

Author's response

April, 2025

1 Editor:

"The manuscript has been carefully evaluated by two reviewers with a background in this field. Both recognize the value of exploring new methodologies for delineating climate states, yet both express the sentiment that the contribution needs to better detail how the breakpoint detection methodology enhances our knowledge of Cenozoic climate. That is, in addition to "Strengthen[ing] the paleoclimate motivation and terminology", as the authors mention in their response, it is also imperative that they discuss the paleoclimatic consequences of their results, comparing them to those obtained from different methods, and discussing the significance of the similarities or differences that arise. As suggested in the response, the idea of applying this method to ice core data from Greenland is also valuable, though will probably require a change of focus/title, and will also need to discuss the paleoclimate implications of the results.

In summary, though there is potential value in applying this method to paleoclimate problems, in this journal the latter take prominence - not the method itself. As such, it may be useful for the authors to team up with or more co-authors with expertise on these questions, to make sure the work is a durable contribution to the paleoclimate literature.

The reviewers further provide very detailed suggestions for more specific improvements."

We appreciate the constructive comments and suggestions. As suggested, we have collaborated with an additional co-author with expertise in paleoclimatology. We believe that this collaboration has significantly improved the paleoclimate interpretations, terminology, and discussions throughout the manuscript. In the revised version, we have strengthened the paleoclimate motivation and expanded the discussion of our findings' implications across the introduction, data description, discussion, and conclusion sections. While a portion of the paper remains focused on methodological development and its broader applicability, as encouraged by Reviewer 1, the paleoclimatic context has been considerably expanded and deepened. We have also provided a more detailed comparison with other methods previously applied to the same record and discussed the similarities and differences with these established techniques.

Regarding the suggestion of applying the method to Greenland ice core data, we agree that this would require a major shift in focus and title. After careful consideration, we decided not to

pursue this in the current manuscript. Instead, we incorporated an additional analysis focused on the Icehouse period using the Westerhold dataset. This revealed breakpoints close to the Mid-Pleistocene Transition (MPT) and the onset of Northern Hemisphere Glaciation, further demonstrating the applicability of our method in shorter samples. Accordingly, we have updated the title to better reflect the broader scope, and to acknowledge the methodological origin. We have also introduced the name "Bai-Perron" for the framework to ease future referencing.

2 Referee 1:

General Comments:

"This paper is a useful contribution to paleoclimate literature. The authors present the usage of a novel breakpoint analysis technique on a record of Cenozoic climate from Westerhold et al. (2020), and adequately support the usage of this technique on said data via sensitivity tests.

My major piece of feedback is that this paper currently lacks substantive earth science/ paleoclimate motivation, and is far too terse. While it's an impressive piece of work on the application of this changepoint technique to CENOGRID, additional context is necessary for broader application in the paleosciences. That is, as a methods paper that is designed to introduce a new approach to changepoint analysis within the paleosciences, further work is necessary to show where and when this method can be used to answer other paleoclimate questions. For example, some discussion of age uncertainty is essential, as this is an issue of fundamental importance in paleoclimate. Additionally, CENOGRID is an interesting dataset, but its length and completeness aren't typical of paleoclimate records, which complicates its being the sole non-synthetic example used in demonstrating the application of a novel technique. Additional explanation of the choice of CENOGRID as well as potential edge cases not covered by CENOGRID needs to be done. As a data scientist I'm left feeling confident that this technique is useful, and I'm intrigued by the idea of applying this method to my own work. However, as an earth scientist I'm left not quite understanding the breadth of problems that it is well suited for, nor what the significance of the additional breakpoints that were identified in the Cenozoic is."

We have strengthened the paleoclimate motivation and provided broader context for the methodology throughout the paper. Specifically:

- We expanded the paleoclimate motivation and terminology across the manuscript.
- We justified the choice of the CENOGRID dataset and discussed its limitations in more detail (Introduction and Section 3.3).
- We added an additional analysis focusing on the Icehouse period to demonstrate broader applicability (Section 3.2).

- We expanded the conclusion to outline potential future applications.

Specific Comments:

“The need for this method in particular within paleoclimatology should be discussed in more detail. This paragraph ‘Our approach contributes to the existing breakpoint detection methods in paleoclimate research by applying well-established econometric tools in the time-domain, developed in Bai and Perron (1998, 2003), to identify climate states in the paleo record. It enables the estimation of multiple breakpoints along with confidence intervals and provides procedures to estimate the number of breakpoints’ should be built upon. I see how confidence intervals might be useful, but what are the other strengths and weaknesses of this approach when compared to other methods? Some discussion of other approaches is offered in the preceding paragraph, but it is somewhat superficial. As a reader, I need better context for breakpoint analysis in paleoclimate studies: its historical usage, current applications, and future potential.”

We expanded the discussion of our method’s advantages and positioning within existing breakpoint detection methods:

- We include a concise overview of breakpoint analysis in paleoclimatology (Introduction).
- We have built on the mentioned sentence and related our approach to established ones more clearly. (Introduction)
- We provide clear explanation of the strengths and limitations of the Bai-Perron framework (Introduction, Discussion, Conclusion).
- We put emphasis on the flexibility of the framework, which for instance allows for including explanatory variables (Conclusion).
- We have added a section dedicated to discussing the limitations of our framework including modelling assumptions, computational requirements, age uncertainty, and irregular spacing challenges. (Section 3.3)

“This sentence ‘The paleoclimate variable $\delta^{18}\text{O}$ measures the ratio of ^{18}O to ^{16}O in the shells of benthic foraminifera obtained from ocean sediment cores, relative to a standard sample.’ is incorrect (or at least misleading), as $\delta^{18}\text{O}$ is not exclusive to benthic forams, which the sentence seems to be suggesting.”

We thank the reviewer for spotting this. We have corrected the statement to ensure accuracy.

“This sentence ‘The weight difference between the oxygen isotopes leads to an inverse relationship between $\delta^{18}\text{O}$ and ocean temperatures; see for instance Epstein et al. (1951) and Shackleton (1967).’ is an inadequate description of benthic $\delta^{18}\text{O}$. Mention of other factors (seawater composition, ice

sheet volume, etc.) needs to be included. CENOGRID is composed of many integrated signals, the makeup of which will determine what detected breakpoints are telling us about the climate.”

We have revised the description of benthic $\delta^{18}\text{O}$ to include additional factors such as seawater composition and ice volume effects. Also, we have clarified that when we write $\delta^{18}\text{O}$, we are referring to benthic $\delta^{18}\text{O}$.

“As the reader, I’m left wondering why only oxygen isotopes were considered. Carbon isotopes are also available, why not include carbon isotopes in the analysis, as was done in the original Westerhold publication?”

We have clarified that our method can be applied to other proxies (Conclusion), including $\delta^{13}\text{C}$. While we focused on Westerhold et al. (2020) $\delta^{18}\text{O}$ time series in this study, we have included an additional analysis focusing only on the Icehouse period and this demonstrates the broader applicability of our method in shorter samples.

“The authors discuss the varying resolution of the time series at length, which is helpful. However, I would be curious as to whether or not the resolution impacted the detection of break points. Is there any correlation between the detection of new breakpoints (discussed later in the manuscript) and the resolution of the time series? For example, just by visual comparison, it seems to me that the breakpoint observed in Coolhouse 1 in later sections might be related to a large change in resolution that occurs nearby. While this breakpoint isn’t heavily interpreted here, this is an important point to understand if other researchers are to apply this method to their own data.”

The resolution issue essentially reflects a bias-variance trade-off, where a higher binning frequency (i.e., higher resolution) reduces variance of the breakpoint estimates but may introduce some bias, whereas a lower binning frequency is more likely to increase variance but will be less prone to bias. This means that the primary effect of changing the resolution is on the constructed confidence intervals rather than on the location of the breakpoints themselves. This is evident in Figure 4 of the new manuscript, where a lower binning frequency generally leads to wider confidence intervals. Additionally, we observe strong stability in the estimated breakpoints across different binning frequencies, suggesting that resolution does not systematically impact breakpoint detection.

Regarding the estimated breakpoint in Coolhouse 1, it is indeed correct that this breakpoint is fairly close to a change in the resolution of the data. We agree that binned data reflect the quality of the original dataset, which can make it more challenging to accurately estimate breakpoints near or within data gaps. However, from a visual perspective, the breakpoint in Coolhouse 1 could also be influenced by the hump-shaped pattern that follows shortly after, between 17 and 14 million years ago.

“This sentence ‘Furthermore, we recommend using binning frequencies 10 and 25 kyr as they re-

sult in the most consistent outcomes.’ *strikes me as rigid and somewhat unhelpful, as many records will not share the general time axis properties of CENOGRID. Is there a different way to describe your binning recommendation that’s more flexible and/or applicable to other datasets?”*

We acknowledge the referee’s concern and have clarified that the choice of binning frequency should be tailored to the characteristics of the time series at hand. We have recommended selecting a frequency that maintains sufficient observations per bin while considering the dataset’s length and resolution, and also quantified these for our application to make it more concrete. We have rephrased the sentence to make clear that our recommendation only refers to the CENOGRID benthic $\delta^{18}\text{O}$ time series.

Certain technical choices need to be better explained given the audience of this journal. For example: ‘To address these issues, we use the autocorrelation and heteroscedasticity consistent (HAC) covariance matrix estimator with prewhitening in our implementations.’. Perhaps this is standard fare in breakpoint analysis literature, but most paleoclimatologists won’t be familiar with this procedure. Some explanation as to why this approach is suitable for this data is warranted here, as is mention of alternatives that were considered. In the same vein, it would be helpful to spend a little bit more time explaining information criteria. That is, expand upon “We use information criteria to estimate the number of breakpoints”. What does this mean, why are they used, have they been used in paleoclimate contexts before, etc.”

We appreciate the referee’s feedback and have made these methodological choices more accessible to a paleoclimate audience. In the revised manuscript, we have expanded on the rationale for using the HAC covariance matrix estimator with prewhitening, explaining its role in addressing autocorrelation and heteroscedasticity, which are common in paleoclimate time series. We have also briefly discussed why our chosen approach is particularly suited for this application. The presence of autocorrelation and heteroscedasticity in paleoclimate time series has been considered in previous work focusing on ice core records, including Davidson et al. (2015) who considers heteroscedasticity, and Keyes et al. (2023) who investigate autocorrelation. Our manuscript focuses on paleoclimate records from ocean sediment cores, where similar challenges arise.

Furthermore, we have clarified the use of information criteria for selecting the number of breakpoint. While their use in paleoclimate research is still limited, studies like Valler et al. (2024) show they can be useful.

“Age uncertainty needs to be addressed somewhere in this paper. It doesn’t need a full treatment, in that the method doesn’t need to be modified to account for it, nor does it need to be included in the analysis, but discussion of how to include it in future studies is essential. Specifically I would be curious as to how this technique might be expanded to include the usage of age ensembles and how

choice of age modeling method might impact breakpoint detection. However, discussion of including age uncertainty directly would also be acceptable here.”

We have included a brief discussion on age uncertainty and age ensembles (Section 3.3). While this is an important issue, we are not aware of any methodology in the paleoclimate literature that explicitly treats a timestamp as a random variable, with its variance representing the uncertainty of the timestamp. This challenge has also not been addressed within our framework, and incorporating the time-stamp uncertainty would require significant methodological developments. Such an extension involves advanced statistical techniques that are beyond the scope of this study. We have stated this clearly in the revised manuscript.

That being said, addressing age uncertainty is a crucial direction for future research. We have highlighted this in the revised manuscript and referenced relevant studies that discuss the issue (Telford et al., 2004; Franke and Donner, 2019) as well as a paper that explores aspects of transition detection in the presence of age uncertainty (Goswami et al., 2018). (Section 3.3)

The simulation study is a particular strength of this work. The authors thoroughly test their method across different data-generating processes, demonstrating its robustness to various forms of non-stationarity and serial correlation. This kind of rigorous testing is essential for establishing the reliability of statistical methods in paleoclimate contexts. I just wanted to make a note of that.

We appreciate this positive feedback.

“When analyzing the possible presence of multiple breakpoints, I’m left desiring some kind of prescription as to how I should set the number of breakpoints. Certainly the claim that there are more than 5 statistically significant breakpoints in CENOGRID seems robust. However, the current analysis feels somewhat hand-wavy, with seven breakpoints being settled upon in a rather arbitrary way. In particular this statement needs to be expounded upon: ‘The estimation results based on information criteria justify dividing the climate states Warmhouse II and Coolhouse II into two substates each at approximately 39.7 Ma and 10 Ma, respectively. This is supported by the presence of breakpoints estimated approximately at these timestamps in the estimations with seven or more breakpoints.’.”

We recognize the need for a clearer justification of the chosen number of breakpoints. As the results are somewhat ambiguous, we emphasize that the information criteria serve as a guidance tool rather than a strict statistical test for selecting the number of breakpoints. In the revised manuscript, we have provided a more detailed explanation of how information criteria inform breakpoint selection and have reframed the discussion to reflect their role in guiding, rather than testing, the choice of breakpoints (Section 3.2).

“The ending of this manuscript is far too abrupt. Potentially new breakpoints are discovered when

varying numbers of breakpoints are allowed, but what do they mean? A few climate events are referenced, but events themselves may or may not justify entirely new regimes. Much context is needed here, interpreting and explaining the presence of these novel breakpoints. While the authors are free to choose how to address this comment, I might suggest including a “Discussion” section, in which the primary results are emphasized, and an explanation/interpretation of these results is offered. Some of my other comments could probably be folded into this section as well.”

We appreciate this suggestion and have included a new Discussion section (Section 4). Here, we discuss the paleoclimate implications of the estimated breakpoints and relate them to relevant literature. We have also softened the language around breakpoints to allow for the identification of both climate events and climate state transitions. While we do not claim to define entirely new climate regimes within the Cenozoic Era, our results statistically indicate that these time points mark shifts in the underlying dynamics of the time series, distinguishing them from other periods. Furthermore, we contextualize our findings within existing literature and provide insights for other researchers working with similar data, ensuring that our methodological contributions are framed within a broader paleoclimate context (Conclusion).

Technical Comments:

“The paper currently is a bit undercited. I suggest the authors go back through with a fine toothed comb and make sure they’re citing existing literature wherever possible. In particular, all sections discussing $\delta^{18}O$ interpretation should be thoroughly cited, particularly regarding ice volume effects, temperature relationships, etc. A couple of other key spots (non-exhaustive) that need citations include:

- ‘The climatic transitions contain important information about variations in Earth’s climate system’ (here Tierney et al. 2020 is referenced, but some explanation of what is contained in that review along with additional citations is called for)
- ‘Our approach contributes to the existing breakpoint detection methods in paleoclimate research’ (cite breakpoint analysis in paleoclimate literature)’
- ‘This breakpoint aligns with the Middle Eocene Climatic Optimum, a known climatic event’ (cite original papers describing this event)
- ‘Some of these breakpoints coincide with other climatic events, for instance, the Latest Danian Event at 62.2 Ma and the onset of the Miocene Climatic Optimum at 16.9 Ma’ (cite original papers describing these events, not just Westerhold 2020)

We thank the referee for this detailed and helpful feedback. We have carefully reviewed the manuscript to ensure that all relevant literature is appropriately cited. In particular, we have strengthened the

citations in Section 2.1 discussing $\delta^{18}\text{O}$ interpretation, now referencing key studies on ice volume effects, temperature relationships, and other important factors.

Additionally, we have taken the following steps to address the referee's suggestions:

- We expanded the discussion of climatic transitions, providing a more detailed explanation of Tierney et al. (2020) and including additional citations highlighting the importance of Cenozoic climate states (Introduction).
- We contextualized our contribution to breakpoint detection in paleoclimate research by citing the reviews of Mudelsee et al. (2014) (benthic $\delta^{18}\text{O}$ time series analysis) and Marwan et al. (2021) (nonlinear time series analysis), along with relevant applications (Introduction).
- We ensured that references to specific climatic events, such as the Middle Eocene Climatic Optimum, and the Miocene Climatic Optimum, are now supported by citations to original research studies rather than relying solely on secondary sources (Discussion). ”

3 Referee 2:

“The authors apply a statistical approach to a reference paleoclimate dataset to evaluate shifts in climate states across the Cenozoic. They review existing methods and highlight the advantages of the breakpoint method and apply this approach on the Westerhold et al. 2020 dataset. They review the model and the impact of model parameterization on determined number of breakpoints. Ultimately, they established a similar number of breakpoints to Westerhold et al 2020 paper using this new approach and summarize their methodological approach and findings.

The application of new statistical approaches to assess time series and breakpoints is an important field of study but this contribution lacks significance beyond the application and usefulness of the method used. Although the paper is organized in structured manner, it is lacking background on other statistical assessments applied to assess Cenozoic or long-term changes in past climates and inclusion of a breath of paleoclimate specific references. Overall, this contribution mostly focuses on method development and application rather than implications of findings. Based on the above, it is difficult to see how it has led to a deeper understanding of Cenozoic climate.

I would suggest that this contribution needs to expands its discussion to outline how its findings enhanced our Cenozoic climate knowledge. Further there is a lack of background information needed to fully characterize the dataset evaluated, its attributes and limitations, and age model considerations. Below I have made some comments to develop the paper and give its finding wider significance.

General comments:

- *Introduction lacks information on Earth's history and significance of these various climate states. A review of existing methods is useful but does not link to wider significance of work and*

importance for paleoclimate field or mention how this analysis could provide new understanding or concepts for workers.

- *Justification for use of Westerhold et al. (2020) dataset needs to be developed further. More information on benthic foraminifera used to construct record, limitations associated with age model differences, and potential interpretation of $d_{18}O_b$ record from past work. There is limited mention of $d_{18}O_b$ and its use to pinpoint climate states and whether these states align when using other similar datasets or depending on inclusion or exclusion of individual/regional $d_{18}O_b$ records does it impact the breakpoint determinations.*
- *The contribution needs to consider its finding and whether it has led to a deeper knowledge of Cenozoic Science. For instance, what framework can the climate transitions of the Cenozoic be categorized and how is this related to dynamics of the climate system? This will allow for the novelty of the method application to be developed in more detail. A thorough review of other approaches used to assess climate change in Cenozoic would be useful to include, going beyond this method, including tipping point analysis and frequency analysis to showcase how this co-eval alongside of the breakpoints .”*

We thank Referee 2 for the helpful and constructive comments. We appreciate the recognition of the importance of applying new statistical approaches to paleoclimate time series and breakpoint detection. As the referee notes, the primary focus of our contribution is methodological, but in the revised version, we have made a serious attempt to expand the paleoclimate context and implications of our results.

To accomplish this, we have added a co-author with expertise in Cenozoic climate history, which we believe has significantly improved the discussions of paleoclimate implications. While the methodological emphasis remains a core focus of the paper, we have made substantial efforts to strengthen the paleoclimate relevance and how our findings led to deeper understanding of the Cenozoic climate.

We have also expanded the motivation for selecting the CENOGRID dataset as our case study, explaining its strengths and suitability for illustrating our method (Introduction). Potential limitations, such as age model differences, regional variations, and other dataset characteristics, are also discussed briefly in Section 3.3, with reference to Westerhold et al. (2020) for more detailed discussions.

In response to the referee’s specific comments, we have revised the manuscript in the following ways:

- We have expanded the Introduction to provide more background on Earth’s climatic history,

the significance of Cenozoic climate states, and the need for statistical methods to characterize climate transitions.

- We have made a serious effort to integrate our findings within the paleoclimate science literature by incorporating key references:
 - Burke et al. (2018) who discuss how Cenozoic climate states serve as analogs for future warming scenarios.
 - Caballero and Huber (2013) who address the concept of state-dependent climate sensitivity.
 - Reviews by Mudelsee et al. (2014) and Marwan et al. (2021) which cover other statistical approaches for detecting climate transitions in paleoclimate time series.
- We have added a "Discussion" section (Section 4) where we interpret and contextualize our primary results, emphasizing their implications for understanding shifts in the dynamics of Earth's climate system during the Cenozoic. We also provide practical insights for researchers working with similar datasets and discuss the broader relevance of our findings (Conclusion).
- We have included a concise discussion of age model considerations and the challenges associated with age uncertainty, referencing key studies (Telford et al., 2004; Franke and Donner, 2019). An example of a transition detection method that considers age uncertainty is also referenced (Goswami et al., 2018). Although explicitly incorporating age uncertainty is beyond the current scope, we have highlighted this as an important direction for future research. (Section 3.3)
- To further demonstrate the flexibility and applicability of our method, we have added an analysis focusing solely on the Icehouse period (3.3 Ma to present) (Section 3.2).

References

- Burke, K. D., Williams, J. W., Chandler, M. A., Haywood, A. M., Lunt, D. J., and Otto-Bliesner, B. L. (2018). Pliocene and eocene provide best analogs for near-future climates. *Proceedings of the National Academy of Sciences*, 115(52):13288–13293.
- Caballero, R. and Huber, M. (2013). State-dependent climate sensitivity in past warm climates and its implications for future climate projections. *Proceedings of the National Academy of Sciences*, 110(35):14162–14167.
- Davidson, J., Stephenson, D., and Turasie, A. (2015). Time series modeling of paleoclimate data. *Environmetrics*, 27:n/a–n/a.

- Franke, J. G. and Donner, R. V. (2019). Correlating paleoclimate time series: Sources of uncertainty and potential pitfalls. *Quaternary Science Reviews*, 212:69–79.
- Goswami, B., Boers, N., Rheinwalt, A., Marwan, N., Heitzig, J., Breitenbach, S., and Kurths, J. (2018). Abrupt transitions in time series with uncertainties. *Nature Communications*, 9(48).
- Keyes, N. D. B., Giorgini, L. T., and Wettlaufer, J. S. (2023). Stochastic paleoclimatology: Modeling the EPICA ice core climate records. *Chaos*, 33(9):093132. Special Collection: Theory-informed and Data-driven Approaches to Advance Climate Sciences.
- Marwan, N., Donges, J. F., Donner, R. V., and Eroglu, D. (2021). Nonlinear time series analysis of palaeoclimate proxy records. *Quaternary Science Reviews*, 274:107245.
- Mudelsee, M., Bickert, T., Lear, C. H., and Lohmann, G. (2014). Cenozoic climate changes: A review based on time series analysis of marine benthic $\delta^{18}\text{O}$ records. *Reviews of Geophysics*, 52(3):333–374.
- Telford, R., Heegaard, E., and Birks, H. (2004). All age–depth models are wrong: but how badly? *Quaternary Science Reviews*, 23(1-2):1–5.
- Tierney, J. E., Poulsen, C. J., Montañez, I. P., Bhattacharya, T., Feng, R., Ford, H. L., Hönisch, B., Inglis, G. N., Petersen, S. V., Sagoo, N., Tabor, C. R., Thirumalai, K., Zhu, J., Burls, N. J., Foster, G. L., Godd  ris, Y., Huber, B. T., Ivany, L. C., Turner, S. K., Lunt, D. J., McElwain, J. C., Mills, B. J. W., Otto-Bliesner, B. L., Ridgwell, A., and Zhang, Y. G. (2020). Past climates inform our future. *Science*, 370(6517):eaay3701.
- Valler, V., Franke, J., Brugnara, Y., Samakinwa, E., Hand, R., Lundstad, E., Burgdorf, A.-M., Lipfert, L., Friedman, A. R., and Br  nnimann, S. (2024). Mode-ra: a global monthly paleo-reanalysis of the modern era 1421 to 2008. *Scientific Data*, 11:36.
- Westerhold, T., Marwan, N., Drury, A. J., Liebrand, D., Agnini, C., Anagnostou, E., Barnet, J. S. K., Bohaty, S. M., Vleeschouwer, D. D., Florindo, F., Frederichs, T., Hodell, D. A., Holbourn, A. E., Kroon, D., Lauretano, V., Littler, K., Lourens, L. J., Lyle, M., P  like, H., R  hl, U., Tian, J., Wilkens, R. H., Wilson, P. A., and Zachos, J. C. (2020). An astronomically dated record of Earth’s climate and its predictability over the last 66 million years. *Science*, 369(6509):1383–1387.