Clim. Past Discuss., 9, C2084–C2104, 2013
www.clim-past-discuss.net/9/C2084/2013/

Climate of the Past Discussions
Open Access

Interactive comment on "Consistency of the multi-model CMIP5/PMIP3-past1000 ensemble" *by* O. Bothe et al.

O. Bothe et al.

ol.bothe@gmail.com

Received and published: 22 September 2013

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

CPD
9, C2084–C2104, 2013

Interactive Comment

Reply to referee comments

Oliver Bothe, Johann H. Jungclaus, Davide Zanchettin

22 September 2013

Dear Editor, dear Referees,

Thank you for your considerate evaluation of our manuscript. In the following we reply point by point to the comments by Referee one and two. The changes following from your annotations clearly improved our manuscript.

Your comments are quoted in italics, our replies are in roman. Blue fonts are used when changes made in the manuscript are reported.

Please see the attached diff-file for all changes to the manuscript.

1 Anonymous Referee 1

1.1 Specific comments

1) I found the introduction to be highly informative as well as clearly and excellently written. I only ask for the sentence starting p3791, l15 "This means..." to be expanded and clarified as I understand the principle but it's a bit awkward to read.

We change this to

Therefore, we have to ask (e.g. Hargreaves et al., 2011; Bothe et al., 2013): Do simulations and reconstructions represent compatible realizations of the unknown past climate? Answering this question implies to establish within confidence margins whether the data samples can be assumed to stem from a common distribution whose shape is constrained by external perturbations and forcings to the climate system.

2) I realise that the Mann et al (2009) provides the only current reconstruction that includes field data and can be used throughout your analysis, but it is a little bit of a shame that the other major reconstructions can't be included to account more evidently for reconstruction uncertainty. They could perhaps give more information on the probabilistic conclusion of differences regarding (MWP-LIA) between the simulations and reconstruction.

We agree that considering the major hemispheric- and global-mean reconstructions would valuably complement the picture we give so far. These would/could include Jones et al. (1998), Briffa (2000), Mann and Jones (1998), Moberg et al. (2005), D'Arrigo et al. (2006), Hegerl et al. (2007), Frank et al. (2007), Juckes et al. (2007) and the area averages from Mann et al. (2009). Furthermore, it would be probably good to include the reconstruction by Ljungqvist (2010). However, from such a point of view it would also be recommendable to assess the consistency of the simulations with the regional reconstructions presented by the PAGES 2K Consortium (Ahmed et al., 2013). These emerging continental-scale reconstructions offer opportunities for further application of the method but are beyond the scope of this paper.

Thus, in preparing the manuscript the question arose at which point such complement, in order to be exhaustive, would have become more unwieldy than valuable. We decided accordingly to concentrate on the single available global field data. We do still think that including results for additional data in the main manuscript would reduce its readability. That is, including (even only some) major hemispheric data would include

CPD
9, C2084–C2104, 2013

additional sections with a limited added value.

We nonetheless follow the Reviewer's suggestion and now provide an additional section in the supplement (Section C) where results concerning a selection of hemispheric-mean temperature reconstructions are presented. These results are summarized in the main manuscript in Sections 3.5 and 3.6 as follows:

The supplement provides figures for a small selection of northern hemisphere mean reconstructions of annual-mean temperature from the recalibration ensemble by Frank et al. (2010). Frank et al. recalibrated nine northern hemispheric mean reconstructions to different periods of observational data resulting in 521 individual reconstructions.

Consistency of the simulation ensemble with the considered subset of recalibrated reconstructions is generally limited to some sub-periods, which generally differ for the individual recalibrated reconstruction series. Over-dispersion is the most common deviation but prominent biases also occur over individual sub-periods. However, we cannot overall reject consistency relative to the reconstruction by Frank et al. (2007).

The assessment of consistency of the PMIP3-past1000 ensemble relative to the members of the Frank et al. ensemble highlights an additional feature. The recalibrated series for a specific reconstruction display different variability dependent on the specific recalibration period. Therefore the simulation ensemble can be consistent with respect to a reconstruction recalibrated to a specific period while lacking consistency with the same reconstruction recalibrated to a different period. This ambiguity highlights the inherent uncertainty in our estimates for the climate of the last millennium and stresses the necessity of increasing the quality of reconstructions, of simulations and of the external-forcing estimates used for the simulations.

The PMIP3-past1000 ensemble and the recalibrated hemispheric mean reconstructions clearly differ in the resolved multidecadal-to-centennial and interdecadal variability. Consequently and similar to the global field data, simulations and reconstructions often differ in their long-term multicentennial trends, e.g., the passage from the MWP

CPD
9, C2084–C2104, 2013

to the LIA and to the industrial warming period.

3) Section 2.1 is finely written and concludes with some very important and informative points!

Thanks

4) Section 2.2 p3798, l2: Are these regions sensitive/robust?

We did consider a number of different definitions for the AMO and PDO regions with differing latitudinal and longitudinal extent. Unsurprisingly, we did find that the results are not identical for all definitions. However, conclusions are robust over all tested regional definitions.

We add a note in Section 2.2

Our later conclusions are robust against different definitions of the regional indices.

5) p3799, l13-l15: good, an important point to stress!

Ta.

6) In Section 3.2 and generally thereafter I became, probably through my own ignorance, confused by the discussions relating to over- and under-disperse in the text. This is in part because I was expecting a simulated or effectively "sample" distribution on the y-axis and a reconstruction target or "theoretical" distribution on the x-axis to be consistent along y = x. However, if this was the case, Fig. 3f or Fig. 2 top-left for example, would show high counts at the tails of the histogram counts distributions, when instead the early, middle and late periods all seem to be less weighted at the tails. I'm sure your arguments are right, I would just need a little more "hand holding" through this section.

We clarify here: Residual quantile-quantile plots visualize on the x-axis the "theoretical" or "observed" quantiles and on the y-axis the deviation of the simulated or "sample" quantiles from the "theoretical" or "observed" quantiles. That is, consistency is given

CPD
9, C2084–C2104, 2013

for y=0.

On the other hand, the old Fig. 3f presents rank-histograms for the sub-periods. Rank-histograms plot the counts of the target data within sortings of target and ensemble data. That is, approximately constant counts, i.e. a flat histogram, indicates consistency whereas high counts in the high/low ranks suggest negative/positive biases in the ensemble. If the target data is too often in the middle ranks the ensemble is potentially over-dispersive. If the target data is too seldom in the middle ranks the ensemble is potentially under-dispersive.

We change the description of the deviations on former page 3794:

In r-q-q plots, biases of the ensemble data are seen as displacements from the $y = 0$ line.

We furthermore modify the text to include more leading. Please see the attached diff-file for the specific changes.

7) I would appreciate p3804, l16-l18 being fleshed out a little to explain which "deviations visualized" are contributing to the conclusion in this paragraph and which types of ensemble characteristics you are speaking of, especially as we don't have the COSMOS-Mill results at hand.

We add to the referred paragraph by changing it to:

We note that deviations in spread and bias are nevertheless pronounced in Fig. 2 and that individual simulations show rather similar climatological residual quantiles relative to the reconstruction. This leads us to the conclusion that the less prominent over-dispersion compared to the COSMOS-Mill ensemble is not so much due to the multi-model character of the past1000 ensemble and to differences in the used forcing data sets, but that it is mostly due to the fact that the analyses of the two ensembles use different temporal resolution.

8) The first line of Section 4.4, I couldn't relate this statement with any evidence in Fig.

CPD
9, C2084–C2104, 2013

5. Is it actually Fig. 6 where any indication is made about the differences between the reconstructed and simulated start of the millennium?

Indeed, we meant Fig.6. Reference is now to the correct figure.

9) Here I am thinking specifically of p3808, l20-l22: where reference is made to a result that is not shown comparing two groups of simulations separated by different volcanic forcing data-sets; and the paragraph from p3810, l9-l20 where multi-model and single-model ensembles are compared (past1000 vs earlier COSMOS analysis). I consider these to be results of great interest in the article but they are rather swamped by way the article is structured in Section 4 and 5. I feel that these results in Section 4 particularly, could be shown in the article explicitly and much of the fragmented results in Section 4 about "sources of uncertainty" could be moved to Section 3 where they appear to me to be complimentary side-by-side rather than two separate analyses. I mean, no mention is made of these rather interesting conclusions in the abstract for example. In the case of the distinct volcanic groups of simulations, I would like more discussion of how the sub-ensemble variations can be compared with the differences in simulations with different volcanic forcings, such as if you use several of the GISS ensemble members for example.

It is of course possible to compare Bothe et al. (2013) and the present manuscript side-by-side. However, discussing results from both manuscripts in more details renders the current document not well structured.

Concerning the GISS-ensemble, we note that detailed analyses of the respective CMIP5-simulations (compare http://data.giss.nasa.gov/modelE/ar5/) are very welcome. An analysis of the consistency of the GISS-E2-R simulations would, however, require ensembles comparable to the COSMOS-Mill simulations.

A previous version of the manuscript discussed the assessment of consistency and the sources of disagreement side-by-side which similarly reduced the readability. Nevertheless, we reorganized the results and discussion sections so that, as shown by

CPD

9, C2084–C2104, 2013

the diff-file, in the new version of the manuscript the parts of section 4 of concern are presented side-by-side with those of section 3.

We agree that the discussion of the external forcing vs. internal variability is worth a more thorough discussion. However, an exhaustive handling of the issue is certainly beyond the scope of the current manuscript. Indeed, the discussion of forced volcanic variability requires a full paper particularly concerning the different implementation strategies of the forcing in the different models. We extend slightly our respective statements and provide additional plots in the supplement. The paragraph has been rewritten as follows [italics are unchanged]:

In view of a possible impact of the choice of forcing inputs on the simulated data, one might think about partitioning the ensemble relative to the various combinations of forcings (Table 1). Only discriminating by the volcanic forcing, this results in two sub-ensembles including, respectively BCC, IPSL, CCSM4 and GISS-R25 (using the data by Gao et al., 2008) and MPI-ESM, CSIRO and GISS-R24 (using the Crowley data, see, e.g. Crowley and Unterman, 2012). Here we exclude the FGOALS data as it does not easily fit into these two categories but considers forcings as presented by Jones and Mann (2004) which are not explicitly included in the PMIP3-protocol (see Table 1 and Schmidt et al., 2011). We refer to the ensembles as Gao-ensemble and Crowley-ensemble.

The Crowley-ensemble generally shows a smaller response to the prescribed volcanic eruptions than the Gao-ensemble for interannual (Supp. Fig. S6) and decadal (Supp. Fig. S7) variations (compare, e.g. 1250s, 1450s, and 1810s). The Gao-ensemble displays more eruptions having an influence on the northern hemisphere mean temperature than the Crowley-ensemble, which highlights the differences between the two volcanic reconstructions. The range of the Gao-ensemble, i.e. largest minus smallest temperature anomaly for each date, is larger than for the Crowley-ensemble for a number of cases (Supp. Fig. S8). Whereas this is mainly due to the generally much weaker response to strong volcanic eruptions in BCC, other individual simulations can

CPD
9, C2084–C2104, 2013

differ strongly from the three other members of the Gao-ensemble at certain dates (Supp. Fig. S6-S8). Thus, the different architectures of the models (see Supp. Fig. 8) can result in differences between the simulations at least as large as induced by different volcanic forcing data sets. That is, the implementation strategy for the volcanic forcing data and the tuning of the model may influence the results as much as the choice of the forcing data (see also discussions by Fernández-Donado et al., 2013). We note that Schmidt et al. (2013) report for the GISS-R simulations and the Gao et al. (2008) data a radiative forcing twice as strong as expected. They attribute this fact to the implementation of the volcanic forcing data in GISS-R. The temperature responses to strong volcanic eruptions from the GISS-R25 simulation considered here, which employs the Gao et al. (2008) data, are among the largest in the multi-model PMIP3-past1000 ensemble but generally not exceptional (see Supp. Fig. 6).

The representation of volcanic eruptions in simulations (but also their assessment in reconstructions) is a highly controversial topic as seen in the ongoing discussions originating from (Mann et al., 2012, see also Anchukaitis et al., 2012). Beyond discussions focussed on the climate of the last millennium, Driscoll et al. (2012) report an apparent lack of skill of the CMIP5 climate models in reproducing the observed response to the well constrained 20th century eruptions, possibly linked to a poor representation of relevant dynamical features. Considering the technical handling of volcanic forcing in climate simulations, Timmreck et al. (2009) and Timmreck et al. (2010) highlighted the influence of the parameterisation of aerosol size on the magnitude of the imposed radiative forcing. Zanchettin et al. (2013) showed for a well constrained top-of-atmosphere radiative forcing, that the simulated climate response to a strong volcanic eruption can strongly vary depending on the background climate state which is defined by ongoing internal climate variability and the presence and magnitude of additional external forcings including their forcing history.

10) p3810, l26 Do the "long-term trends" refer to when the analysis is split into subperiods?

CPD
9, C2084–C2104, 2013

In this context it refers to the fact that the data lacks consistency over sub-periods though the data appears to be consistent over the full period.

11) Table 1 is very useful!

Thanks

1.2 Technical corrections

p3791, l6: "set agrees with the"

Done

general comment the Southwestern North America region of (Wahl and Smerdon, 2012), is it not possible for simplicities sake to refer to this throughout the paper by the authors assignment of 'North American Southwest' NASW?

Generally changed.

p3792, l11: "allows consideration of the consistency"

Done

p3792, l12: "not only with initial and forcing uncertainties (as for the COSMOS-Mill ensemble) but also with different parametric choices, physical models and different structural"

We rephrase to

not only with initial and forcing condition uncertainties (like for the COSMOS-Mill ensemble) but also and especially with structural uncertainties in the models and the different parametric choices

p3793, l7 and l9: perhaps the "see, e.g. Author" can just be replaced with "see, Author" in both cases

CPD
9, C2084–C2104, 2013

l7 changed

p3794, l19: "target quantiles, indicating an over-dispersive data set. "

Done

p3794, l21: "data set, which we refer to, as under-dispersive."

Done

p3795, l26: "reconstructions"

Done

p3796, l3: "re-consideration of"

Done

p3796, l7: "or biased in terms of uncertainty, further"

We change the phrase to "or biased considering the uncertainty, further"

p3796, l10: "that, with the present"

Done

p3796, l25: "reliable during this period"

Thanks, done.

p3799, l4: "We will return to this point"

Done

p3799, l7 and l9: Question to the Editor, should it be "centre", i.e. British English here?

We changed it.

p3799, l10-l12: "full study period, and thereby shift focus on to the comparability of the variability over pre-industrial times only."

CPD

9, C2084–C2104, 2013

Done

p3799, l13: "neither allow the ranking of the various simulated realizations" against

Done

general comment you use the word ”relation” (e.g. p3801, l7) often where ”relationship” is perhaps better suited

We changed it at some places, but not everywhere (e.g. we kept "relation" at the noted occurrence).

p3801, l7: ”between the reconstruction and the simulation’

Done

p3801, l8: ”simulations rarely represent the”

Done

general comment I rather recommend ”see Fig. X” instead of ”compare Fig. X” throughout the article

We changed it at most instances.

p3802, l10: ”for the AMO in this period (Fig. 3f and h)”

Done

p3802, l14: ”the simulation data does not reproduce”

Done

p3803, l22: ”their variability in the simulation ensemble”

Done

p3804, l24: you don’t have to agree with me but may I suggest ”We consider correlation analysis as a universal method in studies comparing simulations and reconstructions”

CPD
9, C2084–C2104, 2013

Done

p3804, l26: "reconstruction indicates that to some extent both data-sets feature a similar signal"

Done

p3806, l25: I was a bit unsure if this was a similar statement to the one earlier about correlation methods. "Again we discuss correlations as example for common practices". Is this a justification of the method?

The statement of concern is necessary to clarify that correlation analysis is performed here only as an example of analyses commonly performed in the field. In fact, based on the found potential lack of consistency between simulations and reconstructions, correlation-based analysis seems not to be an appropriate approach, hence we should have avoided such an exercise if not as example of common practices.

p3809, l5: "Furthermore, it is possible the simulations do not fully"

Done

p3809, l6: perhaps remove the 'e.g. the' so it reads "influence of, solar forcing"

Done

p3809, l11-l12: "also showed that a larger agreement and consistency closer to the present should not be expected."

Thanks

p3810, l22: You can remove "only" from this sentence

Done

p3811, l8: "the necessity for improvements of simulated estimates"

Done

CPD

9, C2084–C2104, 2013

p3819, caption remove "'the range between' on line 4 which becomes "in steps of 0.2 within 0.1 and 0.9"

Done

p3819, caption for clarification, does 'single deviation tests' mean tests done on each individual grid point?

general comment regarding figures with residual qqplots, I realise that axis labels are impractical in Fig. 2, but an explanation of the axes would be useful for the reader. This also relates to Fig. 3 and Fig. 4 where "Simulated Target" is used in the former, whilst "Projection Target" is in the latter. Is there any significance to this difference?

We clarify in the caption of Figure 1 that indeed we perform the tests at each indiviudal grid point.

We add an explanation to Figure 2.

In Figs 3 and 4 "simulated minus target" and "projection minus target" mean the same.

2 Anonymous Referee 2

2.1 Specific comments:

P3792 L11-L15: I suggest provide simple examples of 'different parametric choices' and 'different structural uncertainties' to better illustrate the improvement of current PMIP3-past1000 simulation and the importance of such setups in paleo-simulations.

We rephrase:

not only with initial and forcing condition uncertainties (like for the COSMOS-Mill ensemble) but especially with different structural uncertainties in the models and the

CPD
9, C2084–C2104, 2013

different parametric choices (Mauritsen et al., 2012, Tebaldi et al., 2007). Structural sources of uncertainties include on the most basic level different horizontal and vertical grids in all model compartments (atmosphere, ocean, land) but also the prescribed climatologies (e.g. ozone) and the formulation of, e.g. snow and ice albedo (compare, e.g. Mauritsen et al., 2012). Even if different models share certain components or portions of the same numerical code (e.g. the model genealogy of Masson and Knutti, 2011), the tuning of associated parameters likely differs.

P3793 L11: I found the notation 'larger distribution' confusing; should it be 'the same probabilistic distribution'?

We changed 'larger' to common.

P3794: The Method part presented the motivation and detailed approaches to assess both 'probabilistic consistency' and 'climatological consistency', which was necessary and very helpful.

Thank you

P3795 L18: it is difficult to tell if analysis in this article distinguished the forced condition variability and internal variability from the sentence 'Our analysis : : :'

We change sentence to

Our analysis identifies whether the total variability, i.e., externally-forced and internally-generated together, originates from distributions that are similar enough to be indistinguishable.

P3803 L9, L11: Does the term 'only just' mean that the statistical tests had marginal significance?

Yes. We clarify by changing the text to

Rank histograms (Fig. 4c and d) indicate that the ensemble is probabilistically consistent with the Southwestern North American reconstruction, but the χ^2 goodness-of-fit

CPD
9, C2084–C2104, 2013

test leads us to reject uniformity for the European data rank counts, since the statistics are only marginally significant at the considered one-sided 90 % level (Fig. 4c and d). Similarly, the bootstrapped intervals in Fig. 4a–d do only just (i.e. marginally) result in rejecting consistency for the European data

P3808: Section 4.4 discussed the possible impact of the volcanic forcing on the performance of ensemble simulations, which is very interesting. It will be more intriguing if the analyses to distinguish forcing condition variability and internal/modeling variability would be presented and discussed in greater depth.

See reply to comment 9 of Reviewer 1.

P3819 Fig 1: what does the maximum of p-values for deviation tests represent? We add the following to the caption

The maximum of both p values highlights grid-points where at least one test is significant.

2.2 Technical corrections:

P3791 L7: change 'Comparison of' to 'Comparing between' We changed it to:

'Comparing simulations with reconstructions'

P3791 L14: change 'comparison' to 'comparisons'

Done

P3792 L11: change 'considering' to 'consideration of'

Done

P3792 L23: change 'were performed including' to 'performed included'

Done

CPD
9, C2084–C2104, 2013

P3794 L15: change 'an easy understandable visualization' to 'easy understandable visualizations'

Done

P3795 L10: 'The reconstruction targets uncertainty estimates are used' is confusing.

Changed to

Uncertainty estimates for the reconstruction targets are used to inflate the simulated data.

P3796 L13: change 're-considering' to 'reconsideration of'

Done

P3798 L3 and L10: change to 'The indices are denoted by PDO and AMO' to match the order of indices presented before at L3

Done

P 3804 L4: 'In the worst case' or 'In worse cases'.

Done

P3808 L18: change 'considers' to 'consider'

We think 'considers' is correct.

References

Ahmed, M., Anchukaitis, K. J., Asrat, A., Borgaonkar, H. P., Braida, M., Buckley, B .M., Büntgen, U., Chase, B. M., Christie, D. A., Cook, E. R., Curran, M. A. J., Diaz, H. F, Esper, J., Fan, Z., Gaire, N. P., Ge, Q., Gergis, J., González-Rouco, J. F., Goosse, H., Grab, S. W., Graham, N., Graham, R., Grosjean, M., Hanhijärvi, S. T., Kaufman, D. S., Kiefer, T., Kimura, K., Korhola, A. A., Krusic, P. J., Lara, A., Lézine, A., Ljungqvist, F. C., Lorrey, A. M., Luterbacher, J.,

CPD
9, C2084–C2104, 2013

Masson-Delmotte, V., McCarroll, D., McConnell, J. R., McKay, N. P., Morales, M. S., Moy, A. D., Mulvaney, R., Mundo, I. A., Nakatsuka, T., Nash, D. J., Neukom, R., Nicholson, S. E., Oerter, H., Palmer, J. G., Phipps, S. J., Prieto, M. R., Rivera, A., Sano, M., Severi, M., Shanahan, T. M., Shao, X., Shi, F., Sigl, M., Smerdon, J. E., Solomina, O. N., Steig, E. J., Stenni, B., Thamban, M., Trouet, V., Turney, C. S., Umer, M., van Ommen, T., Verschuren, D., Viau, A. E., Villalba, R., Vinther, B. M., von Gunten, L., Wagner, S., Wahl, E. R., Wanner, H., Werner, J. P., White, J. W., Yasue, K., and Zorita, E.: Continental-scale temperature variability during the past two millennia, Nature Geoscience, 6, 339-346, 10.1038/ngeo1797, 2013.

Anchukaitis, K. J., Breitenmoser, P., Briffa, K. R., Buchwal, A., Büntgen, U., Cook, E. R., D'Arrigo, R. D., Esper, J., Evans, M. N., Frank, D., Grudd, H., Gunnarson, B. E., Hughes, M. K., Kirdyanov, A. V., Korner, C., Krusic, P. J., Luckman, B., Melvin, T. M., Salzer, M. W., Shashkin, A. V., Timmreck, C., Vaganov, E. A., and Wilson, R. J. S.: Tree rings and volcanic cooling, Nat. Geosci., 5, 836–837, 10.1038/ngeo1645, 2012.

Bothe, O., Jungclaus, J. H., Zanchettin, D., and Zorita, E.: Climate of the last millennium: ensemble consistency of simulations and reconstructions, Clim. Past, 9, 1089–1110, 10.5194/cp-9-1089-2013, 2013.

Briffa, K. R.: Annual climate variability in the holocene: interpreting the message of ancient trees, Quaternary Sci. Rev., 19, 87–105, doi:10.1016/S0277-3791(99)00056-6, 2000.

D'Arrigo, R., Wilson, R., and Jacoby, G.: On the long-term context for late twentieth century warming, J. Geophys. Res., 111, D03103, doi:10.1029/2005JD006352, 2006.

Crowley, T. J. and Unterman, M. B.: Technical details concerning development of a 1200-yr proxy index for global volcanism, Earth Syst. Sci. Data Discuss., 5, 1–28, 10.5194/essdd-5-1-2012, 2012.

Driscoll, S., Bozzo, A., Gray, L. J., Robock, A. and Stenchikov, G.: Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions, J. Geophys. Res., 117, D17105, 10.1029/2012JD017607, 2012.

Fernández-Donado, L., González-Rouco, J. F., Raible, C. C., Ammann, C. M., Barriopedro, D., García-Bustamante, E., Jungclaus, J. H., Lorenz, S. J., Luterbacher, J., Phipps, S. J., Servonnat, J., Swingedouw, D., Tett, S. F. B., Wagner, S., Yiou, P., and Zorita, E.: Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium, Clim. Past, 9, 393–421, 10.5194/cp-9-393-2013, 2013.

Frank, D., Esper, J., and Cook, E. R.: Adjustment for proxy number and coher-

ence in a large-scale temperature reconstruction, Geophys. Res. Lett., 34, L16709, doi:10.1029/2007GL030571, 2007.
Frank, D. C., Esper, J., Raible, C. C., Büntgen, U., Trouet, V., Stocker, B., and Joos, F.: Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate, Nature, 463, 527–530, doi:10.1038/nature08769, 2010.
Gao, C., Robock, A., and Ammann, C.: Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models, J. Geophys. Res.-Atmos., 113, D23111, 10.1029/2008JD010239, 2008.
Hargreaves, J. C., Paul, A., Ohgaito, R., Abe-Ouchi, A., and Annan, J. D.: Are paleoclimate model ensembles consistent with the MARGO data synthesis?, Clim. Past, 7, 917–933, 10.5194/cp-7-917-2011, 2011.
Hegerl, G. C., Crowley, T. J., Allen, M., Hyde, W. T., Pollack, H. N., Smerdon, J., and Zorita, E.: Detection of human influence on a new, validated 1500-year temperature reconstruction, J. Climate, 20, 650–666, 10.1175/JCLI4011.1, 2007.
Jones, P. D., Briffa, K. R., Barnett, T. P., and Tett, S. F. B.: High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures, Holocene, 8, 455–471, doi:10.1191/095968398667194956, 1998.
Jones, P. and Mann, M.: Climate over past millennia, Rev. Geophys., 42, RG2002, 10.1029/2003RG000143, 2004.
Juckes, M. N., Allen, M. R., Briffa, K. R., Esper, J., Hegerl, G. C., Moberg, A., Osborn, T. J., and Weber, S. L.: Millennial temperature reconstruction intercomparison and evaluation, Clim. Past, 3, 591–609, doi:10.5194/cp-3-591-2007, 2007.
Ljungqvist, F. C.: A new reconstructions of temperature variability in the extra-tropical northern hemisphere during the last two millennia. Geografiska Annaler: Series A, Physical Geography, 92: 339–351. 10.1111/j.1468-0459.2010.00399.x, 2010.
Mann, M. E. and Jones, P. D.: Global surface temperatures over the past two millennia, Geophys. Res. Lett., 30, 1820, doi:10.1029/2003GL017814, 2003.
Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, P. Natl. Acad. Sci., 105, 13252–13257, doi:10.1073/pnas.0805721105, 2008.
Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C.,

Faluvegi, G., and Ni, F.: Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly, Science, 326, 1256–1260, 10.1126/science.1177303, 2009.
Mann, M. E., Fuentes, J. D., and Rutherford, S.: Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures, Nat. Geosci., 5, 202–205, 10.1038/ngeo1394, 2012.
Masson, D. and Knutti, R.: Climate model genealogy, Geophys. Res. Lett., 38, L08703, 10.1029/2011GL046864, 2011.
Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikoajewicz, U., Notz, D., Pincus, R., Schmidt, H., and Tomassini, L.: Tuning the climate of a global model, J. Adv. Model. Earth Syst., 4, M00A01, 10.1029/2012MS000154, 2012.
Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlen, W.: Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, Nature, 433, 613–617, doi:10.1038/nature03265, 2005.
Schmidt, G. A., Jungclaus, J. H.,Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), Geosci. Model Dev., 4, 33–45, 10.5194/gmd-4-33-2011, 2011.
Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, E., Hargreaves, J. C., Harrison, S. P., Kageyama, M., LeGrande, A. N., Konecky, B., Lovejoy, S., Mann, M. E., Masson-Delmotte, V., Risi, C., Thompson, D., Timmermann, A., Tremblay, L.-B., and Yiou, P.: Using paleo-climate comparisons to constrain future projections in CMIP5, Clim. Past Discuss., 9, 775–835, 10.5194/cpd-9-775-2013, 2013.
Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philos. T. Roy. Soc. A, 365, 2053–2075, 10.1098/rsta.2007.2076, 2007.
Timmreck, C., Lorenz, S. J., Crowley, T. J., Kinne, S., Raddatz, T. J., Thomas, M. A., and Jungclausj, J. H.: Limited temperature response to the very large AD 1258 volcanic eruption, Geophys. Res. Lett., 36, L21708, 10.1029/2009GL040083, 2009.
Timmreck, C., Graf, H.-F., Lorenz, S. J., Niemeier, U., Zanchettin, D., Matei, D., Jungclaus, J. H., and Crowley, T. J.: Aerosol size confines climate response to volcanic super-eruptions, Geophys. Res. Lett., 10.1029/2010GL045464, 2010.
Wahl, E. R. and Smerdon, J. E.: Comparative performance of paleoclimate field and index

reconstructions derived from climate proxies and noise-only predictors, Geophys. Res. Lett., 39, L06703, 10.1029/2012GL051086, 2012.
Zanchettin, D., Bothe, O., Graf, H. F., Lorenz, S. J., Luterbacher, J., Timmreck, C., and Jungclaus, J. H.: Background conditions influence the decadal climate response to strong volcanic eruptions, J. Geophys. Res. Atmos., 118, 4090-4106, 10.1002/jgrd.50229, 2013.