**Reviewer 1**

Summary: This is a generally well written manuscript which makes interesting use of the calendar dating of the Changbaishan Millennium Eruption to attempt to test assumptions behind previous wigglematching of radiocarbon samples for the same event in order to evaluate the potential impacts of "old" carbon contamination. It is an interesting use of these published data to address an important question - the potential impact of volcanic contamination on wigglematching - and also a good way to highlight the differences in approach of using radiocarbon in tree rings to date volcanoes. There is some nice detailed content on the general topic of volcanic old carbon contamination. For all these reasons I'd really like to see this further developed towards publication, but I do have a series of fairly significant concerns that mean that I just can't recommend it for publication at this stage.

***We appreciate the detailed, thoughtful and overall positive review of our MS, and are grateful for the time and effort involved. We respond to their concerns in detail below.

These are my main concerns:

- The published data points are only described in terms of volcanic contamination and this is not sufficient because there are many other interlaboratory or sampling based factors (numbers of rings in each sample, length of time represented, overall age of trees as represented by tree-rings present, changes in the shape of callibration curve relative to these) which could similarly explain apparent outliers. This at the very least should be addressed / discussed and considered within the methodological / decision making processes involved. Without this the whole aim as stated in the title seems fundamentally flawed. How can this study provide "Evidence for old carbon contamination in 14C wiggle-match age series for the 946 CE eruption of Changbaishan volcano" if these other possible explanations are not discussed and ruled out in the manuscript.

***We agree that the absence of discussion of other potential problems with the ages in wiggle match series is a weakness in our MS and will rectify that omission in a revised version. We note in passing that such concerns are rarely dealt with in conventional wiggle match analyses: we agree that they should be addressed in any such study. Our aim was not to explain "apparent outliers" based on the assumption that no contamination affects wiggle match series, but to demonstrate that such series may better fit other sections of a calibration curve if some level of contamination – whatever the source but especially
magentic/geologic carbon – was present. We note that the effects listed also affect conventional age series but are seldom discussed except when invoking a method for identifying and rejecting ages that do not fit the assumed section of the relevant curve. Standard analyses concentrate on forcing a fit (a statistical Procrustes process) to what may not be an appropriate section of the curve.

- Figure 2 "Confidence intervals (95.4%) for wiggle match dates for the Changbaishan Millennium eruption, in relation to the 946 CE date for the eruption (vertical broken line)" is very nice and a logical step in the analysis, but my interpretation of this is that in fact the majority of the previous wigglematches look pretty good with the known calendar date. Yes there are some probability ranges which do have much older possibilities, and some slight off-sets towards older but on the whole this figure gives me confidence in the wigglematch method which can not hope to offer the sort of precision and accuracy that securing 14C on a cosmogenic anchor point can. Generally most studies got somewhere quite close to, or in overlap with, the actual date. The whole 95.4% range needs to be considered.

***We agree with the reviewer’s point, but wish to emphasise that our Fig. 2 is a post hoc representation of the extremely unusual situation of multiple wiggle matches and an independently derived date for an eruption. The main aim of our paper is to show that there can be other, even better, fits of wiggle match age series to a calibration curve which need to be considered if the usual assumption of no contamination from any source is set aside. We agree that standard wiggle matches cannot hope to isolate an exact eruption date that other, independent, dating might provide. We contend that it is possible to get better agreement (within error bars) between wiggle matches series and a known date, when some level of contamination is allowed. In the (usual) absence of a known eruption date, our method of examining the effects of hypothesised non-zero contamination should allow better agreement between wiggle match ages. It provides a systematic methodology for exploration of possible contamination from any source and as a basis for explaining disagreement in ages..

The slightly older age extents for some of the samples could also be explained by interlab variations and differences in models or calibration curves. On page 15, line 12, the authors state: "For example, an earlier wiggle match date of 938 +8/-5 CE, had been favoured from external evidence from dendrochronology (937–938 CE) and varves (912–972 CE) (Nakamura et al., 2007a) until further wiggle match series were measured (Xu et al., 2013)." - Making the point I think that the earlier estimate was somehow wrong, but this fails to recognize that 938 +8/-5 CE is in fact spot on for the later confirmed calendar date of 946 CE.

***We agree that the Xu et al date is “spot on”, we will emphasize this in the new text.But we will also emphasize that this is the last in a series of wiggle match attempts, some of which, although accepted for publication, can now be seen to have been decidedly inaccurate.

For such very slight differences in carbon that appear to being evaluated here there are a multitude of possible explanations, I would have thought, based on what we know of volcanic contamination in other studies, the smoking gun for volcanic contamination would be older dates of c100 or so year or more, not more subtle variations, but here we seem to be focusing on possibilities that are very slight.

***We agree with the reviewer, larger eruptions are more likely to show large systematic offsets (Holdaway et al. 2018). However we are unaware of any such eruptions with an independently known date and multiple wiggle match trees by which to test the hypothesis. The discussion is only possible at Changbaishan because multiple wiggle match series for the eruption and the systematic methodology shows only slight effects
which might be expected in an open forest on a windy mountainside. Our intention was to introduce a methodology grounded on a well-constrained example that can be used by the community to explore the possibility of alternative, perhaps better, fits to a calibration curve. We feel that it is premature to set an arbitrary “smoking gun” value such as 100 years (c. 1.2% old carbon contamination) when such contamination can conceivably affect fits at any level above 0%.

***Contamination from any source can conceivably vary from 0 through “very slight” (0.25 – 1%?) to considerable (>1%) which can occur at different times during the life of the tree leading up to the time of the eruption which is inferred to have killed the tree. Changing levels of contamination could be enough to distort a wiggle match series and create a false match and this is a possibility that must be faced. At present, the standard approach is to seek out and remove ages that compromise the fit at the assumed (no contamination) section of the contamination curve.

(What particularly stands out in the figure to me are the younger outliers which are not explained by old carbon contamination. I like that these are addressed in the text in terms of uptake of younger carbon, but here again but a wide range of other possibilities (like the geological interpretation of the context for the sample) are not really explored (though it seems like the later spatial information might also be very relevant here)... But this is a side issue... something to think about if reframing and refocusing this text for resubmission)

***We agree that the “young” eruption dates are problematic and will expand the discussion on that issue. We note that these are not “outliers” in the conventional sense: each whole age series provides a match with the then current calibration curve that was accepted before later wiggle matches were attempted. The “young dates” resulted in the former name “Millennial Eruption”.

- All other comments aside, I would not feel that this approach could be applied with confidence in other studies especially given that the authors note that it can only be applied where a consistent level of contamination can be assumed over the life of the tree.

***We do not suggest that the method can be applied only where a consistent level of contamination is assumed. We propose the method recognising that by applying a constant contamination is better than applying no contamination at all (a fact which is not recognised in conventional wiggle match dating) as an essential first step in assessing a wiggle match series to see if the assumption of no contamination, consistent or otherwise, is valid. Where there are better matches using the constant contamination assumption then surely the possibility of contamination should be considered. We say that, and will make it clearer in a revised MS that we simplify the analysis by assuming a constant rate. This is analogous to the present situation where no contamination is assumed and only a single fit sought. The number of wiggle match ages is important here, as series of only 4 or 5 ages can conceivably fit several parts of a curve. By introducing the method with this assumption we hope more studies follow that explore non constant contamination models and suspect these will improve dating accuracy further..

Volcanic eruptions and the potential for volcanic gas release and other factors which might impact storage of carbon in sequential tree-ring series overtime are highly variable and very site/eruption specific. Also factors such as growth season of the trees, the period of time covered by the tree-ring sequence involved and the fact that old carbon may be released in some years and not in others all combines to complicate the application of a blanket correction factor for a given study. This sort of blanket correction could equally apply to correct a lab-off set and this would be nothing new.
***We agree, but this is seldom if ever done at the moment, and suggest our methodology would be equally useful for identifying and adjusting for interlab off set.. The comment serves to emphasise what our MS proposes, that wiggle match dating of eruptions (or any other event where systematic contamination/offset might occur) is not simple and investigation of potential methods of refining the analyses are necessary, rather than a statistically Procrustean forced fit by ad hoc exclusion of ages that compromise a no-contamination fit. It is certainly an uncomfortable prospect but one which cannot be ignored.

A few other more specific comments to help with revisions

P1: Lines 1-2: Suggest simplification from: "increasing accuracy as the technologies developed, more refined calibration curves promulgated (Hogg et al., 2020; Reimer et al., 2020), and sophistication of statistical analyses (Crema and Bevan, 2020) to: "increasing accuracy with the use of more refined calibration curves (Hogg et al., 2020; Reimer et al., 2020), and sophisticated statistical analyses (Crema and Bevan, 2020)"

***We will consider the suggested simplification.

P2: Lines 5-6: Suggest change from "as the most likely to give a calendar date" to "as the most likely to give a close to calendar date" (calendar date implies actual calendar date which is only possible with a 14C excursion such as the 774/5 CE event, not through conventional radiocarbon wigglematching, so either make this more subtle change or expand a bit further on making the distinction between the two different ways of using radiocarbon to date samples from eruption contexts, this may be helpful to the reader at this stage as the study rather depends on people appreciating the distinction.)

***We take the reviewers’ point and will expand the text to make the distinction clearer.

P4: Line 27: "cosmogenically-tuned date" - change to "cosmogenically secured" important for readers to understand that this 'radiocarbon date' based on the 774/5 CE event IS a calendar date not subject to variations like wigglematches always will be with new iterations of the curve. Also on line 27 you introduce 3 possible options for why previous radiocarbon dates might be older than the date for the eruption given by Oppenheimer / Hakozaki and then go into a nicely comprehensive list of possible volcanic causes... but you do not introduce the significant range of other possibilities to explain this - for example, were Sun et al's dates based on a sequence of older to younger samples from a single tree? If they were, how old was the tree in total and were the samples blocks of years? Half the dates being older than the date for the eruption given by Oppenheimer / Hakozaki and then go into a nicely comprehensive list of possible volcanic causes... but you do not introduce the significant range of other possibilities to explain this - for example, were Sun et al's dates based on a sequence of older to younger samples from a single tree? If they were, how old was the tree in total and were the samples blocks of years? Half the dates being older in a consecutive sequence from a tree would be predicted, especially if that tree was longer lived. Across how many years is the 14C of a single sample averaged to get the values for the wigglematched blocks and how much does the calibration curve vary (or how much could it vary around the time of the event (i.e. curve plateaus or slopes) - these are all really important to address and provide a thorough framing for the 'contamination' discussion..

***We will explore the implications as to other possible causes (a point raised also by Reviewer 2) and will alter the text accordingly.

P6: Figure 2, line 3, 774 not 74

***We will correct the typo.