Response to the reviewers

In this document we provide detailed responses to the comments made by the referees in the public discussion. Some of our responses coincide with those in the public discussion because they were already very detailed and addressed most of the pertinent concerns/topics. In this document we also detail the changes made to the manuscript according to these comments.

REFEREE 1

General comments

Referee comment #1: *'How do you express a statistical confidence that there is geologically an extra event (or 1 less), because it also affects the confidence of another event next to it?' I guess I am a bit worried that the technique could be using age data to drive geological interpretation. I see that the authors comment on that, but is there a way to show this in a robust fashion?*

Author's response: As we already mentioned in the public discussion, the statistical confidence of the extra event is expressed by the characteristics of the probability density functions (PDFs) involved in the correlation and the rationale behind their shapes. The skewness in the PDFs from DuRoss et al. (2011) is intentional to bias the event's age towards one of the limiting stratigraphic horizons. This condition is founded on geological observations that the authors interpret as indicative of the event occurring closer in time to the age of the limiting layer in question (e.g., the presence of a colluvial wedge or paleosols).

The opposing skewness of the third PDFs in RC vs. GC sites (figure 6 in the manuscript), highlights that the first likely occurred closer in time to the older limiting layer, contrary to the second at the GC site. This is more relevant if we consider that, for both sites, the ages of the limiting units of the third event are similar. This means that, given the same constraints, the geological evidence suggests different timings of the events in relationship to the stratigraphy histories. Furthermore, the higher probability regions of both PDFs are concentrated at significantly different points in time (>500 years apart), which reinforces the two-event interpretation.

PEACH performs an interpretation that is well adjusted to the probability distributions in each site and that, at the same time, are grounded on geological observations and criteria from geologists that conducted the paleoseismic studies. In this sense, we emphasize that the approach does not drive the geological interpretation; rather it is this geological data that drives the dates of the events that are introduced as inputs of the code. With this, we remark that the outputs of PEACH should always be reviewed by the user and not be used as a rigid interpretation, but rather contextualized for the studied region.

Changes in manuscript: In the revised manuscript (lines 423-440) we introduced a better explanation on why the approach and assumptions by DuRoss et al. (2011) can be improved and why we think our extra-event interpretation implies a more detailed analysis of the chronology. The manuscript now details what is the statistical confidence of this extra event based on our method.

Referee comment #2: *'How does this go on to impact ideas/versions of fault rupture segmentation?' e. if you can count x events on a fault system, how does the technique help differentiate whether this will represent multi-section events or smaller section events?*

Author's response: As mentioned in the interactive discussion, the method is designed to facilitate correlation between fault segments, allowing to explore rupture segmentation behavior through the graphical representations generated during the calculations (e.g., figures 5 and 6 in the revised manuscript). Essentially, we can geospatially position each site event PDF based on the site's location along the fault's strike. Then, by analyzing the contribution of each site PDF in the final chronology, we can discern whether an event results from the integration of the records from multiple adjacent sites or not.

In the framework of segmentation, for instance, if two events identified in different sites and segments were to fully overlap in time, PEACH would correlate them as the same event. Such scenario could therefore be interpreted (by the user) as a single multi-segment rupture, given these sites were both located at the tips of the neighboring segments. However, we emphasize that correlation between fault boundaries should be made carefully as there are other conditions at play that our code does not contemplate (e.g., maximum jump distance, stress rupture compatibility, among others).

In the case of the Weber segment of the Wasatch Fault that we discuss, the extra event E4 primarily results from the contribution of the third event at the RC site (as illustrated in figure 6 of the revised manuscript). This suggests that this event likely ruptured only at that specific location in the Weber Segment. The question is whether this event belongs to an across-boundary rupture with the neighboring Brigham Segment to the north or is confined to a smaller sub-segment rupture. To discuss this, we should run the code merging the records of both involved segments, a scope that is beyond our study. But, indeed, DuRoss et al. (2016) already identified that such ruptures across the segment boundary are feasible.

Changes in manuscript: The changes respective to this comment have focused on highlighting the potential of PEACH to evaluate and discuss rupture length and segmentation. In this regard, in the revised manuscript we added a sentence in the introduction (line 78) and section 3.3. (lines 286-288). We enhanced the discussion on segmentation and potential segment-boundary ruptures in the Wasatch Fault (lines 447-457), and we also added two sentences about this topic in the discussion (lines 524-529).

Specific comments

Referee comment #3: *"In reality, how does this deal with multi-peaked calibrated C-14 calibrations?"*

Author's response: The choice of modelling numerical dates as normal distributions was driven by the will to accommodate not only radiocarbon dates, but also luminescence (e.g., OSL), which usually are expressed as normal distributions. In fact, luminescence dates are more prevalent in sites dominated by older, coarser, or less complete deposits – inherently more challenging to date accurately using other, more precise techniques like radiocarbon. It is in these contexts where site-to-site correlation poses a substantial challenge, making our approach particularly valuable.

However, we acknowledge and concur with the referee that shaping radiocarbon dates as normal distributions may be an over-simplification in some cases, especially in multi-peak distributions. We acknowledged this limitation in the discussion (lines 452-457 of the original manuscript) and in the current version of our code we already introduced the option to compute correlation by directly utilizing OxCal chronologies as inputs. OxCal is a specialized software designed explicitly for radiocarbon calibration and is better suited to handle the complexities associated with radiocarbon dating, a scope that goes beyond the objectives of our approach. This addition was made to allow users to choose the most appropriate option of PEACH based on their specific dataset and requirements. Specifically, we recommend to use the OxCal implementation in PEACH in radiocarbon-based datasets.

Changes in manuscript: In the revised manuscript we enhanced the reasoning on why we decided to model numerical dates as normal distributions, its potential implications on the modelling, and recommendation for the users to use the OxCal option of PEACH. We improved this reasoning in section 3.2 (lines 156-160) and in the discussion (lines 501-510).

Referee comment #4: *"Will it always be 4000, why?"*

Author's response: This comment refers to the number of samples (seeds) performed by the algorithm within the numerical date PDF to compute the event date PDF, as explained in section 3.2 of the original manuscript.

The choice of a 4000-sample threshold is based on a sensitivity test conducted across several datasets, including the two shown in Figure 3 of the manuscript, as well as others tested during the development of the code. This value is conservative and is expected to yield good results in most datasets. However, because each dataset is different ideally a sensitivity analysis should be done in each case.

Changes in manuscript: In the revised manuscript (lines 164-175) we enhanced the justification on the use of the 4000 seed value, and we remarked the importance of performing a sensitivity analysis for each dataset. We also added the code to perform such analysis at the Zenodo repository of the publication and updated the User Manual accordingly.

Referee comment #5: *"Can there be an over-interpretation of events?"*

Author's response: The comment refers to our statement in the original manuscript (line 438) "events evident in trenches are always a minimum relative to the actual number of events that occurred".

While we acknowledge that there can be over-interpretation of events in specific trenches, because of the localized nature of paleoseismological studies, overall, the tendency is towards under-detection. This can be due to a variety of factors, including erosion or stratigraphy gaps that fail to record the deformation of events, coarse lithologies-granulometries in sediments that mask deformation, location of the site within the tips of the paleo-surface rupture, among others (e.g., Weldon & Biasi, 2013). This phenomenon is even more apparent in sites with large date uncertainties, as the greater is the date uncertainty of an event, the higher the likelihood that multiple events occurred within that time span. In the cases we illustrate in the paper, especially in the Paganica fault, some sites exhibit considerable event uncertainties even up to a thousand of years (e.g., site TRET in figure 5 of the manuscript). This inherently raises the probability that multiple events might have taken place during that time span.

Weldon & Biasi (2013) interestingly point that the impact of over-interpretation of events in trenches is reduced by increasing the number of trench exposures and correlation among them. That is because true events are likely to be detected at more than one site, as opposed to "fake" or over-interpreted events. In our approach, due to the probabilistic modelling and product-based correlations, genuine events tend to stand out more prominently in the mean distribution and have a higher chance of being detected by the algorithm. Conversely, "fake" events are likely to be downplayed, thanks to the probabilistic framework.

Having said that, we reinforce that the scope of this work is not to engage in a re-interpretation of the paleoseismic findings presented in the studies we use to showcase our approach. Instead, we seek to refine the temporal resolution of such chronologies, while maintaining the integrity of the original interpretations.

Changes in manuscript: We added the discussion of the response in the discussion of the revised manuscript (lines 474-480) and we added the Weldon & Biasi (2013) reference in the reference list.

Referee title change suggestion: *"kind of both say the same thing, what about getting the word statisical' in there?"*

Author's comments: the reviewer is referring to the redundancy on the use of "past earthquakes" and "paleoseismic" in the title, and suggests introducing the word statistical to describe better our work. We agree with the reviewer that the title might sound redundant, although technically past earthquakes and paleoseismic records are not equivalent in definition. In fact, the term "past earthquakes" includes paleoseismic events as well as other events that happened in the past (e.g., historical earthquakes). In the title we wanted to remark that our investigations on past earthquakes come from the analysis of the paleoseismic datasets only. Concerning the second suggestion, we find pertinent to adapt the title in a way that better describes the type of analysis that our approach does. In these terms, the more accurate word is "probabilistic" rather than "statistical".

Changes in manuscript: We changed the title of the article to "Deciphering past earthquakes from the probabilistic modelling of paleoseismic records – The Paleoseismic Earthquake Chronologies code (PEACH, version 1)"

Referee in text grammatical errors / suggestions: We changed all the grammatical mistakes and suggestions raised by the reviewer in the manuscript PDF. These minor changes can be seen in the manuscript with tracked changes.

REFEREE 2

Referee comment #1: "*To add on the comments from the reviewer, I would only advise the authors to better detail and explain what I would refer to "rule of thumbs" decisions on parameters and related assumptions that come along with them (especially those that are hardcoded in their matlab script, as for example the minimum/optimum sample seeding) that any future user of the software tool would/could encounter (and possible fin tune) in his/her study. This would add to the scientific merit of the study, in addition to the examples described, and would also help potential users to better understand the underline granularity of their decisions/assumptions made and their impact on the model outcomes.*"

Author's response: We agree with the comment about the fact that detailing the specifics of the method will enhance its transparency and overall user experience. In the revised version of the manuscript, we find pertinent to address the following points:

a) Regarding the seeding process, and in the line of our response to comment #4 of the Referee 1, we expanded the reasoning behind the determination of the seed parameter (section 3.1), with particular interest on the potential implications of the variations in such value for the model outcomes/quality. In general, lower seeds can lead to poor sampling of the numerical date PDFs and therefore unstable modelling of the event PDFs. This instability can impact the repeatability of a same run and as such we recommend performing a sensitivity analysis for each dataset.
b) Another important assumption (addressed also by Referee 1) is the modelling of radiocarbon dates as normal distributions. The motivation to use this probability distribution as well as recommendations (rule of thumbs) for the user are explained in our response to referee comment #3.
c) Another important specification of the method that we find important to address is the peak detection in PEACH. The peak detection threshold (minprom) is one of the most decisive parameters of our approach and can carry some limitations in the performance and event count of the final chronology if not working properly. In case of poor performance, the final correct detection relies on a series of user checks with the dataset that can help to adjust this parameter. For example, if the number of peaks detected is less than the number of events in an individual site, the peak detection should be reviewed and adapted (e.g., by working with sigma-truncated PDFs). We detail this in the revised manuscript as we describe below.

Changes in manuscript:

- Changes concerning point a): These coincide with the changes explained at the referee comment #4.
- Changes concerning point b): These coincide with the changes explained at the referee comment #3.
- Changes concerning point c): This point was only mentioned in the discussion in the original manuscript (lines 458-467). In the revised manuscript we better highlighted the importance of the minprom parameter in the method (lines 209-212) and we detailed recommendations to the user to check and, if necessary, modify this parameter to ensure optimal performance (lines 239-255).

OTHER IMPROVEMENTS

Aside from the reviewers' suggestions, in the revised manuscript we also did a few other changes. The most important one is the correction of Equation 1 following the error that we detected during the open discussion and that we addressed in our author comment 1 (AC1) at the public discussion. In the revised manuscript this implied the modification of Equation 1 (line 205) and the paragraph explaining its formulation (lines 199-204).

Other changes we made convey formal modifications in some paragraphs to improve readability or clarity in the speech. The main one is moving the general statement of the advantages of our approach vs. DuRoss et al. (2011), which was in lines 408-412 in the original manuscript, to lines 489-495 in the revised version. We also changed the title of section 4.2.1 to “Results and comparison with previous modelling” because a significant part of that section is comparing our results with the results from DuRoss et al. (2011) rather than just result description. Other minor changes can be checked at the revised manuscript with tracked changes.

In terms of figures, we only included the Pearsons Canyon site and Birgham segment names in the inset of figure 6 to allow location of what is discussed in the text.

In terms of assets, we updated the Zenodo respository with a new version that includes the code for the sensitivity analysis. The new doi is: https://doi.org/10.5281/zenodo.8434566 We updated this in the references of the manuscript.

Best regards,

Octavi Gómez-Novell, on behalf of all authors.