Review

In this paper, Zhang et al present a multi-proxy and multi-model investigation of hydrological change at the PETM on the western coast of North America. This is a topic that intrigues nearly everyone studying ancient greenhouse climates and the work represents a compelling variety of methods. Collectively, these provide new insights into the nature of regional (and global) hydrometeorology, including evidence for changes in seasonality and extreme rainfall events. It builds on previous work in exciting ways, especially via the data-proxy comparison. It certainly should be published.

However, I do think the paper could be significantly improved. The data and discussion are sometimes presented too briefly and the interpretations are somewhat unclear. Often a range of explanations are offered (which I appreciate) but with no effort to distinguish them or to make use of the multi-proxy data to integrate them. Overall, I think the interpretations are robust and the caveats duly noted, but the reasoning is not always clearly laid out or explained. Other potentially interesting data are ignored (esp post-CIE data or n-alkane distributions). All of that probably sounds rather critical, but I do not want to discourage the authors! The work done is impressive; I think this is a perfectly adequate paper – but there is probably a more exciting paper that better utilises all of that work.

Below are some comments that hopefully elaborate those thoughts in a constructive way (NOTE that some of these are quite important recommendations whereas others are suggestions and I trust the authors and editor to distinguish the two; but happy to be contacted offline if it is unclear):

Abstract and lines 34-42 of the introduction and lines 371-372 of conclusions: I am not convinced that the framing around California's current or future hydroclimate is necessary or appropriate. It is one thing to treat the PETM as an analogue for the future and another to treat Lodo Gulch as an analogue for California's future. As the authors note, understanding regional responses to warming is essential, and I would focus more on that framing.

Figure 1 is certainly adequate but I think it could be strengthened by linking it to previous studies, i.e. including adding Big Horn Basin sites, and by adding in some atmospheric circulation features that are discussed frequently in the paper.

Line 98: Maybe specify 'bulk organic stable carbon isotopes' in the title for clarity. (And align that title with section 3.1).

Lines 118-119: There is some shifting from past to present tense. It would be useful to check this throughout the Methods.

Line 124: I suggest retitling this as 'Leaf Wax distributions and carbon and hydrogen isotopic compositions.' Note that there are a few chemistry conventions that should be properly sorted – n should be italicized in *n*-alkane and carbon numbers should be subscripts in line 136 (And check throughout the manuscript). In line 140, 'were' should replace 'was.'

Lines 142-159: I am delighted to see the proxy data compared to isotope enabled models. That is a strength of this paper. However, hydrological processes are notoriously variable amongst climate models. It would be useful to briefly draw on DeepMIP (or similar) studies to summarise how CESM compares to other models. Is it 'typical', an 'outlier', etc? This could be a whole paper in itself and I certainly am not suggesting the authors add extensive text, but only enough text to help readers put these findings into context.

The authors should consider flipping the order of figures 2 and 3 to better align with the text.

Lines 162 to 179: I trust the authors, but please include n-alkane CPIs and TARs in the figures and a chromatogram (and proportional abundances) in the SI so that we can be confident that the n-alkanes have a leaf wax distribution. And explain and justify that in the text.

Lines 163 to 168: It might be worth noting that the CIE recorded by the n-alkanes is larger than that recorded by bulk organic matter (as is observed in other records), but also that the bulk d13C values never return to pre-CIE values. Also, the authors write that the top of the PETM body is marked by the truncation of the n-alkane CIE; presumably that means they trust it more than the bulk organic CIE? Also, could the PETM body not have been truncated earlier? And the truncation could also include not just the PETM body but the return. Finally, no information is given on the NP biozones. Without overly reproducing the info in John et al., it would be useful to add a few sentences on the stratigraphy, the uncertainty, the age gap, etc and to label the inferred PETM interval explicitly on the figure. (This will also help with subsequent sections, such as lines 236-237, where the authors discuss the challenges of determining sedimentation rates).

Lines 177-179: I don't think that Results sections should be excessive, but this is a bit perfunctory. The brief negative spikes are very large and merit a few more words, especially as one of those appears to be in a coarser lithology than the other data (and is it 'one' or 'two' brief intervals?). Also, some of those negative values appear to post-date the PETM body (see previous comment) so it is worth describing the stratigraphic occurrence of these data with greater precision. They largely ignore these negative spikes in the discussion and I suspect that could be justified by a more thorough Results section.

The 'slight enrichment' in the main body seems very slight indeed and at the limit of analytical error (6‰; line 140). The negative shift prior to the PETM is recorded by only two pre-PETM data points and that should be acknowledged. Perhaps even more important is the fact that post-PETM d2H values are 2H-enriched relative to those two samples but similar to those of the PETM.

Lines 186 to 191. Great to see clay mineralogy woven into this study. Like my comments on d2H values, this section would benefit from some expansion. In particular, I would note that many of the clay mineral assemblages – especially and intriguingly kaolinite to smectite - never return to pre-PETM values (Although our record is more limited, we see aspects of this at Tanzania as well).

Figure 4: For the published version, please make the text larger and edit the text in the figure caption (there are a number of typos).

Figure 5: Why not show the extreme value index for all months (just for completeness)?

Lines 237 to 239: This text confused me a bit. First, the authors really have not constrained the PETM in the previous text. Second, the CIE thickness (if complete) does not allow for determination of the change in sedimentation rates. I think the authors are trying to briefly explain what John et al. (2008) determined, but as written that is unclear. I think this opening paragraph would be stronger if it clearly explained what has been determined previously in this region and by whom, and then ended with a clear list of how the subsequent discussion sections are going to elaborate on that understanding.

Discussion: A general observation of the discussion is that it treats the data in rather isolated silos (with the exception of using models to interpret leaf wax d2H values). And both sections 4.1 and 4.2 seem less like discussions than extensions of the associated Results. Since the Results have already been presented, then draw from all of them to drive the discussion forward. For example, I would not have a discussion section on clay mineralogy but rather one on extreme rainfall events that draws on the mineralogy and the models. That is just a suggestion, of course, and I am one to give authors latitude in how they want to tell their story! But I think a more integrated approach would ensure that the greatest added value emerges from the multi-proxy study.

Lines 255-257: I don't think these comments quite capture the debate about clay mineralogical change at the PETM. The increase in kaolinite has been attributed to both increased humidity and more deeply erosive events; given the context of the paper, I would make those two interpretations explicit. And then… is there any evidence to distinguish between those? The model simulations (or at least what is included) suggests that extreme events and erosion are more likely explanations than increased humidity. If so, say that. Also, I'd encourage the authors to discuss the post-PETM data and allow that to inform their interpretation.

Lines 301 to 303: See comments above – the description of these records needs to recognize the analytical error and be presented with a wee bit better stratigraphic rigour.

Lines 303 to 304: Ascribing the shift in d2H prior to the PETM to orbital variability seems bold. What is the evidence for this? And why don't we see similar orbital variability during the PETM? Or afterwards?

Paragraph starting line 305: This would be easier to follow if the authors clearly set out what behaviour they are attempting to interpret. I assume (but am not sure) that they are arguing that leaf wax d2H – and by extension local meteoric water, given the caveats they correctly note – does not change much in their record (barring a few anomalies). State that clearly. It will make the rest of the text easier to write and to follow. For example, it will allow the reader to understand why we are discussing different factors that could 'offset'.

In addition, I feel like the authors have said that 'we have some data and there are a lot of explanations for it' without drawing on other data to try to narrow down and distinguish hypotheses. What does the mineralogy say about changes in precipitation? What do n-alkane CPIs say about reworked OM? What do ACLs (in SI but never mentioned) say about changes in vegetation? A stronger structure and a more comprehensive discussion will allow more compelling interpretations.

Lines 338-339: This is a really nice application of the model. But the data are not convincing. I am not convinced that there is an analytically significant shift across the PETM (see line 140). And I certainly don't think it is significant in the context of the entire record. But there is such a compelling story here! Based on other mid- and low-latitude sites, we expect a strong positive d2H shift. Assuming plants record annual precip d2H, then the authors' models also predict that. The fact that this is not seen can be resolved by considering a change in seasonal precipitation d2H and growth. That approach predicts a leaf wax d2H shift that is very small and likely below analytical error, and that is what is observed. That is a really nice finding.

(In fact, it is so nice that I'd like to see the authors validate it a bit – perhaps in the SI by determining if the models can predict leaf wax d2H changes at other sites. If the same approach that yields a minor shift in California also yields a minor shift in Europe and a strong positive shift in Tanzania, then that is very compelling. There has been a big opportunity missed by not using the model to assess global d2H records. Maybe for a future paper…)

Lines 350: I like this inclusion of the d13C record.

End Discussion and Conclusion: The authors have a nice integrated dataset. But they never quite draw it all together into a holistic picture. For example, the d13C record is used to infer lower humidity, but that is not mentioned in the abstract or conclusions. The conclusions mention lower winter precip and slightly higher summer precip but do not make it clear that the overall annual precipitation is much lower in the 6x CO_2 simulation. Picking through all of the data, it seems that there is evidence for decreased overall precipitation, especially in the winter; that the precip also becomes more episodic; that these factors and higher temperatures have combined to yield a more arid climate and that impacted the vegetation as expressed in d13C values. All of these will have contributed to a more erosive sedimentary regime. These interpretations are validated by leaf wax d2H values – but that could only be deduced with careful data-model comparison that allowed the competing controls on plant d2H to be constrained. This is a really interesting suite of data, but it does not quite come together as it could.

The paper has a fairly large number of grammatical errors that should be cleaned up on editing. I caught several in the abstract, but they generally become more common further into the manuscript. There are many of them in some figure captions.