## Response to reviewer 2

Review of "Dynamical drivers of Greenland blocking in climate models"

The present manuscript explores the deriver of Greenland blocking events using a set of multi-model large ensembles. This is a nice way to analyze blocking events as they are rare events. The authors showed that two (MIROC5 and ECHAM6.3-LR) out of five models fairly capture the frequency of blocking events compared to the ERA-Interim reanalysis. Then they used these two models to understand the main driver of Greenland blocking events and found an inconsistency in the dynamics of blocking events between the models. More specifically, the author found a link between CWB and Greenland blocking events in ECHAM6.3-LR model in contrast to MIROC5, which CWB is largely underestimated. They further expand their analysis to future climate and report the changes in RWB and eddy-driven jet to investigate the connection with blocking frequency.

I found the topic and contents relevant, and the paper is falling within the scope of Weather and Climate Dynamics. I have no major comments, and I strongly suggest that for publication after some minor revisions, which I believe are needed to improve the quality of the paper and make it clear.

The authors thank the reviewer for the detailed review and constructive comments. Our point-bypoint responses can be found in blue here below.

L7: The authors mentioned "four out five models" but then throughout the paper, they focused on the two models which they believe have better performance compared to the other models. Even in conclusion (L324) they mentioned "three out of five" and I suggest they change it in the abstract to be consistent with the analyses in the paper.

We agree and will modify the abstract in order to mention only the two models ECHAM6.3-LR and MIROC5.

L42: "have shown"

Will be corrected.

L85: I think it sounds better if the authors rephrase the sentence and mentioned 100 and 501 members instead of "times"

We agree and we will change the sentence accordingly.

L112: Shouldn't it be "Blocking grid points are identified" instead of "Blocking events are identified" as the index identifies each grid point that satisfies the condition, which does not necessarily mean they are separate events.

Yes, we will now use "Blocked grid points" instead of "Blocking events" as suggested by the reviewer, which is also now in better agreement with another sentence, further down, where we also used "grid points".

L115: I am unsure if I can follow the time fraction (%) definition that the authors explained here. Since it is an important point, they need to explain that clearly. Is the fraction defined as (total blocked gridpoints/(total gridpoints\*number of days))?

We agree that the wording was confusing and we will change the sentence.

L116: Have the authors thought about performing clustering analysis to separate Greenland blocks instead of subjective criteria to count the number of blocked days as I think the Greenland region will show up naturally as a region of frequent blocking events base on what the authors have shown in Fig. 1(f).

Despite recent improvements in weather regimes representation by high-resolution climate models (Fabiano et al. 2020), models with lower resolution struggle in representing the weather regimes (Dawson et al., 2012). An alternative would have been to project the geopotential anomalous field on the Greenland anticyclone regime obtained from ERA-Interim but, for consistency, the clustering should be performed with the 10 years or 9 winters of ERA-Interim, which is a short period to obtain reliable regimes. Moreover, this would have attenuated the differences in the blocking pattern itself among the models. In addition, it is not evident that the clustering analysis would give the same Greenland blocking events as the detection method used here. Therefore, we opted for a simple blocking index.

L133: "GB" spell it out

Yes, thank you for spotting it.

L166: Over Greenland, negative bias in the frequency of blocking is not significant in the ECHAM6.3-LR model in addition to MIROC5, and it needs to be added. Also, in the following sentence (L168) the author mentioned "strong negative biases in the ensemble mean." which I think is not true for those two models over Greenland based on what has been shown in Fig. 1.

The reviewer is correct and we will change the sentence accordingly.

L211: I believe that ECHAM6.3-LR is the only one that has the smallest mean-state bias in Fig. 3, and biases in MIROC5 are the same as the other models (specially CanAM4 seems to be even better compared to MIROC5).

To clarify this sentence, we have represented the standard deviation taken at 30°N (south of the mean jet position) for all models (see Fig. R1 hereafter). This figure shows that MIROC5 and ECHAM6-3-LR are the models that performs the best compared to ERA-Interim, especially between 40°W and 20°W. We will add this figure to the Supplement and refer to it in the main text.

L247: Fig. 5d-f not Fig. 5c-e

Will be corrected.

L269: Can weak absolute vorticity gradient be why the frequency of blocking events is captured with less bias in MIROC5? If yes, I think the authors need to expand this part.

We will add more details about the study of Luo et al. (2019) such as: "In an idealized set-up, Luo et al. (2019) showed that a weak mean meridional gradient of potential vorticity at high latitudes



Figure R1: Ensemble mean of the DJF mean daily standard deviation of the 850-hPa zonal wind at  $30^{\circ}$ N (units: m s<sup>-1</sup>) 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).

leads to reduced energy dispersion, enhanced nonlinearity, and more persistent eddy forcing, favouring long and intense blocking. Even though MIROC5 does not exhibit more intense or longer blocking than ECHAM6.3-LR, this mechanism could also trigger blocking thus enhancing its frequency."

L275: Fig. 6m-o not Fig. 5l-n

Will be corrected.

L279: I think the authors mean Fig. S8 instead of Fig. S9.

Yes, it will be corrected.

L293: I understand the numbers are not significant compared to previous studies but are the changes significant compared to each models' current climate?

In this sentence, we actually mean that the changes in blocking frequency over Greenland between the present and future experiments are weak and not significant. In addition, the values are also lower than the values found in previous studies such as shown in the recent paper of Davini and D'Andrea (2020) because the warming between the HAPPI experiments is of  $1.2^{\circ}$ C, which is lower than the warming compared to the pre-industrial in the CMIP5/6 experiments (+2.6 to 4.8°). We will make this part clearer.

L297: It is unclear what the authors refer to by saying "large decrease". Isn't it a weak decrease as they found here, and it is mentioned in L293?

Yes, the decrease is not large. We will remove the word "large".

Last comment: Captions of figures are vague and ambiguous, making it hard for readers to follow. They should be clear as much as possible that readers understand the figures just by reading the captions. Also, if the anomalies are significant (by any t-test) please mentioned that in the caption, otherwise please do statistical significance analysis for those cases that anomalies are shown.

We will improve the captions of the figures and add the information about the significance (e.g. t-test or bootstrap). We will also perform a significance analysis for the composites (Figs. 4, 5, and 6) using a bootstrap method. The lack of significance to a 10% level will be shown with dots. From first calculations, it seems that the anomalies in the composites are all significant (see e.g. Fig. R2).



Figure R2: Ensemble means of the circulation anomalies during GB days for (left) ECHAM6.3-LR, (centre) MIROC5, and (right) ERA-Interim. (a,b,c) 500-hPa geopotential (contours are drawn every 100 m from 5000 to 6000 m) and anomalies (shading, in m). (d,e,f) Cyclonic wave breaking frequency (first contour and interval: 5%) and anomalies (shading, in %). (g,h,i) Same as (d,e,f) but for anticyclonic wave breaking frequency. (j,k,l) Zonal wind at 850 (first contour and interval: 4 m s<sup>-1</sup>, zero contour omitted and dashed contours for negative values) and anomalies (shading, in m s<sup>-1</sup>). Anomalies are deviations from the 10-year DJF climatology and only members with at least one blocked day are used for the composites. Dark gray dots show non-significant areas at the 10% significance level with the significance calculated using a bootstrap method.