We would like to thank Maarten Blaauw (Reviewer 2) for his constructive suggestions and comments, which have helped to improve our manuscript entitled "SCUBIDO: a Bayesian modelling approach to reconstruct palaeoclimate from multivariate lake sediment data." Below, the reviewer's comments are shown in red, and our responses in black.

Inferring fossil climate from calibrating XRF to recent temperature by necessity assumes that the relationship has been stationary over time, and that any other factors which didn't occur as much further back in time, e.g. soil erosion or nutrient pollution, do not affect the temperature reconstructions (or even more importantly, the calibration period). However, humans have messed up many records over the past centuries/millennium, possibly affecting the varved record as well (note that especially Diss Mere is located in a severely human-influenced region). How can these factors be disentangled? This is mentioned in the discussion, but it would be useful to expand more on this problem in the Introduction.

We completely agree with this reviewer that humans have changed the landscape over the last several thousand years. We believe that this is a problem faced by all climate reconstructions whether located in an urbanised area, such as Diss Mere, or more remote, such as Nautajärvi as the calibration period lies in a period where humans have already altered climate behaviour. Unfortunately, our model is not a physical model which can specifically disentangle the role of human activity, though potentially at the price of uncertainty quantification. This is why we emphasise that the user should carry out a preliminary investigation to explore the role of human activity to ensure that there are not significant changes to the geochemical signal. Whilst this information cannot be explicitly fed into the model, it can help with understanding whether the data should or should not be used in this approach. For the example of Diss Mere, Boyall et al (2024) found no significant changes in terms of what the different elements were responding to between 8,000 to present, and rather human activity just influenced the variability. So, whilst human activity likely had a role in the variability shift at ca. 2,000 we believe that the climate signal remains. We think this is a good idea to mention though as the reviewer suggested and thus we will add additional emphasis to the section in the conclusion to make users cautious of the results if there are significant changes to their μXRF-CS records, and will add more information about how human activity can lead to challenges in palaeoclimate reconstructions.

Why did you decide to only use non-biological proxy data? Perhaps this has to do with additional complexities of interacting/competing ecosystem components, but it would be good to spell this out (e.g., in section 2.1).

There are two main reasons: 1) the temporal resolution of the reconstruction and 2) potential seasonal and ecological biases.

The purpose of this approach is to reconstruct climate at an annual to decadal resolution. Because μXRF-CS data has a sample resolution <1 mm, and often as low as 0.2 mm, it means that, for many sediment records, we are able to get several data points each year. For example, in Diss Mere there are about 25 data points on average during the non-varved sections (sedimentation rate = 0.5 cm/yr) and 5 data points within the varved sections (0.4 mm/yr). Such a high sampling resolution for proxies that require to take a sample of sediments (e.g. pollen, diatoms, chironomids) is challenging for different reasons. First, analysis requires a minimum amount of sediments (e.g. 0.5-1 g), which tends to be samples of 0.5 cm^3. In most of the lake records, the sedimentation rate is at mm scale, so annual resolution is not possible. Second, having a continuous record of the Holocene at this resolution is time consuming and expensive.

Nonetheless, the model presented in this study is based on the initial work of Parnell et al. (2015) who built a similar model using pollen. The framework is similar and therefore can be adapted easily to a multivariate biological proxy dataset if a user wishes. For example, new advances in hyperspectral imaging analysis would allow applying this method to a multivariate dataset of biological-related proxies (e.g. pigments, bacterial communities). For this study we choose XRF-CS because it is a state-of-the-art methodology in palaeolimnology and palaeoclimatology, and is relatively fast and cheap, so many other researchers might benefit from this model.

We will add in some more details into the manuscript as we explain the benefits of μXRF-CS and why we chose this data other alternative biological approaches for this study.

Null counts (inferred absences, x_i=0) are frequent in most types of proxy datasets, e.g. pollen diagrams. Would this be a problem for your approach if used in other studies, given the clr transformation? Could the approach be amended to account for null counts (even if just by adding a small constant to all values)?

We agree that for many poxy datasets null counts can be a problem. However, as we specify in the manuscript that the data should be clr-transformed prior to using the SCUBIDO model, we would assume that the user's data does not contain many null values as this is a pre-requisite for clr-transformation (Bertrand et al., 2024). However, we thank the reviewer for mentioning this as we will now add some additional information about the pre-requisites of clr-transformation and that this should be addressed before the modelling commences.

Line 161-2, it is unfortunate that age uncertainty is not considered in this modelling approach. This is problematic, even for this varved lake (note that the varves of Diss Mere don't reliably extend to present-day, causing even more chronological uncertainties). Could age uncertainty not be included as a module of the Bayesian analysis? Lots of work has gone into developing Bayesian age-models, so you could build on existing methods (e.g., https://gchron.copernicus.org/articles/4/409/2022/).

We agree with the reviewer that age uncertainty is important to consider. We do envision that future versions of SCUBIDO will incorporate and model age uncertainty. However, for this current version we were unable to find a way to build age modelling and uncertainty quantification at each timestep into the current model without it being very computationally expensive. We really appreciate the reference suggestion, and we will consider using this or other similar models in a future update of SCUBIDO.

Following this comment and a comment from Reviewer 1, we will now add some more information about the chronology into the Supplementary Information and also remind users throughout that this modelling approach does not include age modelling or account for age uncertainty.

Lines 163-5, See also Blaauw et al. 2010 who produced random-walk 'fake' proxy datasets, some of them steered by (again fake, random-walk) environmental forcers [https://doi.org/10.1177/095968360935518]

We appreciate the reviewer's suggestion; it was a very interesting read. However, we are unsure whether the comment refers specifically to the lines stated here (163–165), or to the subsequent lines (166–167) in the previously submitted manuscript, where we describe the random walk component of the model. As this section of the manuscript is focused on outlining the structure of our inference model, we have opted not to include additional references here. If the reviewer feels that the reference to Blaauw et al. (2010) could be more appropriately integrated elsewhere in the manuscript, we would be happy to do so.

Why did you use a polynomial regression model (equation 4)? Would other models such as a smooth spline have worked?

In our modelling approach we were interested in finding a balance as we did not want to have a model which was overly complex and computationally demanding, especially since there isn't a hugely complicated relationship between the elements and climate. At the same time, we wanted a model that was simple and fast for users to run, but still capable of capturing the main features of the relationship. Given the relatively weak individual correlations between the elements and climate, we settled on a quadratic relationship as it provided a good compromise between simplicity and performance.

The reviewer is correct that more complicated models could be fitted and a similar example of this modelling approach which includes a more complicated p-spline model was included in Cahill et al., 2023 in *Environmetrics*. We did try a P-spline model as part of our model exploration process, but this tended to overfit the XRF data and drastically reduced the speed of our modelling approach.

Vague normal/uniform distributions were used as priors. Did you explore how different/stronger priors would affect the results?

In this modelling approach we have used different types of priors ranging from vague to more informed. The choice behind using a vague prior to, for example, capture the relationship between the elements and climate, was that we did not want to assume that we knew the relationship between the different variables. This was especially important for us as we wanted the model to be easily adaptable to different lakes with significantly different XRF datasets and therefore we did not want to assume the same priors could be applied to all records.

On the other hand, we have also included a more informed prior on the random walk variable for the fossil model as we take this from the modern calibration. We first use a vague prior on the calibration model, but we capitalise on the fact that it is a Bayesian model by borrowing the strength of the modern data where we do know the climate dynamics over time and have applied that into the fossil record. We thank the reviewer for this comment as it makes us aware that we need to be clearer about the choice of priors in the new version of the manuscript.

Figs. 1/2: I agree with the other reviewer that it is very hard to visually spot much of a correlation between temperature (black dots) and the variables (or between true and reconstructed temperature in Fig. 2 - my eyes would tell me that there's essentially no correlation). This is especially because temperature shows very little variability, ranging only from -2 to 2 degC, with the far majority centered around 0 (so, reconstructed temperature anomalies >1 degC would be based on very few calibration data). But perhaps I am misunderstanding these graphs. Could the individual leave-some-out sets be shown with different colours in Fig. 2?

We agree with all the reviewers that from Figure 1 you cannot see a clear direct relationship between the individual elements and temperature. Please see the response to Reviewer 1 on the reasons why we do not expect to see high correlations.

For Figure 2 after all reviewers identified this, we reassessed the validation code and realised there was a slight bug which was misaligning the reconstructed temperature to the true temperature by one year. This was caused by rounding up to the nearest integer. We have since re-plotted the results and have obtained a much clearer set of validation results.

Can you show the structure of the MCMC run's 'energy'? How many parameters were involved? Line 384, how do you mean 4 chains were used - were the results joined afterward? Was any thinning necessary?

We thank the reviewer for this comment though we are a little unclear as to what is being asked for here. We have used standard MCMC diagnostics to assess convergence of the chains (e.g. the Brooks Gelman Rubin R-hat diagnostic). We also evaluated trace plots of the MCMC chains for many of the parameters to confirm convergence, but for brevity we did not include these in the paper. When we say that four chains are used, we mean that four different sets of starting values were used to run the MCMC in parallel. This is the basis for the R hat test mentioned above. To be clearer for readers we will add these references more times throughout the text.

Figure 4 - the temperature reconstruction of Nautajarvi seems close to what is generally assumed for Holocene temperature time-series, with a HTM followed by a late-Holocene cooling. However this is not visible for Diss Mere. This is reported in the Discussion. A devil's advocate might say that the authors were lucky with Nautajarvi's reconstruction as it follows known Holocene patterns, but weren't as lucky with Diss Mere. Did you try multiple runs with different settings/priors, and were the results robust?

We thank the reviewer for this comment. We did multiple modelling runs, with different starting points and settings, and only observed slight changes, mostly in the variability.

We agree with the reviewer that the reconstruction from Nautajärvi has a HTM whereas Diss Mere is warming through the Holocene with no distinct HTM. However, this is still a big topic in the palaeoclimate field leading to the Holocene temperature conundrum. Discrepancies between proxy-based reconstructions and climate model simulations suggest that proxy-based reconstructions might be biased toward the summer season as most of these reconstructions are based on biological proxies (Liu et al., 2014). There is also a complex spatial-temporal distribution of the HTM that might explain the regional differences (Cartapanis et al., 2022), e.g. a higher sensitivity to changes in seasonality at high latitudes, different sensitivities in continental vs maritime climate regions. On the other hand, as shown in Figure 5, there are other recent reconstructions that combine proxy data and data assimilation approaches that show no distinct HTM, which agree with the general evolution of temperature recorded in Diss Mere. Nonetheless, a discussion of the HTM is out of the scope of this manuscript, which rather describes a new method to reconstruct temperature.

The manuscript will need a thorough grammar/punctuation check because at times it is difficult to follow. I am suggesting a few specific changes here:

We thank the reviewer for identifying these and adding in the different lines numbers to locate them. We will make all the grammatical errors mentioned.

It is great that the authors have produced an R package for the approach presented here. Unfortunately, I was unable to actually install and run the R package. JAGS has to be installed in order to run. Of the 23 additional packages that had to be installed on my systems (I tried this on both linux and mac), it was 'rjags' that caused issues. This seems to be caused by the version of JAGS (4.3.2) not coming with 'modules-4', or owing to linking problems. Would it be possible to swap to another MCMC sampler? For example, twalk (Christen and Fox 2010 Bayesian Analysis 5: 263-282) is available as an R package. Even though this twalk package uses pure R and will thus be slower than e.g. c++ versions, at least I expect it would be easier for users to install and run everything without too many issues.

We thank the reviewer for bringing this to our attention. We have now highlighted that JAGS is needed to be installed for SCUBIDO to be ran on both its vignette and GitHub pages and we provide a link to download the software.
We agree that perhaps another MCMC sampler would be easier for some users. However, the computational time for this model is very long given the number of data points commonly present in μXRF-CS datasets. Therefore, we have found that using an alternative sampler would not be beneficial. However, we will be monitoring the SCUBIDO package and in the future will be making it available on other platforms, e.g. Python and thus may make it more accessible for some users. We thank the reviewer again for their suggestion but at present we prefer to remain with the JAGS package in R.