We thank the three reviewers for their good comments and suggestions. Besides improvements in the grammar and text, two major changes are implemented in the revised manuscript:

1. **A seasonal temperature signal** that propagates down through the firn is included in the diffusion and densification model (Sec. 2 and Appendix C). The seasonal temperature variation affects the firn diffusion length nonlinearly due to the saturation vapor pressure’s exponential dependence on temperature. The firn temperature profile is obtained by numerical solution of the heat equation. In the previous manuscript, a constant firn temperature profile was assumed.
2. **Uncertainties in the diffusion, densification and ice flow models** are now included. Previously, the presented precisions were based on the estimation of the diffusion length from data. It was therefore assumed that the accumulation rate, surface density, close-off density, surface pressure, ice flow thinning and ice diffusion were known (they had a constant value for each ice core section). In the revised manuscript, each of these parameters have an uncertainty used in the temperature reconstructions (Table 4.1 in paper). This facilitates a better comparison between the single and differential diffusion techniques in case the single diffusion length methods are more sensitive to uncertainties in the diffusion model.

Although the model used in this study has changed significantly by implementing a seasonal signal and including uncertainties in the diffusion, densification and ice flow model, the main conclusion remains the same; the single diffusion length techniques perform better than the differential diffusion methods. The difference in precision between the techniques has however decreased.

The power spectra of the ice core data are now presented in Appendix F. This allows the reader to see the performance of each diffusion technique in characterizing the spectra. We have also updated a reference in the test with the fractionation factor parameterizations. We now refer to the fully published study by Lamb et al (2017) instead of the 2015 version. The used parameterization is identical so it does not change the results and conclusion.

Specific comments to each reviewer is provided below. Text in bold is the reviewer’s comment.

**Reviewer #1:**

**This manuscript examines a variety of methods for inferring temperature from diffusion of**

**water stable isotopes. The authors use both synthetic and ice core data to assess the reconstructions of temperature. They conclude that despite some theoretical advantages of using differential diffusion lengths, the single diffusion length methods are more robust.**

**This is a strong paper and should be published after minor revisions. The science is sound and well described.**

**This is a technical study and the conclusions are a little underwhelming, but I suppose the authors don't want to negatively bias methods while the field is still emerging. However, I would have liked at least a paragraph of forward looking discussion about how their testing of different methods should impact future measurements of ice cores.**

**[note from GCA Executive Editor: As the primary aim of GCA is to present studies of fundamental significance and broad relevance for understanding geochemical systems, mechanisms, and processes, this is a reasonable request and should be addressed specifically in the revised manuscript.]**

TO DO: Include a discussion part of how it will affect future measurements of ice cores

Future measurements of water stable isotopes in ice cores will likely have similar or better resolution and measurement precision than those present today. Moreover, measurements of the d17O signal will soon be a common output from analyzed ice cores. As we showed with synthetic data, such a signal can also be used to reconstruct temperatures. Especially the differential diffusion length of d17O and dD showed higher precision than that of d18O and dD. We propose that future research should investigate how to utilize the diffusion lengths of d18O, dD and d17O at the same time. For instance by estimating the differential diffusion length between d17O and dD and then use the independent diffusion of the d18O signal, it should in theory be possible to perform temperature reconstructions that could help better constrain the ice flow thinning.

**While I understand choosing climate conditions applicable to Central Greenland and the East Antarctic Plateau because there is ice core data available, I was disappointed that the authors did not take advantage of being able to create synthetic data from warmer and wetter climates. Many ice cores are being drilled closer to the margins of the ice sheet and the work would have greater applicability if it focused on where ice coring was going rather on where it's been.**

Even though we are able to generate synthetic data representing warmer and wetter climates, diffusion-based temperature proxies are most reliable at colder drill sites. Ice core drill sites near the ice margin experience a vast amount of summer melt events. Melt layers will stop the mixing between the air in the firn and the surrounding atmosphere earlier than the natural densification process. Thus, a melt layer will result in a smaller diffusion length estimate which falsely would result in a low temperature. In order to employ the diffusion-based temperature proxies on ice cores drilled at ice sheet margins, a good record of the melt layers with depth is needed. This would make it possible to distinguish between when the diffusion length estimate represents a firn integrated temperature and when it partly represents a melt layer. This has been clarified in Sec.XX line XX.

**The manuscript could also use a quick edit from a native English speaker as some of the idioms and transitions are slightly off (e.g. "wavelengths in the order of 50cm" should be "wavelengths on the order of 50cm" or the use of "nowadays"). I've noted some of these at the beginning, but there are many I didn't note.**

Thank you for your comments and corrections. We have worked more with the grammar.

**A few other general points:**

**1) There was little if any discussion of the initial power spectrum at deposition. This became particularly apparent in the removal of the annual cycle, which struck me as odd the one signal that you know the frequency of just gets thrown away. Is there no way to use the annual signal in a constructive way?**

We have added a small paragraph about the initial power spectrum at deposition (Sec. XX line XX).

We are not aware of a way to utilize the annual signal in the diffusion length estimation, but it is possible to use the annual peak as in the Johnsen et al. (2000) paper. For measured d18O isotopes in the GRIP ice core, they integrated the power of the annual signal over different depths intervals. By doing this, they saw that the amount of diffusion that had attenuated the annual signal was not compatible with the diffusion model. They called that excess diffusion. However, excess diffusion has not been seen in other ice cores.

**2) The use of warm and humid to describe a site with an annual average temperature of 30C**

**and an annual average accumulation of 22cm per year is very confusing. Not even among ice core sites would this be considered warm and humid. Why not just use CG central Greenland and EAP East Antarctic Plateau since this is what you really mean. In addition, CD is not representative of much of Antarctica, just the East Antarctic Plateau.**

We agree that the terminology may seem confusing and we have changed the names to Case A and Case B.

**3) The ability of the single diffusion length methods to reconstruct synthetic data to within 0.5C seems quite good. But as I think about it more, I wonder if this is really something to be excited about. Given that the amount of temperature change within the Holocene is about 0.5C, does this imply that water stable isotope cannot be used to infer anything about temperature in the Holocene? The synthetic tests don't really address the question of what amount of temperature change can be confidently identified. For instance, what percentage of the time can a synthetic data set created with a surface temperature of 31C be distinguished from a data set create with 30C?**

In the new manuscript, uncertainties related to the firn diffusion model, densification and ice flow inaccuracies are included. The uncertainties are included in order to facilitate a better comparison between the single and differential techniques (in case the single diffusion methods are more sensitive to such uncertainties). As a result, the uncertainties of the single diffusion methods are therefore around 1.0C. The test with synthetic data showed that the methods where able to accurately reconstruct temperatures (within 0.2C) despite a precision of 1.0C. We acknowledge that the uncertainty of the temperature estimates might exceed that of variability in the Holocene. However, it should still be possible to reconstruct reliable temperature transitions. This is supported by Gkinis et al (2014) who using the single diffusion of d18O accurately and precisely captured the timing and magnitude of the Bølling-Allerød and Younger Dryas transitions.

TO DO.. Test with the relative temperatures..

**Specific comments:**

**Line 30 add "of the water isotope diffusion methods" so as not to imply that the water isotope diffusion methods are better than other methods, i.e. borehole temperature reconstructions**

**Line 36 change "invaluable". Maybe use unique or some other qualifier, but something with less subjectivity.**

**L43 delete "however"**

**L45 reword this sentence. Don't start with "However"; I'm not sure what "more physical principles" means, should it read "such as"?**

**L51 seminal seems strong for a paper read only by the very small community into water isotope diffusion**

**L57 " More specifically" does not seem like the correct transition**

**L59 " Nowadays" is a bit archaic and informal, no?**

We thanks the reviewer for the comments and all of it has been corrected.

**L62 A reference to the Univ. of Colorado system is noticeably absent here. Jones et al., 2017, JGR.**

A reference to Jones et al. (2017) JGR is now included.

**L76 define what you mean by flavors. I think you can make this word choice work, but not without explaining how "flavor" applies to water isotope diffusion**

We changed it to varieties.

**L83 Sentence beginning "For some cases" does not make sense to me. Also, the Holocene Climate Optimum is a bit nebulous, so you should define what you mean.**

We have defined like Dahl-Jensen et al. (1997) as a warmer period spanning from 5-9 ka with temperatures up to 3 K warmer than present day (Sec. XX, line XX).

**L85 don't call your own work "interesting". Let the reader decide that.**

**L 92-93 If you use "can" then "possibly" is redundant**

**L94 delete "that have dealt with the matter". This is implied**

**L97 I like this approach and feel you have struck a good balance with what information is included in the main, what's in the appendix, and what is referenced**

**L117 use different word than "sensor"**

The required revisions have been made.

**L137 I wonder why the authors only considered such cold sites. Given that much of the paper focuses on synthetic data, why not include coastal sites with warmer temperature and higher accumulation rates. This would make the work more broadly applicable, particularly since the vast majority of ice cores (not necessarily deep ice cores) are being drilled in locations warmer and wetter than either of the CG or EAP situations.**

See previous comment about how melt layers affect the diffusion length estimate.

**Figure 2.1 I favor more detailed figure captions so that a reader can scan the manuscript and know what's going on in a figure without needing to find lots of details in the text. In this case, provide the accumulation rates and temperatures for WH and CD. I'd also like a sentence describing the main takeaway from the figure; something like: The increase in diffusion length with depth from vapor transport is partially offset by compaction of the firn which moves the ice closer together.**

**Figure 2.1 why is there a kink at ~20m depth. Does this have to due with the transition from Zone 1 to Zone 2 densification?**

Yes the kink around 20 m is due to the transition from zone 1 to zone 2 densification. We have clarified that in the text and figure caption (Sec. XX line XX).

**Figure 2.1 make the legend lines bigger**

**L150 " borehole temperature profiles" is the more common description**

This is has been clarified.

**L150 This sentence confuses me. Do you just mean that the temperatures you use in your model are based on borehole temperature profiles? The sentence implies that you are obtaining the diffusivities from the high precision borehole temperature profiles. But really you just need a number to plug into the Ramseier equation.**

This is indeed what we are implying. We used the high precision borehole temperature profiles with depth when we solve the ice diffusion equation. It is possible to just use a constant temperature, but we do this in order to have a good accuracy. It should not have a large influence for the Holocene ice core sections that we use, but it will have a significant influence when you get closer to the warmer bedrock.

**L166 " in" should be "on"**

**L178 this notation comes out of nowhere**

**L232 add comma after f\_lamda**

This has been corrected.

**L232 is there really nothing useful to be gained from the annual signal? This raises the question of why the power spectral density is the way it is.**

See previous explanation.

**L270 is the ice diffusion the same for all isotopologues? What is the evidence for this?**

TO DO: Find reference!!!

**L284 delete the comma and change "can be" to "is"**

**L285 define dxs. This is the first usage**

Dxs was defined in the following sentence.

**Reviewer #2:**

**Review of "Molecular diffusion of stable water isotopes in polar firn as a proxy for past**

**temperatures" by Holme et al. Holme et at. compare several methods of reconstructing paleotemperatures from water isotope diffusion lengths. Based on synthetic and real ice core data, they argue that single diffusion length estimates are more reliable than differential methods, both in terms of their precision and their accuracy. The comparison of these different methods is a good and timely contribution to the literature, and should be performed. However, due to the strong reliance on synthetic data, it is unclear to me whether these conclusions will actually hold for realworld studies. I think additional work is needed to provide realistic estimates of the accuracy of the method when applied to realworld problems.**

**I personally have not worked on the problem of water isotope diffusion, but I've been following the field out of interest. In principle the method is very original and powerful, but I've always had the feeling that several key uncertainties are being ignored, which means that the cited uncertainties are much too optimistic. The 0.5 degree accuracy claimed for the single profile method seems unrealistic to me. The method shows up to 34 degrees offset for modern sites (Fig. 5.2), which seems like a much more realistic and meaningful estimate of the accuracy to me.**

**My main concern is that there is likely a lot of missing physics in the forward diffusion model, which is not accounted for. This is particularly problematic when using synthetic data, as I will argue below.**

**Some missing phsyics includes:**

**1) The work assumes that the initial deposited isotope record has a more or less white noise spectrum. This assumption is made implicitly in the work of Holme et al., and never really addressed. This assumption is baked into Eq. 4.2. We know for certain that the annual cycle has a big imprint, which is briefly addressed (but not included in the synthetic data, as far as I can tell). Besides that, Greenland ice cores are strongly affected by the NAO, and the WAIS core by ENSO/PDO variability as well as the SAM. We know that all these modes of variability have considerable power in the subannual bands. Also, snow is deposited in finite events (up to**

**several cm of precip each), which leads to considerable autocorrelation. How does this influence the accuracy of the methods?**

**There have been studies of isotopes in surface snow and vapor. Can the assumption of whiteness be proven or disproven?**

It is correct that we assume that the initial deposited isotope record has a more or less white noise spectrum. We have clarified that further in the revised manuscript in Sec. XX line XX.

Even though it is correct that the climate modes such as NAO, ENSO/PDO and SAM affects the d18O/dD signals, we can’t see any periodicity in any the power spectra (Appendix F). We acknowledge that such climate modes have the potential to artifactually affect the diffusion length estimate in a similar manner as the annual peak, but no corrections are implemented as we can’t see the impact on such small ice core sections. We have included a comment with that in Sec. 4.1 line XX.

TO DO: How does finite sampling affect autocorrelation?

The assumption of whiteness has currently not been proven, but Gkinis et al (2014) examined how redness in the climate affected the diffusion length estimate. They found no significant influence.

**2) There is considerable uncertainty in the densification model and tortuosity parameterization. There is a large literature from gas studies that looks into tortuosity (in models, measurements, and firn pumping experiments). These suggest that tortuosity is highly sitespecific, as well as scale dependent (for example, cmscale tortuosity is different from mscale).**

**This will skew the reconstructions (it's not clear in which direction…). This bias will not show up in synthetic data tests, because both the forward and inverse model assume the same firn properties.**

It is correct that several studies suggest that the tortuosity is site specific. However, a new study by Schaller et al (presented at EGU, 2017) shows that the trapping of bubbles in a single layer is solely determined by its total porosity and thereby independent of depth. Their results agree well with the Schwander et al (1988) model that we use. We acknowledge that the tortuosity parameterization still is simplified, but by assuming this simple parameterization we reconstruct temperatures similar to present day. This indicates that it still might be a fine approximation when used in cumulated estimates. In the revised manuscript, we incorporate such uncertainties by including an uncertainty in the close-off density. This is done in tests with both synthetic data and ice core data. We have made a note in Sec. XX, line XX that future research should be on making examinations on how the tortuosity parameterization influences the diffusion model.

**3) Does the seasonal temperature cycle matter? This drives both intense grain metamorphism (depth hoar) and influences vapor pressure and diffusivity (exponentially!). Density layering in general is probably important, because the diffusion length is not a single value, but presumably varies within each annual layer. This means that the final PSD is not Gaussian, but a superposition of Gaussians with different sigma values.**

We have included a seasonal temperature signal that propagates down through the firn. This firn temperature profile is used in the diffusion model. A comparison between a firn diffusion length profile using a constant and a seasonal firn temperature profile can be seen in Fig 2.1.

**4) Recently, it was shown that at WAIS Divide the diffusion length changes through time in a very unpredictable way (doi: 10.1002/2016JF003938). Also, field tests suggest large differences between modeled and observed diffusion (doi:10.5194/tc910892015). All these observations argue for missing physics. I do not pretend to know what such physics would be; I just think that these observations should caution us that the physics is not as simple as Johnsen et al. (2000) would have us believe.**

It is correct that the WAIS Divide diffusion lengths have an unpredictable profile as presented in Jones et al (2017). However, an interesting feature with that core is that the diffusion length profile between 11-19 ka (around the irregularity) correlates significantly with the measured total air content of the core (Edwards et al. , 2017, unpublished) This correlation suggests that it might not be the diffusion model itself that is wrong, but that there indeed is some climate related process happening. Such a process could be past elevation changes or thermal gradients in the snow. As the West Antarctic ice sheet currently is very unstable, past elevation changes are not unlikely. But whether it can account for the entire variability is currently unknown. It is indeed something we will look into when the total air content data are published.

In the van der Wel et al. (2015) paper (**doi**:**10.5194**/**tc**-**9-1089-2015)**, the authors estimate diffusion lengths from field measurements at Summit, Greenland. They estimate less diffusion than predicted by the Johnsen (2000) model and they conclude that several aspects of firn diffusion are poorly constrained (in particular the tortuosity). While we acknowledge that the Johnsen (2000) diffusivity parameterization is a simplification of some complex physical processes, we have to point out that we are able to reconstruct realistic temperature estimates when we use the parameterization in our diffusion model. This suggests that the parameterization probably is reliable when used to calculate the cumulated diffusion length at pore close-off. Moreover, a major difference between that study, and what we are doing in this paper is the depth scale of interest. We look at the cumulated diffusion length at pore close-off (60-80 meter), while van der Wel et al. (2015) examines the specific amount of diffusion in the upper 3 meters. Despite possible errors in the tortuosity parameterization, van der Wel et al. (2015) note that the discrepancy between their data and the Johnsen (2000) parameterization cannot be explained by higher tortuosity alone, as the value that would be needed to fit their data is outside the plausible range.

**The use of synthetic data is interesting in principle, because it allows testing in a controlled environment. But it can also be very misleading, because it assumes that the forward model (i.e., how the diffused isotope records are created) is exactly known which I think is a fallacy, see above. Perhaps I misunderstood, but it seems that in the synthetic tests the exact same firn properties (density, tortuosity) are used in the forward model, as in the "inverse" model (with the inverse model I mean the calculations used to go from sigma to temperature). I think this artificially improves the accuracy, and favors the single diffusion length methods (perhaps explaining why they look so good in the synthetic comparison). The differential diffusion methods should be less sensitive to these errors, because both isotopes diffuse through the same firn (tortuosity, density), canceling out some of these uncertainties. What I would advise the authors to do is to build some of this missing physics into their forward model that generates the synthetic data series. For example, they could use random number generators to control annual accumulation rates, temperatures, precip events, firn density layering and densification rates, tortuosity parameterizations, fractionation factors, etc. The isotope input spectrum should have an annual cycle, as well as a "redness" that is randomly selected. These random perturbations are not known to the part of the model that interprets the synthetic data. Then when the authors use the synthetic data to reconstruct T, there are unknown "biases" in their data, which is a much more realistic test that currently performed.**

We have modified the test with synthetic data so uncertainties related to the densification process, diffusion and ice flow model are incorporated. We still generate synthetic data based on an AR-1 data series without an annual peak. We don’t think including an annual peak in the time series will make the synthetic data more realistic. In most cases, there is not an annual peak present in the power spectral densities of the ice core data (See Appendix F).

TO DO: Give more arguments against including an annual peak

**The authors use an AR(1) process currently, but it is clear that the model does not introduce any meaningful level of redness (See Figure C.2). Where do the autocorrelation parameters come from?**

The model does introduce a meaningful level of redness. The power spectrum is only white due to the sampling process masking the redness. We have included an extra figure to show the initial power spectrum before sampling – here the redness of the signal is apparent. The selected autocorrelation parameter of 0.3 is based on experience from power spectrum of ice cores. Moreover, Gkinis et al. (2014) showed that the autocorrelation parameter did not significantly influence the diffusion length.

**My gut feeling is that the single diffusion methods are much more sensitive to such "unknown physics", because they will cancel out to some degree in the differential methods. How does the inverse model deal with accumulation rates? Are these assumed to be known exactly? Here some noise should be added too, since these are never known precisely. The same is true for the thinning function. There are different types of uncertainties: unknown firn physics, fractionation factors, accumulation rates, ice thinning, etc. In general, it would be interesting to know how sensitive these methods are for each of the sources of uncertainty. That could help decide which method is most suitable under which condition.**

Uncertainties related to the accumulation rate, thinning, surface pressure, ice diffusion, surface density and close-off density have been included in the revised manuscript (Table 4.1).

The physical conditions have an impact on which is the most optimal diffusion technique. For instance, when the ice diffusion exceeds that of the spectral estimated diffusion, the single diffusion techniques become invalid. This occurs for really old ice or when the ice is near a warm bedrock. A similar thing can happen when the sampling diffusion exceeds that of the spectral estimated diffusion. This is also likely for old glacial data where the sampling diffusion is estimated from CFA data. In these cases, the differential diffusion methods are the most optimal as they are independent of ice diffusion and sampling diffusion. However, this does not apply on the ice cores used in this study as they are from the Holocene period. So in spite of the differential diffusion techniques potentially being less sensitive to such model parameters, our test with synthetic data shows, that the most accurate and precise techniques are the single diffusion methods.

A more thorough test of how sensitive a technique is to changes in each climatic parameter will not be performed in this study as it would be to extensive for this manuscript. It could be interesting to look into it in future research, but this study focuses on the total precision each technique.

**More detailed comments:**

**Line 44: assuming a linear relationship is of course not the same as using the spatial slope. The temporal slope is also a linear one.**

We have rephrased to a constant linear relationship.

**Line 47: Schwander 1988 is not the correct reference here. Thermal fractionation wasn't discovered until a decade later.**

Thanks for seeing that. That reference has been deleted.

**Line 51: While Johnsen 2000 is undoubtedly a good paper, there were many others that made just as seminal contributions, like Whillans and Grootes 1985, and Johnsen 1977.**

We have updated the text with these references.

**Line 85: HCO is poorly defined, give the age!**

We have defined it and provided an age (see Sec. 1, line 91)

**Eq 2.1: Technically D is a function of depth also, particularly in the presence of layering. Also, this equation is an approximation, because delta is not as mass conserving quantity.**

TO DO:

**Equation 2.2: is this an exact solution, or an approximation?**

This is an exact solution. Confirmation from Vas and Bo needed

**Line 117 and elsewhere: This is confusing. How can sigma be used to reconstruct both T and A? It has to be either one. You cannot reconstruct two independent parameters from a single number.**

It cannot be used to reconstruct both T and A at the same time. In order to reconstruct T, A is assumed known. Similarly, A can be reconstructed assuming the temperature is known from another proxy. We have clarified that in the text (Sec. 2, line 136-138).

**Line 126-127. This is conceptually sloppy. Both drho/dt and D do not care about accumulation rate, of course. drho/dt cares about T and overburden pressure (not A), and D cares about temperature and firn connectivity.**

Our statement was based on the parameterizations that we use in the study. We have changed it in Sec. 2, line 135-136 so the description is conceptually correct.

**Line 128: again, I don't think you could do both T and A, unless you have some other source of info.**

See previous comment.

**Line 129: what are the typical measurement precisions for these various isotopes?**

They are 0.07, 0.50 and 0.05 permil for respectively d18O, dD and d17O. We wrote that in the Appendix D, line 623.

**Line 139: above the closeoff depth ice diffusion also happens of course, within the grains.**

Does it???

**Figure 2.1: Why Fig 2.1? please change to Fig. 1**

**I should change this in the manuscript**

**Eq. 4.2: Shouldn't P\_0 be a function of (k) also? It seems to be a constant here.**

Yes, but we assume that P\_0 is constant for each section due to strong depositional noise. We have clarified this in Sec. 4.1 line 194-197.

**Line 171: I am confused by the use of AR(1) processes. In Eq. 4.2, the noise term represents the instrumental noise floor, correct? Shouldn't this be white noise? That's what's assumed elsewhere on Line 561. Also, this AR(1) is distinct from the AR(1) used in appendix C to estimate the input time series, correct?**

Eq. 4.2 represents the total noise model combining post depositional and measurement induced noise. That is why we fit the noise for each section. We have clarified that in Sec. 4.2 line 199-200.

In the generation of synthetic data we only add white noise, as we don’t have any knowledge about post depositional process. In that sense the synthetic data test and ice core data test differs.

**Line 174/Fig 4.2: Could you show us some more real data PSDs, instead of synthetic ones? Ideally at least one PSD per core site (can be an appendix). Of course the fit to the synthetic data is good, because the synthetic data has the assumption of being Gaussian baked into it. I'm curious to see how well your method fits the actual databased spectrum at various sites. Also, do the spectra at Greenland and Antarctic sites look very different?**

We have included all the fitted spectra used in this study in Appendix F (together with the estimated diffusion lengths said figures). Here you can see that the spectral-based techniques perform well on real ice core data.

The spectra from Greenland and Antarctica do not look very different. The only visible difference is that the noise level looks almost completely white for the Antarctic ice cores.

**Line 179: This is a little unclear. What is the difference between regular P and fancy P? Is regular P the model, and fancy P the observations?**

Regular P was defined as the model and fancy P as the observations. We have clarified that further in order to avoid misunderstandings (Sec. 4.1 line 205-206).

**Line 179: It appears that the background noise level is used as a fitting parameter. Isn't the noise level known from the instrument used?. Do the fitted noise levels agree with the known analytical precision of the instrument?**

The background noise is used as a fitting parameter because the noise model includes both post-depositional noise and instrumental noise. If we fit a white noise line at the highest frequencies it agrees well with the analytical precision of the instrument. This has also been shown by Gkinis (2011).

**Line 196: I assume integrating Eq. 2.7 includes all the parameterizations used for tortuosity, density, thinning, etc? Please be more clear about this.**

We already defined that earlier in Sec. 2, line 131-138.

**Line 205-213: is the single diffusion method more sensitive to this than the differential one?**

No. The same correction is used in both cases. The diffusion length ratio is the only technique independent of densification processes and ice flow.

**Line 218: the annual peak is clearest when there is little diffusion, this is obvious. But the enhanced power in the annual band (in the input spectrum) should be preserved under all circumstances, correct? It may just be smeared out at high diffusion sites, but the power is there. It seems to me that at all sites the annual band should be disregarded from the spectral fitting. Please confirm how this is done. Also, is the annual band avoided in all analyses done here?**

We have only disregarded the annual band from the spectral fitting if it had a spectral signature. The annual band was removed in 3 out of 13 data sets. We have clarified that in Sec. 4.1, line 267 and the power spectra can be seen in Appendix F. The reason why we don’t exclude the annual band everytime is that the amplitude of the annual peak smoothens out too much so that the Picarros can’t measure it. Give example

TO DO: Discuss why we haven’t excluded the annual band in all cases.

**Why was there no annual peak added to the synthetic series?**

Even though the isotopic signals initially have an annual signal in ice cores, the signal diffuses out. As that is the case in most of our ice core data (see Appendix F), we chose not to include it.

**Equation 4.7: What value of df\_lamba do you use?**

We picked a weight function of w(f\_lambda – 0.5 < f < f\_lambda +3 ) = 0 in all cases. We decided to filter out more of the higher frequencies.

**Figure 4.2: please use real data here, synthetic ones are not as interesting.**

We have provided real ice core data in Appendix F. We prefer using the synthetic data as it facilitates a better comparison between the different techniques and power spectra.

Should we change it to real ice core data???

**Figure 4.5: are these real or synthetic data? I guess you know my preference by now :)**

The data in Fig. 4.5 is from synthetic data.

**Section 4.3: since this method performs so poorly, perhaps it can be left out to shorten the MS?**

We would prefer not to remove this technique from the study, as this is the first time the method is utilized on ice core data. Despite its poor performance, the method is still completely independent on densification, tortuosity and ice flow. Future research could be on figuring out a way of estimating the diffusion length ratio directly instead of through the single diffusion lengths method. This is currently not possible. But the technique still has a lot of potential.

**Eq. 4.12: Define alpha**

Done

**Line 321: Here I was confused. Is the noise white or AR(1), as claimed earlier?**

The added noise is white as that represents the measurement uncertainty. We use an AR-1 noise model when we fit data power spectra. With respect to the synthetic data

**Line 324: As mentioned above, I think each time you generate synthetic data you should also perturb the firn physics, because this is an unknown in any realworld problem.**

Done

**Figure 5.1: Do you have any idea where the clear 17Delta^2 bias comes from?**

The bias is not evident in the new simulations. The bias in the previous manuscript was a result of a small offset in the estimate of the sigma\_17 value. That propagated into the differential diffusion technique. That is the reason the bias was not present when using the linear fit technique.

**Lines 368-371: I couldn't agree more! This makes all the methods look too good, and as I argued above, particularly the single diffusion case.**

**Table 5.2: The selection of depth ranges seems somewhat arbitrary. Why were these depths chosen?**

The depth intervals may seem arbitrary, but the used ice core section represent all the usable data we could obtain. So we did not choose the data based on the depth section, but on what we had available in the Holocene period.

**Line 389: 0.5deg C seems too good to be true. Many uncertainties were not really considered, so this number is not all that meaningful. The way of things is that this number will be cited by others in future work without all the caveats needed to understand that this value holds only under very narrow conditions of synthetic data in a perfect world where all physics is understood. I think this should be revised upward by including all the relevant uncertainties, in order for this number to be a meaningful reflection on the accuracy of the method in realworld applications.**

See previous comments.

**Table 5.1 It's not precisely clear how you go from sigma to T. What assumptions were made regarding the accumulation rates, for example? Were they precisely known?**

See previous comments.

**Line 417: Why not just compare the reconstructed temperature with the modern day temperature? I understand there can be differences, but comparing to present day is the only meaningful way to test the accuracy of the method. The synthetic data are artificially unbiased because you assume you understand all the physics. Comparing the reconstructed temp from real isotope data to observed temps is the only meaningful metric to my mind. The abstract claims that the single isotope methods are the most accurate, but this doesn't always seem true when I look at Fig. 5.2, 5.3 and 5.4.**

We don’t agree that comparing to present day is the only meaningful way to test the accuracy of the method. In terms of accuracy, this implies that we know the exact temperature we are trying to reconstruct. Although the temperature has been relatively stable in the Holocene, this study reconstructs temperatures for small ice core sections and it is very likely that we can see higher-frequency temperature variability. That’s why we use the synthetic data test to quantify the accuracy.

**Line 485: I don't believe you have proved that the method is unbiased, because in the synthetic world you know all the physics and parameterizations, which is not true in the real world.**

In the new simulations with synthetic data, uncertainties related to densification, diffusion and ice flow are included (see previous comments). Our results still indicate that the methods are unbiased, especially the single diffusion and linear fit techniques.

We will never be able to exactly know all the physics and parameterizations in the real world. We can only improve our current understanding

**Line 502: Again, I don't think accuracy has been proven given that you didn't compare the temperatures reconstructed from the real data to real site temperatures.**

See previous comments.

**Eq. A7: Note that there have been multiple different parameterizations. This is a very old study. A more recent result to use would be doi:10.5194/tc83192014, or doi:10.3189/172756402781816582.**

**I should acknowledge that in the appendix. We use that parameterization because the temperatures we reconstruct looks realistic. We acknowledge that future work could be to do more research on the influence from various parameterizations on the diffusion proxy**

**Line 556: what are these parameters for the AR(1) input based on? Actual data? The input spectrum of the synthetic data is very white, and does not have an annual cycle.**

The input parameters have been chosen empirically by examining power spectra similar to the warm and cold conditions. The input spectrum is red but it looks white due to the sampling process masking the redness. We have include an extra figure showing the raw power spectrum before sampling.

**Reviewer #3:**

**This manuscript is well written and convincing for the material it chooses to cover.**

**My main reaction, however, is that it leaves the problem only half analyzed and this leads to the final conclusion being a logical fallacy.**

**The Abstract and the Conclusion sections finish (lines 2930, 501502) by claiming "this study has demonstrated that methods based on the single isotope diffusion length result in the most accurate and precise estimates of past temperatures." But the authors have chosen not to analyze the factors that can make the single isotope methods problematic, in particular uncertainties in (1) firn densification and density-depth profiles as a function of climate, and (2) uncertainties in how**

**vapor diffusivity varies as a function of density. These must be addressed!**

**The potential value of the multiple isotope analyses is to find methods independent of these uncertainties. The present analysis does a good job of revealing the inherent weakness of the multiple isotope methods, in turn.**

**This is an original and useful contribution. But ignoring the other side of the problem means that no conclusion can be reached about the relative or absolute merits of the different methods.**

In the new manuscript, uncertainties related to the firn diffusion model, densification and ice flow inaccuracies are included. The uncertainties are included in order to facilitate a better comparison between the single and differential techniques (in case the single diffusion methods are more sensitive to such uncertainties). We have also included a seasonal temperature signal that propagates down through the firn. The seasonal temperature variation affects the firn diffusion length nonlinearly due to the saturation vapor pressure’s exponential dependence on temperature. The firn temperature profile is obtained by numerical solution of the heat equation.

**Another disappointment of the submitted paper is that the authors do not take advantage**

**of the ice core data to quantify the accuracy of the thermometry methods. The reason,**

**apparently, is that temperature varies (with depth and/or time) in the firn column, whereas**

**Eq. 2.7 is for a steadystate (although this is never stated explicitly. I might be misinterpreting this.) Why don't the authors use Eq. 2.5 itself and timedependent calculations of temperature and density to overcome this limitation??**

The diffusion length that we are able to estimate represents the cumulated diffusion length at pore close-off. We can therefore not use Eq. 2.5 to make time-dependent calculations as we only have the final value. As a result, we are only able to reconstruct the integrated firn temperature profile (this has been clarified in Sec. 4.1, line 221).

Since we don’t know the exact temperature at a given depth/time, we can’t quantify the accuracy using ice core data. We are only able to calculate the deviation from the annual temperature.

**The terrain covered by this analysis is already a narrow one for GCA. The authors can**

**probably have this manuscript accepted almost "as is" at a good journal focusing on**

**ice core or glacial studies. It would be much more useful to the readership, and to the**

**authors' scientific reputations, if GCA returns this manuscript to them with instructions to**

**expand the analysis to enable a useful and cogent conclusion. Then it will make a fine**

**GCA paper.**

**A few minor points:**

**1. the authors neglect to cite the Whillans and Grootes JGR 1985 analysis, which is**

**the source of most of our understanding of isotopic diffusivity**

We cite the paper in the revised manuscript (Sec. 1, line 57).

**2. looking at Figure 5.2, is it just a coincidence that the Delta sigma^2 I method**

**(the green circles) does a good job of matching the modern temperatures? A better**

**job than any of the other methods?**

Some of it is could be a coincidence. But it is indeed interesting – discuss more

**3. there is a recent paper by T.R. Jones et al. in JGR Earth Surface (2017) that looks at isotope diffusion in the WAIS Divide core. It is a case in which the relationship between diffusion length and climate does not seem to work as expected.**

It is correct that the WAIS Divide diffusion lengths have an unpredictable profile as presented in Jones et al (2017). However, an interesting feature with that core is that the diffusion length profile around 11-19 ka (around the irregularity) correlates significantly with the measured total air content of the core (Edwards et al. 2017, unpublished). This correlation suggests that it might not be the diffusion model itself that is wrong, but that there indeed is some climate related process happening. Such a process could be past elevation changes or thermal gradients in the snow. As the West Antarctic ice sheet currently is very unstable, past elevation changes are not unlikely. But whether it can account for the entire variability is currently unknown. It is indeed something we will look into when the total air content data are published.

**References**

Edwards, J. E. B., Fegyversi, J. Sowers, T., Alley, R., Severinghaus, J. Buizert, C., Mitchell, L. (2017, unpublished). Determining climatic controls on total air content and elevation changes from the WAIS Divide Ice Core. Manuscript in prep.

Gkinis, V. (2011). High resolution water isotope data from ice cores.PhD thesis, University of Copenhagen.

Gkinis, V., Simonsen, S. B., Buchardt, S. L., White, J. W. C., Vinther, B. M. (2014). Water isotope diffusion rates from the NorthGRIP ice core for the last 16,000 years – glaciological and paleoclimatic implications. *Earth and Planetary Science Letters.* Vol: 405, pp: 132-141.

Lamb, K. D., Clouser, B. W., Bolot, M., Sarkozy, L., Ebert, V., Saathoff, H., Möhler, O., Moyer, E. J. (2017). Laboratory measurements of HDO/H2O isotopic fractionation during ice deposition in simulated cirrus clouds. *PNAS*, Vol: 113, no: 22, pp: 5612-5617.

Schaller, C. F., Freitag, J., Sowers, T., Vinther, B., Weinhart, A., Eisen, O. (2017, unpublished). In-situ observation of bubble trapping in polar firn. G*eophysical Research Abstracts Vol. 19, EGU2017-14721-1, 2017 EGU General Assembly 2017*

Schwander, J., Stauffer, B., Sigg, A. (1988). Air mixing in firn and the age of the air at pore close-off. *Annals of Glaciology*. Vol: 10, pp. 141-145.