Review of the paper entitled "Changes in the high latitude Southern Hemisphere through the Eocene-Oligocene Transition: a model-data comparison "by Alan T. Kennedy-Asser et al.

I. General comments:

There has been a long lasting controversy to assess the major driver of the EOT occurring 34 Ma ago between the opening of the Drake straight and a threshold on the CO2 decrease.

This paper aims to cope with the interesting issue of the comparison of complex coupled low resolution AOGCMs to a multi proxy data base for the period spanning from before the EOT (LE: late Eocene) to after this major event (EO: Early Oligocene).

It is an important issue because there is a gap between the world of data climatic reconstruction including error bars on dating and intensity of temperature changes and the world of the models which are equilibrium climate with prescribed conditions. Therefore building bridges between these two worlds is not an easy task.

This paper is a contribution to bring new constraints in this debate. The authors compiled various annual reconstructions derived from many different proxies over the High Latitudes for 3 periods spanning over several million years (14 sites: 4 terrestrial and 10 oceanic). On the other hand they establish metrics and benchmarking to validate two series of simulations run by low resolution AOGCMs. This is a first attempt to conduct such an approach. As such, it is a pioneer work which indeed depicts strong points and severe limitations.

The major results of this manuscript are

1 to establish the regional pattern of temperatures for high latitudes of southern hemisphere for the 3 key periods: late Eocene, Early Oligocene and during the transition.

2 to diagnose through a series of sensitivity simulations testing different boundary conditions, which are the most appropriate to fit the data. To achieve this goal the authors establish different matrices and benchmarking.

Despite showing new consistent results, there are some caveats in the manuscript.

1 concerning the discussion of the database

2 concerning the model results/data comparison and interpretations which are sometime confusing and unclear

In my detailed review, I will describe these strong and weak points. My general comment is that this manuscript is worth for publication because it is an interesting study, well documented with many supplementary information, but in is current state there is room for improvement. I suggest major modifications and clarifications that are described below.

II. Detailed review:

<u>Abstract</u>

The authors should specify that the database includes 14 sites (4 terrestrial and 10 oceanic) and that two low resolution models are used.

The sentence on the comparison of zonal mean temperature is a bit confusing. As far as I understand, the models are not reproducing the behavior depicted by the data during EOT. Moreover they underestimate the temperature over high latitude for both LE and EO.

The second part of the abstract is consistent with the results depicted in the manuscript.

The major results are not really summarized in the sentence concerning zonal temperature changes.

My reading is that there are clear underestimations of model results for both periods in terms of absolute value but a better agreement when using the relative metric.

The change of latitudinal slope of temperatures pinpointed by the data is not reproduced by both models

1. Introduction

The reference to Lunt et al., 216 is perfectly appropriate, but, as the statement is very general, other references could also be included.

During broad time slices as LE (36.4-34 Ma) and EO (33.2-32 Ma), there are many orbital cycles that produce by themselves a large uncertainty (see PLIOMIP1) when comparing to data. Especially after the EOT when AIS (Antarctic Ice Sheet) is also modulated by astronomical forcing factors. This point has to be better described and discussed.

Moreover the series of simulation have been conducted differently especially concerning the cryosphere, which should be discussed in more detail.

2. Method
2.1. Data synthesis

2.1 data synthesis

Concerning the establishment of the specific database, used for benchmarking, of annual temperature at high latitude of Southern Hemisphere for LE and ETO, these reconstructions are derived from many different proxies and therefore

1. Not sure that these different proxies measure the annual temperature. Because temperatures are reconstructed using various proxies, certainly calibration issues should be discussed.
2. These reconstructions include a contribution of the high frequency (astronomical frequencies at the scale of tens of kyr). But this part is not accounted for in the models which are run with prescribed orbital forcing factors.
3. A good illustration of the difficulty of such a comparison is the time to establish a reliable multiproxy data base useful for LGM as MARGO. This effort of a large community took some time due to the necessity to calibrate the different proxies and intercompare them in a consistent chronological framework. Moreover, similarly to PLIOMIP1, the data/model comparison spanning for this stage over 300ky, suffers from variability due to changes of orbital forcing factors which inhibits accurate model/data comparison.

Therefore, the authors should discuss more clearly these limitations shown in fig1 and discuss how to improve this first attempt.

2.2 Model simulations

The interest of the manuscript is to use a series of sensitivity experiments on the different boundary conditions (ice sheet, Drake opening/closing, CO2). Indeed CO2 and AIS extent are not independent and the impact of each factor is maybe not cumulative, these 2 aspects should be discussed.

2.3 Metrix

Why to establish these absolute and normalized metrics? The authors should clarify this question.

For warm periods, models are generally unable to reproduce the zonal temperature gradient and very frequently, models largely underestimate the temperature reconstructions based on different proxy data at high latitudes.

This is also illustrated in this paper for both periods, therefore another normalized metric which is appropriate to avoid this problem is used in this manuscript which is fine for me but this should be clarified.

Whatever the metric used, the spatial average for each data location over 9-grid box represents a large region around 1000x1000km2 which is anyway huge and should be discussed.

2.5 Benchmarking

Introducing different matrices and developing a benchmarking is appropriate to provide a quantified model-data comparison. What is the benchmarking of both models for present day simulations? It would be interesting to know if some biases may explain the response of both models for deep time simulations.

3 Results

Fig3 and section 3.1 remain unclear. I don't understand whether there is a real robust quantification of the "qualitative agreement" invoked by the authors.

Late Eocene and Early Oligocene.

These sections and Fig4 summarize well the series of model simulations and the comparison with the database through different metrics and benchmarking. Nevertheless, cold biases due to the inappropriate capability of the models to reproduce warm conditions during warm climate are clearly illustrated using absolute metric, but not discussed in terms of consequences. For instance, the large extent of the AIS contributes to amplify the initial cold bias. Therefore, the ice sheet development is not favored in EOT context. Similarly, open Drake is also favored to compensate the original cooling bias. The authors use normalized metric to increase the score and diagnose the best boundary conditions.
The influence of the Drake straight opening, the pCO2 drawdown and the sensitivity to the extension of the ice sheet are expected. More discussion on the fact that the sensitivity of both models to opening of Drake and to the AIS reconstructions should be given, especially the AIS asynchronously

computed in Foam is smaller for early Oligocene. This has for consequence to produce less cooling and due to a cold bias it fits better with data. Therefore, the results have to be discussed not only in terms of benchmarking but more generally accounting for specific bias of each model.

EOT

The transition from late Eocene to early Oligocene is particularly interesting to analyze through all pairs of sensitivity experiments, either to diagnose better scenarios or to disregard some of them. In this section, because we deal with climate variations, the standard metrix has not the same drawbacks than previously. An interesting point of this contrasted section if to show that the poorest correlation is obtained for the Drake passage opening.

4 Discussion

This transition seems driven by many factors. Is it reasonable to modify the pCO2 to get a better fit to data? There is still large room for improvement to correctly simulate reasonable zonal thermal gradient for Late Eocene and early Oligocene which are necessary conditions to understand EOT. Moreover more information quantitatively depicting how different are the equilibriums for both model simulations is needed. What are the criteria used for the spin up phase? Which consequences it may have on the results provided in this study?

5 Conclusion

The build-up of this data set of high latitude southern hemisphere annual temperatures is a first attempt to validate model results of the EOT. Nevertheless these 14 points correspond to reconstructions with different proxies. Moreover, coarse resolution models are used, the spatial averaging includes 9 grid boxes which correspond to a square of around 1000 km2 side. Both features limit the model data comparison.

Nevertheless, these results bring some credit to the respective contributions of AIS development, CO2 decrease and Drake opening to the EOT and conclude to a minor role played by this latter. This conclusion is therefore interesting but remains not completely convincing due to many uncertainties that the authors describe.

The constraints on the development of the AIS when coupling climate and cryosphere during the EOT are also a promising way to capture the evolution of a whole system in which the AIS is active.

Increasing model resolution, coupling GCM and ice sheet models, increasing the number of data through EOT, better accounting for annual temperature variability induced by astronomical forcing factors, all these features are important ways to continue this pioneering study.

Figures

On fig1, for sites that benefited from several temperature reconstructions as Ross or Falklands Plateau, the spread is large which shows that a future work of intercalibration and synthesis like MARGO remains necessary. From fig. 1, it is clear that there is a tendency to lower temperatures for Oligocene, however, the values for both periods are compatible within the error bars. Indeed in fig.2, most of the values for the temperature differences are compatible with zero. However, there is a clearly negative value for U^{K}_{37} , but it is not clear why the error bar for this value is so small.

figure 3: caution it is not North but South on the axis title.

Despite the large uncertainty of the models, the models tend to underestimate the high latitude temperatures (fig. 3 a and b), and the effect is larger for Oligocene. Hence, the models predict in overall a larger cooling at high latitude than the data (fig 3c). The surprising behavior depicted in the text (larger cooling over mid latitude than for high latitudes) is indeed not reproduced by the model. However, the signal for such a latitudinal trend seems very weak and considering the model uncertainty and the error bars on the data, the model results seem to be compatible with each data point.