Reply to the comments by Referee # 1

We would like to thank the referee for the valuable comments and suggestions. In the following, the comments by the referee are listed in Italic, and our reply is provided for each comment in Roman.

(We have made some corrections from the version posted as an interactive coment [npgd-2-C548-2015.pdf]).

Comment:

. . . Indeed different results are to be expected, but it would be good to think of some criterion to evaluate the performance of the models in an objective way. In which way is the proposed model better than previous ones? External validation of the results would perhaps be a convincing way to promote the approach, and here are a few things that come to mind.

- *The prior distribution on the parameters should be very explicitly described (all the results depend on it). Then, the posterior distribution could be compared to the prior, for instance using overlaid kernel density plots. This would give a visualization of how much information is gathered on the parameters from the data. Perhaps some parameters are easier to estimate than others?*

Reply: We omitted to specify the prior distribution. We appreciate the referee for pointing out that.

In this paper, a uniform distribution is used as the prior distribution of each parameter. If we want to use a different prior distribution, the posterior distribution can be obtained as a product between the prior distribution and the histogram in Figure 1. We have added a description on the prior distribution (P. 7, L. 210).

Comment:

- *A simulation study on synthetic data generated from the model would also be informative. How precisely can we identify the model parameters using synthetic dataset (using the same number of observations as in the real dataset)?*

Reply: In this paper, we consider a situation where the model is an approximation of the actual process. If we generate a synthetic dataset under given parameters, the spread of the parameters would be very small, which is highly different from the actual situation. We have no idea how we can appropriately generate a synthetic dataset good for a benchmark of our method.

Comment:

- *The model quality could be evaluated based on its predictive performance: for instance, the parameters could be inferred using the first 80% data points, and the remaining 20% data points could be predicted. Certainly other criteria could be envisioned, such as Bayes factors. In fact the statistical literature is quite rich on this topic (see [6]).*

Reply: We tried the estimation without using the last five agemarkers, i.e., we used the first 80% of the age markers and $\delta^{18}O$ data. Figures A and B are the age–depth relationship and the accumulation as a function of age estimated without using the last five agemarkers. The estimate using all the age markers is also indicated with grey lines. The age–depth relationship was satisfactorily estimated althogh the estimate of the age as a function of depth was very slightly different near the bottom.

The accumulation–age relationship was also estimated without using the last five agemarkers. The difference from the result using all the age markers is visible near the bottom. However, the difference was mostly within the uncertainty between the 10th and 90th percentile. Thus, this difference near the bottom would be acceptable.

In the revised version, this point is discussed in the beginning of Section 6.

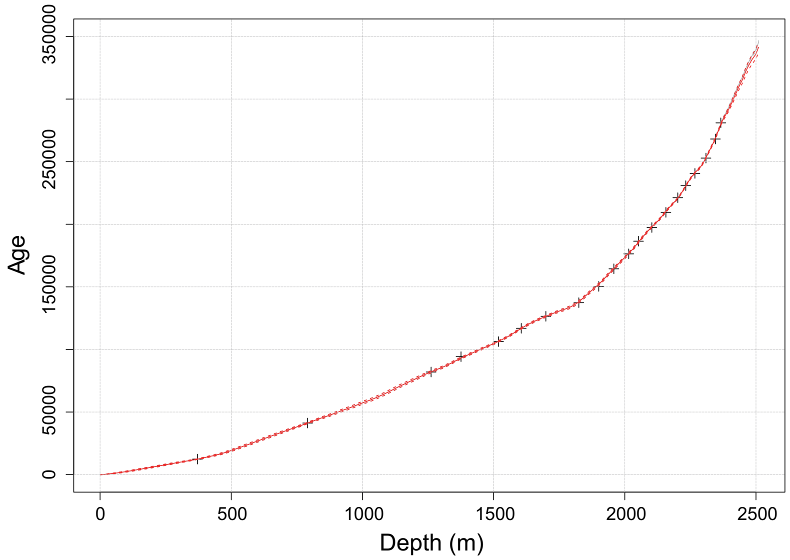


Fig. A. Estimated age without using the last five agemarkers (red lines) and estimate using all the agemarkers (grey). Each solid line indicates the median of the posterior distribution. The 10th and 90th percentiles of the posterior are indicated by dotted lines.

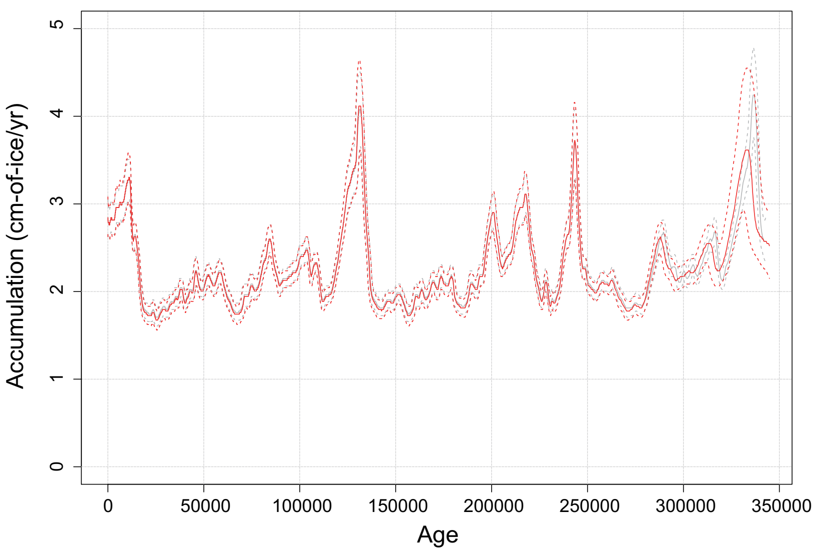


Fig. B. Estimated accumulation rate as a function of age without using the last five agemarkers (red lines) and estimate using all the agemarkers (grey).

In the PMCMC, the likelihood of the parameter vector $\boldsymbol{\theta}$, $p(\boldsymbol{y}_{1:Z}|\boldsymbol{\theta})$, is evaluated at each iteration of the MCMC procedure. In other words, the predictive performance for a given $\boldsymbol{\theta}$ is evaluated at each step of the MCMC. Thus, we think the procedure of the PMCMC would provide a good choice of the parameter in terms of the predictive performance.

Comment:

– *Some aspects of the model seem more arbitrary than others: Gaussian distributions for the noise distributions, the accumulation rate is a random walk process (why not an autoregressive process?). The article could either justify these modelling assumptions in more details, or test various model modifications in practice. The resulting models could be compared, again, using predictive criteria or Bayes factors.*

Reply: It is true we can use various noise distributions and we can consider various models for the accumulation rate. We could select among them by using some metric such as Bayes factors. However, there are a large number of choices, and thus it would take much time to make the selection among those choices. In this revision, we just add a mention about the fact that we can consider various models and we can select among them by using some metric such as Bayes factors (3rd paragraph of Section 7). We will examine the performance of other models in the future works.

Comment:
· The language is clear. The general descriptions of the model and of the methods are fine, but the article should allow readers to reproduce the results; it is not the case here by lack of implementation details (lack of details on the prior distribution), details on the proposal distribution $q(\theta'|\theta)$, etc). Perhaps an appendix could give all the values used in the implementation that are not specified in the main text.

Reply: It is true that we omitted to provide the information on the proposal distribution for the MCMC (Metropolis method) part. In this paper, a zero-mean Gaussian distribution is used as the proposal distribution for each parameter. The variance of the proposal distribution for each parameter is given in the text (P. 11, L. 304). As described above, the prior distributions of the parameters are uniform distributions.

Comment:
· Inconsistent notation: δ^{18}O *or* $\delta^{18}\mathrm{O}_z$ *or* $\delta^{18}\mathrm{O}(z)$ *data.*

Reply: We unify thos expressions into "δ^{18}O data".

Comment:
· State space models are called "sequential Bayesian models" in the article, which is non-standard and a bit misleading, because nothing is really "Bayesian" about them (Bayes formula is just used to obtain the recursion formula for the filtering distributions). "Bayesian" usually refers to inference methods treating parameters as random variables, and does not refer to models. Hence, non-Bayesian approaches could have been applied to the model of the article. Another common term for state space models is "hidden Markov models".

Reply: It is true that the word "sequential Bayesian models" was not appropriate. We refer to it as "state space model" (P. 2, L. 48).

Comment:
· The model description is split into Section 2 & 3, starting in "continuous time" and with the description of Θ_z (Section 2), and then switches to discrete time and to the description of A_z and of the measurement distributions (Section 3). These sections could perhaps be combined in one section.

Reply: Section 2 is intended to review the glaciological model proposed by the existing study (Parrenin et al., 2007). Section 3 is intended to formulate a state space model based on the glaciological model in Section 2, which might be common in the glaciological community. That is the reason why we divided into two sections.

Comment:
· page 945: Equation (14) should read

$$p(\xi_{z+1}|\xi_z,\boldsymbol{\theta}) = \mathcal{N}\left(\xi_z + \frac{1}{A_z\Theta_z},\, \sigma_\nu^2\right), \tag{1}$$

according to Equation (12)...?

Reply: We appreciate the referee for this correction. We have corrected Eq. (14).

Comment:
· page 946: when there are multiple observations δ^{18}O within an interval of one meter, a mean is used (presumably, without modifying the standard deviation σ_w). This seems unfair, as when there are more observations, the uncertainty should be reduced. One simple approach would be to use the mean of the observations at each meter, but with a variance σ_w divided by the number of observations.

Reply: The time sequence of δ^{18}O contains short-term fluctuations. These short-term fluctuations are regarded as noises, which are difficult to model. However, they have short-term auto-correlation, and δ^{18}O within one meter interval usually takes similar values. Therefore, even if a mean of multiple observations within one meter interval is used, it would not be appropriate to divide σ_w by the number of observations.

As a matter of fact, even if a mean δ^{18}O value for each one meter is used, the amount of the δ^{18}O data is still too large. Because of the large amount of the δ^{18}O data, the likelihood for each MCMC step becomes sensitive to the parameter $\boldsymbol{\theta}$. In order to relax the sensitivity, we introduce a relaxation in calculating the likelihood for an MCMC step. We found we omitted to describe about this relaxation and it is now described in the revised version (L. 10, P. 284–289).

Comment:
· Again, the prior distribution on the parameter θ should absolutely be specified somewhere.

Reply: As described above, we use a uniform distribution as the prior distribution of each parameter. We have added a mention on the prior distribution in the revised manuscript.

Comment:
· page 948: why isn't σ_ε included in the parameter θ? More details should be given on this. Does the method fail if this parameter was included in θ? What are the values given to it, in the end?

Reply: The standard deviation σ_ε is provided together with the age marker data by Kawamura et al. (2007). The following table (Table A; Table 2 in the text) shows the depth and age for each age marker (tie point) as well as σ_ε.

Comment:
In the proposed model, the transition is non-linear (because of the term $1/A_z\Theta_z$) but the noise distributions are Gaussian (if we use the parametrization $x_z = (\xi_z, \log A_z)$ instead of $x_z = (\xi_z, A_z)$). Thus, a "locally optimal" particle filter approach could be implemented, that is, instead of propagating the particles using $p(x_{z+1}|x_z,\theta)$ and weighting using $p(y_{z+1}|x_{z+1},\theta)$, one could sample from $p(x_{z+1}|x_z,y_{z+1},\theta)$ and weight the particles using $p(y_{z+1}|x_z,\theta)$; these two distributions are Gaussian. This is called the optimal proposal scheme in [1]; it could reduce the variance of the likelihood estimator.

Depth	Age	Uncertainty of the age ($2\sigma_\varepsilon$)
371.00	12390	400
791.00	41200	1000
1261.61	81973	2230
1375.67	94240	1410
1518.91	106263	1220
1605.27	116891	1490
1699.17	126469	1660
1824.80	137359	2040
1900.74	150368	2230
1958.31	164412	2550
2015.01	176353	2880
2052.23	186470	2770
2103.14	197394	1370
2156.67	209523	1980
2202.02	221211	890
2232.45	230836	780
2267.28	240633	1230
2309.35	252866	1160
2345.32	268105	1980
2366.01	280993	1600
2389.31	290909	1210
2412.25	301628	880
2438.37	313205	840
2462.36	324774	1110
2505.4	343673	2000

Table A. The depth, the age, and the uncertainty (2σ) of the age at each tie point.

Reply: As mentioned in Concluding Remarks, we are considering to improve the proposal scheme in the future. However, we are also considering to extend the model and possibly we may use non-Gaussian distribution for the noise distribution. That is the reason why we have not yet tuned the proposal scheme for the SMC part.

Comment:
page 951, line 11: "using the SMC" + algorithm ?

Reply: Right. We meant the SMC algorithm.

Comment:
page 952, line 5: perhaps give the formula for the likelihood estimator, since it is quite central in the particle MCMC method?

Reply: We thank the referee for the suggetion. We approximate $p(\boldsymbol{y}_z|\boldsymbol{y}_{1:z-1},\boldsymbol{\theta})$ as follows:

$$
\begin{aligned}
&p(\boldsymbol{y}_z|\boldsymbol{y}_{1:z-1},\boldsymbol{\theta})\\
&\quad=\int p(\boldsymbol{y}_z|\boldsymbol{x}_z,\boldsymbol{\theta})\,p(\boldsymbol{x}_z|\boldsymbol{y}_{1:z-1},\boldsymbol{\theta})\,d\boldsymbol{x}_z\\
&\quad=\int p(\boldsymbol{y}_z|\boldsymbol{x}_z,\boldsymbol{\theta})\,p(\boldsymbol{x}_{0:z}|\boldsymbol{y}_{1:z-1},\boldsymbol{\theta})\,d\boldsymbol{x}_{0:z}\\
&\quad\approx\frac{1}{N}\sum_{i=1}^{N}\int p(\boldsymbol{y}_z|\boldsymbol{x}_z,\boldsymbol{\theta})\,\delta\left(\boldsymbol{x}_{0:z}-\boldsymbol{x}^{(i)}_{0:z|z-1}\right)d\boldsymbol{x}_{0:z}\\
&\quad=\frac{1}{N}\sum_{i=1}^{N}p(\boldsymbol{y}_z|\boldsymbol{x}^{(i)}_{0:z|z-1},\boldsymbol{\theta}),
\end{aligned}
$$

where we used the assumption introduced in Page 947:

$p(\boldsymbol{y}_z|\boldsymbol{x}_{0:z},\boldsymbol{\theta})=p(\boldsymbol{y}_z|\boldsymbol{x}_z,\boldsymbol{\theta})$.

We then approximate the logarithm of $p(\boldsymbol{y}_{1:Z}|\boldsymbol{\theta})$:

$$
\log\hat{p}(\boldsymbol{y}_{1:Z}|\boldsymbol{\theta})=\sum_{z=1}^{Z}\log\left[\frac{1}{N}\sum_{i=1}^{N}p(\boldsymbol{y}_z|\boldsymbol{x}^{(i)}_{0:z|z-1},\boldsymbol{\theta})\right].
$$

These are described in the revised text (P. 10).

Comment:
page 952, not much details is given on the tuning of the proposal distribution $q(\theta'|\theta)$. How is the variance tuned? Using preliminary runs?

Reply: We performed some preliminary runs to find out the landscape of the posterior distribution. Then, the width of $q(\theta'|\theta)$ was taken to be small enough in comparison with the width of the target posterior distribution.

Comment:
page 952, Equation (35) and onwards: it is not very clear that only the likelihood estimator $\hat{p}(y_{1:Z}|\theta^*)$ on the numerator is calculated at each step, and that the one in the denominator is kept fixed. The method would not be valid if both the numerator and the denominator estimators were drawn at each step.

Reply: The denominator is kept fixed at each step. We have modified the desciption (P. 11, L. 308).

Comment:
page 953, line 13: "this greatly reduces the computational cost": does this refer to the memory cost instead of the computational cost? Is the memory cost a problem here?

Reply: We agree we should revise the description more concretely. It reduces the computational time because it can skip some procedures for handling the whole sequence of 2510 steps ($Z=2510$ in this paper) for 5000 particles (P. 11, L. 319–321). However, it is also true that the memory cost is also essential. If 5,000 particles for the whole sequence of 2510 steps of two variables are retained for all of the 50,000 MCMC steps, a TB-sized memory would be required.

These situations are described in the revised version.

Comment:
page 953, line 13: "$p(y_{1:k}|\theta)$" should be $p(y_{1:z}|\theta)$?

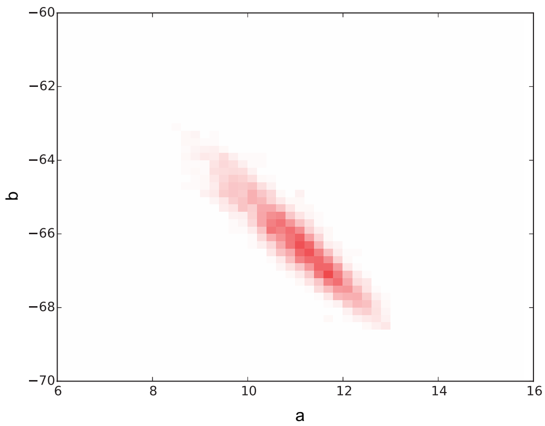


Fig. C. Two dimensional histogram for the joint posterior distribution of a and b.

Reply: We thank the referee for the correction.

Comment:
End of section 4: perhaps mention other particle MCMC methods. In particular, some variations such as particle Gibbs, and particle Gibbs with ancestor sampling (see [3]), would be applicable here and could significantly improve the performance.

Reply: We have added a mention on other particle MCMC methods (P. 10, L. 292). We appreciate the referee for the comment.

Comment:
Section 5: some comments could be made on the correlations between the components of the posterior distribution. If they are not close to zero, perhaps some pairwise scatter plots would be informative.

Reply: It is true some of the parameters are closely correlated with each other. For example, the two parameters for Eq. (19), a and b, have an anti-correlation as shown in the following figure (Figure C) We have added some two-dimensional histograms (Figure 3).

Comment:
Section 5: some indication that the Markov chains have mixed would be appreciated, for instance using traceplots instead of histograms. There is no indication in the text that multiple chains, with the same tuning parameters and starting from various points, lead to similar results. By the way, how were the Markov chains initialized? And how long was the burn-in period? Why was a sample kept every fifth iteration and not at every iteration?

Reply: As described above, we performed some preliminary runs to find out the landscape of the posterior distribution. The initial point of a Markov chain is determined around the center of the posterior indicated by the preliminary runs.

We kept every fifth iteration in order to reduce the computational time. Actually, since each sample typically has a high correlation with some subsequent MCMC samples, the estimate would not get worse even if four samples are discarded for every five steps.

Comment:

On the dataset: how large is it? Can it be plotted in some way? Can it be downloaded somewhere? It seems that the maximum depth is $Z = 2,500m$, and that there are a few dozen age markers (from Figure 2); it should be described in the text.

Reply: We use 25 age markers as shown above. The δ^{18}O data are published by Watanabe et al. (2003). We will add a plot of the δ^{18}O data (P. 4, L. 101; Figure 1). As mentioned above, Z is taken to be 2510 (m) in this paper (P. 11, L. 320). It is also described in the revised manuscript.

Comment:
page 956, on the computational cost: there should be some mentions of parallel computing, which could make 250,000 iterations with 5,000 particles much faster to run than 1,250,000 iterations with 1,000 particles. There is a rich literature on how to implement particle filters on parallel computing hardware.

Reply: It is true that the computation with 5000 particles can be much faster than that with 1000 particles if we use a parallel computer having larger than 1000 processors. We have added a description on this point (P. 15, L. 437).

Comment:
Figure 7-9 could be replaced by traceplots of the chains, starting from a few initial points, and plots of the average acceptance rates against number of particles, for a fixed proposal $q(\theta'|\theta)$.

Reply: The average acceptance rates were 7.6%, 26.4% with 1000 and 5000 particles, respectively. The average acceptance rates is now described (P. 15, L. 432).

We found the experiments with 3000 and 1000 particles are not so meaningful because a result with 3000 particles or 1000 particles is sometimes bad but sometimes good. We therefore removed the results with 3000 and 1000 particles in this revision except that the result with 1000 particles and 1 250 000 iterations is mentioned. Thus, the acceptance rate is mentiond only about the cases with 5000 and 1000 particles.

References

– Watanabe, O., Jouzel, J., Johnsen, S., Parrenin, F., Shoji, H., and Yoshida, N: Homogeneous climate variability across East Antarctica over the past three glacial cycles, Nature, 422, 509–512, 2003.

Reply to the comments by Referee # 2

We would like to thank the referee for the valuable comments and suggestions. In the following, the comments by the referee are listed and our reply is provided for each comment.

(We have made some corrections from the version posted as an interactive coment [npgd-2-C549-2015.pdf]).

Comment:

The developed technique rests upon two major glaciological assumptions, on both of which I would like to see an expanded discussion: . . . While the authors of this paper do mention the steady-state assumption, it lacks a thorough discussion, and if possible an investigation, of how this impacts the resulting age scale. I will even suggest the authors to consider to include the changes in elevation over time as another hidden variable to be estimated using the PMCMC technique. The thinning factor is thus calculated based on a steady-state assumption, assuming e.g. a constant accumulation over time. Yet, the authors subsequently assume the accumulation rate to be related to past temperature, and thereby oxygen isotope values from the Dome Fuji ice core. . . . It should also be mentioned that recent research has shown that using such relationship is oftentimes a poor assumption (e.g. WAIS Divide Members, 2013). While accumulation rates are indeed very affected by climate, there are sudden periods during which the relationship between accumulation rates and isotopic values does not hold. Please discuss this aspect. Indeed, to some extent the model does allow deviations from the steady-state thinning function and expected accumulation rates based on the isotope profile, and exactly this is one of the major forces of the described technique.

Reply: As the referee points out, the steady-state assumption for the thinning is inconsistent with the assumption for the accumulation rate. We also agree with the referee that it is not guaranteed that the accumulation rate and δ^{18}O have the same linear relationship over the whole period. We assume the regression coefficients a and b, which represents the relationship between the accumulation rate and δ^{18}O, do not depent on age. However, even if we can accept the linear assumption between the accumulation rate and δ^{18}O, a and b might change due to the variation of climatological conditions other than temperature.

As the referee points out, our method allows errors in estimates of thinning and accumulation. An uncertain variable η_z in Eq. (13) represents the variation of accumulation rate including not only the variation related with δ^{18}O but also the variation due to other unknown factors. Thus, errors in our assumption in the relationship between the accumulation rate and δ^{18}O are partly compensated by η_z. In addition, ν_z in Eq. (12) allows errors which might affect the age-depth relationship including the errors in the thinning function and the misestimation of the accumulation rate. (The meanings of ν_z and η_z were not correctly described in the text and we have revised the description on it. See P. 5, L. 128–133 and P. 6, L. 173–181.)

Because of the limitation of the available data, it is difficult to distinguish the effect of thinning and that of accumulation. This is the reason why we consider the uncertainty of age ν_z rather than improving the model of thinning and the model of the relationship between accumulation and δ^{18}O. However, our framework can be extended to consider such long-term changes by augmenting the vector $\boldsymbol{x}_z$ with some of the parameters for accumulation–isotope relationship and thinning. If other more relevant proxies become available in the future, we would be able to resolve the effects of these long-term changes.

Comment:

a) In both cases, the difference between the estimated and true values will be very strongly correlated with depth. In the paper, the error estimates are described as white noise, i.e. independent with depth (P. 495).

Reply: The errors in our model of glaciological processes are represented by ν_z and η_z in Eqs. (12) and (13). In this paper, the prior distributions of ν_z and η_z are given independently of depth.

However, their posterior distributions do depend on depth. (Note that the estimate for each z is given by the posterior distribution conditioned by the measurements over the whole ice core depth.) For instance, the posterior mean of ν_z shows long scale variations with respect to depth. Thus, the posterior distribution of ν_z and η_z satisfactorily represents the correlation with depth.

Comment:
b) Further, given that 1 m of ice contains substantially more years in the deeper part of the core, is it reasonable to expect that the discrepancy from the expected values as accumulated over 1 m are the same in top and bottom part of the core? (I would suspect these to be significantly larger in the bottom part). As this method is developed to allow more flexibility in the error structure, it is unfortunate that the assumptions of the underlying errors are sub-optimally chosen. Does the model have sufficiently flexibility that these error structures can be changed into more appropriate ones?

Reply: We agree that ν_z and η_z should be set to be larger for a deeper part of the core. As a matter of fact, we assumed ν_z and η_z to be larger for a deeper part of the core by multiplying ν_z and η_z by a factor proportional to $1/(A_z\Theta_z)$. Eqs. (12)–(13) did not agree with what we actually did.

However, in order to ensure that the evolution of the deviation from the glaciological model per year would not depend on Δz, it is more appropriate to multiply by $1/\sqrt{A_z\Theta_z}$. We have therefore revised the estimation program. Eqs. (10)–(13) have been modified accordingly.

Comment:
Further, the paper should include a description of the age markers used in the model, and their associated uncertainties in terms of depth as well as age. There are many kinds of age markers, with very different properties in terms of their uncertainty. It appears that the authors use O2N2 markers, which have age uncertainties of maybe 2000 years. Which values and how are these uncertainties accounted for here? How many tie points are used? How are they spaced? Could other age markers (such as volcanic horizons) be used in addition to these? It would also be an idea to select a subset of these age-markers, repeat the analysis, and compare the resulting ages at depths corresponding to the age markers omitted for age-scale construction. This would allow another estimate for the validity of the corresponding timescale.

Reply: Table A shows the value of age and the uncertainty (2σ) for each tie point. The first two points were given by Parrenin et al. (2007). The other points were determined from O_2/N_2 by Kawamura et al. (2007). Table A is added as Table 2 in the revised version.

The tie points are also indicated with black crosses in Figure 4, which shows how many tie points are there and how they are spaced.

Comment:
Finally, the technique relies on prior distributions for the involved parameters. But nowhere in the text is it described how these are obtained, or which values are used.

Reply: We appreciate the referee for pointing out that we omitted to describe about the prior distribution. In this paper, a uniform distribution is used as the prior distribution of each parameter (L. 209–211).

Comment:
It would be very helpful for the reader if the authors provide a table with definitions of the many variables employed.

Reply: We have added a table of the definitions of the variables (Table 1).

Depth	Age	Uncertainty of the age ($2\sigma_\varepsilon$)
371.00	12390	400
791.00	41200	1000
1261.61	81973	2230
1375.67	94240	1410
1518.91	106263	1220
1605.27	116891	1490
1699.17	126469	1660
1824.80	137359	2040
1900.74	150368	2230
1958.31	164412	2550
2015.01	176353	2880
2052.23	186470	2770
2103.14	197394	1370
2156.67	209523	1980
2202.02	221211	890
2232.45	230836	780
2267.28	240633	1230
2309.35	252866	1160
2345.32	268105	1980
2366.01	280993	1600
2389.31	290909	1210
2412.25	301628	880
2438.37	313205	840
2462.36	324774	1110
2505.4	343673	2000

Table A. The depth, the age, and the uncertainty of the age at each tie point.

Comment:
P: 940, line 17, P. 941 line 14-20: Without assuming linearity or Gaussianity - of what? Please make sure that this is clear throughout the text. The technique assumes Gaussianity of age markers etc.

Reply: As indicated in Eq. (12), the relationship between A_z and ξ_{z+1} is nonlinear. Accordingly, ξ_z can not be represented using a Gaussian distribution.

Note that the PMCMC assume neither linearity nor Gaussianity anywhere. As the referee says, we choose Gaussian distributions for $p(\xi_{z+1}|\xi_z, \boldsymbol{\theta})$ and $p(\tau_k|\xi_{z_k})$ in this paper. However, we can choose other probability distributions such as log-normal distribution for $p(\xi_{z+1}|\xi_z, \boldsymbol{\theta})$ and $p(\tau_k|\xi_{z_k})$. It is not necessary to choose Gaussian distributions for them. We have added comments to remark that (P. 15, L. 466–472).

Comment:
P. 941, line 3-7: Please expand on these earlier approaches where Bayesian and MCMC methods are used for estimating the depth-age relationship.

Reply: Klauenberg et al. used Bayesian and MCMC methods for estimating the age as a function of depth based on the estimation of accumulation for each ice slice, although their method was not designed to make use of the constraints of age markers to estimate the age for the entire ice core. Parrenin et al. used Bayesian and MCMC methods for estimating the parameters in the glaciological process model. However, they did not consider the deviation from the glaciological process model, and they did not estimate the magnitude of the deviation. We have revised the description on the earlier approaches (P. 2, L. 24–35).

Comment:
P. 944, line 18: It might be worth a mention that recent research (Freitag, 2013) has shown that thinning may also be affected by impurity content.

Reply: We are grateful for the suggestion. We will add the mention on their result.

Comment:
P. 945, line 21: Are the O2N2 tie-points assumed to have no depth uncertainty? If so, is this a reasonable assumption?

Reply: The tie points are assumed to have no depth uncertainty. We think the depth uncertainty would make no essential effects on the estimate of the age for each slice of the ice core labeled with a depth value, even if its true depth is uncertain. The estimates of accumulation and thinning might be affected by the depth uncertainty. But the estimates of accumulation and thinning would not be sensitive to the depth uncertainty because accumulation and thinning are related with the increment of depth rather than the absolute depth from the surface. In addition, the uncertainty in age would compensate the possible effect of the depth uncertainty on the estimates of accumulation and thinning.

Comment:
P. 946, line 8-13: Discuss why this type of equation is chosen to translate from isotope values to accumulation values. Provide reference(s) for previous usages of similar equations.

Reply: This is the same assumption as used by Klauenberg et al. (2011). It is true we should add the reference to that. We have added it (P. 6, L. 170).

Comment:
P. 946, line 18: How often does it happen that isotopic data is missing for a 1 m section? (My guess would be that it is very rare)

Reply: The isotopic data are densely available for the deeper part of the ice core. However, near the surface, the isotopic values provided to us were smoothed over the depth larger than 1m to reduce the noises. For example, isotopic values are provided for only 17 segments above 50 meter depth.

Comment:
P. 948, line 8: Please explain how we would know that the uncertainty is " too large "

Reply: The posterior of θ in Eq. (23) provides a metric to evaluate whether the uncertainty is too large or not. One advantage of the use of the Bayesian approach is that it provides a framework to objectively determine the magnitude of the uncertainty.

The reason why σ_v should not be taken too large can also be explained in another way. If σ_v was taken too large, large variations of the age ξ are allowed. Thus, the result could be sensitive to the noises contained in the data. We have improved the explanation on why σ_v should not be large (P. 8, L. 217).

Comment:
P. 948, line 10: It is not the uncertainty of the d18O data that gives rise to the deviations described by σ_w; it is the flaws in model used for predicting the accumulation rates based on the isotope values. Hence this parameter does not have any significance in terms of standard deviation of the isotope values.

Reply: The referee is right. In Eq. (19), w_z just represents the discrepancy between the accumulation in the model and the measured δ^{18}O value. Thus, σ_w just gives a typical magnitude of this discrepancy. It can not necessarily be attributed to the uncertainty of the d18O data. We have corrected the description (P. 8, L. 219).

Comment:
P. 948, line 15-20: Describe the advantages of using the hybrid method.

Reply: The SMC can be used only for obtaining $p(x_{0:Z}|y_{1:Z},\theta)$ under a given θ. It can not be used for obtaining $p(\theta|y_{1:Z})$. In principle, the MCMC could be used for obtaining any probability distribution including $p(x_{0:Z}|y_{1:Z},\theta)$, $p(\theta|y_{1:Z})$, and $p(x_{0:Z}|y_{1:Z})$. However, it would require prohibitive computational cost for high dimensional problems. In practice, the MCMC is not applicable to obtain a high dimensional distribution like $p(x_{0:Z}|y_{1:Z},\theta)$ and $p(x_{0:Z}|y_{1:Z})$. Combining the SMC and the MCMC, we can obtain $p(x_{0:Z}|y_{1:Z},\theta)$, $p(\theta|y_{1:Z})$, and $p(x_{0:Z}|y_{1:Z})$ with acceptable computational cost.

This point is also described in the revised oversion(L. 8, P. 219–220).

Comment:
P. 949, equations 25-30: I do not understand these equations. What is meant by the notation $\{X_{o:z-1|z-1}\}$? Define delta and N used in equations.

Reply: We indicate one sample from $p(x_{0:z-1}|y_{1:z-1})$ by $x^{(i)}_{0:z-1|z-1}$, and a set of N samples are denoted by $\{x^{(i)}_{0:z-1|z-1}\}$, where N is the number of the samples. The function $\delta(\cdot)$ denotes the delta fuction. We had added the definition of them (P. 8, L. 245). We appreciate the referee for pointing out the flaw in our explanation.

Comment:
P. 950: It would facilitate understanding if the authors included a figure illustrating the method.

Reply: The illustration of a past paper by one of the authors (Nakano et al., 2007) might be helpful for understanding the procedure described in this page. The instruction on the particle filter (which is the same as the SMC) was also privided by (van Leeuwen, 2009). We have added the references to these papers.

Comment:
P. 954, line 3: Why is every 5th iteration retained? Is this number based on a correlation analysis of the MCMC samples?

Reply: Since each sample would be highly correlated with some subsequent MCMC samples, it is not necessary to retain all the iterations. However, there is no particular reason why we retained a sample every 5th iteration. Maybe it is enough to retain one of 20 samples or one of 30 samples. But, it would not make any essential effects on the results.

Comment:
P. 954, line 6: Which values were used as priors for θ? Surely, the result will be very dependent on what is used for priors?

Reply: In this paper, a uniform distribution is used as the prior distribution of each parameter. The result will be dependent on the prior distribution. However, if we use a different prior distribution, the posterior distribution can be obtained as a product between the prior distribution and the histogram in Figure 2 which was based on the uniform prior. We think Figure 2 is informative enough to guess the posterior with a different prior.

Comment:
P. 954, line 17-19: I suggest the authors to spend a little more time reflecting on the difference in the results obtained here relative to those in Parrenin 2007. If, as suggested by the authors, the obtained velocity profile in the ice sheet (reflected in parameters p and s) really depends so significantly on the isotope-modelled accumulation rate - which in best case is a rough approximation - this is not very encouraging for how well an age-model can be constructed away from age markers.

Reply: We have found that one reason was the problem with the setting of ν_z in Eq. (12). As described above, we multiplied ν_z by a factor $1/(A_z\Theta_z)$. This allowed too large variations for a deeper part, and therefore the thinning function was sensitive to the measurement errors for a deeper part of the core. That seems to be one reason why p was estimated to be large. We have modified the setting of ν_z. The mode of the posterior of p is now similar to that obtained by Parrenin et al.

The shape of the posterior of p is still not similar to that obtained by Parrenin et al. It might be caused by the different setting of the accumulation. As the referee points out in an earlier comment, the difference in the assumption for the thinning function might also cause the difference in estimates of the parameters for the thinning function.

Comment:
P. 954: How well does the resulting age scale match the age markers? Does it correspond to what was expected?

Reply: In the following figure (Figure A), the differences between the age markers and the medians of the posterior distribution are compared with the difference of the 10th and 90th percentiles of the posterior distributions from the medians of the posterior; i.e., the median is subtracted from each line or each point of Figure 4. The grey line will be explained later. The age markers are seen within the range of the uncertainty with a few exceptions. Figure 5 has been replaced by this figure in the revised version.

Comment:
P. 954, line 6: 5 trials were performed starting from random seed; could the results from these be combined for the final results?

Reply: The results shown in this paper is one of the 5 trials. We have added the description on that (P. 13, L. 432).

Comment:
P. 954, line 20: How does the timescale compare to the one obtained by Parrenin, 2007? (this should also be added to figure 2)

Reply: The following figure (Figure B, Figure 4 in the text) is the comparison with the result by Parrenin et al. 2007, and In Figure A, the difference of the result by Parrenin et al. from the posterior median of the age obtained using the proposed method is shown. Our method tends to rely on the age markers more confidently than Parrenin et al. 2007. The difference between the two results is more than 3000 years at largest.

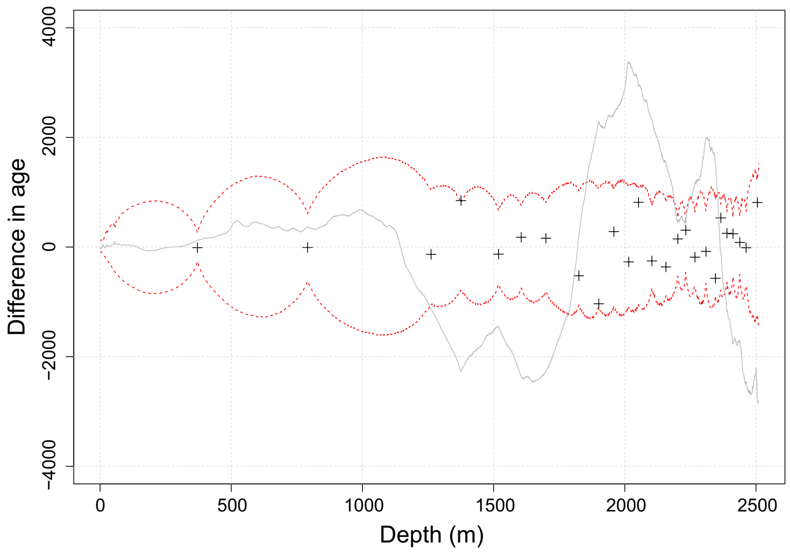


Fig. A. The difference of the 10th and 90th percentiles of the posterior distributions from the medians of the posterior (red dotted lines), and the differences between the age markers and the medians of the posterior distribution (black crosses). The grey line indicated the difference of the result by Parrenin et al. (2007) from the median of the posterior.

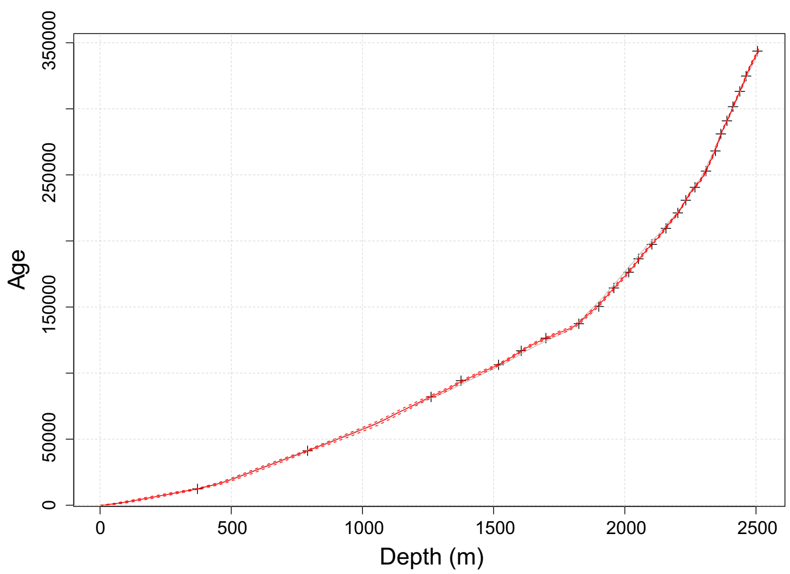


Fig. B. Estimated age as a function of depth. The red solid line indicates the median of the posterior distribution. The 10th and 90th percentiles of the posterior are indicated by red dotted lines. The grey solid line indicates the result by Parrenin et al. (2007). The black crosses indicates the age markers.

Comment:
P. 955, line 6: The thinning factor shown here is significantly different from the one obtained in Parrenin 2007. Why is that? What would be the thinning function simply based on the initial non-steady-state version with parameters as e.g. given by the mode of the obtained posterior distributions?

Reply: Parrenin et al. (2007) estimated the thinning factor down to the depth of 3000m. However, the tie points and $\delta^{18}O$ data are available only above the depth of 2510m. Accordingly, this paper estimated the thinning factor above the depth of 2510m. This might be the reason why the thinning factor apprears to be different from that in Parrenin et al. (2007).

However, as the referee points out, our steady-state assumption would be also one of the reasons of the difference from the estimate by Parrenin et al. (2007). The difference in the assumption on the thinning factor will be mentioned in the revised version (P. 13, L. 385–387).

Comment:
P. 955, line 11: Similarly, how does the initial estimate for accumulation-rates look based on the simple isotope model that forms the basis for the analysis? And compared to the estimate from Parrenin 2007? It would be great, both in figure 4 and 5, to include the results from Parrenin 2007, so that it possible to evaluate the difference between the two.

Reply: As described above, we use a uniform distribution as a prior for each parameter. Thus, our method does not use any initial estimate.

The followings (Figure C) are the comparison with the results of Parrenin et al. (2007) for accumulation, which is to be added in the revised version.

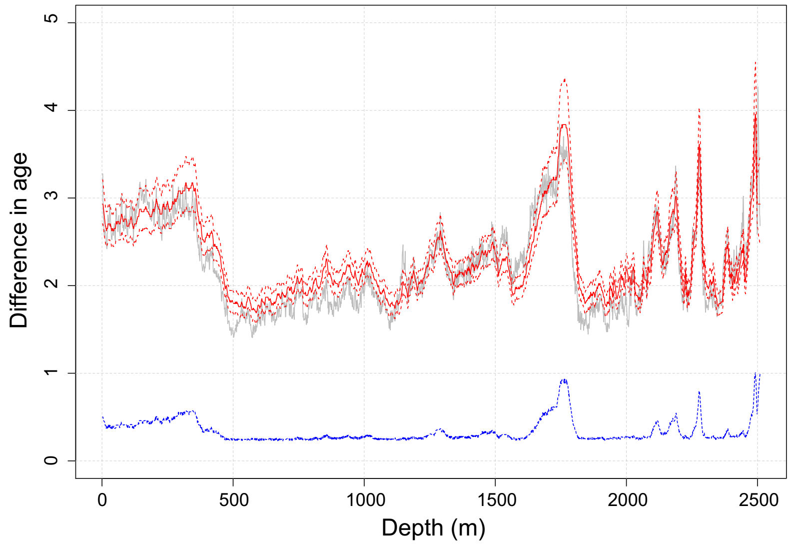
Comment:
P. 955, line 17-P. 956 line 11: These are not really results, but rather a sensitivity study.

Reply: We have separated this part as an independent section named "Discussion". However, we found the experiments with 3000 and 1000 particles are not so meaningful because a result with 3000 particles or 1000 particles is sometimes bad but sometimes good. We therefore removed the results with 3000 and 1000 particles in this revision except that the result with 1000 particles and 1 250 000 iterations is mentioned.

Comment:
Technical corrections: P. 940, line 2, P. 942, line 7: Remove "mainly" and "primarily": Below the uppermost zone, where snow is compacted into ice, a depth-age relationship can be calculated directly from the initial accumulation rates and the thinning rates; this is the definition of the thinning rate. Of course, we can only aim to estimate this function.
P. 940, line 5: Except for the uppermost zone where snow is turned into ice, ice is not compressed, since its density remains constant.

Reply: We thank the referee for the correction. We have corrected those.

Comment:
P. 942, line 12-14: This sentence is awkward. A is the accumulation at time corresponding to the age at z, i.e. it is actually a function of time, not depth. Thinning factor is not defined. Any reason not to use (the usual) t for time instead of ξ?



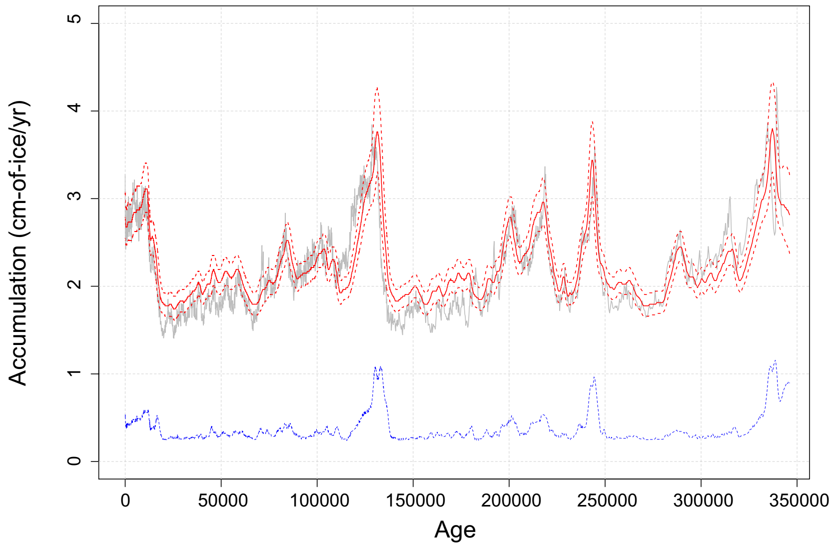


Fig. C. Comparison with the results of Parrenin et al. (2007) for accumulation. The red lines show the result using the proposed method. The grey line indicates the result by Parrenin et al.

Reply: We treat A as a function of depth, not a function of age. The accumulation with respect to age is estimated after considering the uncertainty of age as:

$$p(A|\xi) = \int p(A|z)\, p(z|\xi)\, dz \tag{1}$$

where we assume $p(z)$ to be a uniform distribution in obtaining $p(z|\xi)$:

$$p(z|\xi) = \frac{p(\xi|z)p(z)}{\int p(\xi|z)p(z)dz}. \tag{2}$$

Probably, we should describe how we obtained $p(A|\xi)$ in Figure 6 in P. 955.

The SMC is usually applied to time series data where time t is given. On the other hand, in this paper, age is unknown. We denoted age by ξ to avoid this confusion.

Comment:
P. 942, line 21: Define H, and facilitate the reader's understanding by describing these two equations in words

Reply: We denote the thickness of the ice by H as described in the 2nd line of P. 943. The variable ζ is a rescaled vertical coordinate which becomes 0 at the bottom and 1 at the surface, and u indicate the velocity in the ζ coordinate.

Comment:
P. 943, line 16: Θ is a function of ζ, not z

Reply: It is corrected.

Comment:
P. 944, line 16: Z is not used.

Reply: We appreciate for the correction. Eq. (10) should be modified as:

$$\xi_{z+\Delta z} = \xi_z + \frac{\Delta z}{A_z \Theta_z} + \nu_z \sqrt{\frac{\Delta z}{A_z \Theta_z}} \quad (z = 0, \Delta z, 2\Delta z, \ldots, Z),$$

i.e., the upper limit of z is indicated. We have also found that the third term of the factor was wrong and it is corrected here.

Comment:
P. 944, line 19: Add "in a steady state"

Reply: It has been added in the revised version.

Comment:
P. 945, line 10: Equation is missing "$+1/A_z\Theta_z$"

Reply: We appreciate for the correction.

Comment:
P. 945, line 20: " The tiepoints .. depths ": This should obvious; sentence can be removed.
P. 946, line 15: Same as above

Reply: Those sentences have been removed.

Comment:
P. 947, line 14: Define Z here.

Reply: Z is the depth at the bottom, which is to be used in Eq. (10) in the corrected version. But, we recall the meaning of Z here (P. 11, L. 320).

Comment:
P. 948, line 1 (and various times later): "accumulation at the surface": A0 is present accumulation (accumulation always occurs at the surface).

Reply: We thank the referee for the correction.

Comment:
P. 948, line 13: Provide value.

Reply: The value of σ_ε for each tie point is given in the table above, which will be added in the revised manuscript.

Comment:
P. 954, line 8: What is the unit of A0? Clearly, it's not kg/m^2/yr?

Reply: It is cm(of ice)/year.

Comment:
Figure 1: "Accum at surface" -¿ A0. It is clear from figure what shows the various parameters, and hence their names do not need to be repeated in figure text.

Reply: We agree with the referee on that. We have revised that figure.

Comment:
Figure 2: Missing solid line and red dotted lines.
Figure 2 and 3 can easily be combined to a single figure.

Reply: The caption of Figure 2 was wrong. The 10th and 90th percentiles of the posterior were indicated by blue dotted lines. However, it is difficult to discriminate between the 10th and 90th percentiles in Figure 2. That is the reason why the difference between the 10th and 90th percentiles is shown in a separate figure, Figure 3, where the vertical scale is expanded.

Comment:
Figure 3: This figure shows the width of the 80% confidence interval. This should be stated somewhere.

Reply: We will add a comment on Figure 3 according to this suggestion.

Comment:
Figure 4: it is shown as function of depth, not age.

Reply: We have corrected that. We appreciate for the correction.

Comment:
Figure 5: There is no need to show accumulation as function of depth as well as age. Accumulation rates as function of time makes most sense.

Reply: As described above, the accumulation with respect to age is obtained after considering the uncertainty of age as:

$$p(A|\xi) = \int p(A|z)\, p(z|\xi)\, dz. \tag{3}$$

In our opinion, $p(A|z)$ provides slightly different information from $p(A|\xi)$.

Comment:
Figure 7-9: The coloring makes it hard to compare the two distributions. I would suggest instead to e.g. only show the outlines of the distributions (i.e. without vertical lines) in different colors. This would allow to combine at least figure 7 and 8, and possibly also figure 9.

Reply: We thank the referee for the suggestion. We have edited the figures according to this suggestion.

References

- Nakano, S., Ueno, G., and Higuchi, T.: Merging particle filter for sequential data assimilation, Nonlin. Process. Geophys., 14, 395–408, 2007.
- van Leeuwen, P. J.: Particle filtering in geophysical systems, Mon. Wea. Rev., 137, 4089–4114, 2009.

Reply to the comments by Referee # 3

We would like to thank the referee for the valuable comments and suggestions. In the following, the comments by the referee are listed in Italic, and our reply is provided for each comment in Roman.

Comment:
(1) P. 940, lines 22-23 Before this sentence, please put a brief explanation why the age-depth relationship is created because statistical scientists may not follow the present description.

Reply: We add a sentence saying, "In order to make use of the chronological records from each slice of an ice core, it is crutial to accurately determine the age for each slice."

Comment:
(2) P. 942, line 12 Units? Although I think the statistics does not need them, the authors put the physical image here. For example, z [cm], $A(z)$ *[cm/year],* $\Theta(z)$ *(no unit),* ξ *[year]*

Reply: We add the information on units. We thank the referee for the suggetion.

Comment:
(3) P.944, line 17 The "denoted by by Az" should be replaced with " denoted by Az ".
(4) P.946, line 16 The "the δ^{18}O *data* δ^{18}O *are" should be replaced with "the* δ^{18}O *data are".*

Reply: We have corrected them. We thank the referee for the corrections.

Comment:
(5) P.954, lines 7-9 Is A0 here estimated purely by this model without any observational information of accumulation such as Kameda et al. (2008)? In this statistical model, do some conditions under the surface affect the estimation of the present surface accumulation, A0? While A(z) is estimated by this model, is it right that A0=A(0)?

Reply: We used only the δ^{18}O data and the age markers. No other data were used for the estimation of A_0.

In our method, $A(0)$ becomes A_0.

Comment:
(6) Figure 2 Each line is not distinguished in the figure because the difference between the 10th and the 90th percentile is too small, which is found in Fig.3. But I think it is better this situation is explained here.

Reply: Exactly, Figure 3 is added because it is difficult to discriminate between the 10th and 90th percentiles in Figure 2. But, according to the suggestion, we add a comment on the situation happening in Figure 2.

Comment:
(7) P.955, lines 4-5 I think that the difference between 10th and 90th percentile should be zero at each tie point because the tie points indicate accurate date. But why do the differences not become zero?

Reply: The age at each tie point has some uncertainty. Therefore, the uncertainty can not be reduced to zero even at a tie point.

Comment:
(8) P.955, lines 6-10 As for Fig.4, what is the reason that the uncertainty gets smaller toward the bottom?

Reply: The uncertainty gets larger toward the bottom. At the surface, the effect of the ice sheet deformation (thinning) is negligible, and thus the thinning factor can be assumed to be 1 with no uncertainty. In the deeper core, the effect of the ice sheet deformation becomes significant but it is unknown. That is the reason why the uncertainty of the thinning factor gets larger toward the bottom.

Comment:
(9) P.956, line 15 The " pvovides " should be replaced with " provides ".

Reply: We have corrected it. We thank the referee for the correction.