does not decrease monotonically. Ice of 30 ka age has thinned significantly less than ice of 20 ka age. This is a puzzling scenario and one that needs to be discussed by the authors. While Parrenin et al. (2004) showed that such variations could occur at Vostok, this was because of the large changes in the ice thickness and convergence or divergence along the flow line. Neither of these situations is applicable to NorthGRIP which is located on the divide and the slight flow down the ridge to the core site has no significant ice thickness variations (Dahl-Jensen et al., 2003). The authors hint that the cause might be impurity driven; however, softer (to shear) LGM ice will not simply cause the LGM layers to be thinned more. To do so would violate continuity. For the derived thinning function to be acceptable, the authors must show that it can be recreated with an ice flow model and conditions of the NorthGRIP site.
The unphysical thinning function appears to be the result of the inappropriate constraint on the thinning function. If I'm following the description in the appendix, the thinning functionThe authors are incorrect to conclude that “our study confirms the overestimation of GICC05 accumulation.” They have only shown that if you remove constraints based on the physics of ice flow, you can infer an accumulation rate and timescale consistent with the firn-based accumulation reconstructions. This work sheds no light on why the ice-flow-based and firn-based accumulation rates reconstructions disagree.

→ The whole discussion on the glaciological parameters and hence on the thinning function has been removed. Actually, many tests were already performed during the realisation of AICC2012 to study the behaviour of the thinning function in the different sites and avoid as much as possible the strange behaviour as described by the reviewer. Such behaviour when using multiple ice cores with numerous stratigraphic links is however unavoidable: we necessarily obtain some fluctuations in one or several ice core thinning function. This means that (1) either the stratigraphic links are questionable (some of them were removed after this observations when constructing AICC2012, see papers by Bazin et al., 2013 and Veres et al., 2013) or that (2) the thinning function may indeed show some non monotonic evolution with depth at a certain depth level. In order to prevent such behaviour of the thinning function at shallow depth, we impose a small variance to background scenario for the thinning function for small depth (see SOM of Bazin et al., 2013 and Veres et al., 2013 + appendix of this work). A next and more physical step will be to directly implement glaciological model within a model like Datice or Ice Chrono to deal more properly with this aspect. This is however beyond the scope of this study.

Specific comments:

→ We have taken into account the specific comments that were relevant with the new manuscript dealing much more with methodological issues than application for a new chronology. See below

Introduction -The introduction is very focused on problems in implementing GICC05 in AICC2012. I think the focus is too narrow as this should be about improving ice core timescales, not Datice. It might also be a wise idea to discuss some of the other advances in ice core dating that are occurring and how this work will complement them.

→ It was difficult to really extend the introduction with the new manuscript much more focused on the method itself but still, we refer much less to AICC2012 and particular glaciological parameters of NorthGRIP and present the method improvement as something to be used in general for different ice core dating.

-The reference of Cutler et al. (1995) for a 30% uncertainty in accumulation and thinning function seems overly simplified. The uncertainty in the thinning function is depth dependent. The uncertainty at the surface is essentially 0, while uncertainty at the base of the ice sheet is essentially infinite. In addition, the uncertainty between accumulation rates for ice with similar ages is much reduced.

→ We agree that the uncertainty in the thinning function is depth dependent. This is the way the variance associated with thinning function has always been implemented within Datice (Lemieux-Dudon et al., 2010 ; Bazin et al., 2013 ; Veres et al., 2013) as depicted also in the second annex of the manuscript. Because of mathematical problems linked to inversion, it is not possible to have strictly 0 at the surface but we impose a very small value.

- The accumulation inferences from d15N rely on accurate firn densification modeling. This uncertainty needs to be discussed as the PIRE firn project has shown that there is great uncertainty in the firn models even in steady-state modern conditions, let alone transient glacial conditions.

→ We do not use the accumulation rate deduced from firnification model and d15N in this version. We have also added a precision and two references in the introduction concerning the uncertainty in the firn models : the uncertainty essentially applies to Antarctic firn and not Greenlandic ones.

Methodology

- I don't understand what constraints the Antarctic ice cores provide to the NorthGRIP chronology. It seems like there is so much uncertainty in the ice and gas timescales from these low accumulation rate East Antarctic ice cores that it would be better to just run NorthGRIP by itself

→ In the new version, validation and sensitivity tests to provide guidelines for future dating exercises have been done using NGRIP only.

- The references to Buiron et al. 2013, Veres et al., 2013, and Bazin et al., 2013 reminds me that the thinning functions for EDML and Talos Dome also produce reversals in the thinning function for ice in the upper half of the ice sheet. I suspect that thestruggles with the thinning function at NorthGRIP apply to the Antarctic cores as well.

→ See answer to the technical comment above

-There is something unsatisfying about starting with background scenarios that produce timescales that we know are inaccurate. But I guess I have a larger question: are the background scenarios self-consistent? If I understand correctly, the background thinning function, accumulation rates, and lock-in-depths are all independently derived and would not produce a realistic timescale.

→ Indeed, using independent background scenarios for thinning, accumulation and LIDIE can produce a non realistic timescale as a background. Integrating all constraints within a Datice run then enable one to converge toward a realistic timescale with the more probable glaciological parameters associated.

Chronological and climatic implications

- I am surprised that there is no discussion of the gas ages associated with GICC05. While I understand that GICC05 is technically just a timescale for the ice, it seems like the gas timescale will be more affected than the ice timescale.

→ This is now beyond the scope of this methodological paper.

Appendix L21

- these numbers of cT2 don't have any meaning. What is this correction physically?

→ cT2 is the parameter describing the increase of the variance of the thinning function with depth.

L24- is this just saying that the uncertainty on the thinning function at the surface should be zero? If so, why write "the 0 variance hypothesis" ?

→ As explained above, it can not be 0 at the surface because of mathematical problems.

References -The references seem a little short on timescale work from non-European countries.

→ Reference to the WAIS paper including layer counting has been added

Table -Add the event name and approximate age for each event so readers can readily understand where the new delta-depth markers are being applied.

→ Done

-What is the justification of the uncertainties

→ The uncertainties come from the resolution of the measurements and different assumptions to calculate the Ddepth (either the differences between the mid-slopes of d18O and d15N increases or the differences between the peaks of d18O and d15N increases). This is now added in the text.

Figures

- Figure 4 has too any lines which are not visible. I'm not sure what the main point of the figure is, so I can't suggest a better presentation. - Figure 5 has too many lines as well.

- Figure 7 is unintelligible.

→ All figures have been modified and figure 7 removed.

Reviewer 3

The manuscript describes a methodological extension of the Datice model used to create the AICC2012 ice-core time chronological framework. Although mainly of technical interest at this point, it is a methodological advance of potentially very significant importance that addresses one of the major shortcomings of the previous model: the lack of ability to include the results of annual layer counting in the model in a way that respects the nature of layer counting. Therefore, I believe the results are appropriate for publication in CP or possibly even more appropriate for publication in GMDD. However, as the AICC2012 papers were published in CP, it makes sense to publish the next step here too. However, the proposed methodological extension seems to have some weaknesses that seriously limits my confidence in the results. If these weaknesses can be addressed satisfactorily, the method will a valuable addition to the Bayesian ice-core chronological modelling framework, and I therefore encourage that the authors try to address these issues.
In addition to the comments on the content below, the manuscript is in need of significant improvements in language and clarity. I have not attempted to comprehensively mark up language errors or minor issues in the presentation, but encourage the authors to make sure that the English is significantly improved before submitting a revised version.

Detailed comments:
Page 3587
Line 7: “Tuning” is an understatement. AICC2012 was forced to fit GICC05 back to 60 ka.

→ OK, this has been changed

Line 10: I do not think that “markers” is a good description of a duration estimate from layer counting. Similarily, “age-difference” is an unclear term to use for durations.

→ We understand the request of the reviewer. Still, we are not referring to any particular event that can have a duration. Actually, in the GICC05 approach, we better have age markers given every 20 years than duration of events for fixed depth levels. This is the reason why we stick with age markers and avoid to use duration in general.

Page 3588
Line 2: “1–8years for counting of 20 annual layers” Please refer to where this comes from or revise (8 out of 20 years is an order of magnitude more imprecise than the published estimates I am aware of).

→ This has been corrected to 0-4 years for counting of 20 annual layers.

Line 3-4: “Since the layer counting is not independent from one interval to another, the final uncertainty on the GICC05 chronology cumulates the counting error”. Both parts are true, but one does not imply the other: uncertainties in counted chronologies cumulate down core regardless of whether the counting is independent between intervals.

→ As we explain now in section 2.3 on the maximum counting error and its implementation in Datice, the counting error indeed cumulates but not in the same way if the errors are correlated (addition of the errors) or non correlated (addition of the squares of the errors).

Line 6-7: GICC05 odes not strictly speaking have a 1-sigma uncertainty. Setting MCE/2 = sigma is an assumption.

→ This is indeed true and we have devoted this section 2.3 to really explain it better.

Line 23: I believe that using naïve "first guesses" would be a bad idea, as the inversion cost function depends on the distance between the solution and the background scenarios. The model therefor "prefers" to give solutions as close to the background as possible. The background scenarios should be independent from the stratigraphic constraints in order to satisfy the underlying assumptions, but must represent fairly good guesses of the accumulation, thinning and LIDIE in order for the model to produce good results. Figure 4 clearly shows the non-trivial influence of feeding the model with different background scenarios (e.g. at depths around 2300 m).

→ The reviewer is right again and this is a very important issue. This is the reason why in some sensitivity tests and validation of the method, we have discussed different cases for the relative uncertainty associated with the background scenario and uncertainty associated with the observations. Appendix E has been devoted to this issue « Balance between background and observation error and impact on the analysis ».

Page 3590
Line 8-9: "In this paper, we propose an improvement of Datice to better implement the chronological uncertainties". I believe the aim is to better represent the results of layer counting (and the associated uncertainties).

→ Section 2.3 addresses now fully this problem of uncertainties.

Page 3591
Line 6-7: . . . provided that the uncertainty of all constraints and background scenarios are properly quantified (and that any correlations between the uncertainties are also properly described and modelled adequately). This may sound trivial, but I think it is far from that . . . I agree that the model is useful because it provides the best compromise between a lot of different information under a certain set of assumptions, but the assumptions are not trivial, are most likely not fully met, and thus cannot be ignored.

→ We agree and this is the reason why we removed the discussion here on glaciological parameters and refrain even more to propose any new chronology.

Page 3592
Equation 4 and text around it: Does this mean that all types constraints are added with equal weight (apart from the weighting that depends on the uncertainty/confidence), or is some normalization performed so that, for example, the cost function contribution of all "gs" and all "ad" type constraints are the same?

→ The terms of the cost function are only weighted according to the error covariance matrices R and B (i.e., uncertainties of the markers and background). No extra normalization factor is introduced between terms.

Page 3593
Line 15: The term “analysed chronology” confused me. I believe it is the model output, but it is not fully clear to me.

→ Indeed, it is the model output. We have made it clearer in the new text.

Page 3594
Line 4. “Largely” is used in a wrong meaning.

→ Text has been changed

Line 4-6. This illustrates how difficult it is to argue that the background scenarios and their uncertainties are well-represented in the model input, and thus that the result is the “best solution”.

→ See above comments on the relative weight of background and obervation uncertainties. Indeed, the model remains a tool and a lot of subjective choices are left to the user. Here, we only concentrate on the methodological development and validation.

Page 3595
Line 8 and 13: “time interval” sound like an absolute time period and “periodicity” sounds like something with an oscillation. I believe that “resolution”, “spacing”, or another similar word would more clearly convey the message.

→ OK

Line 13-16: The assumption of absence of correlation between the uncertainties of the layer count in neighboring sections is most likely not justified. When counting layers, certain typical features that may or may not represent an annual layer occur repeatedly. If such recurrent features are misinterpreted in the layer counting procedure, they will give rise to correlated errors. And why 2xMCE? It seems like a rather random choice that must be justified.

→ These issues have been addressed in section 2.3

Line 16: “We decide in this paper to treat the error on individual annual layer as normally distributed. On this assumption, one can apply the MCE to age-difference markers as a Gaussian error”. Sorry, but I do not follow the argument.

→ It is now explained in section 2.3

Line 18-20: “The periodicity of markers of age-difference in GICC05 is 20 years with a 1 to 8 years associated uncertainty (Rasmussen et al., 2006)”. Firstly: The GICC05 data files are released in 20-year resolution but GICC05 is an annually resolved time scale. So the 20 years are not reflecting any intrinsic resolution. Secondly, the 1-8 years uncertainty for each 20-year interval seems very large. Are you sure? The GICC05 MCE is on average around 5% or less, very far from 8 out of 20.

→ The 8 years have been corrected (4 years indeed). And the reviewer is right about the sampling issue which is fully subjective. Section 2.4.1 now better addresses this issue.

Line 21: How do these inconsistencies occur? Or is it more a numerical limitation as the next sentence indicates?

→ The inconsistencies linked to the sampling at too high frequency are explained in section 2.4.

Line 25 and Fig 2: One thing is that a short duration (ie. 20-60 years) of counted intervals leads to numerical problems (or inconsistencies . . . see previous comment), but it is very worrying that the results seem to depend critically on the duration/resolution or the annual layer counting constraints used. Roughly speaking (by comparing the blue and cyan curves of figure 2), the deviation of the newly modelled time scale from GICC05 doubles when the number of constraints is reduced to half. As no a priori "correct" choice of marker/constraint resolution exists, and the model seems to produce a different result for different choice of marker/constraint resolution, the solution is not unique, but depends on a user-chosen arbitrary value for the marker/constraint resolution. The issue relates to the question to page 3592: How do the results depend on the number of markers used, and is some kind of normalization required (or desirable)? This issue needs to be further explored. It is possible that a similar issue could exist for other model constraints or for the resolution of the background scenarios. To summarize: At this point, I get the impression that the "best compromise" solution obtained depends critically on the "age-difference" marker resolution (and possibly on the resolution of the background scenarios), which has no intrinsic "best value". The "best compromise" thus depends on a rather arbitrarily chosen parameter.

→ The reviewer is right and this is also one of the reasons why we have fully rewritten the paper to addressthis issue. We now present several sensitivity tests regarding the sampling, the correlation between errors and propose intermediate solutions like adaptative sampling (high sampling at shallow depth and low sampling at higher depth) associated with finite range correlation. As the reviewer says, there is not any obvious best choice and this is the reason why we only provide some guidelines from our sensitivity tests at the end of section 2.4.

Page 3596

Line 18: Now sigma = MCE. It is still double the quoted uncertainty and still a rather arbitrary choice. Please argue for your choice.

→ section 2.3

Line 19: ". . . the error correlation value has been varied between its minimum and maximum possible values." What are these values and how are the limits determined?

→ see section 2.4.2 fully devoted to the error correlation.

Line 21: "As for the chronology built without correlation, the maximum difference between the new chronologies and GICC05 is of 130 years (Fig. 3)". There is a problem with the sentence structure.

→ removed

Line 26-28: Why is this expected?

→ text has been fully changed so that this comment does not apply to the new version.

Line 10-28: It is great that authors try to test how different ways to add up counting uncertainties influence the results. This is a weak point of GICC05 and many pother counted time scales. However, the way this is done needs more explanation (especially what the physical meaning of the error correlation value is). Also, sensitivity studies should be performed to show how the results depend on the choice of 100-year spacing of the markers in the case of correlated errors. As an example, consider a period of 1000 counted years with 50 years MCE evenly spread over the interval. In case the section is split up into 20-year sections each with MCE 1 year and the errors (for simplicity) are assumed to be fully correlated between sections, the total uncertainty would be 7 years (sqrt(50)). In case the splitting is made into 10 100-year long intervals each with MCE 5 years, the total uncertainty would be 16 years (sqrt(10*5ˆ2)), while the total uncertainty would be 35 years if the section is split up into 2 sections each 500 years long and with MCE 25 years. The difference is probably smaller in your case as you do not assume full correlation, but I would guess that the effect would be there (again emphasizing that the somewhat arbitrarily chosen 100-year marker spacing leads to non-uniqueness of the solution).

→ This is now addressed in section 2.4.1.

Line 6: The statement "This result is not unexpected for two reasons" has some very complex logical implications :o) Please clarify. I also have a hard time understanding why the absence of a direct link between MCE/2 and a Gaussian sigma (partially) explains the result. The explanation could also simply be that MCE is a conservative uncertainty estimate.

→ Does not apply to the new text.

The remaining comments (below) concerned the glaciological applications that we prefer to separate from this methodological paper. They should better be discussed in a future coherent dating exercise involving a larger « ice cores » community

Line 13: It seems to be a stretch to claim that the method have been tested for correlation values larger than 0.6 when the full model does not run with these values.
Line 19-22: Can you back this statement up with data?
Line 23: Are the results with error correlation 0.5 with sigma = MCE (as above) or 2MCE? If sigma = MCE, I find it hard to compare the two sets of results.
Page 3598
Line 26-28 and Fig. 4: Yes, except around 2300 m depth where the PK2014 background scenario has very low accumulation rates and drags the solution away from the other curves. It could indicate that the uncertainty of the PK2014 scenario is too low across this section.
Page 3599
Line 10: "This conclusion is mainly valid for the last 60ka". I would say "only", not "mainly"
Line 31: "The #depth constraints from #15N measurements cannot be questioned" is an overstatement. High confidence does not imply that a result cannot be questioned. For example the argument relies on a pretty large assumption, namely that #18O directly reflects surface temperature. Rephrase and add modesty, please.
Page 3600
Line 11-26: The procedure is worrying to me: The LIDIE results are too variable, so they are forced to disappear by reducing sigma_B,L. Is there a physical meaning of this parameter adjustment? What are the possible reasons for these problematic results? Could it be an indication that some of the data constraints are internally inconsistent?

If you decide to keep the current solution where you force LIDIE to have smaller variability, I think you at least should discuss more carefully whether the observed dips in LIDIE are indeed too large to sensibly be climate-related (they sit around 2050, 2200,
C1708
and 2300 m, corresponding to GI-8, 12, and 14, which are the longest interstadials in that part of the record).

Page 3601

Line 16-18: These two lines and Fig. 7 seem detached from the rest of the manuscript. The chronological adjustments are not of large importance for a bipolar seesaw discussion, and in any case, the sentence and Fig. 7 must either be extended to an entire paragraph or removed. I suggest removal as the bipolar seesaw discussion seems outside the scope of the manuscript.

Line 22: Sentence unclear.

Page 3602

I agree that the GICC05free results should not be recommended for use over AICC2012 because of the small scale of the differences (i.e., the possible chronology improvement does not outweigh the hassle caused by having many parallel and almost identical time scales) and because the results seem to be dependent of model implementation choices that are somewhat arbitrary (see above).

Caption of Fig 5

“. . . over the glacial period” is not correct

“AICC2012 + #depth” is shorthand and not easy to understand for non-specialists.