Dear editor,

We have revised our manuscript according to the referees' comments and your suggestions. In particular, we provided additional data and corrected all typos. We believe and hope that, whenever we chose not to modify our manuscript, the provided rebuttals are satisfactorily justified.

Regarding Figure 7, on which you commented: it is standard practice in radiocarbon dating to publish such figures. Actually, the original article in which the data were published contains three of these figures (Figs. 2, 8 and 10 in Bayliss et al., "Getting to the bottom of it all: a Bayesian approach to dating the start of Çatalhöyük." Journal of World Prehistory 28.1 (2015): 1-26). It would be very awkward in our view to change the way they presented things in the original work and we are convinced that readers would be confused if we changed the standard way of presenting OxCal radiocarbon-based chronological models.

On behalf of all authors,

Guillaume Guérin

Reviewer 1

Dear Authors,

thank you for the thorough revisions, I agree with most of them, but not all.

Your manuscript explains several aspects of age-depth modeling/sorting in a very useful manner, supplemented by examples showcasing effects of sample density and -properties.

Figures seem inconstant to some extent – units are not systematically added, and ages are given in different units. Also, age axes are on the left, right, up, down. Please homogenise these.

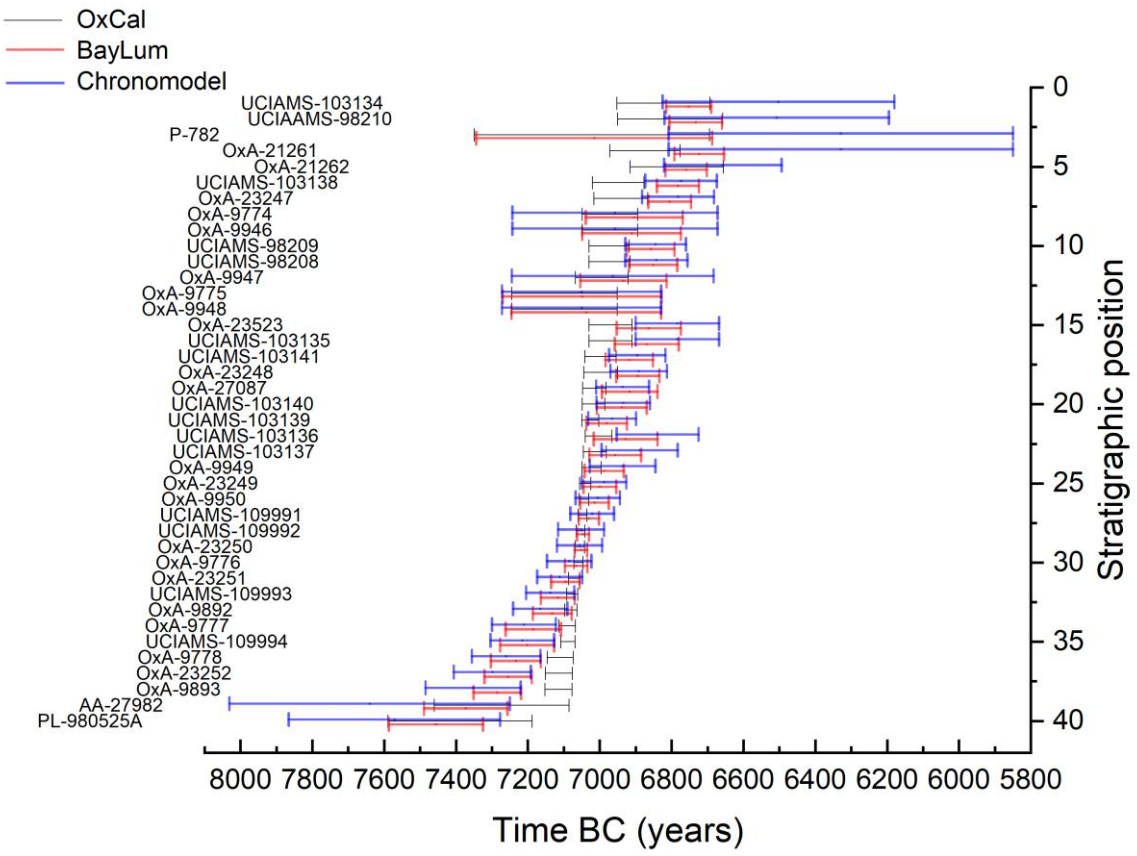
This issue appears to be really problematic in our view, because conventions are not the same in the OSL and radiocarbon communities: in radiocarbon, the x-axis goes from past (left) to present (right), whereas in OSL the x-axis is age, so increasing age (going right on the x-axis) goes back in the past. Thus, by convention the Bayesian tools developed in either community use different standards. The same applies to time units: radiocarbon uses calendar dates expressed in years BC, whereas in OSL the standard physical quantity is age, expressed in ka. We have added this paragraph in the introduction, but we see no justification for choosing one or the other convention and system of units.

Here is the added paragraph (l. 151-158): 'Before presenting the method, we should warn the reader that radiocarbon and OSL communities use different conventions, both in terms of time axis orientation and system of units: in radiocarbon, the x-axis goes from past (left) to present (right), whereas in OSL the x-axis is age, so increasing age (i.e., going right on the x-axis) goes back in the past. Thus, by convention the Bayesian tools developed in either community use different standards. The same applies to time units: radiocarbon uses calendar dates expressed in years BC, whereas in OSL the standard physical quantity is age, expressed in ka. As a consequence, we used the OSL conventions for the study of OSL ages – and the radiocarbon convention for the C-14 study. The only exception to this rule is when only BayLum (developed in the OSL community) is used to model a sequence dated with radiocarbon only (Fig. 9); for this illustrative figure, the OSL convention was adopted for a radiocarbon-based study.'

In addition, units are given on every figure – we double checked.

In my last review I asked you to add mean data values to some figures. Having a standard software output as manuscript figure is not ideal, I trust software puts out data which can be plotted together with data. Please

The referee is referring to their comment on Fig. 10 here. Here is Fig. 10:



We are really sorry, but on this already quite dense figure we do not see which data could be shown: uncalibrated ages? 95% C.I.? If the latter, obtained with which model – and why? All the data and the codes used to generate this figure are provided in Supplementary material are provided, we honestly do not understand this query.

Further: Line 31/following formula are inconsistent. The formula suggests older or equal age, where the text states a clear sequence.

In these probability density-based models, the probability of ages being equal is zero – so the text and the formula are in fact consistent.

102ff: I suggest to refer to Steier and Rom here- they explained this effect best in literature to my knowledge

Agreed.

In my last review I suggested to place the data behind figures in Supplements – please consider this, so that your figures and arguments can be better followed.

Agreed, we have added the OSL dataset in Supplement (the radiocarbon data and all codes were already provided).

Review of Revised

The conflict between sampling resolution and stratigraphic constraints from a Bayesian perspective: OSL and radiocarbon case studies

20th Jan 2026

Overview

I think this version is much improved compared with the previous version. I understand much more clearly the points that the authors are trying to make:

1) That when you have a set of samples that are very similar in date then, if fitting joint Bayesian models with the specific priors described in Eq (1), the posterior will seemingly tend to spread out these dates (which may be unwanted). This is what they call the spread effect.
2) That through modelling of phases (in OxCal in particular) one can concentrate dates (i.e. shrink them into narrow intervals) – the concentration effect

In my view this is likely an innate property of the prior in Eq 1 (or priors in the case of the phase model). No prior is actually uninformative in any Bayesian model, and perhaps the Eq 1 prior actually prefers some spacing between dates? This is not necessarily a failing but a property of the prior that perhaps people do not realise. As a simple example, in the case where we only have two dates in Eq (1) then you can directly work out the equivalent prior on the difference in ages, $d = t_2 - t_1$, and it is not uninformative (although interestingly I might have expected it to shrink the dates towards one another).

Key Message

I certainly agree that the fundamental message for readers should be that all users of Bayesian models must check their assumptions and the suitability of the specific model and prior that they are fitting (and not just treat them as a black box).

I do also agree that in the Jianbing examples provided it seems (visually) to be the case that the model posterior has overly spread the dates at depths 200-300m. However, I feel that readers should maintain some caution when reading the broad interpretations provided which, IMO, are quite strong in places regarding where the model posteriors are wrong. The authors have a tendency to perhaps present their interpretation of the posteriors (and their failings) as definitive whereas, as we do not actually know the right answer, some of these inferences are perhaps a little more subjective.

This OSL example (and to some extent the Catalhoyuk example too) are interpreted/criticised based on the authors implicitly assuming that they know what the right answer is, e.g., for the OSL ages that effectively they seem to think there is a large and sudden accumulation of sediment from the same calendar age. This is entirely fine/legitimate, but it might be considered simply to be a different subjective prior (quite possibly a better prior if people think there are such fast accumulation processes). Importantly, readers should note that they do not show any examples where the models actually might work to improve inference.

We thank the referee for this feedback; and we are happy to see that they summarized perfectly the key messages we wanted to convey.

I noticed a few minor typos that need revising (and I have suggested a few further minor comments that IMO would be more equitable in terms of accompanying interpretation). Once the typos are fixed then I would accept the paper (I will leave decisions on the more subjective commentary suggestions to the editor/authors):

Typos/Lack of Explanation

1) Line 47 – 54 (Baylum model) – No explanation given of what e is. I would also argue this is not the same prior as Eq (1) due to the introduction of log, assuming t are the actual dates of interest (rather than e)?

$e_i = \log(t_i)$ and this explicitly defines e_i (l. 48). Indeed, due to the OSL equation form, the Jeffreys (non-informative) prior for the age is given by $(e_1, \cdots e_n) \sim \mathcal{U}^*(\log(P), \log(A))$; this was demonstrated by Combès and Philippe (2017; appendix A).

2) There is some odd bullet pointing tat extends into the margin throughout my version (e.g., line 50, 95, …)

In our version as well, we hope this is just a reflection of the template we used.

3) Line 213 – Lost capitalisation in OxCal (not oxCal)

Agreed.

4) Figure 2 caption refers to section 3.1 below (when discussing outliers). The only mention of outliers I could find was on line 214 (i.e. Section 2.2 and not 3.1). Does this Fig reference need updating?

Agreed, thank you for spotting this mistake.

5) Line 507 – a spurious line break has been introduced in the middle of a sentence

We double checked this issue.

Comments (not required for acceptance but I would personally reword):

1) I would reword (tone down) the text c.a. line 427 about the start of the mound not being credibly before the date of the earliest sample and that one should not extrapolate. To me this is a highly subjective statement, and implicitly assumes substantial extra information about the sampling, i.e., that this mound is so heavily sampled that the oldest object must in fact be from the very date that the mound was created. I think it is fine to say something similar to what is there currently, but without the definitiveness that suggests extrapolating is wrong.

   We have checked our wording carefully: we state that there is no data suggesting a start of the mound as early as OxCal suggests. Therefore, the OxCal inference is an extrapolation of the data. This statement does not mean that OxCal is wrong, but that it has to be somehow speculative (or, to put it differently, a reflection of the model priors, which – by definition – are independent of the data). Our feeling is that, too often in the literature, OxCal is taken to deliver 'the truth'.

   In many instances, I think it is quite plausible (indeed would be correct inference) to have an estimate for the date of the start of a phase (e.g., mound creation) that extends before the first/oldest sample on the basis that you probably haven't dated a sample from the mound creation. This also goes in the other direction, where an extinction/mound end can be considerably after the most recent bone/sample has been found, i.e., the Signor-Lipps effect: https://en.wikipedia.org/wiki/Signor%E2%80%93Lipps_effect

   Agreed – this is our case: to answer a specific question, adequate sampling to address this issue must be undertaken. We find it dangerous to rely on modelling to correct inadequate sampling.

   In fact I actually think this is something that a (good) model might actually be able to inform you on – if you have very dense dates for samples in a mound (i.e., all close to one another)

one might (??) think that they do mark the full use of the mound and so the earliest and latest dates should be v near the mound creation/end. However, if you only have a few samples (that are widely spread in calendar age) you might worry that the start of the mound could have been considerably earlier and you just haven't sampled much.

Agreed – but we see no need to modify our text.