Reviewer 2
Main points

This study utilises the ensemble mean of 6 CMIP5 models that have the required set of experiments. The authors argue that, while they acknowledge each model is not free of biases, model bias is not a prohibitive issue for investigating the temporal stability of teleconnections [p. 1588 (25)]. I tend to agree with this argument, but it is still important to provide an indication to what extent the multi-model mean (MMM) represents the entire 6 samples. This is particularly necessary as model selection have not been extensively conducted, perhaps given the limited models available. Even one or two models could exhibit severe bias that may skew the MMM, especially in the climatology that can affect teleconnection patterns. A severe cold tongue bias for instance can spuriously shift rainfall teleconnection. In this case, rainfall in the western Pacific or Maritime Continent and the Nino3.4 could be positively correlated, while in reality or in more realistic models they should be negatively correlated. Averaging these teleconnections across models would result in a weak correlation as seems to be the case in Fig. 8. It would be a better approach to present the results for each model or present a confidence interval over each of the ensemble mean. For example, in Fig. 8 a confidence interval (or even better each model correlation value) should be added over the MMM.

- We have now included details of each model, as recommended. Furthermore, we now show each model correlation and statistical significance in Figures 7 and 8.

Li et al. 2013 (see their Fig. 2) found that their paleo proxies in the west Pacific (Maritime Continent) and east Pacific generally correlate quite well with Nino3.4, in contrast to those suggested by Fig. 8. It is necessary to comment in Section 4.2 the possible reasons for this mismatch (e.g., due to certain model biases, as per above points). Also, it would be good to put the results of Section 4.2 in the context of other existing studies.

- We have now discussed this key difference with Li et al 2013 and discussed our results in a broader context in section 6.

In light of the above comment, it is actually necessary for the readers to get a better sense in how each of the 6 models performs in terms of the ENSO characteristics and in terms of the mean climate. In section 3 (page 1587) it seems that the authors tried to do this: “Here, ENSO was examined through 6 metrics....” but there are no figures that show some of these metrics (e.g., seasonality, Nino3 vs Nino4 amplitude, etc). There are by now a number of studies evaluating the fidelity of ENSO simulation. In terms of climatological bias, at present the authors are comparing the multi-modelmean vs observations in Fig. 4 where climatological bias is severe. Perhaps one or two models are significantly contributing to the westward bias (e.g., IPSL-CM5A-LR, HadCM3, possibly the GISS-E2-R as well; see Taschetto et al. 2014 J. Climate, their Fig. 3c). IPSL-CM5A-LR for instance cannot simulate the nonlinear response of rainfall to Nino3 SST anomalies that underpin extreme El Nino (Cai et al. 2014 Nature Climate Change).

- The model evaluation section has been altered to clarify this point. First, it is now made clear that the discussion of ENSO metrics refers to the Bellenger et al study. Next, we now include plots of each model in multi panel figures (Figures 1-4, 7 and 8). Finally, we also discuss biases in section 6 in further detail, with reference to Cai et al 2014. (“This study highlights several avenues for further research. … Several models have known difficulties simulating aspects of ENSO, such as the nonlinear response of rainfall to extreme El Niño episodes (e.g., Cai et al., 2014). Additional targeted experiments within a single climate model would provide further insight into the apparent complexity of ENSO impacts through time.

Shouldn't HadCM2 be HadCM3? HadCM2 is an old model used in IPCC 2nd assessment report, definitely does not contribute to CMIP5.

- In the original manuscript, this model was mistakenly referred to as HadCM2 in the text and HadCM3 in the figures. This has now been corrected throughout to HadCM3.

The authors chose to use Nino3.4 to represent ENSO based on the similarity between the MMM of EOF1 surface temperature in historical and past1000 runs. However, this does not take into account the fact that the temperature pattern changes through time. At certain epochs (of say 30 years), Nino4 can better capture the predominant ENSO characteristics over that particular epoch (E.g., after the 90s – McPhaden et al. 2011 GRL; possibly mid Holocene – Karamperidou et al. 2015, Paleoceanography), and at other epochs, Nino3 could be better. Spatial changes in ENSO pattern are not discussed in this present paper, but it is an important aspect as far as teleconnection is concerned. This should be discussed to a certain extent in the manuscript. Section 4.3 for instance should mention recent results by Karamperidou et al. (2015 Paleoceanography) in which they used CCSM4 model that the mid Holocene involves a change in the spatial pattern of ENSO from eastern Pacific to central Pacific. See also Carre et al. (Science).

- We now include a supplementary figures addressing spatial changes and Nino3 and Nino4 indices. In addition, we now compare the Last Millennium simulation with a 100 year historical period (1906-2005). Furthermore, we explicitly discuss changes in spatial patterns in section 6, including the references recommended.

p. 1584 (5): Is it air temperature or sea surface temperature for the Nino3.4? It is also not clear how composites are calculated here. Is it for the 6 consecutive months or just annual mean?

- This has now been clarified ("El Niño episodes were defined based on simulated surface air temperature anomalies in the NINO3.4 region, with events defined in the models when NINO3.4 temperature anomalies were >0.5 K for at least six consecutive months (Trenberth, 1997). Conversely, La Niña episodes were defined when NINO3.4 temperature anomalies were <-0.5 K for at least six consecutive months. Spatial patterns are examined by compositing monthly temperature and rainfall anomalies into positive (El Niño) and negative (La Niña) phases using these definitions for all CMIP5 models analysed.")

Various studies (e.g.., Li et al. 2013) have used other locations more remote than those used in this present paper for proxy reconstructions. As the authors argue that it is important to link remote proxies with those in central Pacific, why not include far more remote regions as well (e.g., North Pacific, Central Asia) to better illustrate their argument.

- We agree that it would be interesting to next investigate more remote regions. However, the current study specifically focuses on the tropical Pacific and we now identify this as an avenue for future work in section 6 (" In addition, various studies have linked remote proxy variability to the tropical Pacific (e.g., Li et al., 2013) and hence it would useful in the future to investigate regions remote from the Pacific basin, such as in North America or China.")

Fig. 3: It's worth mentioning that the higher power at low frequency in the past 1000 yr runs is also likely attributed to the much longer time series than the historical (30 yrs), better resolving the low-frequency variability.

- This figure has now been modified to included extended historical data from 1906-2005).

P1591 (10), The first sentence implies variability in control simulations is similar to that in past1000, but the subsequent sentences contradict that. It would be easier to compare with Fig. 2 if the power spectra in Fig. 6 are computed using 100-yr samples. It is not clear whether the differences between the two simulations are due to the different length of time series.

- We have now shortened this section and have clarified where information appeared to be contradictory. The range of spectra shown in (revised) Figure 4 and Figure 6 are for 100 –year samples.

p.1592 (20), the difference in the magnitude of the teleconnection patterns in Fig. 1 between past1000 and historical should more likely be due to the averaging of more samples in the past1000 simulations (compared with only 30 years in historical). Actually statistical significance can be added in Fig. 1 by constructing confidence interval based on 30-yr chunks in the past1000 across the 6 models. Do the same for the 30-yr historical. This will then allow determination whether the ensemble means between the historical and past1000 yr are significantly different.

- We now include details of each model and also compare Last Millennium data to the historical period of 1906-2005, rather than the last 30 years of the experiment.

Minor points

The line specs in Fig. 7 are confusing. Why are there 3 different colours for the fitted line? On the left panel the black lines seem to match the blue dots better, so I'm not sure which one is for which. I think Fig. 7 can be culled since the same information can be found in Fig. 8. The historical values can be added in Fig. 8 instead.

- This suggestion has been included, with original Figure 7 culled and historical values included on the same figure as the Last Millennium values.

Again, what does Fig. 8 look like in each model? Insert horizontal lines indicating statistical significant level in Fig. 8.

- The original Figure 8 has now been modified to included details of each model and statistical significance. In the revised manuscript, the precipitation and surface temperature are now shown as separate figures.

This manuscript is well written but it could be shorter as there appears to be a lot of repetitions, e.g.,: p. 1592 (15) first, second, and third sentences basically convey the same message. It need not be stated three times, especially in the same paragraph. A lot of information stated in Section 2 is again repeated in Section 4 (e.g., p. 1593 (10)). P1587 (20, 25): "Models that have.....In addition, ...The MIROC-ESM....(Fig.3)" is a long unnecessary repetition from Section 2 (page 1584). Could consider shortening Section 2 and integrate it to the other sections or move it to an appendix.

- We have made substantial changes to section 2, in order to remove repetition and removing the discussion on various forcings that did not contribute to understanding the stability of teleconnections.

The literature review is lacking on ENSO behaviour response to greenhouse warming, and model-based studies on the sensitivity in the relationship between ENSO and background climate state (p. 1581). A number of recent studies beyond the Collins et al. 2010 (Nature Climate Change) have found that there is indeed inter-model agreement in the response of ENSO to greenhouse warming. Apart from the Power et al. (2013) paper, the other studies show that this response appears to be in the form of an increase in the frequency of extreme El Nino and La Nina (Santoso et al. 2013 Nature; Cai et al. 2014; 2015, Nature Climate Change). It would be good to mention these studies in the introduction to provide a more updated background literature. The model projected change toward more extreme ENSO occurrences under greenhouse

warming can provide an interesting avenue for paleo studies to investigate.
1st paragraph on page 1581: "observed changes in the character of ENSO since mid-70s towards a dominance of El Nino" is not accurate, since late 90s the mean state has changed toward a La Nina-like (e.g., England et al. 2014, Nature Climate Change; Hu et al. 2013 J. Climate 26, 2601-2613).

- This review has now been expanded and updated to reflect more recent model-based studies ("While changes in ENSO behaviour may occur under future global warming (Power et al., 2013), there is a large dispersion in global climate model (GCM) projections of changes in ENSO characteristics (e.g. Collins et al., 2010; Vecchi and Wittenberg, 2010), and hence the sensitivity of the coupled ocean-atmosphere system to future changing boundary conditions may be uncertain (DiNezio et al., 2012). Alternatively, model-based recent studies demonstrate projected changes toward more extreme ENSO occurrences under greenhouse warming (Cai et al., 2013; Power et al., 2013). Investigations of the sensitivity of ENSO to anthropogenic climate change are also restricted by the relatively short instrumental record, which provides us with limited guidance for understanding the range of ENSO behaviours. For example, the observed changes in the character of ENSO in the 20th and 21st centuries (including dominance of El Niño, rather than La Niña, episodes from the mid-1970s and La Niña-like mean state since the 1990s (England et al., 2014)) are difficult to evaluate in terms of a forced response or unforced variability given the limited observational record almost certainly does not capture the full range of internal climate dynamics.")

Page 1589 on inter-decadal modulation of ENSO behaviour, one relevant paper is Borlace et al. 2013, J. Climate that demonstrate how this can arise naturally via vacillation of the internal ENSO dynamics.

- This reference has been added.

Fig 1. : The Y-axis ticks do not look correct, and the western and central boxes are not centred about the equator.

- This has now been corrected in the revised multi-model Figure 1 and Figure 2.

P1584 (20, 25, etc.) "past 1000" should be "past-1000" or "past1000". Otherwise 'past 1000 simulations' could be mistaken as one thousand simulations in the past, while it should mean past 1000-yr simulations.

- This is an artifact of typesetting, not error.

P1584 (25) Refer to Fig. S2 and Fig.... How about showing observations as well?

- A reference to Supplementary Figure 4 has now been included. As this figure now shows past1000 time series, observations have not been added.

10 'categorised' should be 'categorise'

- This has been corrected.

p.1585 (15), "experiments in was" delete 'in'

- This has been corrected.

"For the GISS-E2-R (Schmidt et al., 2014) and IPSL-CM5A-LR (Dufresne et al., 2013) models...." It is better to insert these reference in Table 1 for all of the models.

- These references have been removed, but not added to Table 1, as no specific model is discussed here, but rather the combined contributions to CMIP5.