Response to reviewer 1

Review of "Dynamical drivers of Greenland blocking in climate models" by Clio Michel et al.

This manuscript investigates the drivers of atmospheric blocking in the Greenland region using several large ensembles of climate models. The representation of blocking and other dynamical features in each of the models is compared to ERA-Interim. Based on this analysis two models (ECHAM6.3-LR and MIROC5) are selected based on their performance in simulating historical blocking in the region of interest. The study continues to investigate several atmospheric features such as Rossby wave breaking and and wind speed in present and future in order to identify the drivers behind the expected changes in winter blocking in the Greenland region. Interestingly, the authors find that different dynamical features seem to be driving blocking in the two models considered.

This is relevant and timely work on the underlying dynamics driving atmospheric blocking for one of the main blocking regions in the northern hemisphere. I think the study fits well into the scope of Weather and Climate Dynamics and generally I think it should be published and I have no major comments. However, I found the manuscript to contain a large number of smaller issues, inconsistencies, and ambiguities that made it very hard to follow the authors points and sometimes even impossible to be sure what the authors were referring to. I would urge the authors to carefully address these and particularly focus on a consistent description of their methodology as well as their figures.

We thank the reviewer for the careful review of the manuscript and detailed comments. Our point-by-point responses are in blue here below. We will carefully revise the manuscript with a particular focus on the description of the methodology.

Comments

35: CMIP actually stands for Coupled Model Intercomparison Project as far as I know

True! Will be corrected.

84: As additional information the author could state how much warming that is approximately compared to the present decade used in the paper.

We will change the sentence.

Figure S1: I have to admit I've trouble interpreting that figure for several reasons:
- The caption is unclear to me: "Difference of blocking frequency where blocking is detected". Do both panels show blocking frequency or some kind of difference?
- I assume that the figure actually shows absolute blocking frequencies and not any kind of difference. So to show if there are "substantial changes in the results" (line 114 of the main manuscript) the author should plot differences (or even relative difference even though I realize that his might be problematic as well).
- It is unclear to me what the data-basis for this plot is. 'All-Hist' in the title lets me assume that it refers to all 125 members of NorESM1 in the historical period but this does not seem to be the case

as the blocking frequency looks too patchy for that – so is it possible that it is only one member?
- The high 'blocking' frequency in the Atlantic below 40 degree north is not normally defined as blocking (at least not in the classical sense), right? The authors focus on blocking over Greenland so I would suggest optimizing the colorbar to the blocking frequency there.

Figure R1 below (Fig. S1 of the original supplement) represents the DJF mean blocking frequency for the first member of the present-day experiment only. Following the reviewer's suggestion, we have added a third panel to the figure that shows the difference between the two blocking frequencies (see Fig. R1 below). Moreover, the colorbar of the top two panels has been adjusted to better highlight the blocking frequency north of 40°N, the unit is added to the figure, and the caption has been rewritten to make it clearer and add the third panel's description.

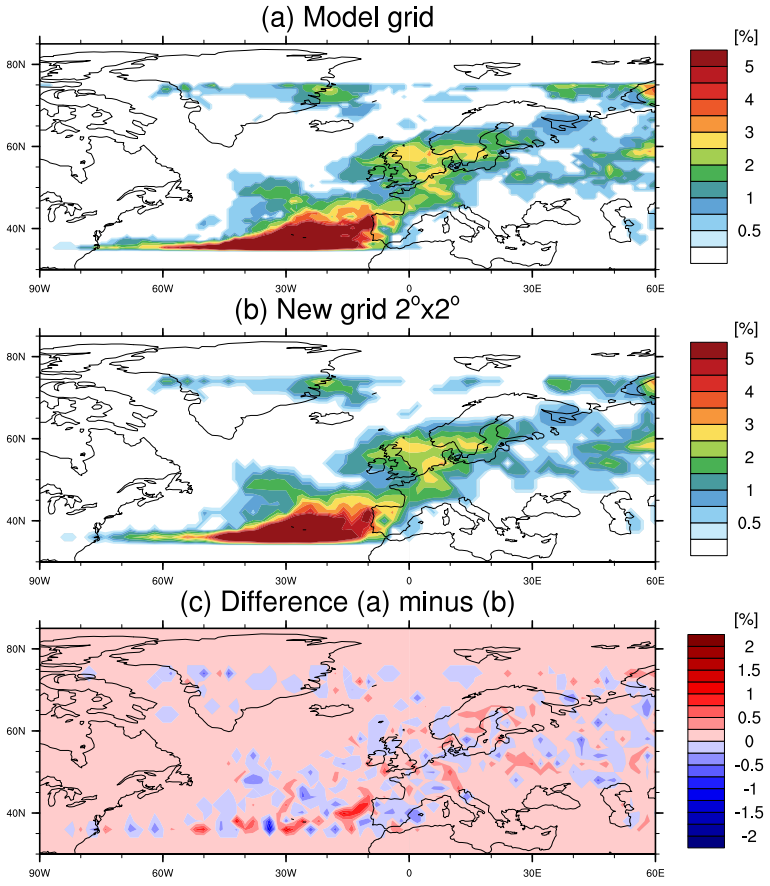

Figure R1: DJF mean blocking frequency (unit: % of winter days) for member 1 of the NorESM1-HAPPI ensemble, with blocking detected using the 500-hPa geopotential on (a) the model grid and (b) a grid of 2° × 2°. (c) Difference between the two mean frequencies (a) minus (b) where (a) has been remapped on the same 2° × 2° grid as (b).

Section 2.3: Did the authors calculate the annual blocking frequency and then select Greenland blocking days in DJF from that or was the blocking frequency already calculated only for DJF? In the second case the first and last 4 days of each season will have a lower probability of blocking due to the 5 day persistence criterion I assume?. This should be made clear.

We detected blocking with a 5-day persistence using the whole 10 years of each member. Then, we selected the days occurring during the 9 winters (January, February 2006 and December 2015 are excluded to only keep full winter seasons) such that there is no issue with the first and last 4 days of each DJF season. We will rewrite this part of the method section.

133: I assume GB stands for Greenland blocking?

Yes, we will change it to *"Greenland blocking"*.

134: Why is there a difference between the blocking and the anticyclone definition? Should blocking not just be a stationary anticyclone?

Yes, blocking can be considered as a stationary anticyclone. However, the blocking detection method used in this study relies on the overturning of a geopotential contour, which does not necessarily happen when there is an anticyclone. That is why we use an objective detection algorithm that is identical to a cyclone detection except that it detects highs in the 500-hPa geopotential anomaly field instead of lows in the mean sea level pressure field. Looking at the anticyclone frequency can tell us if there are anticyclones over Greenland that are not linked to an overturning of a geopotential contour and without any minimum persistence. We will add a short paragraph in the method section (2.5) to better justify our use of this anticyclone detection.

150: Figure S2 does not show that "none of the 30 climatologies of 9 consecutive winters [...] is significantly different from the total 40-year [...] climatology". as the authors seem to indicate here it just shows the blocking climatologies. How was this established? Was that done on a grid-point level or somehow averaged? Would we not even expect about 10% of cases to be outside of the 90% range?

As written in the caption of Fig. S2, we conducted a t-test that assessed the difference between each of the 9-winter blocking frequency climatologies and the 40-year (or actually 39-winter) climatology of the blocking frequency. Therefore, each grid point is tested and nowhere the two climatologies were found to be significantly different.

Figure 1:
- I personally think that there should be an indication of what is shown as well as the units directly in the figure and not just in the caption.
- "Black lines show the 2, 4 and 6% contours for ERA-Interim (2006-2015)." For me these lines appear grey but that might just be my viewer.
- In panel f the black/grey contours do not seem to match the respective shading. This might just be due to different interpolation but it maybe should be clarified.
- I assume that the white areas without dots just have zero blocking frequency (rather than significant biases)?
tt

The units will be added next to the colorbar not only to Fig. 1 and also on all the other figures of the revised manuscript and supplement. The contour lines are indeed grey. The contours do not match the shading because they have been smoothed. Yes, the white areas without dots do not have blocking. All these details will be added to the caption.

165: Could the authors rephrase this? I did not count but it looks to me as if for all models there are more dotted than un-dotted gird cells even when only focusing on 'blue' areas so I don't think the concluding that "Most of the negative biases are significant" (except for MIROC5) is supported as it is.

We will rephrase the sentence.

168: This seems to partly contradict the earlier statement that 9 years are enough to establish a robust blocking climatology, right? (even though the argument there was only made for observations). I would argue that for a sufficiently long climatological period the blocking frequency should be a model property and not depend on the realization, right?

As written by the reviewer, we showed for ERA-Interim that the overlapping 9-winter blocking frequency climatologies are not significantly different from the 39-winter climatology. This was performed to justify the fact that we will use ERA-Interim's period 2006-2015 (the period of the simulations) to calculate the biases. In line 168, we look at the ensemble spread. Therefore, it is not surprising that some members better reproduce ERA-Interim's climatology than others. Earlier at the end of Section 2.1, we state that the use of large ensembles overcomes the fact that nine winters might be too short to obtain realistic blocking frequency climatologies, even though Davini and D'Andrea (2016) found that nine winters should be enough to get relatively accurate blocking statistics.

Figure 2: I am again not sure what is shown in this figure. The caption suggest to me that it shows the area average of blocking frequency over the GB region: "mean Greenland blocking frequency" but this is not correct, right? Frequencies are too high for that, so I assume it actually shows Greenland blocking days as defined in line 116? Please clarify.

The figure shows the percentage of Greenland blocking days during the nine winters of each decade, hence a Greenland blocking frequency. We will rewrite the caption of the figure.

Figure S3: The unit of the shading should be somewhere in this figure.

We have included the unit of the shading in both the figure and the caption.

Figure 3: I assume that for ERA-Interim not the ensemble mean but the mean over the different periods is shown?

True, the caption will be rewritten.

211: "MIROC5 and ECHAM6.3-LR are the models with the largest variability on the equatorward side of the mean jet and the smallest mean-state biases with respect to wind variability." Can the authors go into a bit more detail here? Looking at figure 3 I am not sure if I can see what the authors mean with this statement. In particular the standard deviation from MIROC5 does not seem to be higher southwards of its climatological jet than any other models'.

We meant that the larger standard deviation of the zonal wind at 850 hPa shows that the eddy-driven jet can shift south of its climatological position more often for ECHAM6.3-LR than MIROC5. Figure R2 represents more clearly that the models with the largest standard deviation of the 850-hPa zonal wind south of the mean jet at 30°N and most similar to ERA-Interim are ECHAM6.3-LR and MIROC5. We will add this figure in the supplement. The sentence will be modified for clarity.

216: "Climatologies show that AWB is more frequent and located on the equatorward side of the jet at both low and upper levels (solid contours in Fig. 4 left)" What is meant by "low and upper levels"

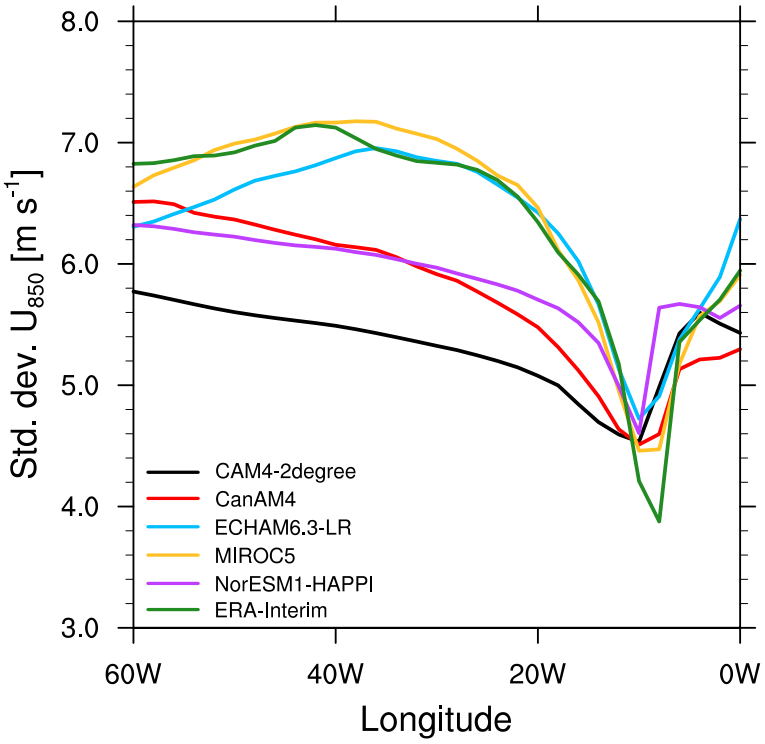

Figure R2: Ensemble mean of the DJF mean daily standard deviation of the 850-hPa zonal wind at 30°N (units: m s^{-1}) as a function of the longitude (covering the North Atlantic domain) for every HAPPI ensemble. The green line shows the DJF mean daily standard deviation for ERA-Interim (2006-2015).

and where is this shown in figure 4? This sentence somehow seems to suggest that the solid contours in figure 4 are an indication of the jet but to know the position relative to the jet the reader actually needs to compare figure 4 to figure 3f, right?

The sentence is indeed not clear and will be changed.

Section 3.3: It would be interesting to discuss these results in the light of this recent study as well: https://onlinelibrary.wiley.com/doi/10.1029/2020JD034082

See response to comment about Figure S8a below.

l251: "ECHAM6.3-LR [...] slightly underestimates the Greenland blocking frequency" I'm not sure if I would call a mean underestimation of about 50% (if I read figure 2c correctly) 'slightly'.

The reviewer is right if we consider Fig. 2c (or Table 1). However, Fig. 1c shows an underestimation of the ensemble mean Greenland blocking frequency of -2 to -3%. As Fig. 1c is the type of figure usually shown in the literature dealing with blocking frequency and where a 'more regular' blocking detection is used (not using the blocking index defined in this study), we will refer to Fig. 1c in this sentence. Therefore, we will change the sentence.

Figure S8a: Again it would be interesting to see if the results are consistent with Drouard et al. (2021) when split by AWB and CWB.

We thank the reviewer for this comment. Several studies, including the recent Drouard et al. (2021), showed or mentioned that Greenland blocking are driven mainly by CWB. Therefore, we do

not think that splitting the blocking events in CWB and AWB type would yield additional insight into the problems addressed in the present manuscript. Also, previous studies showed that Greenland blocking lasts longer than other blocking (e.g. Scandinavian blocking), in accordance with the study of Drouard et al. (2021), a fact that we will mention in the introduction.

291: CMIP5 was published around 2013 if I'm not mistaken. Can the authors check if the cited studies really all look at CMIP5 data?

Indeed, the studies of Sillmann and Croci-Maspoli (2009) and Barnes et al. (2012) used CMIP3 simulations and we will add the CMIP6 results as well.

292: "we note a decrease in the ensemble mean blocking frequency over Greenland, in particular for ECHAM6.3-LR (up to -0.6% of the time)..." Is this referring to Greenland blocking days as defined in line 116 or to blocking frequency directly as I would assume from this statement?

The percentages refer to the responses (Future minus Present experiments) of the ensemble mean blocking frequency as seen over Greenland in Fig. R3, figure that will be added to the Supplement and cited in the main text. The figure shows the ensemble mean percentage of blocking days relative to the total number of winter days (90*9=810 days). We will remove the confusing "of the time" in the parenthesis but add the reference to the figure.

295: It seems to me that not the change relative to pre-industrial is important but the change relative to the historical period?

We refer here to the pre-industrial period as the this is the reference for the +2°C target. Following this comment, we noticed that the temperature increases given for the end of the century were relative to the historical period 1986-2005 and not the pre-industrial 1850-1900. We will change the sentence to keep the reference of the pre-industrial period.

302: "The composites over the blocked days are very similar between the present and future experiments (Fig. 7)" Figure 7 does not show the present?

The sentence is indeed not clear. The reader has to compare Fig.7 to the left and middle columns of Fig. 5 . We will add this information.

Figure S11: I assume this figure shows change between historical and future?

Yes, the caption will be changed accordingly.

311: "Over the North Atlantic, CWB do not seem to change much (very weak values and noisy field; not shown)" What does 'not shown' mean in this context? CWB is shown in figure 8 and S11, right? There just happen to be no contours in the frame for the case of figure 8b? Or do I misunderstand something? What does 'seem to change' mean? In a given model it either changes or not, right?

As guessed by the reviewer, the response of the CWB frequency is small and does not appear with the contour interval chosen. We will change the text.

336: Except for AWB in MIROC5, right?

Correct, the sentence will be changed accordingly.

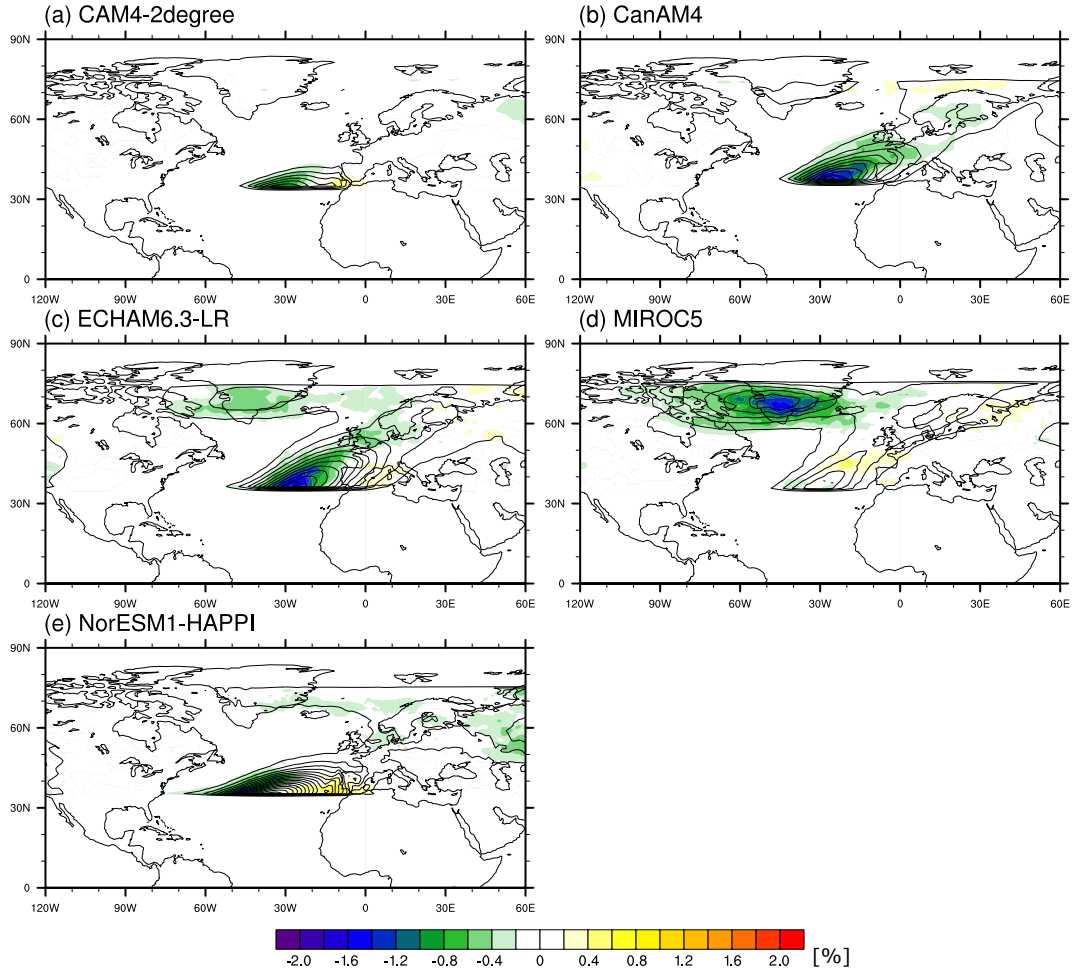

Figure R3: Ensemble mean response (shading, Future minus Present experiments) of the DJF mean blocking frequency (percentage of blocking days relative to the total number of winter days, in %) for the five HAPPI models. The contours show the ensemble mean DJF mean blocking frequency (first contour and interval: 1%).

Technical comments:
42: 'have showed'
Figure 6: "(g,h,i) Same as (c,d) but for anticyclonic wave breaking frequency"
288: "CWB are"

Thank you, they will all be corrected.