Interactive comment on "Evaluation of PMIP2 and PMIP3 simulations of mid-Holocene climate in the Indo-Pacific, Australasian and Southern Ocean regions" by Duncan Ackerley et al.: Responses to Editorial Review – A. Lorrey

I openly declare ahead of this review that I know almost all of the authors of this paper personally. I've worked directly with many of them in the past on other published work and with some of them on previously funded projects. I have no personal interest conflicts with any of them.

An additional review was requested of me by CPD because we have been unfortunate to have only one solicited review for this paper. As such I submit this review here, but would like the authors to understand that final adjudication will be undertaken in consultation with the editors of CPD.

Response: Firstly the authors would like to thank Drew for undertaking the review of this manuscript, given the circumstances mentioned.

Overall, I thought this was a useful manuscript to put forward and it will be a welcome addition to the regional literature when completed. Attempts to reconcile Australian palaeoclimate data with simulations are not numerous. The region has a great number of complexities of climate drivers and processes that could be evaluated using model proxy comparisons.

Response: We agree that this region has largely been overlooked in global compilations and we hope this contribution goes some way to addressing this.

However, I have some concerns about how this paper has been presented though, or at least how a few things are couched, and with some of the visual content that is presented (or data not shown). I suggest it could be publishable in CP, but only if many minor and some major revisions were undertaken. Please see specific comments in the attached PDF.

Response: These are addressed individually below.

First, the balance of the paper strengths lean heavily toward the analysis and descriptions of the climate model simulations. I'm elated that some basic physics of the climate system has been brought to bear with the incorporation of the discussion on the latitudinal gradient and geostrophic wind equations. There are some very interesting findings here for the model results, but I also cannot determine if any bias corrections were actually undertaken for the circulation patterns or if there is simply reference to identifying them using the post 1750CE data sets. Please make this clearer.

Response: No bias corrections were undertaken on the circulation, precipitation or temperature patterns in the models. We are actually looking for those biases and therefore any correction applied to the model data would negate that. In reference to the cold tongue bias and the Southern Ocean temperature biases specifically, we use the post 1750 datasets to identify processes that have been described elsewhere in the literature (which has been cited); however, our study shows that:

1. The biases are present in the pre-industrial control runs (and not just the historical simulations, which are the focus of the cited literature).
2. The biases are present regardless of the time period under consideration, which is very clear when compared to the paleo-proxy reconstructions too.
3. Most important of all, the biases can actually be made worse in the 6 ka simulations as the changes in insolation appear to enhance those errors further.

We do state in section 2.3 that, "As neither of these datasets is representative of the climate at 1750 C.E. (as in the 0 ka simulations), they are only used to highlight known biases in the GCM simulations that may cause discrepancy relative to the proxies". The post 1750 CE datasets are therefore extra confirmation only, in order to draw attention to the processes that drive the well-known model biases. The authors therefore do not see how this can be made clearer than it already is in the paper.

Ahead of the following comments, I would also suggest the authors simply recast the use of the palaeo proxy data network (from Reeves et al., 2013a) as supporting field-based information that the models can be compared against (details why are stated below). Section 2.1 is also poorly written; it leads off with a description of INTIMATE and then Reeves et al 2013 - and it seems very odd to me that for a SHAPE issue that there is no mention of

that initiative anywhere, which has superseded INTIMATE in the Southern Hemisphere. I think it would be more contemporary to refine the aims of the model-proxy intercomparison in light of the stated goals of the SHAPE IFG - which are similarly stated on the SHAPE project website.

Response: The authors apologise for this oversight and this will be rectified in the revised manuscript with the following text:
"This research forms a contribution to the Southern Hemisphere Assessment of PalaeoEnvironment (SHAPE) program, an INQUA International Focus Group (project 1067P). SHAPE is the successor to the Aus-INTIMATE program, focussing on the Southern Hemisphere. One of the key remits of SHAPE is to develop model-proxy comparisons to both help understand the dynamic mechanisms behind regional palaeoclimate proxy-based reconstructions and test the robustness of the palaeoclimate models."

Second, there appear to be no real surprises to me in terms of the findings – the proxy data-model comparisons are elementary (mostly descriptive, but still very useful and clearly-written). They are divided into sections that essentially show where the proxy-model comparisons work, and where they don't. I would greatly appreciate if the listing of proxy data derived from Reeves et al., 2013 (fundamental to supporting this work) was tabulated, including all metadata about location, type of archive, dating controls, seasonal sensitivity and signature for climate during the 6k interval are stated.

Response: We agree that this would be a useful contribution and would be willing to undertake this. However, we would request an additional 3 months to allow someone to be employed to undertake this task. Much of this material is available in the regional papers, however having these accessible via database would be of benefit to the broader community.

The Reeves et al., 2013 paper is also mentioned as providing 'a method' but it does not do that in terms of integrating the data or providing a dynamical understanding of past variability or change with reference to a mean climate state. That particular work collected climate proxy records under certain criterion, and binned them into different geographic

regions for Australia. If you adopt the spatial division of Reeves et al., 2013, and the data series used there in, it would be best to simply say so.

Response: We will amend our text accordingly as the editor suggests.

There in, those geographic regions are somewhat arbitrarily ascribed; but I temper this comment by saying in reality there is good reason to have made those divisions. Just a bit more support and justification from modern climate studies that indicate there is a strong reason for the geographic divisions would go a long way to informing the readership.

I believe that information can be easily obtained, and cited in the revised work. And better recognition that the real strength of Reeves et al., 2013 is the pre-selected proxy data that are 'regionally-representative'.

Response: There is a substantial description of the modern regional climate provided in Reeves et al., 2013 and the regional-specific papers of the same volume. We would be reluctant to repeat too much of this, but can direct readers to this justification and the updated Köppen-Geiger classification system (Peel MC, Finlayson BL & McMahon TA (2007), Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633-1644), which is broadly comparable. The actual delineation of boundaries (using straight lines) is somewhat arbitrary, however it is the clustering of records with the regions that is more robust. As mentioned in the response to the other reviewer (and to be added to the text) these records were selected by many of the Australian palaeoclimatic community as being the most regionally representative.

Third, the Reeves et al., 2013 depictions did not compare the past climate change signals to a common modern interval, but rather assessed the direction of change from one time step to the next. This limits meaningful comparisons of the past patterns that are shown in Reeves et al., 2013 to the climate model simulations shown in this study.

I realise there were probably data limitations in Reeves et al., 2013 that sent those authors down such a path, but it was identified as problematic early on (in discussions in Aus-

INTIMATE). In this paper, it (and the pictures showing signals for different time slices) is advocated as 'presenting a new opportunity to integrate models with data'. At the risk of repeating myself, it does not: What it does is supply a series of pre-screened data and climate signals for the mid Holocene where assessing PMIP2 and 3 model signals may be undertaken. The authors have largely done this in a point-by-point fashion; if the data from Reeves et al. has been further transformed, it is not clear how it was done. Deeper understanding from data integration would have been more meaningful; so I feel justified in mentioning this specific point here.

Response: We acknowledge that the Reeves et al., 2013 method of comparing directions of change is not to everybody's taste. However, this approach was applied to help understand the 'on-ground' transitions in climate, particularly from a biogeographic perspective. This manuscript has gone back to those records and focussed in on the 6ka timeslice, in comparison to pre-industrial modern, where possible. We will make this statement more clearly in the revised manuscript. Furthermore, we also believe this does provide an opportunity to integrate models with the proxy data for two reasons:

1. It provides a way to diagnose model output relative to the proxy archives over a spatial scale they can resolve (i.e. many grid points are used for the spatial averaging).
2. It provides a regional overview of the state of the climate that *could* be used in data assimilation i.e. it may be more useful to integrate the regional-scale temperature anomalies rather than individual point measurements, for example. This is something that could (and probably should) be attempted in the future for proxy system modelling.

While we do see the editor's point about this not being a model-proxy integration method (which it is not), there is definitely a case for this. Nonetheless, we will change the words "integration" to "comparison" within the text, as that is more appropriate language.

Fourth, I would also strongly encourage the authors to submit the data from Reeves et al., 2013 along with this paper, or provide a supplement with stable URLs where the data may be obtained. Sub-issues related to the points of viewing and assessing those data are: a. mapping of proxy signals onto the PMIP simulation outputs shown in Figures 3, 4, and 5. b.

being able to observe the time series for each c. seeing how the 6k signatures compare to modern or pre industrial times.

Response: As mentioned above, we agree this would be an important contribution, but require time to compile this. We would like to request 3 months. Nevertheless, the authors also feel that this request is getting beyond the scope of the paper. Such a process should really be considered as a long-term investment by the SHAPE initiative beyond just the mid-Holocene (and this study). Furthermore, the work is intended to provide an important, modelling perspective on mid-Holocene climate over Australasia. This paper therefore, already provides a timely modelling-centric synthesis, which has been a significant omission in the literature thus far. It also directly points the way to areas that could easily become the focus of future work.

Fifth, the scaling of the proxy signals so that they are compatible with the GCM signals is still unclear to me. This relates to point number 3. In using a tercile-based evaluation system of the proxy data, one needs to create a distribution for the data, with reference to a common interval (also the same interval used in the control run for the model simulations), then establish what the thresholds are for the terciles to obtain meaningful signals (warm, wet, cold, dry etc.). That has not been clearly shown anywhere here ... and it cannot rely on antecedent work. Seeing the data and the new analysis are required for the descriptions of the proxies to be understood as factual.

Response: The spatial coverage of quantifiable, high-resolution palaeoclimate records across the Australian continent is sadly lacking. Whilst recent efforts in the palaeoclimate community are seeking to address this, we can currently only work with the records we have. However, the records included in this compilation are those that have been selected by many of the Australian paleoclimate community to be most representative of the respective regions discussed. As noted above, what is attempted here is to compare the relative climatic change as described by the model and proxy records between 6 ka and modern. Even at this coarse scale (warmer, colder; wetter, drier), meaningful comparisons between the proxies and models have been made and clearly point to areas (both geographical and scientific) that warrant more targeted research. We are happy to make the

metadata available in supplementary material for future analysis; however, comparisons with peer reviewed, antecedent work are important and necessary steps for **any** piece of research and has been undertaken very carefully (and in-depth) in this study. Such comparisons must therefore remain in the paper otherwise all context for the work will be lost.

Addressing the above comments, the more minor grammatical issues in the text, and recasting the paper toward the main strengths (modelling results and forcing mechanisms, supported by point data, rather than proxy-model intercomparison) would see this through. I'd also like to encourage the authors to evaluate their future work section and to try to be broader with regard to proxy development, chronology evaluation and integrative approaches that could help future efforts bring models and proxies together - please see if that can also be done in a more refined manuscript.

Response: As per the response to Reviewer 1, yes we will recast the future work section, particularly with regards chronology and uncertainty of the proxy-based reconstructions. However, the regional model-proxy comparison is hugely important to this. Comparing point-for-point model and proxy data is nonsensical, as the models cannot resolve the necessary processes that occur at such a fine scale. Also, the proxies are clearly representative of a wider area than the individual records represent—and on scales the models **can** resolve (i.e. regionally). For example, a model grid box may cover 22,500 km^2 (i.e. 150 km x 150 km) over a topographically varied region, e.g. the Southern Alps. Within that box, there could be many individual proxy records that are airflow direction dependent (e.g. leeward versus windward slopes). Comparing such varied data within a GCM is nonsensical and impractical, as parameterization does not explicitly resolve such complexities. By looking at regionally coherent signals over multiple reconstructions (as done here) we begin to build up a picture of the overall climate throughout a region that the GCMs may be able to represent. Only at this point is it worth comparing the two datasets and not on a point-by-point basis as suggested. The strength of this work is the regional comparison (as Ackerley et al., 2011, did for their New Zealand study) and absolutely needs to be maintained as such. We will make sure a statement outlining this point is included in any revised paper, as we believe it is very important.

Response to specific comments in manuscript (given as a supplement to A. Lorrey's review). Other minor comments will be addressed subject to the decision on whether we are asked to submit a revised manuscript.

Note: in general we accept the grammatical and stylistic suggestions put forward and are happy to incorporate most of them in an updated version of the paper.

P1l1 We will tone down the references to INTIMATE and restructure in view of the contribution to SHAPE – particularly at the outset of this abstract.

P2L16 Yes – we will redefine the Southern Ocean to be in keeping with the area in reference.

P2L18 – Regional divisions and temporal reference addressed above. We suggest modifying the text to read:

Over Australasia, the Maritime Continent and Southern Ocean, such an upscaling/downscaling approach has not yet been attempted to integrate proxy and model data; however, the synopsis of the OZ-INTIMATE initiative Reeves et al. (2013a) presents an opportunity to do so, which is explored here. In particular, there is the move beyond the descriptive and begin to test some of the dynamical mechanisms represented by the proxy response, providing a deeper understanding of some of the key drivers of change.

P2L24 – Addressed above

P2L30 - Addressed above

P4L7 – Addressed above – the 6k palaeodata is compared to the present

P4L10 – there are no watermelons – maybe "Granny Smiths vs. Gravensteins", but we understand the point being raised fully. We will make it clearer in the text that this paper is

comparing 6k to 0k, using the INTIMATE suite of papers, but not the same approach (that is trends through time). This contribution uses the SHAPE approach of comparing past conditions to modern conditions.

P6L12 – The 1870–1899 data were used, as they were the closest we could get to observational data from actual instruments. Furthermore, we wanted as much of the modern-day (and overall 20th Century) warming signal to be removed so as not to exacerbate the already existing cold tongue bias (with respect to the Tropical Pacific). We know the bias is there (abundantly clear in the literature) and we wanted to use some reference dataset to give it context, in this case we chose HadISST. We are reluctant to extend the averaging period as it could artificially make the biases appear to be worse than they are due to recent warming. Nevertheless, we will state in the paper that we used this time period to minimise the impact of climate change on the temperature record. N.B. when undertaking our proposed analysis with responses to P23L15 and P26L16 below, if it becomes apparent that we should mention the use of more modern SSTs then we will also include it in Section 2.3.

P9L17 - As per comment above, we agree that tabulating the source data and making it available through supplementary material would be a valuable contribution, if we can be permitted the extra time this would entail to compile. In addition, we can provide a simple annotation to the maps provided to indicate direction of change (see example Figure EX1 below). We will also re-structure the wording to make sure that it is consistent with the figure (and adjust the figure further if necessary—the Fig. EX1 below is just an initial draft for the purposes of this response).

P23L15 – This is a good point. While there were likely to have been ship observations across the Pacific during 1870 to 1899 (and land-based observations to infer the polarity of the Southern Oscillation Index and an inference of El Nino/La Nina), it is unlikely that such observations would have been taken over the Southern Ocean. We will therefore include the value for the modern (1980 to 2009) SSTs in this region and include the reasoning behind this here. Again, given the Southern Ocean cloud and temperature biases are large

(and discussed in-depth in the cited literature) it is unlikely to alter our conclusions but we fully agree that this is a necessary checkpoint here.

P26L16 – We completely see the "incompatibility" argument here between using the 1870 to 1899 HadISST dataset and the 1979 to 2008 ERA-Interim dataset. The reasons for this are:

1. To avoid overemphasising the cold tongue bias by including the climate change signal in the 1979 – 2008 period.
2. ERA-Interim (and therefore any measure of the circulation over the Pacific) does not run back to 1870 – 1899 and so it is our only observationally constrained estimate of the mean circulation.

It is important to note that we are not presenting a case for a new process, merely highlighting that something that is known as an important modelling problem is also seen in the mid-Holocene simulations. Furthermore, the error is made worse in the 6 ka simulations because of the error in the base state, which is clearly important for any 6 ka GCM study. That said, we would be willing to include a figure in the supplementary material showing the current Figure 11 with SST and circulation from the same reference time period, to validate the editor's point. We would then state in the text that using the modern period does not change the overall result.

P30L24 - We see that the use of a tool such as PICT could provide a positive way forward – but is outside the scope of this publication. The issues with MAT in Australia are particularly fraught with the uncertain and varied impact of a long human presence in the landscape, bringing into question some of the assumptions of climatic stationarity – even during the Holocene.

P32L9 - this has been addressed above and will be more clearly detailed in the text.

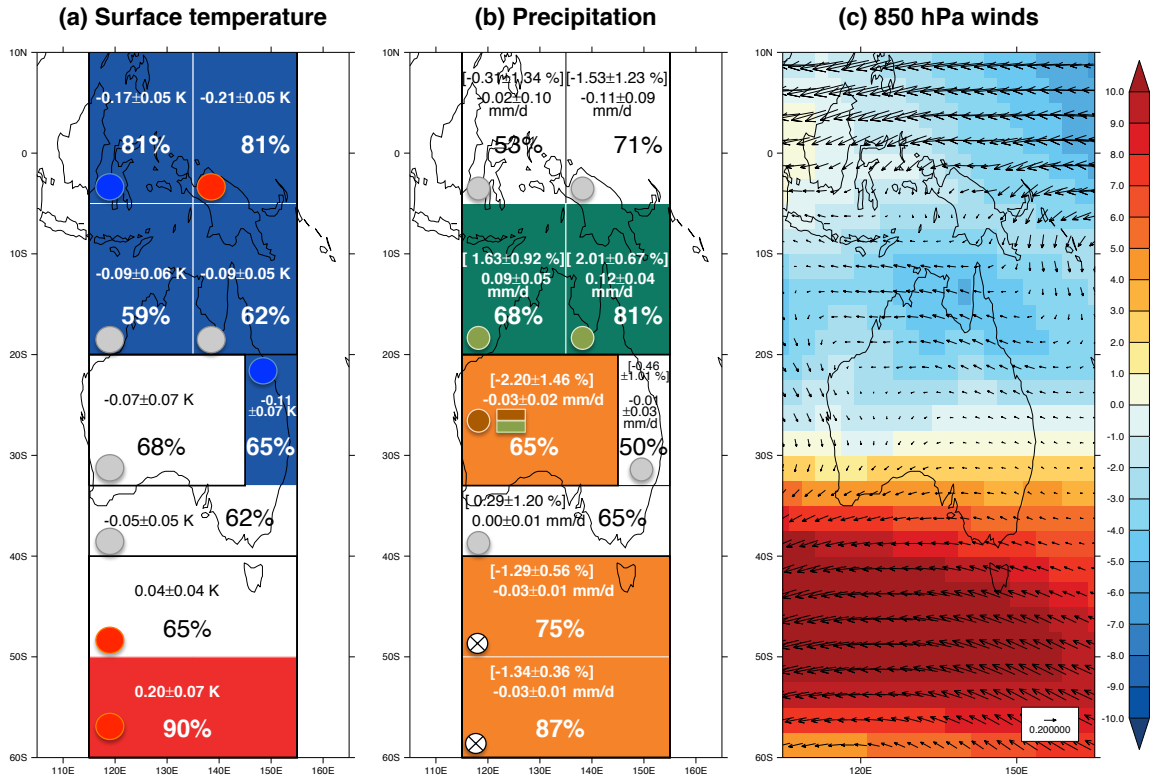

Figure EX1: The ensemble and regional annual mean differences in (a) surface temperature (K), (b) precipitation (mm day−1 and [%]) and (c) 850 hPa circulation (m s−1) for the 6 ka simulations relative to the 0 ka simulations. In (a) blue shading represents lower area-averaged surface temperature and red indicates higher at 6 ka. Red circles indicate higher, blue circles indicate lower and grey circles indicate little change for temperatures at 6 ka relative to 0 ka from proxy estimates. In (b), orange indicates lower area averaged precipitation and green indicates higher at 6 ka. In both (a) and (b) the values of the ensemble mean changes are given in white and the percentages of models that agree on the sign (positive or negative) of the ensemble mean temperature or precipitation differences are given by the white numbers. Amber circles indicate higher, green circles indicate lower and grey circles indicate little change for precipitation at 6 ka relative to 0 ka from proxy estimates. White circles with an X inside indicate not data available (both temperature and precipitation). In (c) shading indicates the direction and strength of the ensemble mean 850 hPa zonal wind (blue colours = easterly and red colours = westerly) in the 0 ka simulations. This figure is purely an example here and will be refined and checked for any revised version of the paper.