Anonymous Referee #1

Guitian et al., 2024 fills a gap in our understanding of Cenozoic CO2 concentrations, and highlights that the long-term decline in Ɛp in the mid-Cenozoic previously described is global in nature. The study aims to identify global shifts in Ɛp, that are likely to be related to CO2 concentration, by using sites from several contrasting locations to produce a combined Ɛp record. However, this goal is constrained by limitations in the sites' respective age models, which make detailed comparison of the timings of Ɛp changes challenging. The study is well written and illustrates its point well, but would benefit from further integrating its data with existing published records, and with a greater acknowledgement of the limitations of modern calibrations to geologic data. The study does not reconstruct CO2 directly, but reconstructs Ɛp and applies transformations derived from culture studies to account for temperature effects, generating a semi-quantitative CO2 reconstruction. This approach is valid, given the current literature, but a more robust comparison of the semi-quantitative reconstructions with existing δ11B CO2 reconstructions would be of great relevance to its conclusions and place its results better in context. As-is, it is difficult to say on reading this paper whether it agrees or disagrees with existing CO2 reconstructions produced using a technique generally considered robust. The proposed large drop in CO2 through the study interval has profound climate implications if true, as described in the text, but seems at odds with δ11B data, to the best of my knowledge. I would like to see greater integration with existing data from alternative archives and, as the claims of a four-fold CO2 drop and large-scale decoupling of CO2 from temperature over the interval are, as the text admits, paradoxical. It seems as likely to me that the modern culture studies from which the temperature deconvolution is derived are not directly applicable to Oligocene/Miocene alkenone producers, that some other change in algal biology occurred, or that nutrient dynamics shifted more profoundly than biogenic silica records suggest.

We sincerely appreciate the suggestions from the Anonymous Referee #1. As detailed below we propose to incorporate some additional discussion of the existing published records which were plotted in Figure 8. Our manuscript is indeed seeking to call attention to the paradox and we have extended the final paragraph to further emphasize the additional steps that will be needed across the full set of CO_2 and climate proxies to reconcile these apparent differences.

Specific points:

• Line 48: there are a lot of boron estimates for the younger half of the interval – I don't think it's accurate to say that most of the existing estimates are marine phytoplankton-derived.

We propose to revise the paragraph of the introduction to clarify:

The long term pCO_2 trends from the Oligocene to early Miocene are derived from the sensitivity of marine algae to pCO_2 , while published $\delta^{11}B$ based CO_2 estimates cover the latest Oligocene into early Miocene (younger than 24 Ma) (Rae et al., 2021).

• Line 201 - I think you need to discuss the difference in results from using benthic and bulk δ13C for your carbonate measurements. Since your bulk and benthic-derived Ɛps are reconstructed with 2 different values that are ~2‰ apart in Figure 2, have these both been converted to δ13CDIC? That's a large offset. The fact that the monte carlo simulations of the bulk- and benthic- derived Ɛp don't overlap in figure 2 or 4 suggests the uncertainty is much higher than it's been calculated as.

In the methods section (line 205) we described that, similar to previous studies, we adjusted the benthic *$\delta^{13}C$* by adding 2 permil to correct for the surface-deep *$\delta^{13}C$* gradient and estimate the surface ocean *$\delta^{13}C$*.

We propose to rearrange this paragraph to clarify.

We calculate the $\delta^{13}C$ DIC from the $\delta^{13}C$ of the bulk carbonate, which is dominated by Reticulofenestra coccoliths (Guitián et al., 2020). Because there is no divergence of vital effects between small and large coccoliths in the late Oligocene to early Miocene (Bolton and Stoll, 2013), we propose that the offset between coccolith $\delta^{13}C$ and DIC is likely to remain constant. We subtract 0.5 ‰ from the $\delta^{13}C$ bulk to calculate $\delta^{13}C$ DIC, based on average alkenone-producing coccoliths cultured at DIC <4 mM compiled in Stoll et al. (2019). Support for estimating photosynthetic fractionation from coccolith $\delta^{13}C$ is provided by recent culture studies of G. oceanica (Torres Romero et al., 2024). In previous studies, the $\delta^{13}C$ DIC has also been estimated from the $\delta^{13}C$ of calcium carbonate of benthic foraminifera with the assumption of a constant and known offset between the $\delta^{13}C$ DIC of the deep and surface ocean. Site 1406 and 925 features sufficient well preserved benthic foraminifera, mainly epifaunal Cibicidoides spp. larger than 200 µm. For an additional sensitivity test to evaluate the significance of the method of DIC estimation and facilitate comparison to other published Ep records calculated from benthic $\delta^{13}C$, we also estimate surface ocean DIC by adding a constant offset of +2 ‰ to the $\delta^{13}C$benthic measurements, following previous Miocene and Oligocene studies (Guitián et al., 2019; Pagani et al., 2011; Zhang et al., 2013).

• A single rapid 3‰ Ɛp drop is stated to occur at multiple time intervals in the text – 26.5-25.4 on line 246 (the only drop in that interval is site 1406), 26-24.5 on line 281 (occurs at 925 and possibly at 1168, but not 1406), and then at 27-24.5 in the conclusion. With the age model uncertainty and low resolution in several records, it seems difficult to say whether these were a synchronous event or if the timing differed between sites. At site 516, it looks like Ɛp increases rapidly in (or close to, depending on how it's defined) the same interval, and decreases later. This makes it seem less likely to me that it was a single event – the text states that it's still possible within age model uncertainty, but that's a lot of age models that don't match up well. Are there any potential explanations for a staggered Ɛp drop over several millions of years? Either way, I don't think it's entirely accurate to say it's "resolved" (as in the conclusion line 467).

We propose to revise the text to emphasize the uncertainties imposed by: sedimentation gaps in some sites making it impossible to test the reproducibility of some events in the new records presented here, and the uncertainty in age models of sites which have not been synchronized to a Sr isotope stratigraphy. We also propose to revise the text to refer to the broad interval over which the most abrupt change is observed.

Specifically, we propose to edit lines 281, 467, and 246 to describe the core time interval of this transition with some uncertainties, as 27 to 24.5, which is similar in sites 925 and 1406. The new version of the text will highlight that the lack of Ep measurements prior to 25 Ma in 1168 makes it difficult to evaluate if the transition also occurs in this site.

In 516, we had noted in line 284 " *The late Oligocene at DSDP 516 features a 5‰ peak in Ep between 24.5 and 24.9 Ma, which is not reflected at 1406, or 1168 sites."*

Regarding this record we propose to clarify that:

With current information, we cannot assess if this difference reflects age model uncertainty, potential analytical uncertainty from GC-IRMS chromatography, or aliasing.

• Figure 4 – what happened to site 925? It's on figure 2 but not here.

Following the reviewer suggestion we propose to add 925 to this figure in the new version.

• Figure 6 – that's a lot of correlations that all show a lot of different things, and in many cases have very few data points associated. The time bins are inconsistent lengths of time apart, and contain inconsistent numbers of data points. Several bins contain too few data points for a Pearson's correlation to accurately and precisely constrain trends. Consider excluding bins below a certain number data points, and looking at Cook's Distance, DFITS, or some other measure of influential observations – e.g. the (slightly) positive correlation in Panel (a) at 16-19 Ma probably wouldn't be present if it wasn't for the youngest point. If you're going to show correlation lines then you should show the points they refer to, meaning the temperature-detrended points and temperature/size-detrended points should also be shown, rather than just their trendlines. I'm sure this makes the plot much messier and may require a separate or much larger figure to show properly, and part of it could potentially be moved to the supplementary, depending on what it shows. The dataset relationships shown in Table S3 illustrate the point better, if less interestingly than a nice figure, though they need a description of what the colours mean. I'm also not entirely sure what the grey lines for the Torres et a. (2024) relationship between temperature and Ɛp mean here. All the Ɛp values recorded seem a long way away from this relationship – is this because CO2 was higher in the geologic data than in the culture experiments? Clearly these samples are behaving very differently from the cultures, which makes me wonder if any temperature response data taken from them is applicable. I suspect I'm misinterpreting this and the intended interpretation is more complex, but this needs clarifying in the text.

We appreciate the reviewer's suggestion to simplify this Figure 6. We propose to retain the colors to highlight the individual time windows, but we will remove the individual linear fits from the figure. Then Table S3 will remain as a reference to assess the relationships within each time interval.

Since we feel it is important to show the measured data distribution as well as the trends when the temperature-growth rate effect on Ep is accounted for, we propose to add a separate panel for each site plotting the points with the growth rate-corrected Ep. We agree that this will more transparently illustrate the key points.

With this illustration, it would no longer be required to include the Torres *et al.,* (2024) slope in the figures (in the current version this relationship is deliberately offset from the other data to illustrate the slope).

• Similarly with figure 7 – this one is better, but I'd still prefer it if there was a point representing the temperature-removed reconstructed values, rather than just a line that looks like an error bar at a glance. Is the R2 for the temperature-corrected points similar to that for the raw values?

We appreciate the suggestion of the reviewer and would propose to present Figure 7 then with two sets of panels, the upper set of three with the measured Ep and a lower set of three with the growth-rate corrected Ep (rather than the lines which could be confused with error bars). In this way there will be separate R2 values and their origin will be clear.

• Figure 8: I'm not quite sure how the compound axes with the CO2 doubling work – either there's a direct conversion that can be made to CO2, in which case the doubling CO2 axes aren't needed, or there isn't, in which case the axes shouldn't be compared on the same y axis.

We propose to describe in the figure legend that the pCO2 refers to the boron and leaf gas proxies (in fact, red dots from 925 published CO_2 will be omitted). Additionally we suggest clarifying this visually by moving the atmospheric CO_2 axis to the right hand side of the graph and could be vertically displaced.

• Please can you add in a paragraph about how your record compares to the boron (and leaf gas exchange) in Figure 8? With the two records overlaid, it's difficult to see if they match well. On close examination, they don't seem to, which is an important finding.

Following the suggestion of the reviewer to further describe the new records relative to the other proxies, first of all, we propose to modify Figure 8 to better reflect proxy records from the same time interval - specifically truncate the included records at 16.1 Ma (the youngest alkenone point) so that it is easier to compare the boron and leaf gas proxies with the alkenones in the period when they overlap and provide more visual clarity. We also propose to add small symbols to the new 1168 and 1406 Ep records to accurately reflect their resolution and to remove the connecting line when there are hiatuses >1 m.y.

Then, as suggested, in section 4.4. after line 429 we propose to add the following paragraph:

"Similar to phytoplankton proxy records, the available low resolution leaf gas CO_2 records suggest a decline in CO_2 from the mid to latest Oligocene. However, in contrast to phytoplankton proxy records for a significant long term decline in CO_2 from the early Oligocene through mid-Miocene, leaf gas CO_2 proxies suggest higher CO_2 in the early Miocene than the Oligocene due to a positive shift across the OMT. Boron isotope-based CO_2 records from 24 to 18 Ma show significant variability with no clear trend, although the higher density of data around the OMT suggests a CO_2 rise from 23 to 20 Ma which may be consistent with the trend observed in the Ep record at Site 1406, which has the highest resolution for this time interval."

Additionally, we propose to clarify last section of the discussion in line 461:

"However, the Oligocene paradox is not easily resolvable from updated calibration of the εp-CO2 relationship. The late Oligocene paradox arises from an inverse correlation between εp and SST reconstructions in regions other than the Southern Ocean such as the North Atlantic, and a lack of correlation between εp and the global climate signal in benthic $\delta^{18}O$ trends. The discrepancies between alkenone and published TEX86 at ODP 1168 suggests continued reevaluation of SST proxy interpretation are needed, along with evaluation of the potential influence of changing surface ocean circulation on SST in some locations such as the North Atlantic. Additionally, the divergence of CO_2 trends among εp and boron isotopes suggest that further interrogation of ocean chemistry biogeochemical cycles potentially affecting the growth and physiology of alkenone producers and the calculation of CO_2 from boron isotopes, are crucial to reconcile climate sensitivity to CO_2 in the Oligocene to early Miocene."

Minor points and typos:

• Line 63: 1406 is a bit far north to be subtropical – it's referred to as midlatitude later in the same paragraph as well.

Corrected as suggested in the new version of the text

• There are a few places where references are in the wrong brackets – i.e. (Guitian et al., 2024), rather than Guitian et al. (2024) where they're referred to in the text. I spotted ones on lines 94 and 96, and 226, but there might be more.

We will correct this wrong referencing in the new version of the text

• At line 174 it says that samples 23.1-29./1 Ma in 1168 could not have UK'37 resolved (and therefore, presumably not good enough resolution for alkenone δ13C measurement), but in figure 2, the 1168 record goes up to around 26 Ma – are the dates right?

Most of the samples from 23.1 to 29.1 had not well resolved C37, however some of them this interference was found in the C37.3, and therefore 37.2 was resolved well enough for a δ13C measurement. Detailed description can be found the supplementary table

We propose to clarify the text:

"The RTX-200 column provided substantially improved resolution of C38 peaks, allowing quantification of C38:2 and C38:3 ME peaks, but for samples between the ages of 23.1 and 29.1 Ma in ODP 1168 it did not perform well enough for all the C37 peaks. Therefore, for this set we provide temperatures estimated from the $U^{k'}_{38ME}$ *ratio applying the Novak et al. (2022) core top calibration (Table S1)."*

• Line 184: grammar – "GC oven was set at 90C…"

Will be corrected in the text.

• Figure 4 – the axis lines are sometimes missing, and the tick marks aren't placed right on the age scale at the bottom.

We thank the reviewer for noticing these errors. Are due to the low resolution of the graph file placed within the word text. Will be corrected in the new version.

• Line 380 – GDGTs plural.

Will be corrected in the text.

• Line 419-411 – R2 should be superscripted

Will be corrected in the text.