Anonymous Referee #2
Received and published: 7 September 2016

Response of the authors
Thank you very much for your review and your comments.

Radiocarbon dating of glacier ice is an important asset to ice core sciences in mid-latitudes when classical methods to derive age-depth models (layer counting, ice flow modelling, tephra-chronology) fail. However, suitable material for radiocarbon dating of macrofossils is sparse in the ice making it desirable to date other organic carbon compounds. Water-insoluble organic carbon (WIOC) has been shown to be a suitable candidate for radiocarbon dating of mid-latitude ice. Uglietti et al. review the efforts to develop the analytical methodology for radiocarbon dating of WIOC in ice, and test its accuracy. The group of Bern/PSI has been instrumental in developing the method so it seems only natural that they provide a review thereof and is certainly a valuable contribution to the literature of this topic. The paper is generally well written, thorough, suitable for the Cryosphere, and should be published. However, I have some comments that the authors might want to consider for the sake of clarity in the paper and some details the authors may want to check before publication.

1. My main comment is that in the context of analytical precision, I find it slightly misleading to discuss calibrated ages instead of ^{14}C ages or fraction modern. The calibration of radiocarbon dates is a second methodology, introducing additional uncertainty due to the uncertainty of the calibration curve itself. Hence, 2 significantly different radiocarbon ages may lead to statistically indistinguishable calibrated ages. However, if the radiocarbon dates and their uncertainties are reliable, then 2 different radiocarbon ages are indicative of different calendar ages. This applies for example to Table 2 and 3 and Figure 1. These experiments refer to analytical accuracy of radiocarbon preparation and measurement, independent of the calibrated age and it is hence worth discussing (and showing) the differences in ^{14}C years or F^{14}C instead of calibrated years. A comparison of the ^{14}C measurement results could for example be indicative whether the pure analytical precision does reflect the true uncertainty of the method. I would assume that the true uncertainty is somewhat larger, due to the inhomogeneous distribution of WIOC over an ice sample. This can for example be seen in Table 3 where the samples JUV 0_5/6 as well as JUV 0_7/8 yield significantly different radiocarbon ages, despite being from the same ice block. This questions whether these samples can be summarized to an error weighted mean and standard error as done in table 3. The final uncertainty of 9 ^{14}C yrs. for JUV 0_3-8 seems very small given the scatter of the individual measurements. A reduced Chi2 statistic of the sample pools in Table 3 could be used to assess and discuss whether the uncertainties of single measurements are realistic. In a second step, it can then be discussed whether these uncertainties matter in terms of the absolute chronology, given that the calibration adds additional uncertainty.
We understand that the final uncertainties of the calibrated ages are different than those of the ^{14}C ages and also include the additional (method unrelated) uncertainty of the calibration curve. Nevertheless the main purpose of ice core dating is to provide a final age range and therefore we consider it more important to discuss and compare the uncertainties on the calibrated ages. In any case, Tables 2 to 4 do contain ^{14}C ages and F^{14}C, so the information about the method related (analytical) precision is also available (also see cited literature).
We further agree that the uncertainty of a single measurement is indicative for the analytical uncertainty only and does not consider variation due to potential inhomogeneity in the ice sample (as indicated by different WIOC concentrations). Therefore, even if replicate results for the same sample do slightly differ from each other, considering the analytical 1σ uncertainty we think they can still be combined (in case the thinning does not suggest otherwise which of course has to be checked). In

fact, replicate measurements or measurements in high spatial resolution are thus preferable to avoid dating bias due to these (small) variations caused by inhomogeneity (averaging out of variations). This is feasible and a strength of the method described. After thorough consideration we however concluded that the OxCal combine tool is not appropriate to average such samples because it is not intended to take such possibility of inhomogeneity into account. We therefore revised our results by using the averages of the ^{14}C ages (or equal $F^{14}C$ values) with the uncertainties estimated using the standard error of the unbiased standard deviation (i.e. accounting for number of replicates) (Tables 2 and 3). Results of combined samples thus changed slightly but changes are negligible considering the uncertainties which now increased to likely more realistic estimates. We would like to emphasize, that for all the age-depth modeling performed in the past and summarized here, the model uncertainty estimates have always been selected very conservatively, one reason being exactly the consideration of potential inhomogeneity in the ice.
The unreasonably small uncertainty you mentioned (Table 3) was a typo. Thank you for spotting it. For JUV 0-B the standard deviation was 79 instead of the given value of 9 (it now increased to 168 years using the new approach). There was another typo: for JUV 0_7 the ^{14}C age uncertainty was indicated with 18 instead of 218 years (now 219, because we now consider 3 instead of 2 digits for $F^{14}C$). Using 3 digits now generally resulted in small changes of the results (usually in the order of 1 to 5 years).

2. Please check the data in the Tables. I was confused seeing that the samples Bel2_THEODORE and Bel2_Sunset yield significantly different $F^{14}C$ values while their ^{14}C ages agree. Using an $F^{14}C$ of 0.425 for Bel2_THEODORE I obtain a radiocarbon age of 6874 ^{14}C BP, as compared to 7329 ^{14}C BP given in Table 2. So unless I missed something either the $F^{14}C$ or the ^{14}C age of this sample is erroneous which might also impact on the calibrated ages shown in Figure 1. Please check.
We apologize; this was indeed an error. Thank you for spotting this. The Bel2_THEODORE $F^{14}C$ is 0.402, but was given as 0.425 which was picked from the wrong column in the data files and is the value before procedural blank subtraction. The correct value is given now (Table 2) and all other data in the manuscript have been cross-checked for correctness. In addition the Sunset values have slightly changed because of the blank correction. For the THEODORE – Sunset comparison the samples measured with the Sunset system were intended to be blank corrected with the corresponding blank values (1.21 ± 0.51 µg of carbon with an $F^{14}C$ of 0.73± 0.13)) but in the excel file there was an automatic link to the new combined procedural blank value (1.34 ± 0.62 µg of carbon with an $F^{14}C$ of 0.69± 0.13) which is used for all the other samples, but not intended for the comparison, as also already stated in the main text (Lines 223-225). Therefore in the new version the values appear slightly different because we used the correct blank values (1.21 ± 0.51 µg of carbon with an $F^{14}C$ of 0.73± 0.13). Moreover, we changed the sample names in Table 2 and Figure 1, to be consistent with Table 3 for the Juvfonne samples. For the Belukha samples we also changed the names. For example Bel2_THEODORE is now 4_THEODORE (BEL 2).

3. Throughout the manuscript the term "conventional" ^{14}C dating is used to describe the dating of macrofossils. However, in the radiocarbon literature "conventional" ^{14}C dating refers to ^{14}C measurements using liquid scintillation and gas proportional counting techniques as opposed to AMS measurements. Please either use a different term than "conventional" or add a sentence defining how it is used in this paper.
Yes, good point. We included a sentence to define this term in the introduction at Lines 60-61:
"In the following we refer to dating of ice with macrofossils as conventional ^{14}C dating".

L 45 *(now line 48)*: please replace "nuclear" with "radiometric"
We normally use the term "nuclear dating", which is common in the radiochemistry community (see for example textbook Nuclear- and Radiochemistry Vol. 2 (https://www.degruyter.com/view/product/41711).

L 179-182: This is a very long sentence and a little unclear. Maybe divide it up into 2 sentences. Are the sample background corrected using OxII? I suppose the standard is used for normalization and not background correction? Please rephrase.

We agree that the sentence is quite complicated and long but we think it is better to give only a short and easy explanation. We tried to improve the phrasing though (Lines 189-191):

*"*With the current setup, the $^{14}C/^{12}C$ ratio of the samples is background subtracted, normalized and corrected for mass fractionation by using fossil sodium acetate (^{14}C free, NaOAc, p.a., Merck, Germany), the reference material NIST standard oxalic acid II (modern, SRM 4990C) and the $\delta^{13}C$ simultaneously measured in the AMS, respectively (Wacker et al., 2010).*"*

L 183: Please insert "relative" before deviation, as the samples are normalized to the standard.

We realized that the term "deviation" is not correct in this context and therefore confusing. We rephrase as the following (Lines 192-193):

"…which is the $^{14}C/^{12}C$ ratio of the sample divided by the same ratio of the modern standard"

L 183: "BP" is not explained at this point yet, but only in line 185. Please explain it here instead.

We corrected (Line 196):

"^{14}C ages (before present (BP), i.e. before 1950)) are calibrated using OxCal v4.2.4 (Bronk Ramsey and Lee, 2013) with the Northern (IntCal13) or Southern Hemisphere (ShCal13) calibration curves (Reimer et al. 2013, Hogg et al. 2013), depending on the sample site location. Calibrated dates are given in years before present (cal BP)".

L 227-228: Are the uncertainties given in ^{14}C years here? If so, please write "^{14}C yrs" instead of years to be clear.

We here talk about an assumed age of the ice sample and thus refer here to the "true" age and not the ^{14}C age. Accordingly the uncertainties denote the final overall uncertainty of the dating method which as such also includes the calibration so they refer to the calibrated age. To clarify we changed the sentence to "As an example, for hypothetical samples with a WIOC mass of 5 or 10 µg, the resulting uncertainty of the finally calibrated ages for 1000 year old ice would be around ± 600 yrs or ± 250 yrs and for 8000 year old ice around ± 1600 yrs or ± 700 yrs, respectively (Line 249-253)

L 249 and following: See comment number 3. Please either define what you mean by conventional or use a different term throughout the manuscript instead.

See reply above (comment 3.) and see Lines 60-61 in revised version.

L 252: This may be nitpicking, but AD 1258 – AD 1050 = 208 years, not 200 years.

With an error of 70 years, rounding to down to 200 seems appropriate. However, we changed to "around 200 years" then omitting the need for indication of the uncertainty which could be derived anyhow with the value of AD 1050±70 being provided (Line 274).

L 259-271: Several times it is stated that the WIOC dates "agree well" with the macrofossil dates, while the ^{14}C ages are indeed significantly different. I am not arguing against the general agreement but it would be great if you could add 1-2 sentences to make this more precise. Are the differences due to sampling differences (i.e., different ice layers have been sampled for macrofossils and WIOC)? If so, are the results in stratigraphic order?

Yes, the differences arise from the sampling procedure. The samples in the Juvfonne ice patch were obtained extracting clear ice just adjacent to the organic remains layers where the macrofossil ^{14}C ages are from. Thus a 1:1 agreement cannot be expected. Here, the samples from Ødegard et al. (2016) are simply given as a reference age range to compare our samples with. To allow for the most reasonable comparison we always considered the organic layers closest to our ice layers. In this

version we thus changed for example the organic layer Poz-37879 with Poz-37877 which is closer to the ice sample JUV 2 and was not available for the original study by Zapf et al., 2013 which is why we did not consider it initially but now decided to. The same for JUV 1 and the corresponding closest organic layers Poz-36460 which we kept and Poz-37878 which was replaced with a closer layer Poz-56952 only sampled in 2012 and therefore also not available in the study by Zapf et al. 2013.
We now describe the issue more carefully by including the stratigraphic order in Table 3 and an explanation in the related caption text and also in the main text (Lines 292-294).

L 351: Please write "climate wiggle matching" instead of just "wiggle matching" which could also refer to the wiggle matching of radiocarbon dates.
Yes, done, thanks (Line 395).

L 357: Please add a reference to [Godwin, H. 1962. Nature, 195 (4845)] for the half-life of radiocarbon.
Yes added. Thank you (Line 402).