Response to review 1 (anonymous)

First, we thank this reviewer for a close reading and a comprehensive review. This was extremely helpful and we appreciate it.

The main text of this review focuses on the issue of the amplitude of variation in reconstructed Be-10 fallout fluxes. Ideally, if one wants to argue that these really are variations in fallout fluxes, it is not enough to show that the pattern of variability is the same as seen in other fallout records – it is also necessary to show that the magnitude of variability is compatible, or at least not physically impossible. This is important, and in this paper we have not shown that the relative magnitude of variability in our reconstructed fallout fluxes is the same as expected from independent records. This reviewer then goes on to propose that because of this, the correlations between NAVC and other fallout events are speculative.

Basically, we agree with this reasoning, with one important exception. The problem is that the relative amplitude of our reconstructed fallout variability is not reliable because of its dependence on the assumed inherited ratio Rs. Figures 1 and 2 below show this for the Glastonbury (GL) decadal record shown in Figures 16-17 in the paper. In Figure 1, we calculated the fallout flux $Q_{10,a}$ for three different assumed values of Rs, which shows that these different values yield a similar pattern of variability in the reconstructed fallout flux, but different values for the mean fallout flux. Similar absolute variability imposed on different mean values results in different relative magnitudes of variability. Figure 2 shows this effect: higher values of Rs predict higher relative variability in the reconstructed fallout.

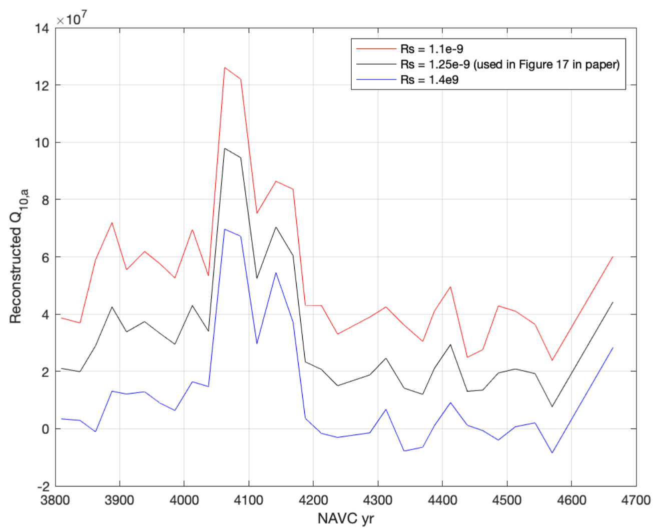

Figure 1: Fallout fluxes for Glastonbury (GL) decadal record reconstructed with different values of R_S. Of course, the highest value shown is unphysical because it predicts some negative values, but it makes the point clear.

This effect is just a property of the linear model when applied to data sets whose intercept is close to the origin. As can be seen from, e.g., Fig. 13 in the paper, an uncertainty in the intercept (which should give the mean fallout flux under simplifying assumptions as described in the paper) that is small in absolute terms is large in relative terms just because it is close to zero.

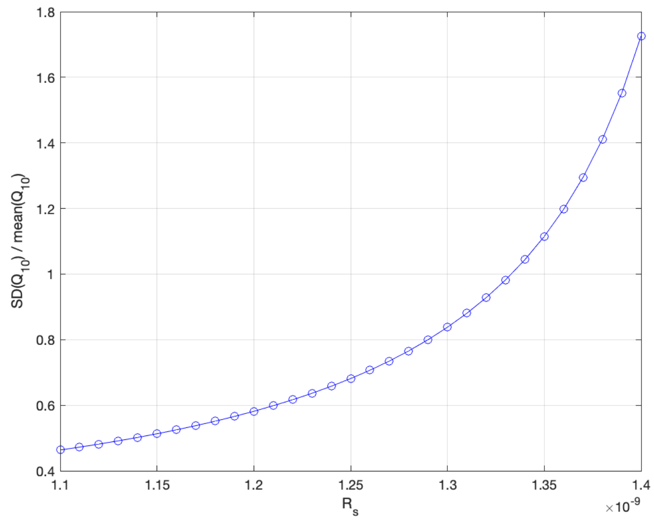

Figure 2: Effect of different assumptions for R_S on relative variability of the fallout flux for the Glastonbury (GL) decadal record.

This becomes a serious problem for us because for the decadal records, we really have no way to reliably estimate Rs, as we discussed on p. 30-31 in the paper. If we can't estimate Rs accurately, we cannot accurately estimate the mean fallout flux and therefore also cannot accurately estimate its relative variability. From the perspective of the reviewer's point, therefore, we cannot use the relative magnitude of our reconstructed fallout flux as a point of comparison with independent records of Be-10 fallout. In other words, we cannot disprove the hypothesis that the relative magnitude of variability in our reconstructed fallout records is similar to that in independent records. We also can't disprove the hypothesis that they're different.

So, because of this problem, in discussing potential matches between our records and independent fallout records, we have proceeded from the assumptions that our inability to estimate Rs accurately leads to a situation where we may have correctly reconstructed the pattern of variability, but we cannot assume that we have correctly reconstructed the amplitude of variability. This is why we subsequently rely only on scaled and centered records in the discussion of correlation between records.

To summarize, we agree with the reviewer that convincingly arguing that we have reconstructed fallout variability would require showing that both (i) the pattern and (ii) the amplitude of variability are consistent with independent records. We cannot determine whether the amplitude is consistent or inconsistent, so we have relied only on the pattern of variability. This is, in fact, an inherent weakness of our situation in which most of the Be-10 is inherited and very little is fallout.

To address this in a revised paper, we propose the following:

(1) Add additional discussion in the paper highlighting the difficulty of reconstructing the amplitude of fallout variability given a lack of knowledge of Rs. We did discuss this in section 4.1, and Figures 14 and 15 show examples of reconstructed fallout fluxes with different values of Rs that highlight the difference in amplitude but similarity in pattern between the resulting reconstructions. However, we did not repeat this discussion in section 4.2, where it is of course more directly relevant to the later attempts to correlate the

decadal records. We can add additional discussion of this.

(2) Add additional discussion of why we relied on scaled and centered records for combining the KF and GL decadal records and also in comparing them with ice cores. We agree that the existing paper doesn't give a clear justification for this.

(3) Add additional material to the "weaknesses" section of the conclusions emphasizing that the difficulty in reconstructing the amplitude of variability is a problem for evaluating whether correlations with other records are or are not valid.

It would be possible to add additional figures like Figs. 1 and 2 above in this review, but on the other hand an advantage of the open review system is that this review response remains available to readers. As there are already a lot of figures in the paper, we are inclined not to add more, but we would welcome editorial guidance on this question.

That completes our response to the major issue in this review. We now deal with the minor comments.

NAVC varve count error: In the entire discussion there is no mention of the varve counting (relative) error of the NAVC. Is this really zero? It should at least be noted in the appropriate sections that differences in the GICC05/NAVC offset may well in part be due to errors in the NAVC, not just GICC05.

P6, L1: See earlier: Could you please add a sentence on the counting uncertainty of NAVC? How valid is the approach of using a single value for the offset?

Basically, the NAVC is believed to have negligible counting uncertainty below approximately NAVC 7200, where nearly all parts of the sequence are replicated in multiple records. The counting uncertainty in stratigraphically higher paraglacial varves that are much thinner and occur only at the Newbury section is significant, and has been discussed in the Ridge (2012) and Ridge and Toll (1999) references. We agree this should have been mentioned in section 1.1 or 1.2, and we will add it. However, note that it's not relevant to the overall conclusions, because we don't consider any Be-10 data above NAVC 7200 for purposes of correlation.

9Be/10Be extraction: Typically, 9Be and 10Be are measured on the same leaching fraction of sediment. Here, the authors employ different extraction techniques for both isotopes. For this to work, the extraction efficiency of each method must be constant. Is this a problem? Maybe the authors could comment on whether this has the potential to introduce variations in 9/10Be.

Assuming that using two different methods does not bias the results relies on two assumptions. One, each method is equally efficient on all samples, i.e. that the leaching method for Be-9 extracts all Be-9 that is adsorbed and not mineral-bound, and that the fusion method extracts all Be-10, period. Both of these methods have been tested for complete extraction in a variety of step-leaching experiments for the Be-9 method (described in the Greene reference) and by a variety of tests for the Be-10 fusion method (described in the Stone reference). Thus, there is no indication that the efficiency of these methods is sample-dependent. The second assumption is that the total fusion method does not extract a significant amount of non-adsorbed Be-10 – the aim of the analysis is to determine the amount of adsorbed Be-10 and Be-9, so if we added Be-10 in minerals we might introduce a bias. The only possible source of mineral-bound Be-10 would be in-situ-produced cosmogenic Be-10 present in sediment that had been exposed to the surface cosmic-ray flux prior to glacial erosion and transport. Existing measurements of in-situ-produced Be-10 in subglacial sediment of the Laurentide Ice Sheet indicates in-situ-produced Be-10 concentrations on the order of 10,000 atoms/g, which is four orders of magnitude smaller than observed total Be-10 concentrations in this study. The highest in-situ-produced Be-10 concentration observed anyhere on the present land surface in New England is slightly less than 1 Matom/g, which is still two orders of magnitude less than total Be-10 observed in this study. Thus, it is not possible for in-situ-produced Be-10 extracted by fusion to cause any significant bias in the present study.

From the perspective of paper revisions, we would prefer not to recapitulate internal reliability tests already described in the Greene and Stone papers. However, we can add a more explicit discussion of the minimal importance of in-situ-produced Be-10 to the methods section....this is already referred to briefly on page 7, but it would probably be good to remind the readers in section 2.5 in the methods as well.

P 1, L2: replace ?calendar year timescale? with ?ice core timescale?

Well, really what we are trying to do is synchronize it with both. We'll clarify this.

P 1, L11-14: long, convoluted sentence. Consider dividing.

P 1, L20-24: difficult sentence. Similar but different?

P1, L20: ?not consistent with independent evidence? ? the previous sentence states the opposite?

We'll try to clarify these areas of the abstract.

P2 L15-16: Please indicate what this uncertainty refers to, i.e., how many sigma.

As discussed in the Ridge (2012) and Thompson (2017) references, this uncertainty is probably not symmetric or normally distributed, so characterizing it as having a standard error σ would be inappropriate. We'll clarify this.

P2, L28-29: ?generally globally synchronous?. See my comment on an additional sec- tion for the introduction. This is only true under ideal conditions and when averaging sufficiently over meteorological variability. In many settings, meteorological influences will be dominant over the production rate changes.

The purpose of this section is to give a simplified overview of the concept...that's why we said "generally globally synchronous" instead of "globally synchronous." We will add some clarification here.

P2, L30: Heikkilae et al. 2013 is not a good reference here, as they do not reconstruct solar variability. Please find more appropriate references such as: (Bard et al., 1997; Raimund Muscheler et al., 2007; Steinhilber et al., 2012).

Point taken. We will clarify this.

P2, L32: Again, Heikkilae at al and Adolphi et al. do not contain new measurements. Please give credit to the authors that produced these long records: (Adolphi et al., 2014; Baumgartner et al., 1997; Finkel and Nishiizumi, 1997; Raimund Muscheler et al., 2004; Vonmoos et al., 2006; Yiou et al., 1997)

Our aim is certainly not to disregard the contributions of the original researchers, but as a matter of scientific citation practice we think it is better to refer to these review papers, which provide extremely thorough and comprehensive reviews of past work, instead of attempting a weaker and less comprehensive review in this paper. The purpose of the present paper is not to review past work, so we have predominantly cited the review papers.

P3, L22-24: If there is a transient transition over 100-200 years in the lake system: How can a synchronization of local climate records to Greenland be attempted? The result would depend on which varve record was chosen? Could you please comment? What does this mean for the achievable uncertainty of climate synchronizations?

Correct: this would indicate that it would be a bad idea to try to correlate a composite record that spans the glacial-interglacial transition at one or more sites. We haven't tried to do this – our attempted correlations involve records that are entirely glacial (KF/GL) or span the transition at a single site (Newbury). At Newbury, of course, we later show that there doesn't seem to be any fallout signal in the paraglacial varves, so data from within and after the transition do not seem to have any value for correlation. Thus, this turns out not to be an issue.

P6, L16-26: Recently a new radiocarbon calibration was released (IntCal20) which contains significant changes from IntCal13. I understand that it is beyond the scope of this study to redo all the calibrations, but maybe it is worthwhile mentioning that this adds extra uncertainty to the (by looking at the data, in my opinion very optimistic) 200 years.

Agreed. It is awkward that INTCAL20 came out just as we were working on this paper. As a practical matter, I am not sure we can do anything about this without incurring major delays.

P6, L29-33: Please do not use the term ?climate events? here, when the event stratigra- phy by the ice core community does not contain any events during this time. Generally, the climate wiggle matching in this section should be cross-checked by not only using 1 ice core, since these minor wiggles in d18O may be noise. They should be verified using GRIP, NGRIP and NEEM d18O.

Here we were using "climate events" in a general sense to refer to any climate variations recorded in ice cores, not with reference to any specific published "event" stratigraphy. We can easily use "climate changes" or "climate variations" instead.

Another important point, which is also discussed below, is that we are specifically not trying to redo or re-examine the previously proposed varve thickness-oxygen isotope correlations. We are only taking note of previously proposed correlations and comparing them to possible Be-10 fallout matches. This paper is already long enough with just the Be-10 data.

P9, L15-16: Please add a citation to the figure for the fallout flux.

Agreed.

P9, L22: remove the ?-? in the 0.6m.

Matter for copy-editing.

P12, L1-3: This is only true if these inaccuracies are systematic. Are they?

They are certainly systematic in the sense that they can only lead to an underestimate of the density, not an overestimate. However, the reviewer's statement is not strictly accurate. Suppose we measured the density on a number of samples and the results showed a broad distribution that was skewed to the low side, indicating an underestimate of the density that had both random and systematic contributions. If we then took the average of all the density measurements and applied the average to convert concentrations to fluxes, we would, as is stated in the text, systematically underestimate the total Be flux. In other words, it would not be the distribution of the density measurements themselves, but the use of a simple average to summarize them, that would result in the systematic underestimate of Be fluxes. The point of collecting the data shown in Figure 3 was to to a better job than this.

P12, L9: Looking at figure 3 the +-0.05g/cm3 seem very optimistic. How would the regression in figure 3 change if the two summer varve measurements were excluded from the analysis?

The result would be similar, as long as the low-skewed residuals were also excluded. However, the purpose of making the summer and winter measurements was specifically to make this regression more accurate by spanning a wider range of grain sizes, so it is not clear why we would then exclude them.

P12, L20: replace ?between? with ?around??

We'll try to clarify this section.

P14, L9: Is there a +1 missing from the right hand side of the equation? Otherwise, if C/Cs equals 0.9, the right hand side becomes 0?

Yes, this is an error. Apologies. We will correct it.

P16, L1-2: Earlier you write that S may be variable? Does this correction come with an uncertainty that is propagated throughout the study?

We will clarify this. Overall we conclude from the comparison of data between AMS measurement periods that S is most likely constant. However, this is not actually relevant to the uncertainty estimate for the Be-10 concentrations, because we also have an internal criterion for how well the correction procedure performs that is provided by large numbers of replicate measurements. The uncertainty we eventually apply to the Be-10 concentrations is not derived from error propagation from the individual measurements, but instead from analysis of the corrected replicate data. Thus, it includes any uncertainty contributed by variable S.

P18, Figure 9: Around 6700 NAVC years, there are large changes in MAR and 9Be, but not in 10Be. While it could of course be, that simultaneous and large production rate changes (however, of likely unphysical amplitude, see major comments) ?counteract? this the changes in the delivery of inherited 10Be, this seems unlikely. Rather, this may highlight variable 10/9Be ratios in the inherited Be. Please add a few sentences discussing this feature.

Figure 9 shows grain size, not MAR. However, this is true and we have discussed it somewhat in section 4.2.2 that deals with the Newbury data that cover that period. This is an interesting point, though, because it is true that a large peak in reconstructed fallout flux around this time, which later becomes important in the correlation section, is associated with large variations in total deposition rate. As the reconstructed fallout peak is only represented by one sample, we agree this looks suspicious. However, this is exactly where we expect a large peak in Be-10 fallout based on ice core records and previous, independent, age calibration of the NAVC. There is no evidence that the Be-9 or Be-10 measurements in this area are inaccurate or anomalous, so we think we have to assume all the data are valid, apply the linear model to reconstruct the fallout rates, and then proceed with the results without second-guessing them. Disregarding a few data, no matter how suspicious they look, without evidence that they are spurious would be bad practice. So I think we are kind of stuck with this. We do highlight at the end of the paper that some of the apparent correlations are based on very poorly characterized peaks and it would be very helpful to replicate them at higher resolution.

P22, L8-9: ?constant or normally distributed?: But neither of these assumptions is true. It is obviously not constant (otherwise this study wouldn?t be possible) and it is also not symmetric because i) if solar variability was normally distributed, the non-linear production rate relationship would still cause a skewed distribution of production rates, and ii) the transport and washout of aerosols from the atmosphere causes a logarithmic distribution of aerosol deposition even if the production rates were normally distributed. It is still ok, to use the model as is, but it should be highlighted, that these assumptions are not true, but sufficiently correct to not affect the validity of the results.

Correct. We can clarify this. Basically, what we are arguing here is that short-term periodic variations should be close enough to symmetric that we can make progress with simple assumptions.

Figure 13: it would be interesting to see this figure in relative units, i.e., both 9Be and 10Be relative to their respective mean. If the assumption of a constant 10Be/9Be of the inherited Be was true, all data-points should scatter around the 1:1 line?

Here they are in Fig. 3 below, normalized to standard deviations from the means. However, I think this argument is only strictly valid, though, if the data are really normally distributed? Anyway, they are pretty close.

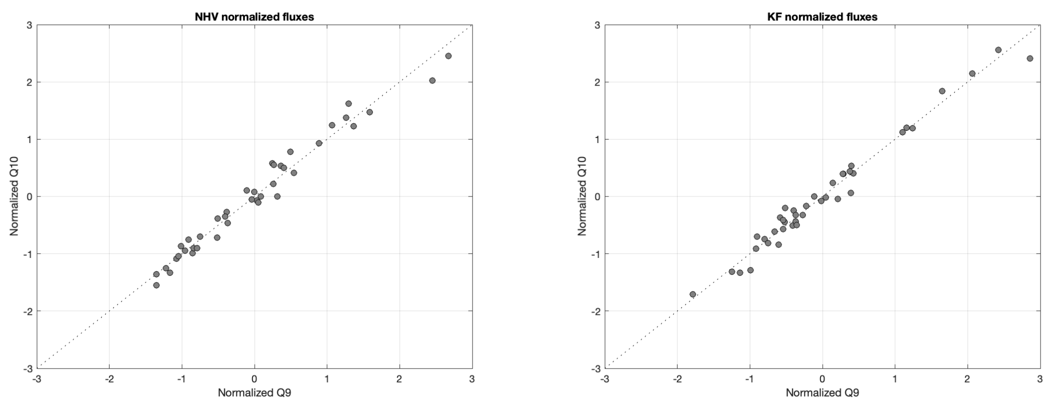

Figure 3: Relationship of normalized Q_{10} and Q_9 for NHV and KF biennial data. The dashed line is 1:1.

P26, L18: ?factor of 2? it should be noted that also ice cores are affected by transport and deposition effects of 10Be, especially during periods of such variable climate. As mentioned earlier, there are physical constraints of what amplitude of changes we can expect.

We agree.

Figure 18: ?95% confidence? How is this determined? Do you take autocorrelation of the records into account?

This is just a simple comparison against the expected distribution of the correlation coefficient for independent random data. Specifically, we used the MATLAB implementation described here:

https://www.mathworks.com/help/matlab/ref/corrcoef.html

P31, L3-4: I?d argue that the agreement is not that good? The amplitudes are different, and due to the similarity of both MAR records, the similarities in 10Be fluxes may well be cause by the flux calculation. Is it worthwhile discussing this option?

We agree that this is not established with extremely high confidence, but the two records have similarities not only in MAR but in the measured 10/9 ratio, which should theoretically be independent of MAR. It's the variations in the ratio that are most important to the flux estimate. If you compare MAR to the flux estimate in Figs. 16-17, it is evident that there are both (i) areas with fairly constant MAR but variable reconstructed flux (NAVC 3600-3800) and also (ii) areas with variable MAR but constant reconstructed flux (4300-4500). Given this, it seems unjustified to argue that only some variations in MAR create spurious flux variations. Having established how to do the flux estimate, we think we have to follow it through and not second-guess particular features later.

P31, L4-5: The general problem with matching these records is, that one will always find a match due to the periodicity of the signals. Hence, the amplitude discussion is important.

Figure 19: The doubling of 10Be/9Be between 6800 and 7200 NAVC cannot be pro- duction (unphysical amplitude). Please discuss.

Figure 21: See above

As discussed above at the beginning of the review, the amplitude of our reconstructed fluxes can't be taken to be reliable. Thus, although the reviewer is correct here, we aren't strictly able to conclude anything from the amplitudes.

P34, L24-35: I?d argue that the sediment and ice core 10Be records simply don?t look alike. Aligning a single peak can easily lead to erroneous results. Please find a mea- sure for the similarity of the records that can be used to quantify the uncertainty in the match as well.

Certainly we agree that aligning a single peak can lead to erroneous or at least non-unique results, and we said this in lines 30-35. We also said this in section 5.5 (which actually has the title "Weaknesses."). So we agree completely. Also, we think the correlation metric does show this, specifically in the lowest panel of Fig. 22 that clearly shows identical correlations between the records for different offsets of 20920 and 20675. Obviously, the uncertainty in this situation is not characterized by a continuous uncertainty distribution around some most likely value, but instead by a variety of distinct values that are equally possible, and we think Fig. 22 communicates this.

Figure 22: Please do not display the data as zscores. A lot of crucial information is lost that way, and it causes deviations between the ice core datasets that are merely due to resolution affecting the standard deviation of each record.

We think we disagree with this assessment. As we discussed at the beginning of the review, our reconstruction of the amplitude of fallout variability is not reliable and should not be used as a criterion for comparison. Thus, we specifically used a centering and scaling procedure to remove amplitude as an element in the comparisons. As noted above, we can add additional material clarifying this.

Figure 24: Please specify on which timescale the GISP2 d18O record is shown. GICC05 I suppose?

Correct. We can fix this omission.

Figure 25: Consider plotting NGRIP, GRIP and NEEM d18O as well to get an idea about the robustness of those features.

In writing this paper we were specifically trying not to break new ground in correlating varve thickness records with ice core climate records, because we are trying to focus on the Be-10 data. The proposed varve thickness-oxygen isotope correlations are reproduced exactly from the Ridge (2012) reference, and we have intentionally not added any new discussion or evaluation of the previously proposed correlations – our aim is just to benchmark the Be-10 data against previously proposed varve-ice core correlations, without reopening the subject of the validity of the previous correlations. We think this is the best approach, because the present paper is already very long, and although we agree that it would be interesting to bring in the GRIP and NGRIP data, we continue to think it is off topic from the perspective of this paper. Certainly we can add some text to clarify that we are not proposing or endorsing the event correlations in this paper, just presenting them as previous work.

P40, L7: According to the ice core event stratigraphy (Rasmussen et al., 2014) there are no ?events? in the ice cores around this period.

Again, in several areas of this paper we are using "climate events" in a more general sense to indicate any climate variations recorded in the ice cores. We can clarify this section to make it more clear whether we are talking about generic events or specific named ones.

Apologies for the lengthy review ? I hope some of these comments are helpful to im- prove the manuscript and the efforts of dating the NAVC.

Don't apologize. This review is extremely helpful. Frankly, this has been a new area of research for all of us and we are continuing to learn a lot as we put together this paper.