General Response

Dear Editor McClymont,

We thank you, Dr. Rattanasriampaipong, and Dr. Naafs for taking the time to evaluate our work, which greatly benefitted from your feedback. The revised manuscript addresses the outstanding points, which are outlined below.

Additionally, we apologize for omitting the response to your first round of comments in our previous submission. We are perplexed as to how this happened on our part. Our responses to your first round of comments are also attached below.

Editorial Comments (Round 2)

In relation to the second comment from reviewer 2, the inclusion of the east-west North Pacific comparison is valuable in this iteration of the manuscript. In that comparison, you discuss that as U1417 used CIMS but ODP 887 did not, you have used the ODP 887 data. I previously commented in the first round of reviews that for some sites, GC-FID may not be able to identify the alkenones sufficiently well where there is complexity in the sample. For U1417 this was certainly the case, even after sample clean-up, due to it's proximity to sources of petrogenic carbon in the St Elias mountains, which is unlikely to have affected ODP 887; GC-CIMS provided an opportunity to ensure isolation of the alkenones which might not otherwise have been achieved. Clarifying this utility of GC-CIMS analysis in some settings, albeit with the caveats you have shown, would give some nuance to the assessments made here and perhaps address the note of reviewer 2 about the implications of your work for other sites.

This is an important point, as we do not want to diminish the usefulness of the GC-CIMS technique in such situations. We added the following text to address this point (new text is bolded):

The same methodological differences (GC-FID vs. GC-CI-MS) exist between the late Neogene and Quaternary alkenone SST datasets from ODP Site 887 and IODP Site U1417 in the subpolar northeast Pacific, but we are unable to determine whether these methodological differences lead to the same systematic biases between the datasets because there are very few proximal samples measured by both methods. The alkenone record from IODP Site U1417 was analyzed by GC-CI-MS (Sánchez-Montes et al., 2020), while the record from the nearby ODP Site 887 was analyzed by GC-FID (Dowsett et al., 2017; Herbert et al., 2016). ***The selectivity of the GC-CI-MS method permits quantification of alkenones in complex mixtures that cannot be resolved by sample preparation (Rosell-Mele et al. 1995). At***

IODP Site U1417, the GC-CI-MS method enabled the isolation of alkenones from a complex mixture that resulted from the introduction of petrogenic carbon from the nearby St. Elias mountains (Sánchez-Montes et al., 2020), which was not an issue at the more pelagic ODP Site 887 (Dowsett et al., 2017; Herbert et al., 2016). *There are only 11 samples from these two records that overlap in time (Fig. 6a). Although the mean $U^{K'}_{37}$ values from mid-Piacenzian subsets of these data are statistically indistinguishable (Fig. 6b, 0.008 $U^{K'}_{37}$ or 0.24°C, p = 0.66), the systematic differences between the GC-CI-MS and GC-FID $U^{K'}_{37}$ values from ODP Site 882 and the well-documented systematic offset between the GC-CI-MS and GC-FID methods give us reason to suspect that the IODP Site U1417 and ODP Site 887 alkenone datasets may not be strictly intercomparable. For this reason, we exclude the IODP Site U1417 alkenone dataset from our discussion of the implications of the proposed correction to the ODP Site 882 GC-CI-MS alkenone record for the paleoceanogrpahic history and paleoclimate data-model evaluation in the subarctic North Pacific.*

Dear authors - thank you again for your careful replies and incorporation of the feedback into the response and updated version. I am recommending minor revisions because I am satisfied that this does not need to be re-sent to the reviewers. In your final submission can you check that the comments I made on the first version are also addressed, as I could only see your responses to the reviewers here.
Thank you for your submission and your responses so far.
Best wishes, Erin

Our apologies for missing the editorial comments in our previous response. Our responses are included below.

Editorial Comments (Round 1)

- As with the other reviewers, clarity on the adoption of linear vs non-linear corrections is needed. Given the second-order polynomial relationship between GC-FID and GC-CI-MS UK'37 values identified by Chaler et al. (2003), the rationale for the assumption that a correction for ODP 822 should be linear is unclear.

We selected a linear model because not all of the terms of a quadratic fit of the GC-CI-MS to GC-FID $U^{K`}_{37}$ values from Site 882 are statistically significant. This is now stated in lines 145–148. We also discuss the exploration of alternative models in much greater detail in section 3.1 lines 184–235.

- Line 100 indicates that a simple correction was made between "adjacent samples" from Haug 1995 and Studer et al. (2012), so not measurements made on the same sediment sample? If this is the case, is there not additional uncertainty introduced to

the comparison of the two approaches, because comparing samples of different ages will also be expected to have different SSTs at ODP 882?

Yes, but these uncertainties are incorporated in the standard error of the regression between the GC-FID and GC-CI-MS data. In particular, the standard error would likely be lower had we made new GC-FID measurements at the exact depths sampled by Haug (1995). Furthermore, given that the sedimentation rates at ODP 882 are quite high (~10 cm / kyr in the interval sampled by Studer et al. 2012 per the Tiedemann & Haug 1995 age model), interpolation on the scale of centimeters should not result in differences in SST greater than the prediction uncertainty of the alkenone SST proxy. This can be seen in the comparison between the two datasets shown in Figure 8, for example.

- I'm still unclear why the constant values from "Instrument A" were used rather than "Instrument B" (Line 86-87). The text says that the "masses better corresponded to the UK'37 values in the chosen sites", but Figure 3 in Chaler et al. (2003) doesn't show UK'37 values, only the masses of the individual alkenones. Can you clarify the reasoning here? The Yamamoto and Kobayashi (2016) paper seems to indicate that at least for some samples, the masses of alkenone present could have been equivalent to Instrument B according to Figure 3 in Chaler et al (2003).

The main reason we did this is because the equations derived for "Instrument A" are based upon a much more heavily sampled "trace" range of analyte masses in the detector compared to "Instrument B." This can be seen in Figure 2 of Chaler et al. (2003).

The choice of which set of equations we used to show the theoretical GC-FID vs. GC-CI-MS $U^{K`}_{37}$ relationship is inconsequential since the same sigmoidal relationship is present in both instruments investigated by Chaler (see Figure 3 of Chaler et al. 2003). We included the statement about "Instrument A" for the sake of transparency to explain how we derived the black line in our Figure 3, *which is used for illustrative purposes only.* Furthermore, the observed relationship between GC-FID and GC-CI-MS $U^{K`}_{37}$ values in ODP 882 samples is analogous to what is shown in Figure 3 of Chaler et al. (2003); this can be seen in our Figure 4a, where samples with a larger GC-CI-MS peak area show the largest deviation from the 1:1 line.

- Results on Figure 3: If the corrections are mass dependent as well as UK'37 dependent, should there not be uncertainty around the solid black line of the theoretical GC-CI-MS / GC-FID relationships given that masses in the 5-30 ng range

were used? Given that the Yamamoto and Kobayashi (2016) paper also highlights concentrations exceeding 200 ng g-1 (e.g. Figure 4), should the impact of a wider range of concentrations have been tested?

Yes, there are uncertainties in this relationship, but we cannot plot them as the uncertainties in the coefficients were not reported by Chaler et al. (2003).

The impact of varying analyte mass in the GC-CI-MS detector was explored and is discussed in lines 210–235 and is shown in Figure 4. The proposed correction is restricted to the range of GC-CI-MS peak areas where we were able to reasonably characterize the relationship between GC-CI-MS $U^{K`}_{37}$ and GC-FID $U^{K`}_{37}$ (lines 233–235).

- The equation used by Chaler et al. (2003) requires the mass of the analyte to be provided as an input to the constant, but you acknowledge that such data is not available from Haug (1995). However, there is alkenone mass data available from the Studer et al. (2012) paper and from Yamamoto and Kobayashi (2016): have you tried to predict the GC-CI-MS values using measured masses and UK'37 data from GC-FID, to test whether the results align with the Haug (1995) paper?

No, we did not try this because it requires the mass of sediment that was extracted, the injection volume, and the dilution volume in order to back out the mass of analyte in the detector, which is information that we unfortunately do not have.

- As noted by the Chaler papers, the GC-CI-MS technique has been used to isolate alkenones for quantification where concentrations are very low and/or where the samples are complex and co-elutant removal is challenging. Under those circumstances, the GC-FID data would not be representative of the actual UK'37 value either, because the uncertainty on quantifying one or more low abundance alkenones would be large, or there may be over-estimation of the alkenone signal due to co-elution. Whilst acknowledging the caveats of the GC-CI-MS approach there could be a note here to the benefit that this technique has offered (perhaps with a recommendation in the Conclusions that undertaking the instrument-specific calibration of Chaler and co-authors ensures a translation of the data to "GC-FID equivalent").

The following text was added to the conclusions section to address this point:

Our findings underscore the need for careful evaluation of the datasets incorporated into paleoclimate data-model comparison exercises (Fig. 9). The advantages of the GC-CI-MS technique in both sensitivity and selectivity can be particularly useful in

samples with trace alkenone abundances or complex chromatographic backgrounds. In these cases, we emphasize the need for instrument-specific calibrations to calculate GC-FID-equivalent $U^{K'}_{37}$ values from GC-CI-MS measurements so that the data produced in these instances can be used in future paleoclimate data-model comparisons.

Reviewer #1

Dear Editor,

Thank you for the opportunity to review the revised manuscript by Novak et al. The revision is substantially improved, and I enjoyed re-reading it. The statistics are robust, and the new discussion of the implications for regional paleoclimate data–model comparisons is informative and valuable.

I have one suggestion that I believe would further strengthen the manuscript: a brief discussion (e.g., 2–3 sentences in the Conclusions) of how the authors' methodological framework for correcting UK'37 data between GC-FID and GC-CI-MS could benefit future work. In particular, it would be helpful to note how many existing datasets the authors expect may be affected by the same issue they identify at ODP Site 882.

I have no further comments and congratulate the authors on a fine paper.

Best regards,
Ronnakrit Rattanasriampaipong

We are reluctant to state a specific number of datasets that may have this same issue since we do not have GC-FID data to make the same sort of comparison to back up what the authors of those works may perceive as an accusation. In the introduction, we cite all of the studies we could find that used the GC-CI-MS method, which we think is a more diplomatic way of documenting where else to look for this issue for those interested in tracking it down.

Otherwise, we included the following to round out the conclusions (L379–385):

Our findings underscore the need for careful evaluation of the datasets incorporated into paleoclimate data-model comparison exercises (Fig. 9). The advantages of the GC-CI-MS technique in both sensitivity and selectivity can be particularly useful in samples with trace alkenone abundances or complex chromatographic backgrounds. In these cases, we emphasize the need for instrument-specific calibrations to calculate GC-FID-equivalent $U^{K'}_{37}$ values from GC-CI-MS measurements so that the data produced in these instances can be used in future paleoclimate data-model comparisons. Limiting data-model comparisons to alkenone datasets produced by GC-FID, or with GC-FID-

equivalent $U^{K'}_{37}$ values, will ensure that the proxy data under consideration are intercomparable.

Dear Editor,

The authors have done an excellent job in addressing my original concerns. I especially appreciate the statistical details given in the revised manuscript, as well as the extended calibration dataset using data from Yamamoto and Kobayashi (2016). This revised manuscript shows a thorough reassessment of approximately 30-year-old data and also discusses the implications of the proposed correction.

I have a few minor comments left (see below), but this manuscript is now ready for publication.

Best wishes,

David

Line 66: There is an authentic alkenone standard, and the OGU still has this available. See Rechka and Maxwell (1988).

Thank you for the correction. The phrase "*Because there is not an authentic alkenone laboratory standard…*" was removed from this sentence.

Line 99/100: Please state here the modern SST at both sites so readers can see that the modern difference in SST between the two sites is small.

The following text was added to clarify this point (Lines 110–111):

Modern sea surface temperatures at the two sites are essentially identical (5.82°C at ODP Site 882, 5.75°C at ODP Site 883, Fig. 1).

Figure 9c/d: Can the site names be added to panels c/d for clarity?

Added.

References:

Rechka, J.A., Maxwell, J.R., 1988. Characterisation of alkenone temperature indicators in sediments and organisms. Organic Geochemistry 13(4), 727–734.