Weather Clim. Dynam. Discuss.,
https://doi.org/10.5194/wcd-2019-19-RC2, 2020


# *Interactive comment on* "The representation of Northern Hemisphere blocking in current global climate models" *by* Reinhard Schiemann et al.

**Anonymous Referee #2**

Received and published: 25 February 2020

Review of "The representation of Northern Hemisphere blocking in current global climate models" by Schiemann et al.

This paper examines the current state-of-the-art, and the most recent prior state-of-the-art, in blocking simulation by climate models. The study is well designed, examining both CMIP ensembles and complementary experiments by PRIMAVERA models to test the impact of model resolution changes. It also usefully tests the robustness of the results to the choice of blocking index. The paper is very well written, well organized, and concise. I'm not sure if my comments & questions should be considered major or minor revisions. At any rate I think it should be very feasible for the authors to address these and I hope to see this paper published.

Main comments:



1. Overall I'm surprised by how little change there is in blocking biases from CMIP5 to CMIP6. The small size of the currently available CMIP6 ensemble (13 models) is a concern (although there's not much the authors can do about this). But it seems to me that a main conclusion from this study could be that the pervasive blocking biases documented in previous generations of models are more or less unchanged in these new ones. The authors do note a sizeable change in the median DJF Atlantic blocking frequency bias, but overall most of the models seem to look the same. However, having only 13 models for CMIP6 could mean that this (or other) changes aren't robust. Perhaps a test could be to see what spread in CMIP5 results is obtained by drawing random samples of 13 models from the CMIP5 ensemble - do the CMIP6 models consistently beat these?

2. Only a single ensemble member is used for each CMIP model, and a number of the PRIMAVERA experiments have one member. Are these samples big enough for robust results? More ensemble members are available for many of the CMIP models. On the line plots for PRIMAVERA models (left panel of each plot in Figs 3-6) could the uncertainty in each metric be shown? Perhaps a bootstrap test would work, randomly sampling years from each experiment to get a confidence interval for the metric. Many of the changes with resolution seen in these plots are marginal or differ widely among the models. These could represent genuine inter-model differences, but also sampling variability. Re. the boxplots of CMIP models (Figs 3-6), there are many cases where it looks like the CMIP6 distribution could be drawn randomly from the CMIP5. Hence it would be useful to do significance tests of the mean changes for these metrics.

3. For PRIMAVERA models, it's interesting that in many cases the AMIP runs show little or no change with resolution but the coupled runs do. One interpretation is that the mean state doesn't change as much with resolution in the AMIP runs because the SSTs are fixed, whereas they can change in the coupled runs. Can this be tested by looking at the mean states, e.g. by adding a mean-state metric to the boxplots? Perhaps RMSE of time-mean Z500 or similar.

Comments & suggestions by line number:

8: "and in" –> "during"

62: Should state if "high-resolution" refers to changing just horizontal resolution, as Table 1 seems to suggest, or if some of the PRIMAVERA models also increase their vertical resolution.

75: I'm surprised one historical simulation is enough for looking at persistence, given that some long-duration blocking events will be quite rare occurrences. Multiple historical ensemble members are available for many CMIP5 models and most CMIP6 models (some modelling centres have even submitted "large ensembles" of these). These extra members could be used to get more robust results, or at least to test if one ensemble member per model is sufficient.

Somewhere in Sec. 2.1: it would help to comment what are the ranges of resolution spanned by the CMIP5 and CMIP6 ensembles. Are the changes in blocking frequency between CMIP5 and CMIP6 roughly what would be expected, based on the PRIMAVERA analysis, given the change in resolution from CMIP5 to CMIP6?

115: "throughout" –> "at each gridpoint throughout" (if this is what you mean)

117: "Z500 anomaly" –> "Z500 anomaly at each gridpoint" (again, I presume this is what you mean)

148: In the Pacific high latitudes in some cases (in Fig 1) there's not much stippling, so this seems to be one region where models don't share the same bias.

150: My initial reaction to panels (e) & (f) of Fig 1 was that there's essentially no difference in DJF blocking frequency between CMIP5 & CMIP6, for both Atlantic and Pacific. Looking more closely I do see that CMIP5 has a larger area of negative bias in the Atlantic basin, but it's not a very big difference. It seems hard to believe this is a reduction of 1/3 in the bias, as the abstract says. Is that because the change in median (referred to in the abstract) is a lot more dramatic than the change in mean



(shown in Fig 1)? Fig 3, top left, seems to show a bigger shift in the median than the mean between CMIP5 and CMIP6, assuming the triangle is the mean. Perhaps it's worth plotting the difference in ensemble-mean blocking frequency between CMIP5 and CMIP6, and calculating its statistical significance.

151: "Fig. 2e,f" –> "Fig. 1e,f"

152: As with the CMIPs, I don't see very big differences between the different PRIMAVERA cases, so plotting differences & their significances could help. Given there aren't too many PRIMAVERA models, robustness could be a concern. However there is another issue with the H,L groups used to split up the PRIMAVERA models: in some cases the "low" resolution of one model group is similar to the "high" of another model group - e.g. the low-res CMCC models (64 km atm. grid) compared to the high-res AWI model (67 km atm. grid). So these composite groups, as plotted in Figs 1 & 2, are mixing together models of different resolutions. If changes in blocking frequency with resolution follow a similar pattern in all the models, then maybe this is ok. However Figs 3-6 suggest the models can differ widely in how blocking changes with resolution. What do the plots look like if instead models were grouped just based on their horizontal grid spacings?

157: It's hard to see much difference between CMIP5 and CMIP6 for summer blocking. The changes over the Baltic and Siberia are so small I wonder if they're just sampling variations. As with DJF, it would be useful to plot the difference & its statistical significance.

165: For DJF I see some differences in the Pacific during DJF among the PRIMAVERA models.

182: This could also be done with the HadGEM3 and EC-EARTH groups, is a similar uncertainty range obtained?

200: "mean and variability" –> "mean" (shouldn't ANOM focus on the variability?)

204: "underestimates" –> "underestimate"

236: "interval" –> "internal"

236: "comparatively large": as compared to the median?

240: "Moving on to summer (Fig. 4)" –> "For summer (Fig. 6)"

242: Refer to Sec. 2.4 to remind reader what a "blocking survival function" is. Also 2nd instance of "survival" misspelled.

243: I'm not sure that smaller ensemble spread indicates an improvement, given the CMIP6 models number less than half (13) than the CMIP5 models (29).

251: In Fig. S13, CMIP5 & CMIP6 look essentially the same to me.

254: "and improvement" –> "an improvement"

254: Again, I'm not sure that smaller ensemble spread for a smaller ensemble indicates improvement. Can this be quantified somehow - perhaps by drawing random groups of 13 models from the CMIP5 ensemble and seeing how likely it is to get a spread the size of CMIP6's by chance?

263: insert comma after "blocking"

271: Although blocking frequency has improved in CMIP6, given how similar the CMIPs look in Figs 1 & 2 I think calling it a "very clear" improvement may overstate the change.

271: "blocking" –> "blocking frequency" (since blocking persistence shows little change)

287: This sounds plausible, but is it actually happening in the PRIMAVERA models? Given that they aren't retuned at the higher resolution, it's possible that the opposite is true, that their mean states are actually worse at the higher resolution. This could be checked.

294: "blocking" –> "blocking frequency"



295: "increase in resolution in and of itself" –> "increase in resolution over the range considered here in and of itself"

300: This last sentence is kind of long, and speculative. Perhaps instead just say the results suggest that further model development is expected to reduce blocking biases.

304: I agree this is possible, but could a simpler interpretation might be: coupled models show larger sensitivity to resolution that AMIP models because the SST mean state improves with resolution in the coupled models (whereas it can't in the AMIP ones, because it's prescribed). This suggests the blocking changes with resolution are more to do with mean-state changes than changes in the variability. Do you think the results support this conclusion?

Figs 1 & 2: perhaps change contour levels so that zero isn't shown, e.g. begin at +/- 0.005. It might help reduce the appearance of noisiness in the plots.

Figs 3-6: define the different elements of the boxplots (is "box-whisker" plot a more typical name for these?). What is the triangle, quantiles for the box, the circles, etc. Panel labels (a,b,...) could be useful.

Captions in general: since so many of the plots are of the same type, it would be useful for just the first plot of each kind to have a full descriptive caption, and then for subsequent plots just say, "As Fig. X, but for [JJA, ANOM, etc]". Otherwise the reader has to go through each caption to find out if much has changed from the previous plot.

Interactive comment on Weather Clim. Dynam. Discuss., https://doi.org/10.5194/wcd-2019-19, 2020.