Clim. Past Discuss.,
https://doi.org/10.5194/cp-2018-164-RC2, 2019
© Author(s) 2019. This work is distributed under
the Creative Commons Attribution 4.0 License.

Climate of the Past Discussions
Open Access
EGU

Interactive comment on "Identifying teleconnections and multidecadal variability of East Asian surface temperature during the last millennium in CMIP5 simulations" *by* Satyaban B. Ratna et al.

Anonymous Referee #2

Received and published: 9 April 2019

General Comments: Ratna et al. examine the influence of transient external forcing (volcanic eruptions) on PDO and AMO variability and teleconnection patterns as they relate to East Asian surface air temperatures (SAT) in three PMIP3/CMIP5 past1000 simulations and paleoclimate reconstructions. This is an interesting study, and the results have interesting implications for how external forcing can impact internal variability and teleconnections. However, more work is needed to compare model output to observations and expand the study to other models.

Main Concerns:

Printer-friendly version

Discussion paper

CC BY

1) There are at least ten CMIP5/PMIP3 past1000 (Last Millennium) simulations available on ESGF that span the 850-1849 CE time period (BCC, CCSM4, CSIRO, FGOALS, GISS, Had, IPSL, MIROC, MPI, MRI). The authors exclude several of these simulations (MIROC, FGOALS, GISS) due to spin up/model drift/trend issues and cite Atwood et al for why they exclude these simulations. However, the authors choose not to use the output from CSIRO, HadCM3, IPSL, or MRI (some of which are included in the analysis of Atwood et al). The results therefore seem incomplete and selectively presented- why the exclusion of these other simulations? Please include analyses of these other Last Millennium simulations or at least provide a reason for why these other Last Millennium simulations have been excluded (the data have been available for at least 8-12 months online, so I hope it's not a data availability issue?). As the manuscript is currently written, 1/3 of the models show a completely different result, but this is only one model- is this really 1/3 of all CMIP5 Last Millennium models, or just one outlier in the CMIP5 Last Millennium simulations?

2) The authors concatenate the Last Millennium (∼850-1849CE) and the Historical simulations (∼1850-2005CE) after removing the linear trend from each of these time segments separately. Removing a linear trend from either instrumental or CMIP5 data over the entire 1850-2005CE time period can be problematic if the main component of the 'warming trend' is in the 20th century. Multiple papers choose to remove the linear trend over the 20th century only (e.g., Deser et al., 2010; Messie and Chavez, 2011; Franzke, 2014, Nature Climate Change; Ji et al., Nature Climate Change, 2014). Similarly, many CMIP5 historical simulations appear to show much of the global warming trend starting in the 20th century, so removing a trend over the full historical simulation period (1850-2005) may add in decadal-centennial variability. To avoid this detrending and concatenation problem, could the analysis just be conducted over the 850-1849CE time period (especially because it seems the authors are mostly focused on the impacts of volcanic eruptions on the PDO and AMO in the pre-1850CE time period?). Some recent work even suggests that the dynamics of the system change once GHG forcing becomes dominant (e.g., Song and Yu, 2015, J Clim; Brown et al., 2017, Nature

CC BY

Climate Change), so including this time period could be arguably problematic.

3) There is no comparison between the spatial patterns of the AMO and PDO in instrumental-based reconstructions and the three models used here- perhaps some of these simulated spatial patterns are more realistic than others? The authors state that the model results are realistic, but never show this in the manuscript. The Climate Variability Diagnostics Package (http://webext.cgd.ucar.edu/Multi-Case/CVDP_ex/CMIP5-Historical/) shows that the spatial expressions of the AMO (and PDO) can be quite different in the various CMIP5 Historical simulations. Interpretation of the model results may be viewed through a more informed perspective if the models are compared to instrumental-based observations.

4) Varying significance levels are used in the paper (90% vs 95%). Please use a consistent 95% or 99% significance level- as the paper stands, it appears that the significance level has been lowered to show 'significance spectral peaks' (e.g., Fig 10), but the spectra barely surpass this 90% level- why not use 95% or 99% everywhere? At least please include some discussion of the sensitivity of the results to significance level if the results don't pass this higher threshold (significance levels are admittedly is arbitrary, but the current, inconsistent use of 90% runs the risk of appearing selectively low to attempt to present a 'significant' result).

General minor issues: Many authors abbreviate pre-industrial Control as PI (e.g., Atwood et al., 2016, J Clim, among others)- in an effort to maintain some sort of standard abbreviation that may be quickly recognized, I would encourage the authors to employ more commonly used acronyms (e.g., PI or piControl).

Also, when reading through the figures, it is difficult to interpret the acronyms used in each figure without searching through the other figure captions or the text for the definitions of the acronyms- please define the acronyms used in each figure in each figure caption (or at least reference where they are defined) so readers can quickly understand the figure without searching for what they mean.

CC BY

Specific comments: Page 1, Line 12-13: The simulated PDO and AMO spectra and spatial patterns are never compared to instrumental-based patterns or spectra (or even to proxy-based spatial patterns). Please include figures/analysis that support this statement in the main text or remove it.

Page 2, ∼line 10: The previous paragraph critiques the instrumental and proxy-based records, but little attention is paid to potential model deficiencies- can you at least briefly discuss or cite a few papers that may critique or even acknowledge that CMIP5/PMIP3 models have their own biases and problems as they relate to low-frequency SAT variability (e.g., Laepple and Huybers, 2014; Parsons et al., 2017 J Clim;) or 'modes' of internal variability (or their responses to stratospheric aerosol loading from volcanic eruptions)? Alternatively, directing the reader to where these model deficiencies, and their implications for your results, are going to be discussed later in the paper could be helpful.

Page 3, lines 5-6: please see general comments in previous section- why were the bulk of the CMIP5/PMIP3 Last Millennium simulations excluded? Analysis of results would appear much more robust if an attempt is made to present more than 1/3 of the Last Millennium simulations, or if reasoning can be given why the other simulations were excluded. Also, what is the cutoff used for a drift that is 'too strong'? Is this a global or local drift? All the CMIP5/PMIP3 past1000 simulations appear to show some sort of trend/drift at many grid points- the question is what is too much for the purposes of this AMO/PDO teleconnection study. Please clarify.

Page 3, Line 20: please see general comments in previous section- removing one linear trend over the full 1850-2005CE time period seems like it may add in low-frequency variability, and I am still not even sure why the historical simulations have been included if the focus is on the impact of volcanic eruptions in the pre-historical simulation time period.

Page 4, Line 8: 'we don't see much differences'- this is a subjective statement. What

Printer-friendly version

Discussion paper

CC BY

criteria are used? Perhaps something like a pattern metric or Euclidean distances metric could be used to say something more quantitative?

Page 4, Line 10: Please explain how the TAS time series is made- I assume annual mean (Jan-Dec?) temperature at each grid box, latitude-weighted, and masked ocean grid boxes? Over what latitude and longitude range is this area average made (is it the whole region used in the maps in the figures showing East Asia?)? Please provide more details in the text.

Page 5, line 5-7: There are other PDO reconstructions- fine to not include them, but can you state why this one is selected over others?

Page 5, lines 11-15: As far as I can tell, the model-based PDO indices are made from monthly data, and the paleo-based PDO indices are made 'annual' data (or seasonally sensitive proxy records)- would a better comparison be to make annual means of SAT for the model data, then construct the PDO index, so the index is more comparable to the annual proxy-based index? (or can you show that the annual and monthly model-based PDO patterns and time series are similar?)

Page 5, Line 21, line 25: The 90% significance level seems oddly low, and arbitrarily used in only certain cases- do your results consistently pass a 95% significance test (both the regions in the maps and the spectra)? For example, the 'significant' spectral peaks in Figure 10 appear quite close to the 90% significance level- if you made this a 95 or 99%, are these 'significant spectral peaks' at all significant?

Page 8, line 26: the authors discuss a weak response in BCC to volcanic eruptions- is this a finding that has been noted previously (e.g., Driscoll et al., JGR, 2012, or some sort of similar CMIP5 comparison to observations?)? How realistic is this model's response relative to the other models' responses to volcanic eruptions (especially compared to observations of more recent eruptions and their impacts)? I ask because this difference seems to be important to the results- for example, should the BCC changes (or lack thereof relative to the other models) in PDO, AMO, and associated teleconnec-

CC BY

tions with E Asia be viewed as just as realistic as the other models' responses? Or is it an outlier because it doesn't respond at all to volcanic eruptions when it should?

Page 9, Line 25-26: It would be helpful to show results from the other four CMIP5 Last Millennium simulations here to put these results in context- right now, 1/3 of the models show a completely different result, but this '1/3 of models' is just the BCC model.

Page 9, Line 26-29: Would this result imply that the models show an unrealistically large response to eruptions? Or that there is too little internal, low-frequency variability (e.g., Laepple and Huybers)? Or does this suggest both, or something else?

Page 10, line 2-3: the authors state that 'all models display red spectra'- in the methods (and in the time series in the figures), it seems that the data have been low-pass filtered, so by definition, the high-frequency variability has been reduced relative to the low-frequency variability (thus reddened)- I'm not sure that 'redness' really means anything in this case. If 'redness' does mean something after the data have been filtered, or if the data have not been low-pass filtered before spectral estimation, please clarify/explain- for example, if the authors mean to say that one model has more low-frequency variability than another, that may be more accurate.

Furthermore, the 'pronounced multidecadal variability' barely surpasses the 90% significance threshold, as do most of the 'significant' peaks referenced in this paragraph.

These power spectra (AMO and PDO power spectra figures) are shown without any error bars- when the spectra are compared and declared similar/different, some sort of spectral estimation confidence bound/error bar on the figure could show if these differences fall within the confidence bounds of the spectral estimates.

Page 10, Lines ∼5-15: Perhaps this is the first time that this analysis has been done, but I would be surprised- has anyone else compared the power spectra across these simulations before? For example, Cheung et al., (2017) compares instrumental-based AMO and Pacific variability to CMIP5 historical simulations (and also how the spatial

CC BY

patterns associated with these modes can change through time). Parsons et al., 2017 (J Clim) compares instrumental, AR1, and CMIP5 Last Millennium, and CMIP5 Control spectra over the North Pacific and North Atlantic, and Fredriksen and Rypdal (2016, J Clim) compare spectra over ocean basins in CMIP5 models.

Page 10, Line 23: the authors claim that the spatial patterns of AMO and PDO are similar to the patterns from observations. I see no comparisons among modeled and observed spatial patterns of variability. In fact, it would be helpful if the authors would show the spatial patterns from observations (of course acknowledging that the instrumental-based data have their own limitations) in Figures 1 and 2- this would help put the model results in context.

Page 10, line 25: again, it's unclear if the data have been low-pass filtered before spectral analysis. Also, see my above comments- saying the spectra are 'red' seems meaningless if the data have been low-pass filtered. Again, the significant peaks barely surpass a 90% threshold- please discuss or mention if this significance is sensitive to threshold level.

Also, as stated above it would be good to include error bars/lines on the spectra to know if the 'significant' differences from the background spectrum significant given uncertainties in the power spectral estimation?

Page 11, ∼Line 25: good point.

Page 12, ∼line 4: OK, so other recent methods have been used to reconstruct SAT fields (e.g., Last Millennium Reanalysis from Hakim et al., 2016, JGRA and Tardiff et al., in review at CP)

Figures: Figures 1, 2, 3, 5: please include panels showing similar analyses from instrumental-based data products.

Figure 4, Figure 6: it is interesting to see the PDO-E Asia and AMO-E Asia differences, but it would be nice to see some confidence bars on the control run values. For exam-

CC BY

ple, Coats et al. 2013 show that teleconnections can change from century to century. Could you do some sort of running correlation or subsample the control run to see how variable this E Asian relationship is (or is there enough data?)

Figure 9: Is there a way to put these results in context? For example, if you include the post-Pinatubo response in these models, could you show how the models compare to obs? Which models are more realistic? (CCSM4/MPI or BCC?)

Figures 10 and 11: inclusion of instrumental-based spectra could be helpful here too- how realistic are these reconstructions?

Compact listing of purely technical corrections (typing errors, etc.). Page 1, Line 13: 'and their spectral characteristics'- remove 'their' Page 3, line17: change sentence to: 'Each model version was the same across all the simulations.' Page 4, line 8: 'much differences'- please re-word (e.g., 'A pattern correlation statistic shows minimal differences among . . .' Page 5, Line 7: 'largely suffer from the influence of external forcing' Page 6, Line 4: 'no time-varying (transient?) external radiative forcing' Page 6, line 31: 'This situation is equivalent to (that?) of Fig.' – there appears to be a missing word here Page 7, Line 13-15: 'in the southern parts'. . .'in all three models'. . .'with the strongest correlation in the northeast region' Page 7, line17: 'though it varies' Page 9, Line 10: the sentence starting with 'Despite' appears a bit awkward- suggest rewording

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2018-164, 2018.

CC BY