Dear Joost and co-authors,

First, my apologies for the long delay in reaching a decision. Your revised version was re-evaluated by one of the original reviewers, and while your revisions have addressed several earlier comments, they still find that insufficient support for the current conclusions is present; mainly due to the uncertainties in the dates of your core-top samples.
This is a valid concern which indeed remains an important caveat of the manuscript- although I understand there is no easy way to resolve this and provide dates with confidence.
-Consider if it would be possible to constrain sedimentation rates for the sites more quantitatively, and in particular then also whether large differences in sedimentation rates are present.
-as noted also by this reviewer, the supplementary figures show that many of the relationships are driven by 3 data points (relationships with pCO2 and/or nutrients); yet it is not clear to the reader which sites these data correspond to.

I feel your data are valuable and use cutting-edge methodologies and therefore merit giving an additional opportunity for revisions; but leave it to the authors to decide whether they feel they can go further in accommodating the reservations of Reviewer #2. In case not, you could consider reformulating your discussion and conclusions in a more cautious way and add some discussion on the way forward to explore the controls on dinoflagellate cyst d13C in future studies.

With best regards
Steven Bouillon

Author response:
We once again thank the editor and reviewer for their constructive views on our work.

Indeed, we agree it is unfortunate that we have no means of constraining sedimentation rates for these sites. But as mentioned, even if we did have sedimentation rates, that would still be an ‘imperfect’ approximation of cyst-production age as the age-distribution of the dinocysts themselves cannot be constrained. This is a challenge unique to the here presented data and we fully agree that these challenges should be discussed in the text and that, in absence of constraints, some additional sensitivity analyses are needed and these will be added to the revised manuscript.

As suggested by reviewer 2, we explore the potential impact of an offset between pre-industrial $p\mathrm{CO}_2$ and potentially ‘modern-like’ $p\mathrm{CO}_2$ cysts. As suggested, we propagate a 45±15 ppm error (a normal distribution with a 3-sigma range between ~0 and ~90 ppm) and the analytical errors (5% of the $p\mathrm{CO}_2$ value). In a second scenario we explore the errors associated with a random draw $p\mathrm{CO}_2$ -change from pre-industrial, assuming cysts were produced between 1800 – 2000 CE. These $p\mathrm{CO}_2$ values are strongly non-normally distributed, and might be considered a more realistic scenario, as the exponential $p\mathrm{CO}_2$ rise over the last decades implies there is a greater likelihood of cysts being produced during times of only moderately elevated $p\mathrm{CO}_2$ conditions. Both scenarios result in an offset of the absolute $p\mathrm{CO}_2$ values, but regression parameters are fairly robust with regard to these errors. Still, the offset and slight changes in the regression parameters do imply that the $p\mathrm{CO}_2$ estimates resulting from the original 1850 CE assumption are more likely to underestimate both absolute values and $p\mathrm{CO}_2$ variability, which will be added to the revised manuscript.

Furthermore, we include a secondary dataset spanning 0 – 1500 CE from the North Atlantic offshore Ireland (Feni Drift; Richter et al. 2009). The measured $\delta^{13}C_{DINO}$ in these sediments are broadly similar to those of the nearest three core-top samples. When the core-tops are grouped together, they fall exactly on the drift-sediment average, as do the data distribution and ^{13}C-variance within the cyst population. This shows that an anthropogenic imprint on those three core-top samples cannot be detected. Clearly, however, the potential for added uncertainty for other localities can still not be fully dismissed.

The above challenges and statistical exploration are now discussed in a separate section "*4.5 Challenges of age-control and potential caveats associated with anthropogenic carbon*". In addition, Figure 6 now includes a plot that should ease identification of the localities used in this study.

Unfortunately, my primary concern with this paper remains, and the additions and alterations do not go far enough for me to be confident that the data support the conclusions. The addition of Supp Figure 1 (which should be in the main paper) does demonstrate the relatively minor impact of the removal of “outliers” however it also highlights the more substantial concern.

Firstly, on the removal of “outliers” the selection of these is based on an a priori assumption that CO2 does control ep, and that there are some modern contaminating cysts in the samples. However the a priori nature of this correction means it is arguably inappropriate, and the selection process fairly arbitrary.

Author response to point 1:
We are happy to see reviewer agrees the impact of outlier omission is very minor. However, we are surprised that the reviewer believes CO_2-dependency is an a priori assumption to outlier detection or omission; this is not the case. We further clarify that outliers are identified from the $\delta^{13}C_{DINO}$ data populations, for each sample and species individually, as indicated in the manuscript and also described in our reply to the previous review of Reviewer #2 (lines 238-240):

“This final step of data-treatment removed positive and negative measurement outliers from the sample- and species-specific $\delta^{13}C$ *population (outside ±2.5 IQR), after eliminating the extremely low-signal intensities (<0.2 Vs) and correcting for the drift induced by background C in the system.”*

While negative (*i.e.* $\delta^{13}C_{DINO}$ below the sample- and species-specific population average) outliers ($n = 21$) clearly outnumber positive outliers ($n = 3$), both are identified. After outlier omission a greater number of distributions are indistinguishable from a normal distribution; which is a common effect of omitting outlier values. In fact, the data correction, exclusion of measurements with low signal intensity and outliers leads to somewhat poorer, not better, correlation to pCO_2. We have carefully revisited our explanation of the selection process so that it is now clearer that this is not arbitrary nor biased (line 221-227):

“Based on typical deep ocean sedimentation rates in the range of centimetres per kyr, the core-top samples are expected to contain a mixed assemblage of dinocysts produced mostly within the last centuries to millennia but could also include cysts produced during the last few decades that are likely affected by anthropogenic influences. It is particularly relevant to consider because a steep $\delta^{13}C$ *decrease (~2‰ since 1850 CE of which >1.5‰ occurs after 1950 CE) (Francey et al., 1999; Keeling et al., 2017) accompanies the* pCO_2 *rise (>130 ppmv since 1850 CE, of which >100 ppmv after 1950 CE). So even if enhanced carbon isotope fractionation at higher* pCO_2 *(Freeman and Hayes, 1992; Hoins et al., 2015; Brandenburg et al., 2022) would not play a role, the most recent specimens are likely to be impacted by decreasing* $\delta^{13}C_{DIC}$*.”*

And lines 249-254:

“Distinctly non-normally distributed $\delta^{13}C$ *values were not previously observed in recent pollen and ancient dinocyst species analyzed with the same method (van Roij et al., 2016; Sluijs et al., 2018). The here presented down-core pre-industrial* $\delta^{13}C_{DINO}$ *show a similar mean, variance and data distribution to the nearby core-top samples (Supplementary Fig. 1), suggesting that, at least for these nearby localities, the analysed core-top specimens represent pre-industrial conditions. We find an influence of Suess-effect and increased* pCO_2 *impacts on*

the $\delta^{13}C_{DINO}$ data is the most likely factor to explain the appearance of a small number of predominantly ^{13}C-depleted outliers and resulting (subtle) negative skewing of the $\delta^{13}C$ distributions (Fig. 4)."

Secondly the acknowledgement that these core tops are of mixed age both in the need to remove "outliers" and the statement at the start of the response: "core-tops (the top-most 2 cm of sediment) contain individual sedimentary components with a range of ages. While it might be possible to 14C-date carbonate, bulk organic matter or even specific chemical components, these materials will derive from different times in the past not necessarily the same as the dinocysts analyzed here. Hence, a single measurement will not show the range of ages of the individual dinocysts (i.e. the age-distribution of our individual dinocysts)." highlights that the age of the material analysed is highly uncertain.

> **Author comment:** Please note that we already acknowledge the age uncertainty in previous versions of the manuscript and have expanded this discussion based on the valuable comments of reviewer #2 in the previous and the current version of the manuscript.

In many cases in oceanography this is the case but not consequential, as the potential difference mixed ages causes in an individual core top sample is much less than the signal that is being reconstructed. This is not the case here. Whilst it is not unusual (although strictly wrong) to assume that a core top is "present day" in this paper it is accepted that the core top is a time-integrated slice, but then assumed that it all represents an arbitrary date of 1850.

> **Author response:** This is not our assumption. In our analyses, we use the date of (pre-)1850 simply to distinguish cysts affected by fossil fuel combustion and those older than that for our sensitivity studies.

As Supp Figure 1 and Figure 1 show, the signal being reconstructed – the different between lowest and highest CO2 at the different sites, is on order 150 ppm, whilst the difference between modern and 1850/pre-industrial CO2 is close to 90 ppm. As these are similar order of magnitude, the tightness of the age control becomes a major concern as any contamination of modern specimens in "low CO2" high latitude sites could overwhelm the signal.

At present it is impossible to assess how great a problem this could be. Estimates of site sedimentation rates are not provided (beyond "typical deep ocean sedimentation rates in the range of centimetres per kyr" lines 314-5) and the sites are not sufficiently documented or cited in a way that allowed me to find out whether sedimentation rates for the sites are available, and whether they are particularly low, high, or variable across the calibration set. At present, the 2 cm core-tops may represent 200 years of time, during which the variable they are trying to reconstruct has changed substantially. This uncertainty is not sufficiently dealt with in the calibration, with a 5 % uncertainty added but no explanation as to how this value was reached.

> **Author response:**
> The 5% error derives from the analytical error on the CO_2 measurements, and this has now also been clarified in the main text (line 148-149):

> *"We employed a Monte Carlo simulation to assess the potential impact of the pCO_2 correction by propagating (1) the 5% analytical error on pCO_2 values [...]"*

A more robust treatment of these data would be to assume a 90 ppm (present to 1850 adjustment) uncertainty in the pCO2 concurrent with cyst formation and rerun the regression analyses with this uncertainty included. Better still, for each site an assessment of the age uncertainty for the core top could be made (based on sed rate) and a CO2 uncertainty for the population of individuals estimated. This should then be made used in the regression analysis.

Without this step, it is impossible to know whether the regressions proposed are plausible.

Author response to point 2:
We thank the reviewer for challenging us to think more critically about this step in our analyses. The sedimentation rates for the core material are unfortunately unconstrained but, using explorative error analyses and new data (see below), we now show unlikely of importance to the main conclusions ($p\text{CO}_2$ and nutrients control ε_p in dinocysts).

To accommodate the suggestion of the reviewer, our revised manuscript includes two new scenarios assuming the top sediment represents 200 years where we explore 'worst-case' conditions and their potential effects on the calibration. The first scenario adds a $p\text{CO}_2$ value from a single random draw on a sample-level, from a normal distribution that spans 0 – 90 ppm to the analytical error (5%). The second scenario is identical but uses a random resampling of $p\text{CO}_2$ values above pre-industrial $p\text{CO}_2$ levels for the period 1800 – 2000 CE. The assumption that the core-top records 200 years of sedimentation is equivalent to a 2 cm core-top slice at a very high-accumulation-rate site (10 cm/kyr sedimentation rate).
These analyses show that, while the reviewer is correct to state that the CO_2 difference between the date of measuring or collection and the hypothesized bulk of the data (1850 CE and older) is substantial, it is unlikely that "*any contamination of modern specimens in "low CO_2" high latitude sites could overwhelm the signal*". When adding a 0 – 90 ppm error to all data or resampled $p\text{CO}_2$ values, with exception of the absolute values, the regression parameters are fairly robust.

In addition to the ~90 ppm error scenario, we also simulate the uncertainty in sedimentation rates by assigning a $p\text{CO}_2$ value (above pre-industrial) sampled from the atmospheric CO_2 between 1800 and 2000 CE. The reason for doing this is that the $p\text{CO}_2$ values from 1800 to 2000 follow a non-normal distribution and therefore strongly biased (anthropogenic) values are considerably less likely to occur. Errors are included in the regressions through a simple resampling of atmospheric $p\text{CO}_2$ for each datapoint. Both analyses and error distributions are included in the new Figure 7 (included below).

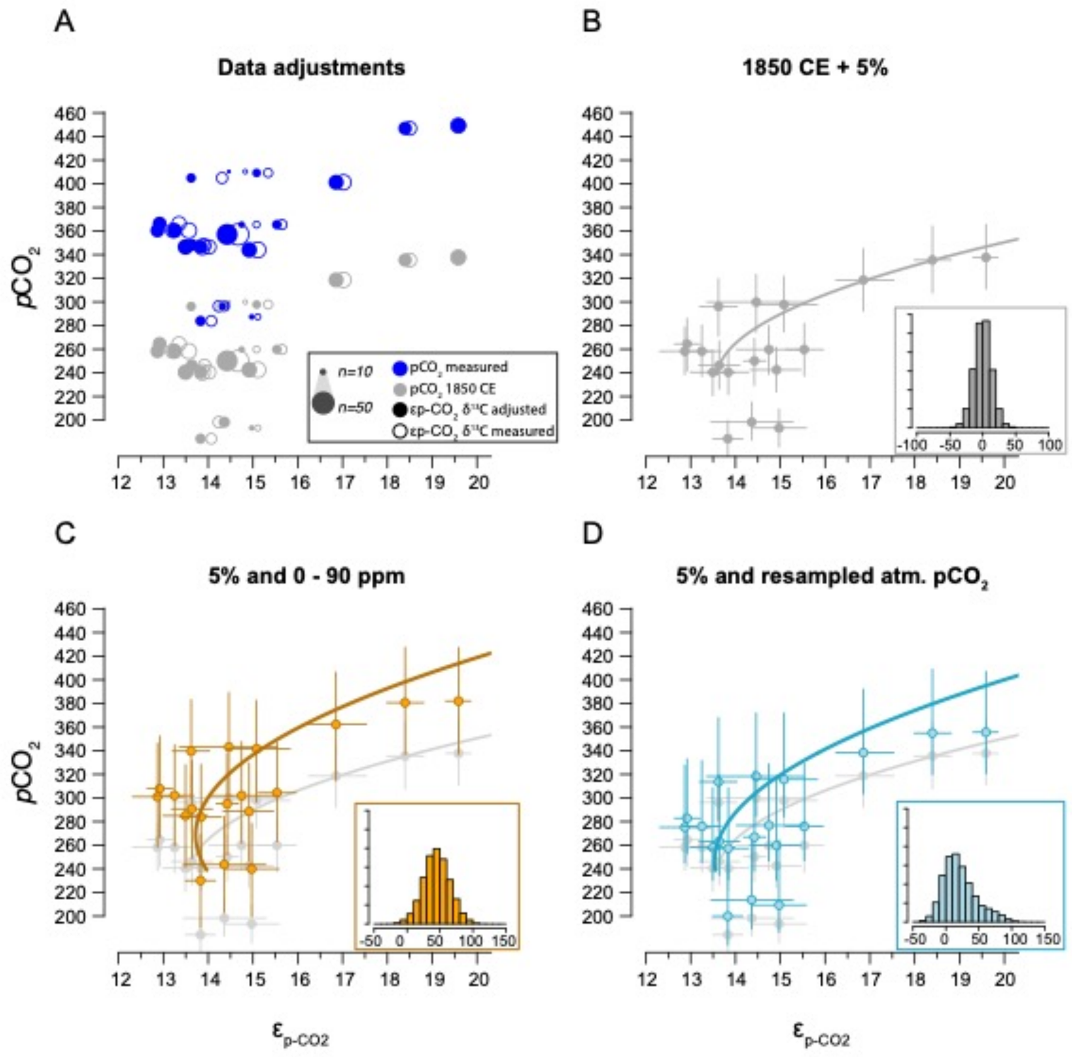
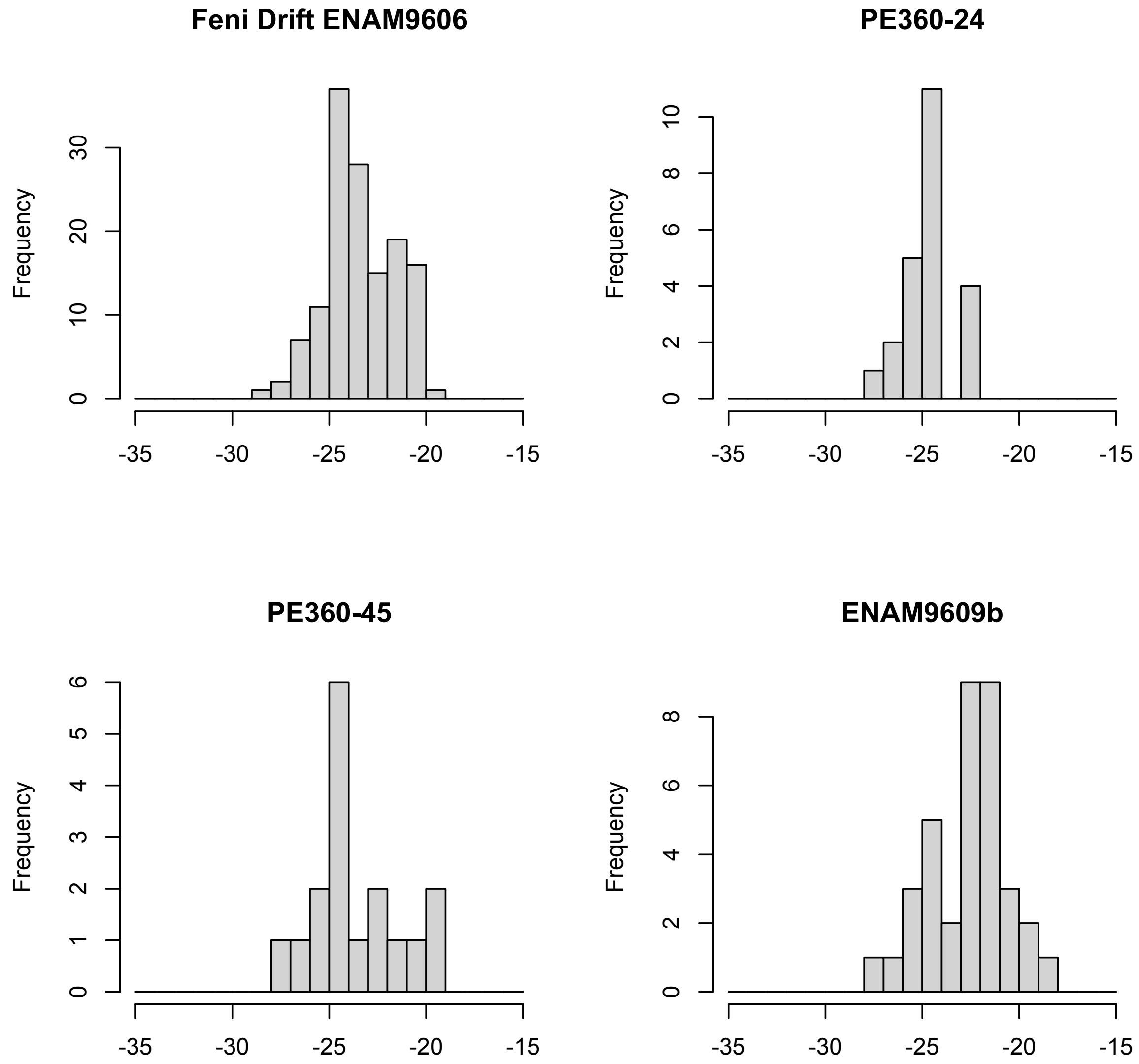


Figure 7. Data treatment and potential effects of anthropogenic carbon emissions. A. *Effects of data treatment on the difference between measured and adjusted* pCO$_2$ *and* δ^{13}*C* (*εp-CO*$_2$) *(same as Fig. 3). Open symbols indicate measured* δ^{13}*C, closed symbols represent data after eliminating small signals (<0.2 Vs) and outliers. Blue dots represent measured CO*$_2$ *values and grey dots indicate the CO*$_2$ *around 1850 CE.* ***B.*** *Quadratic regression (red line in Figure 6B) with propagated analytical error on* pCO$_2$ *and* δ^{13}*C only, using CO*$_2$ *values around 1850 CE (grey filled symbols in panel A.* ***C.*** *As in B but with addition of a 45 ± 15 ppm error to reflect potential impact of anthropogenic CO*$_2$ *in orange. Grey dots and curve of panel B are added as a comparison.* ***D.*** *As in B but with addition of the CO*$_2$ *increase relative to pre-industrial in the period 1800 – 2000 CE. Insets (bottom right) in panels B, C and D show the combined error distributions (in ppm) imposed on pCO*$_2$*. All error bars in panels B – D on pCO*$_2$ *and* $\varepsilon_{p\text{-}CO2}$ *are 2.5 – 97.5% percentile ranges from Monte Carlo simulations (n=1000).*

In addition, in our revised manuscript we now include a new down-core $\delta^{13}C_{DINO}$ *(O. centrocarpum)* dataset spanning 0 – 1500 CE and compare this data to $\delta^{13}C_{DINO}$ of the same species in three nearby core-top samples. The $\delta^{13}C_{DINO}$ distributions of the pre-industrial cysts match those of the core-top samples and provides circumstantial evidence that the nearby core-tops (still) include mostly cysts unaffected by anthropogenic sources (line 250-252). Histograms of these data are included below and as a new Supplementary Fig. 1.



Supplementary Figure 1. Histograms of down-core $\delta^{13}C_{DINO}$ (O. centrocarpum) for ENAM9606 in the North Atlantic compared to three nearby core-top samples (PE360-24, PE360-45, ENAM9609b). *ENAM9606 (55.650 °N, -13.985 °E) represents down-core $\delta^{13}C_{DINO}$ for ~0 – 1500 CE (Richter et al., 2009), whereas PE360-24 (55.496 °N, -15.801 °E), PE360-45 (55.539 °N, -15.845 °E) and ENAM9609b (57.160 °N,-10.26°E) represent nearby core-top $\delta^{13}C_{DINO}$. Frequency (y-axis) indicates the number of measurements for each of the 1‰-wide $\delta^{13}C$ bins. All $\delta^{13}C$ distributions are background-corrected values, without outliers.*

Finally, we agree with the reviewer that further details and error propagation would benefit further assessment of this work and utilize the above analyses to illustrate where the main uncertainties lie to inform future efforts. Accordingly, we have added a paragraph in the discussion "*4.5 Challenges of age-control and potential caveats associated with anthropogenic carbon*" (lines 429-448). This includes a brief outline of the challenges and the results of the statistical exercises. We hope to have sufficiently illustrated the potential for age-dependent errors to play a role in the calibration offered in our work and following the reviewer's suggestion, we have added further nuance where needed.

Overall, we find the various ‘error’ scenarios indicate that the calibration based on pre-industrial $p\mathrm{CO}_2$ is fairly robust, though the absolute values and perhaps variability may be underestimated, if a significant number of cysts was produced (long) after 1850 CE.

Our explorative error analyses yielded the following changes to equation 2a, which have been included in our revised manuscript (lines 291-302):

Equation 2 quadratic (only suitable for use > 240 µatm)
$\varepsilon_{p\text{-}CO2} = 40.8 \pm 7.2 - 0.23 \pm 0.055\, p\mathrm{CO}_2 + 4.88 \pm 1 \times 10^{-4}\, p\mathrm{CO}_2{}^2$
(Adjusted $R^2 = 0.79$, $p < 0.001$, RSME = 1.13 ‰) (Figure 6B,F)

Equation 2b quadratic (Monte Carlo constrained errors – analytical for $p\mathrm{CO}_2$ and $\varepsilon_{p\text{-}CO2}$) (Figure 7B)
$\varepsilon_{p\text{-}CO2} = 35.6\ {}^{+5.8}/_{-5.6} - 0.19\ {}^{+0.045}/_{-0.045}\, p\mathrm{CO}_2 + 4.1\ {}^{+0.91}/_{-0.88}\ 10^{-4}\, p\mathrm{CO}_2{}^2$

Equation 2c quadratic (as 2b with additional 45 ± 15 ppm $p\mathrm{CO}_2$ error) (Figure 7C)
$\varepsilon_{p\text{-}CO2} = 39.3\ {}^{+11.5}/_{-8.8} - 0.19\ {}^{+0.058}/_{-0.076}\, p\mathrm{CO}_2 + 3.4\ {}^{+1.3}/_{-0.95} \times 10^{-4}\, p\mathrm{CO}_2{}^2$

Equation 2d quadratic (as 2b with resampled $p\mathrm{CO}_2$ rise 1800 – 2000 CE) (Figure 7D)
$\varepsilon_{p\text{-}CO2} = 29.8\ {}^{+11.0}/_{-8.0} - 0.13\ {}^{+0.061}/_{-0.084}\, p\mathrm{CO}_2 + 2.6\ {}^{+1.5}/_{-1.1} \times 10^{-4}\, p\mathrm{CO}_2{}^2$

Minor comments.

L 43-45 “However, many of the organic compounds used for CO2 reconstructions such as alkenones (e.g. Pagani, 2013), phytane (e.g.Witkowski et al., 2018), porphyrins (e.g. Freeman and Hayes, 1992) or bulk organic matter (e.g. Hayes et al., 1999) are not related to a single species, genus or even group of organisms.” It is incorrect to say that alkenones are not related to a single group of organisms.

Author response:
Thanks for pointing this out. We have rephrased to “*are not unique to a single species, genus and sometimes not even a group of organisms.*” (lines 44-45)

L 132 Is there a confusion between Ocean Data Viewer and Ocean Data View here?

Author response:
Yes, this should read Ocean Data View and not ‘*viewer*’ (link also corrected) (line 134).

L 140 Assumption of moderate to low sedimentation rate of < 10 cm kyr given. Is this at all reasonable for the sites used? Are any sites particularly low or high sedimentation rate? Note that at this sed rate the 2 cm core tops span 200 years of accumulation.

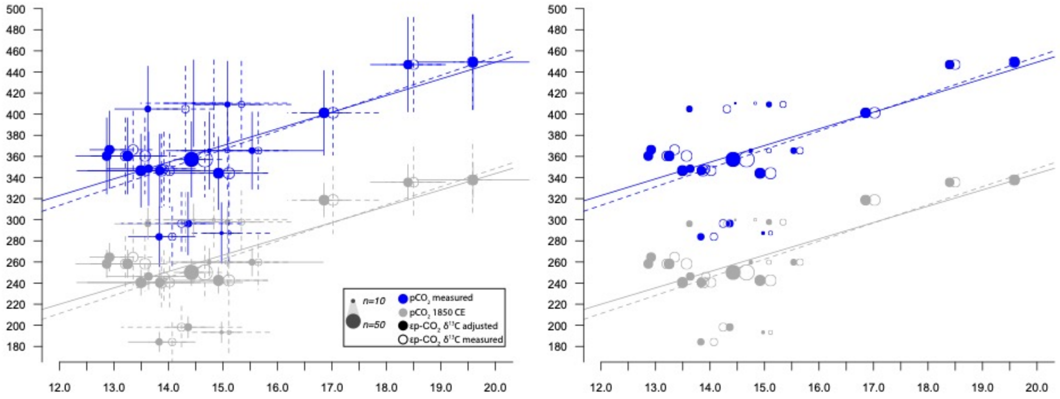
Author response:
As the sedimentation rates are unconstrained, we unfortunately must rely on generalisations. As all our sites are reasonably far offshore and not deposited in drift sediments, we find an upper limit of 10 cm/kyr is reasonable (line 141-143).

L 145 “this correction has only a small impact on the patterns in the CO2 data (Supplementary Figure S1).” This is not a fair assessment of what Figure S1 shows. It shows that the difference in regression between adjusting and not adjusting for “eliminating small signals and outliers” is small, but the impact of assuming all 1850 CO2 vales is very large.

Author response:
We have clarified that the offset is potentially large (line 145-147) but the regression parameters are fairly robust (lines 320-325).

What do the difference sizes of symbol on Supp. Fig 1 mean? Can error bars please be added.

Author response:
The symbol size represents the number of measurements included from each location, similar to those in Figure 6 – this has been added to the figure key (note that SI Fig. 1 is now included as Figure 3 and 7A). Error bars can be added but will make this plot very crowded; in addition to the 76 data points with different sizes and fill, it will then include x- and y-error bars of both different color and type (see Figure below, left). We understand the value of included error bars but find the version without (Figure below, right) is a clearer illustration of how the data treatment affects the data used in the regressions. Of course, we are willing to optimize a version with error bars if that is deemed preferable.



Supplementary Figure 1 and Figure 5 also highlight that most of the regression is driven by three samples with high estimated CO2. As a full dataset is not provided it’s difficult to know for sure, but from Figure 1 I estimate these are all from the Mediterranean. Is there a difference between Mediterranean and Atlantic types? Is there any reason to think that the correction may be more or less valid in the Mediterranean? What are the sedimentation rates for these and all sites? Not enough information is provided in the manuscript or provided files for me to chase down the individual core documentation.

Author response:
We thank the reviewer for this suggestion - to facilitate identification of individual samples in the fractionation plots, site numbers are now included in Figure 6A, which correspond to the now numbered sites listed in Table 1 and Figure 1.

The regression is most dependent on the cluster of data points as well as the three data points referred to by the reviewer. The two highest fractionation values derive from the Mediterranean while the third highest fractionation value is observed in the core-top sample from the equatorial Atlantic. The other Mediterranean sample plots within the larger cluster of values.

No changes in cyst morphology were seen between samples and we thus have no reason to assume that the cysts from the three localities with the highest fractionation values or those from the Mediterranean sites differ from the majority of the North Atlantic localities.