REVIEW: Lunt et al. "A model-data comparison…"

Scientific Significance
This paper is a major contribution to paleoclimate research. The authors are to be commended for coordinating efforts to model early Eocene climate and producing a remarkable coherent ensemble model that compares well with the compiled proxy data, at least for the upper end of CO_2 simulations. This paper will be an important reference for future research on greenhouse climates of the past.

Scientific Quality

Modelling approach: The approach to modelling is valid and differs from other multi-model comparisons by comparing five models that were developed independently with differing boundary conditions, which has the added benefit (as noted by the authors) of widening the range of climate predictions.

Proxy compilation: My concerns about how the proxy data have been compiled are outlined below. My overall concern is that the strategies taken tend to result in cooling down the "too warm" SST proxies (by choice of time-slice and inclusion of suspect d18O records) and warming up the "too cool" MAT proxies (by using a "provisional" LMA calibration), thus effecting a better but possibly spurious data-model fit. This is particularly evident in treatment of high-latitude datasets.

Time slice: I do not follow the logic of the argument that the approach is consistent with the EoMIP simulations. The latter have been run according to a specific set of CO_2 levels, resulting in a separate climate prediction for each CO_2 level. However, the authors decide to be surprisingly non-discriminating for the proxy data. A cursory examination of the early Eocene climate record (e.g. Zachos et al. 2008; Bijl et al. 2009, Cramer et al. 2009) reveals that there are 3-4 distinct climate intervals: the PETM, the EECO, the interval between the two, in which additional hyperthermals may or may not be present, and the interval above the EECO, which is distinctly cooler than the EECO in some records (Zachos et al. 2008; Cramer et al., 2009) but not much different from the EECO in others (e.g. Pearson et al. 2007; Bijl et al. 2009; Hollis et al. 2009).

Thus, there are two distinct climate states included in the proxy data sets: the cooler climate of "background" early Eocene and the warm climate of the PETM and EECO (+/- hyperthermals). In the modelling sense, these may be likened to lower and higher CO_2 simulations. However, these two climate states are jumbled together in the proxy data set. There are EECO only records (Hollis et al., 2009), entire early Eocene records (e.g. Bijl et al. 2009), records from an unknown part of the early Eocene (most of the terrestrial records), middle Eocene records (at least 6 in the Huber and Caballero data set, probably also Seymour Island - see below), and latest Paleocene records ("pre-PETM"). The PETM is explicitly excluded from the proxy data for reasons that are not adequately explained. Yes, it is a transitory warming event but the maximum temperatures are not significantly greater than those recorded for the EECO at sites where both events are documented. In this exercise, the authors are not concerned with the process that generates the temperatures but simply the resulting temperature, i.e. how warm can it get at specific locations around the globe.

This compilation of proxy data for differing time slices introduces spurious proxy-proxy mismatches and obscures an issue that is central to the proxy-model debate: can models and proxies be reconciled for the warmest period of the Cenozoic.

For example, using TEX_{86}^{L}, I derive the following median values for ODP Site 1172 (Bijl et al. 2009): pre-PETM = 22°C, PETM = 25°, pre-EECO = 23°, EECO = 26°, all early Eocene including the PETM = 25°, same excluding the PETM = 25°. If only pre-PETM data were available for this site, early Eocene SST would be underestimated by 3°C. Therefore, the use of pre-PETM values to improve geographic coverage results in a spurious cool bias to the proxy data for 5/13 (38%) of the marine sites. I recommend that for better consistency with the rest of the data, PETM values should be included for the five sites in question. Indeed, I would consider including all PETM records to improve geographic coverage.

This issue is particularly evident when comparing the New Zealand data to other records. As the New Zealand data is limited to the EECO, there is automatically a warm bias when compared to non-EECO records, such as the five pre-PETM records and Seymour Island (which is now thought to be of middle Eocene age, A. Sluijs pers. comm.)

<u>Marine SST proxies</u>: I question the inclusion of data from ODP sites 690 and 738. Given increasing evidence for diagenetic alteration of planktic tests recovered from deep sea cores (Sexton et al. 2006; Pearson et al. 2007), caution should be exercised in using planktic $\delta^{18}O$ as a guide to SST in southern high latitudes unless the planktic tests from these sites can be shown to be “glassy”. Hollis et al. (2009) highlighted this problem for DSDP site 277 (Campbell Plateau) where the lack of a significant thermal gradient between planktic and benthic $\delta^{18}O$ during the early Paleogene makes the planktic $\delta^{18}O$ record questionable. Very low gradients (or none or reverse) are also evident from benthic and planktic $\delta^{18}O$ data from sites 690 and 738 (Stott et al. 1990; Barrera and Huber 1991). A cool bias caused by seafloor diagenesis at these high latitude sites is also supported by Liu et al. (2009) who used TEX_{86} and $U^{k'}_{37}$ to derive SSTs that were ~20°C warmer than the SSTs derived from planktic $\delta^{18}O$ at Site 277. I understand that the argument to use $\delta^{18}O$ records from high latitude sites may be that greater vertical mixing at high latitudes should result in reduced offset between surface and seafloor temperatures (although I have yet to see where this is stated). I contend that this argument is a little presumptuous when considering the early Eocene ocean.

Thus, the repeated reference to anomalously warm “New Zealand” data needs to be corrected for two reasons: (1) it is the entire Southwest Pacific dataset (ODP 1172, DSDP 277 and Waipara River) that is warmer than the models predict not just the single New Zealand record, (2) too little reliable marine data are available from other high latitude sites to consider these data anomalous. Similarly warm temperatures have now been recorded from the Antarctic margin (Bijl et al. 2011), the local terrestrial record (Carpenter et al. 2011) and the Mesozoic Southern Ocean (Jenkyns et al. 2012). The Arctic SST record (Sluijs et al. 2006, 2009) is also consistent with these Southern Ocean records. I respectfully suggest that the modellers are making life too easy for themselves by not recognising that Seymour Island may be the anomaly!

I have further concerns with the way the marine proxy data have been manipulated:

Mg/Ca: I can accept the strategy taken of averaging values based on low and high estimates for Mg/Ca[sw] except that (1) I can find nothing in Lear et al. (2002) to suggests an Mg/Ca[sw] value of 3.5 for the early Eocene (Table 7 suggests a range of 3.3 to 4.6 is possible based on the poorly calibrated *Oridorsalis*); (2) Sexton et al. (2006) show why an early Eocene value of <5 is hard to reconcile with $\delta^{18}O$, and (3) there is enough uncertainty surrounding the planktic foraminiferal sensitivity to Mg/Ca[sw] and carbonate ion saturation state (Lear et al. 2010; Cramer et al. 2011) that it seems warranted to use a Mg/Ca[sw] value that allows Mg/Ca and $\delta^{18}O$ records to be reconciled (i.e. as in Sexton et al. 2006). For this reason, I would recommend using a single higher value such as 4 or a range of 4-5. Note that Lear et al. (2002) conclude that early Eocene Mg/Ca[sw] was >3.35 [65% modern], so there seems no basis for considering a value lower than this. A higher Mg/Ca[sw] value results in cooler SSTs – helping to reconcile Mg/Ca and TEX86L SSTs in the SW Pacific but potentially cooling down the equatorial Pacific (Site 865) too much for the high CO2 models.

$\delta_{18}O$: For these calculations I am mystified by the use of two approaches to estimate d18O[sw]. Why use only two, when you could use three, i.e. Roberts et al. (2011)? I would prefer to see site-specific comparisons of these three approaches and a case made for using the preferred approach.

TEX_{86}: I am completely mystified by the approach taken here, which is to average the values derived from three calibrations. There are two serious problems with this approach:

1. TEX_{86}^{L} is recommended for records where SST is likely to be below 15°C (Kim et al. 2010) and therefore should not be applied to any of the records included here without careful consideration. Its inclusion in low-latitude records in particular introduces a spurious cool bias. There has been some discussion about treating this proxy as a high-latitude proxy rather than a low-temperature proxy (Hollis et al. 2010, 2011, MS; Bijl et al. 2011), so there is justification for utilising the proxy in high latitude records with due acknowledgement of this discussion

2. In most records included here, there is very little difference between TEX_{86}^{H} and $1/TEX_{86}$ (the reciprocal equation of Liu et al., 2009). Therefore, an average of the three proxies introduces a warm bias. My recommendation would be to use an average of TEX_{86}^{H} and $1/TEX_{86}$ for low latitude records and average of TEX_{86}^{H} and TEX_{86}^{L} for high-latitude records.

<u>Terrestrial proxies</u>: There are many questions relating to how these data have been compiled that were not addressed by Huber and Caballero (2011) and need to be considered now that the data are being uncritically transferred to this paper. Why are middle Eocene records included? How can use of a "provisional" LMA calibration, based on only 10 reference samples, be justified (Kowalski and Dilcher 2003) when the only point in its favour appears to be that it's a better match for the models. Why is CLAMP not used for MAT determinations where available when Huber & Caballero (2011) state it is used in most cases to generate CMM? Methodologies used to calculate errors, to correct for paleo-altitude, to estimate % entire margins and to derive average values from multiple assemblages are not adequately explained. A more

detailed critique of this dataset will follow. It is also important to note that in contrast to the marine data, PETM records appear to be included in this dataset.

Additional minor points

1. [2.3] CCSM3_H was developed for Hollis et al. (2009) not for Liu et al. (2009), which has a late Eocene model.
2. [3.1.1] Bemis et al. (1998) showed that the Erez and Luz (1983) equation suffers from a warm bias of up to 3.5°C and introduced alternative equations, which give values very similar to Kim and O'Neil (1997). Why were these equations not used at least as an alternative to Erez and Luz. I cannot find reference to a standard error of 1.43°C in the latter paper, on p. 1028 they list sources of error that total 2.15°C.
3. [3.1.3] Line 18 - "high" and "low" should be reversed. Line 27 – delete (TEX86 and TEX86L)? Line 3(p. 1239) – Hollis et al., 2012 not in refs (replace with Hollis et al., 2011 – see below). Line 11- The error on each of three proxies is different, 2.5 for TEX86H, 4 for TEX86L and 5 for 1/TEX86. Why is the minimum error of 2.5 used here?
4. [3.2] Give simple explanation of what LMA and CLAMP are – physiognomic analysis of leaf fossils … Note the second paragraph is out of place, should go at the end of [3] or the start of [4].
5. [Figure 1] Model labels overprinted
6. [Figure 4] "temporal uncertainty (black bar) and calibration uncertainty (grey bar)"
7. Global find and replace "New Zealand" with "southwest Pacific"!

Presentation Quality:
The results and conclusions are presented in a clear and well-structured way. The figures are all appropriate although in several cases they suffer from being too small to be legible on a standard screen and are certainly too small for A4 printing. It is especially hard to resolve the colours for proxy data in Figures 2 and 3.

The marine proxy table lacks references and would benefit from comments on individual records where the quality of data (ODP sites 690, 738; Hatchetigbee TEX_{86}) or the age of the record is questionable (Seymour Island). The terrestrial proxy table is superfluous as it is simply an unlabelled extract from the table in Huber and Caballero (2011)

Additional references

Bijl, P.K.; Bendle, J.; Pross, J.; Schouten, S.; Roehl, U.; Stickley, C.E.; Olney, M.; L., T.; Bohaty, S.M.; Brinkhuis, H.; Escutia, C.; Scientists, E. 2011. Integrated stratigraphy of the Eocene Wilkes Land Margin, Antarctica; preliminary results from IODP Expedition 318: dinoflagellate cysts and TEX_{86} results. *In:* Egger, H. (Editor), *Climate and Biota of the Early Paleogene (CBEP 2011)*, p 40. Geologische Bundesanstalt, Salzburg, Austria.

Carpenter, R.J.; Jordan, G.J.; Macphail, M.K.; Hill, R.S. 2012. Near-tropical Early Eocene terrestrial temperatures at the Australo-Antarctic margin, western Tasmania. *Geology 40*: 267-270.

Douglas, P.; Ivany, L.C.; Keating-Bitonti, C.R.; Pagani, M.; Affek, H.P. 2011. Eocene sea surface temperature reconstructions from bivalve clumped isotope measurements. *In:* Egger, H. (Editor), *CBEP 2011*, p 66. Geologische Bundesanstalt, Salzburg, Austria.

Hollis, C.J.; Taylor, K.W.R.; Handley, L.; Pancost, R.D.; Creech, J.; Baker, J.; Schouten, S.; Kennedy, E.M.; Crouch, E.M.; Huber, M.; Ackerley, D. 2010. Paleoclimate data-model comparisons for early Paleogene New Zealand, Abstract GC42C-04. *AGU Fall Meeting, 2010, 13-17 December*, San Francisco, CA.

Hollis, C.J.; Taylor, K.W.R.; Pancost, R.D.; Creech, J.B.; Kennedy, E.M.; Strong, C.P.; Morgans, H.E.G.; Crouch, E.M.; Neil, H.; Ackerley, D.; Huber, M. 2011. An update on paleoclimate data-model comparisons for the southwest Pacific. *Climate & Biota of the Early Paleogene, 5-8 June 2011, Salzburg, Austria : conference program and abstracts*. Geologische Bundesanstalt, Wien, Austria.

Sexton, P.F.; Wilson, P.A.; Norris, R.D. 2006. Testing the Cenozoic multisite composite $\delta^{18}O$ and $\delta^{13}C$ curves: New monospecific Eocene records from a single locality, Demerara Rise (Ocean Drilling Program Leg 207). *Paleoceanography 21*: PA2019.