Response to Reviewers:
On the role of moist and dry processes for atmospheric blocking biases in the Euro-Atlantic region in CMIP6

We would like to thank the reviewers for taking the time to review this manuscript, we sincerely appreciate your constructive feedback. Your recommendations have been addressed in the updated manuscript. The text below (in blue) documents the exact amendments that we have made in response to your suggestions. The figure and line numbers refer to the original manuscript.

Reviewer 1

This paper studies blocking biases in the wintertime Euro-Atlantic sector in CMIP6 models,. They argue that mean state biases and inadequate representation of moist processes within the models contribute to underestimation of Euro-Atlantic blocking frequency. This paper is well written and executed and the subject matter is consistent with the scope of WCD. My comments are minor. I enjoyed reading this paper and think that it adds to the literature on this topic. Please find my comments below.

My primary concern is that the results presented are consistent with the hypotheses suggested (blocking biases due to lack of WCBs and overly zonal jets), but that their causal relation is more difficult to establish and will likely require carefully designed model experiments to test. For example, one could imagine that a lack of blocking can contribute to a southward shifted mean jet. That is, correlation does not equal causation. I therefore suggest some discussion of this nuance in the paper. The authors could even suggest experiments to test these ideas.

We agree that establishing a causal relationship is not simple, as there are always feedbacks between the different synoptic features analyzed in this manuscript. Therefore, we try to approach this topic with caution using terms such as "linked" or "associated". However, we argue the existence of a causal link from mean state to the blocking bias, namely that the mean state is not only influenced by blocks but also by the representation of the topography and the land sea distribution, tropical convection. Coarse resolution models cannot capture these factors with needed accuracy. We make this clear in the discussion section as follows:
"Establishing the direction of causal relationships is challenging due to the complex feedbacks between the various synoptic features analysed in this study. However, we argue that a causal connection exists between the mean state and the blocking bias. Specifically, the mean state is not only influenced by atmospheric blocks but also by factors such as the representation of topography, land-sea distribution, and tropical convection. Several studies show that coarse-resolution models struggle to accurately capture these elements (e.g., Scaife et al., 2011; Pithan et al., 2016; Kleiner et al., 2021). Carefully designed sensitivity studies would be needed to quantify the causal relationship between diabatic processes and Euro-Atlantic blocking biases. A first step could be numerical experiments testing the overall sensitivity of the large-scale flow to WCB activity and its associated latent heat release. A rather subtle approach would be experiments with and without stochastic perturbation schemes in the region of WCBs. For example, Dawson and Palmer (2015) and Christensen et al. (2015) show that stochastic perturbations in general improve the representation of blocking in the Atlantic-European region. However, we still lack an understanding of exactly how diabatic

processes contribute to this improvement."

Minor comments:

- Lines 93-99: Is there a reason that only models with an adequate representation of blocking are studied and do you expect these results to generalise to all models? This is quite a small subset of the full CMIP6 ensemble.

  We acknowledge that the sample size is relatively small. Although running ELIAS2.0 requires fewer computational resources compared to the traditional trajectory approach, it still necessitates access to multiple variables across various pressure levels. Therefore, we selected models based on their performance in the Euro-Atlantic region, prioritizing those that both perform well and provide the necessary 3D variables. Additionally, we focused on models with independent components. We anticipate that expanding the sample to include more models would yield similar results, as biases in background flow and dry processes have been consistently identified in previous studies.

- Lines 127-134: Is there a reason that meridional gradients of Z are used for the background flow rather than simply zonal wind? Could the authors explain this as surely these will be roughly equivalent under geostrophic balance?

  We selected the gradients of Z because the flow reversal used to identify blocking, is derived from these gradients. This is mentioned in line 134.

- Figure 1 caption – the blocking frequency for the ERA5 climatology seems quite low, only around 3% at most for blocking over northern Europe. I assume that this is a typo and should be contoured every 5% as typical blocking frequencies are more like 10-15% (e.g. Woollings et al 2018, figure 2)?

  We acknowledge that the blocking frequency is lower than reported by Woollings et al. (2018). However, the caption is accurate. This discrepancy arises because we applied the method proposed by Brunner and Steiner (2017), which introduces a third gradient to exclude slow-moving, low-latitude ridges from being classified as blocking. As a result, our blocking frequency is lower compared to Woollings et al. (2018). We have added a sentence in section 4.1 to clarify this point.

References

Woollings, T., Barriopedro, D., Methven, J., Son, S.W., Martius, O., Harvey, B., Sillmann, J., Lupo, A.R. and Seneviratne, S., 2018. Blocking and its response to climate change. Current climate change reports, 4, pp.287-300.

Reviewer 2

In the manuscript, the authors investigate the representation of atmospheric blocking in CMIP6 models and explore the role of moist and dry processes in the climate model biases. The authors consider the representation of the background flow, transient wave activity and warm conveyor belts (WCBs) to elucidate the climate model biases relating to both dry and moist processes. WCBs are identified using a machine learning technique as data is not available from the CMIP6 models to calculate the Lagrangian trajectories typically used to identify WCBs. It is shown that there are consistent biases in the climate models that together result in an underestimation of block frequency over Europe: the jet stream is located too far south in the models and extends further downstream; eddy activity peaks further equatorward in the models and act to accelerate the flow in the peak blocking region; and the models have a strong negative bias in WCB outflow across the North Atlantic/European region.

I very much enjoyed reading this paper: it is very well written, well structured and the figures are clear and support the main conclusions drawn by the authors. I believe the manuscript is a nice addition to the literature relating to climate model biases in atmospheric blocking and will be of interest to many readers of WCD. I have a couple of relatively major comment and some smaller comments that I would appreciate the authors response to. After addressing these comments I think the paper will be suitable for publication in WCD.

Major comments:

1. This comment relates to the identification of WCBs in the climate models.
   The machine learning technique used to identify WCB inflow/outflow was built/verified using ERA interim, i.e. trained to identify WCB trajectories that have the same characteristics (such as total scent, ascent rate) as those in ERA interim. As I understand, we do not know if the technique can accurately identify WCBs in the climate models. For example, if the WCBs in climate models tend to ascend slightly slower or do not ascend by as much as in ERA interim then these will be missed and contribute to the large negative bias in WCB outflow in the CMIP6 models found here. Of course, this would still be a bias that is related to moist processes in the WCB and will contribute to the blocking biases but is nonetheless a different bias to just a gross underestimation of WCB outflow. I assume it is not possible to verify that the machine learning method identifies similar WCB characteristics to a trajectory tool such as LAGRANTO in climate model output, which would be the best option. So, I do not expect the authors to do this, but some more discussion about the potential caveats of using the machine learning method on the climate models is needed.

   Thank you very much for this comment. We agree that a broader discussion of ELIAS2.0 limitation and the interpretation of the results is needed. Ideally, ELIAS2.0 would provide not only information about the presence of WCB inflow, ascent and outflow but for example would give information about the density of WCB trajectories at a certain location or about the outflow height. Unfortunately, ELIAS2.0 was not trained for such types of analysis and a retraining would be necessary. Accordingly, we now include a more detailed discussion of the results and also add to the interpretation that it may be a too low outflow height that contributes to the identified blocking biases (see Section 5 of the manuscript). In general though, we are quite confident concerning the interpretation as we have shown in previous studies that biases identified with ELIAS2.0 match biases identified with LAGRANTO (e.g., Quinting and Grams, 2022; Quinting et al., 2024).

2. My second major comment relates to causality statements.
   The main argument proposed by the authors is that the climate models have a too-zonal background flow and this drives a too far equatorward maximum in eddy activity which reduces the diffluent flow associated with blocking events in the main blocking region, as well as driving a reduced WCB outflow in the region. Could the background flow biases not be a symptom of too-zonally propagating eddies/cyclones and not necessarily the cause? The jet position may be biased in climate models for a variety of reasons and a southward/zonal bias in eddy-driven jet latitude would result in cyclones being steered across the Atlantic with the same bias. The background flow would then appear to

have a southward and zonal bias, as it is made up of the average behaviour of the eddies. Some more discussion relating to this causality is needed I think.

We thank you for your comment. This is similar to the primary concern of the Reviewer 1. We have expanded the discussion as follows:
"Establishing the direction of causal relationships is challenging due to the complex feedbacks between the various synoptic features analysed in this study. However, we argue that a causal connection exists between the mean state and the blocking bias. Specifically, the mean state is not only influenced by atmospheric blocks but also by factors such as the representation of topography, land-sea distribution, and tropical convection. Several studies show that coarse-resolution models struggle to accurately capture these elements (e.g., Scaife et al., 2011; Pithan et al., 2016; Kleiner et al., 2021). Carefully designed sensitivity studies would be needed to quantify the causal relationship between diabatic processes and Euro-Atlantic blocking biases. A first step could be numerical experiments testing the overall sensitivity of the large-scale flow to WCB activity and its associated latent heat release. A rather subtle approach would be experiments with and without stochastic perturbation schemes in the region of WCBs. For example, Dawson and Palmer (2015) and Christensen et al. (2015) show that stochastic perturbations in general improve the representation of blocking in the Atlantic-European region. However, we still lack an understanding of exactly how diabatic processes contribute to this improvement."

Minor comments:

1. L10: you write "we define the background flow as the most frequent value of the latitudinal gradient of the geopotential..". I'm not sure what is meant by this. Do you mean the most frequent value of the maximum latitudinal gradient?

   First, we calculate the rate of change of Z in the meridional (latitudinal) direction. Then, we determine the mode (the most frequent value) at each grid point. Specifically, for each grid point, we analyze the entire daily time series from 1979 to 2014, grouping the data into bins of 5 mkm^{-1}. The most frequent bin is then identified as the mode. However, we agreed it needs more details, thus, we have have replaced Line 10 with:
   "..., we define the background flow as the most frequent value of the daily time series of the meridional gradient of geopotential height at 500hPa."

2. L59: some readers may not be familiar with "anomaly blocking indices" or even blocking indices in general, consider adding a brief description here.

   Thank you for pointing this out. We have added a brief description as follows:
   "There are two commonly used types of indices: absolute and anomaly indices (e.g., Barriopedro et al., 2010). Absolute indices rely on absolute fields, such as flow gradients, to identify the circulation associated with blocking events. In contrast, anomaly indices detect deviations in variables like geopotential height (Z) or potential vorticity (PV) from a given climatological baseline."

3. L75-85: you mention how WCBs/diabatic processes are important for blocks in ERA5 and that this has not been assessed in climate models much yet (partly due to the difficulty in identifying WCBs in them). I wonder if this area has been studied in numerical weather prediction models and whether we should expect similar in climate models based on that. Some discussion around this could be interesting.

   Thank you for this comment. Indeed, blocking biases in numerical weather prediction models have been linked to WCB biases. For example, Wandel et al. (2024) show that subseasonal forecasts that miss the onset of a blocking over Europe typically underestimate the WCB activity over the North Atlantic several days before. Vice versa, successful forecasts of European blocking at extended-range lead times go in line with accurate WCB forecasts over the North Atlantic several days prior to the blocking onset. Interestingly, and this relates to this study, Wandel et al. (2021) found a systematic underestimation of WCB outflow frequency in the subseaonal prediction system of ECMWF which is similar to the biases identified here. We now refer to their study in Section 5 of the manuscript.

4. L94: it might be beneficial to briefly describe how Palmer et al. (2023) do their selection. To save the reader having to look up the paper themselves if they are unfamiliar.

   We will add the following details: “Palmer et al. (2023) used a combination of quantitative metrics—such as RMSE, bias, variance, and correlation—and qualitative assessments, including the examination of circulation wind patterns. Based on these evaluations, the models have been categorized into four classifications: inadequate, unsatisfactory, satisfactory, and not available. We analyse only models with adequate performance, excluding the inadequate models, as suggested by Palmer et al. (2023).”

5. L121: do you include all latitudes between 75S and 75N?

   Yes, all the latitudes are included but the third gradient ($\Delta Z_E\left(\lambda,\phi\right) = \frac{Z(\lambda,\phi-2\Delta\phi)-Z(\lambda,\phi-\Delta\phi)}{\Delta\phi}$)) introduced by Brunner and Steiner (2017) helps to remove low-latitude ridges.

6. L138: is this a common way to compute the storm tracks? Does it differ much from approaches that track cyclones and then construct the storm tracks from the cyclone counts?

   Yes, the Eulerian approach used in this study is a common metric to identify the stormtracks (e.g., Hoskins and Hodges, 2002; Greeves et al., 2007; Davini et al., 2017; Harvey et al., 2020). There are notable differences between Eulerian and Lagrangian metrics. For example, the Eulerian metric tends to miss weak cyclones in the Mediterranean region, and it places the maximum intensity near the eastern coast of North America. In contrast, the Lagrangian metric positions this maximum intensity south of Greenland (see Hoskins and Hodges, 2002; Greeves et al., 2007; Zappa et al., 2013; Dolores-Tesillos et al., 2022). We discuss this in line 220-223.

7. L220-225: how large is the spread in storm track/jet biases among the CMIP6 models? You indicate where the majority agree on sign of the bias but it would also be interesting to see how the magnitude/spatial structures vary among the models.

   Yes, this a good point. We have prepared a new figure (Figure 1), where we show the stormtrack biases of each model. This will be part of the new Supplementary material. Figure 1 shows that the model overall capture the two zonal storm tracks.

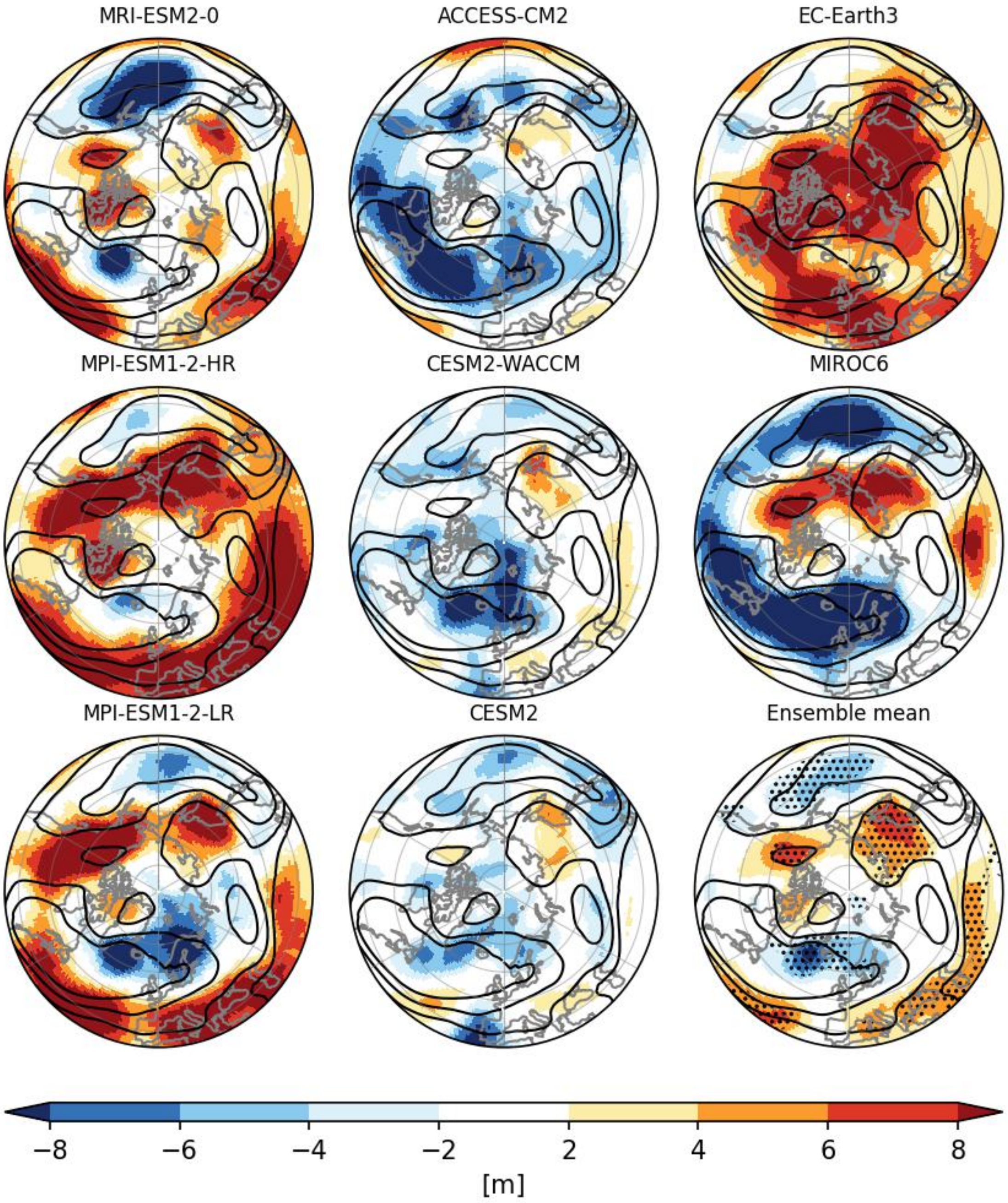

Figure 1: a-h) Storm tracks (standard deviation of the band-pass filtered Z at 500 hPa in units of m) biases of individual CMIP6 models and i) multi-model mean against ERA5 (1979-2014). Black contours depict the ERA5 storm tracks (every 10 m, starting from 40 m). Stippling denotes regions where more than 80% (6 models) of the ensemble members indicate a bias of the same sign.

8. L280: could these Lagrangian WCB trajectories from Joos et al. (2023) be used to verify the machine learning algorithm works okay on climate model data.

   We believe it is possible to use CESM-LENS to validate ELIAS2.0 and we actually mention this aspect when discussing the caveats of ELIAS2.0 (Section 5). However, this would not affect our main conclusions, as our focus is on identifying the sources of blocking biases rather than verifying ELIAS2.0 itself.

9. Fig. 4 caption: what are the relative biases compared to?

   We thank you for noticing this. They are relative to ERA5, we have added it in the caption of Figure 4.

Technical correction:

1. L93: sub-set —¿ subset

   Thank you, we have replaced “sub-set” with “subset”.

Reviewer 3

In this study, the role of biases in the background flow, dry and moist processes for the misrepresentation of atmospheric blocking in a subset of CMIP6 models is investigated. The misrepresentation of blocking in Climate models is of high scientific relevance and this study is a very valuable contribution to the scientific literature which contributes to an increased understanding. The data and methods which are used are suitable, the paper is well structured and well written. I therefore recommend the acceptance with Minor revisions.

Major comments:

1. Performance of ELIAS2.0 with climate model simulations.
   Of course, I fully understand that it is not possible to calculate Lagrangian trajectories to identify WCBs in different climate models due to the high computational costs and because the necessary output of wind data on model level and a time resolution of 6h is often not available. However, it would be beneficial for the paper to discuss more carefully possible biases in the WCB climatology in CMIP6 which might arise from the calculation of WCBs with ETLIAS2.0. For example, from Joos et al., 2023 we know that WCB trajectories in ERA-Interim ascend faster than in CESM1. In case this would be similar in the CMIP6 models, would that have an impact on the predicted ascent and outflow regions? This could be added to the discussion (line 311 onwards or extend what you already wrote in paragraph lines 332. . .).

   Thank you for this suggestion. We absolutely agree that different ascent rates and resulting lower outflow heights could affect the interpretation of the results. For example, if the ascent rate is similarly underestimated as found in Joos et al., 2023 this would result in lower outflow heights which are not necessarily captured by ELIAS2.0. In order to identify the outflow height with ELIAS2.0, this would require a full retraining of the machine learning model which is beyond the scope of this study. Still, we very much acknowledge your input and modified the discussion following lines 369.

2. Discussion of causalities.
   In your manuscript you state that the background flow and jet is too zonal and too far to the south and that the underestimation of the eddy activity in the North Atlantic is associated with WCB inflow and outflow biases. However, could you also argue, that due to a misrepresentation of the WCB inflow and ascent frequencies, WCBs do not disturb the upper level waveguide in a reasonable way (for example not often enough), such that in consequence, the amplification of upper-level ridges or the initiation of Rossby waves and/or blocking (or more general, a poleward shift of the dynamical tropopause) is too weak and therefore the jet and the time mean flow gets too zonal? I think these questions can not be disentangled from your study, but I would appreciate a more precise discussion of these aspects.

   Thank you for this comment. We agree that the interactions between different synoptic features complicate drawing definitive conclusions. We address this further in the discussion section as follows: “Establishing the direction of causal relationships is challenging due to the complex feedbacks between the various synoptic features analysed in this study. However, we argue that a causal connection exists between the mean state and the blocking bias. Specifically, the mean state is not only influenced by atmospheric blocks but also by factors such as the representation of topography, land-sea distribution, and tropical convection. Several studies show that coarse-resolution models struggle to accurately capture these elements (e.g., Scaife et al., 2011; Pithan et al., 2016; Kleiner et al., 2021). Carefully designed sensitivity studies would be needed to quantify the causal relationship between diabatic processes and Euro-Atlantic blocking biases. A first step could be numerical experiments testing the overall sensitivity of the large-scale flow to WCB activity and its associated latent heat release. A rather subtle approach would be experiments with and without stochastic perturbation schemes in the region of WCBs. For example, Dawson and Palmer (2015) and Christensen et al. (2015) show that stochastic perturbations in general improve the representation of blocking in the Atlantic-European region. However, we still lack an understanding of exactly how diabatic processes contribute to this improvement.”

3. Introduction
   It would be nice to include very clear research questions / objectives at the end of the introduction section and to make very clear what is new in this paper. It helps the reader to understand even better what the questions are that will be answered in this publication.

   We thank you for this idea, indeed it would help the reader, thus, we have included the next research questions in line 87:

   (a) Role of dry processes: How do biases in the representation of jet stream positioning, storm tracks, and eddy-mean flow interactions in CMIP6 models influence the development and maintenance of atmospheric blocking in the Euro-Atlantic region?

   (b) Role of moist process: How well do CMIP6 models represent WCB activity and what is the link to blocking biases in the Euro-Atlantic region?

   (c) Drivers of biases in moist processes: which processes contribute to the biases of WCB inflow and outflow in CMIP6 models, and how might this affect the development of Euro-Atlantic blocking?

4. Section 4.3
   Here you discuss how WCB can influence the upper-level waveguide. I would appreciate a little bit more detailed discussion on what determines the impact on the waveguide and which processes could lead to the observed differences. This includes (i) bias in the frequency and or location of the outflow, (ii) a bias in the PV anomaly that is produced by the WCB. Here, the representation of microphysical processes or the simulated outflow height could have an impact on the PV value with which the WCB reaches the upper level. If this PV value is not correct, also the impact on the waveguide will be wrong, even if the predicted frequency and location is correct. So the questions is, what exactly determines the PV value in the WCB outflow. (iii) It could also be that the location of the WCB outflow, the outflow height and the PV value in the outflow is simulated correctly, but that the climatological tropopause height in CMIP6 models is not correct. Thus, even if the WCB is perfectly represented, its impact would be wrong just because the tropopause is e.g. too high or too low. A more detailed and careful discussion of all the effects that might play a role would strengthen the manuscript further. These aspects could also be discussed in the discussion section.

   We appreciate your comment, we will discuss further the three points you have suggested in the discussion section as follows:
   “Furthermore, we have discussed that WCBs can significantly influence the upper-level waveguide. Nevertheless, it is essential to highlight the key factors that govern the WCB’s influence on the waveguide and the processes that may introduce biases in this interaction. These factors include: i) bias in the frequency and/or location of the WCB outflow: As discussed in this study, inaccuracies in the predicted frequency or positioning of the outflow can lead to incorrect assessments of the WCB’s impact on the waveguide. ii) Bias in the potential vorticity (PV) anomaly generated by the WCB: The representation of microphysical processes and the simulated outflow height are critical in determining the PV value at the upper levels. If this PV value is inaccurately represented, it can distort the WCB’s impact on the waveguide, even if the frequency and location are correctly simulated. iii) Climatological tropopause height bias in CMIP6 models: Even if the WCB outflow’s location, height, and PV value are accurately simulated, an incorrect representation of the climatological tropopause height could still lead to errors. For instance, if the tropopause is modelled too high or too low, the impact of an otherwise well-represented WCB on the waveguide would be misestimated.”

Minor comments:

L3: strong bias in frequency $\Longrightarrow$ only frequency or also location of these features

We have now added “and location” in line 3

L12,13: gradients are equatorwards shifted $\Longrightarrow$ are shifted equatorwards

We have replaced “are equatorwards shifted” with “are shifted equatorwards”

L29-30: connect the two sentences with e.g. ..., however in recent years, research has shown. . . .

We have connected the two sentences with “..., however in recent years, research has shown...”

L34: . . . undergo diabatic processes. ⟹ . . . .undergo diabatic processes whereas the influence of these processes differs in different regions of the world.

Thanks for pointing this out, we have added in line 34 “whereas the influence of these processes differs in different regions of the world”

L34: Moist diabatic processes which are linked to the formation or dissipation of clouds contribute prominently to block development. . . .

We have added the sentence in line 34 “ which are linked to the formation or dissipation of clouds”

L38: new line for “The impact of moist diabatic processes. . . .”

We have started the sentence in a new line.

L58: applying anomaly block indices ⟹ what exactly do you mean here? Can you explain?

Here, we refer to the study by Attinger et al. (2021), which demonstrated that the misrepresentation of tropopause height leads to blocking biases. However, this bias was observed when using anomaly indices, specifically the anomaly of the vertical average of potential vorticity in the upper levels. We have added the following sentence to clarify this:
“There are two commonly used types of indices: absolute and anomaly indices (e.g., Barriopedro et al., 2010). Absolute indices rely on absolute fields, such as flow gradients, to identify the circulation associated with blocking events. In contrast, anomaly indices detect deviations in variables like geopotential height (Z) or potential vorticity (PV) from a given climatological baseline.”

L61: . . . the finer horizontal resolution could improve ⟹ improves

We have replaced “could improve” with “improves”

L71: ..by increasing horizontal resolution on a scale of storm resolution. Please clarify what you mean here.

We refer to kilometer-scale Earth system models, also known as storm-resolving models. Schemm (2023) demonstrated that increasing the resolution to 5 km reduces biases in storm tracks by more accurately resolving moist processes. We have replaced “storm resolution” with “kilometer resolution”

L80: mention somewhere that in Joos et al., 2023, WCBs have been calculated in a climate model based on Lagrangian trajectories.

We have added in line 81 the following sentence:
For instance, in Joos et al. (2023), WCBs have been calculated in a climate model based on Lagrangian trajectories.

L85: ELISA2.0 ⟹ ELIAS2.0

We have replaced “ELISA2.0” with “ELIAS2.0”

L 98: please clarify: Of the possible qualitative levels for model performance .... ⟹ what exactly do you mean here?

We refer to the qualitative levels describe in Palmer et al. (2023): inadequate, unsatisfactory, satisfactory and not available. Those levels refer how good are the models representing atmospheric blocking in the Euro-Atlantic region. Palmer et al. (2023) recommend to exclude models with inadequate performance. We mention this in Table 1. However, we agreed that this is confusing, we will rephrase in the text as follows: "Palmer et al. (2023) used a combination of quantitative metrics—such as RMSE, bias, variance, and correlation—and qualitative assessments, including the examination of circulation wind patterns. Based on these evaluations, the models have been categorized into four classifications: inadequate, unsatisfactory, satisfactory, and not available. We analyse only models with adequate performance, excluding the inadequate models, as suggested by Palmer et al. (2023)."

Table 1: I don't understand what the label adequate (satisfactory/unsatisfactory) means here. Can you please clarify?

Please, see previous answer. Furthermore, we have removed the fourth column in Table 1 to avoid confusions.

L110: Marco Rohrer and Wild, 2019 $\Longrightarrow$ Rohrer and Wild, 2019 , check reference

We thank you for noticing this. We have corrected the reference.

L121: lambda denotes longitude from 180W-179E and phi the latitude from 75S to 75N

We have modified line 121 to:
λ denotes longitude from 180° W to 179° E and ϕ the latitude from 75° S to 75° N.

L125: ...featuring a Rossby wave breaking. What do you mean here? Please clarify.

What we intended to convey is that the flow reversal index captures blocking associated with Rossby wave breaking, typically during the mature stage. This differs from anomaly indices, which can identify blocking at earlier stages or in other configurations. We have rephrased this sentence as follows:
"This method captures blocks in their mature stage associated with Rossby Wave Breaking (RWB) (e.g. Pinheiro et al., 2019)."

L160: "…probabilities of WCB inflow, ascent and outflow" . I would appreciate a lot a more detailed description on how these probabilities are calculated. Do the inflow, ascent and outflow regions correspond to a specific pressure range? How do you differentiate between these three categories? This section would also be a potential place to discuss the reliability of the ELIAS WCBs when calculated with CMIP6 data.

Many thanks for this suggestion. In the revised section 3.4 of the manuscript, we now provide more details concerning the definition of the three WCB stages inflow, ascent and outflow. The section now reads: "ELIAS2.0 is trained on the basis of ERA-interim data (Dee et al., 2011) remapped to a regular latitude-longitude grid spacing of $1°\times1°$. Accordingly, any input data to ELIAS2.0 need to be remapped to this same grid spacing. The predictor variables are derived from T, q, Z, u and v on seven pressure levels comprising 1000, 925, 850, 700, 500, 300, and 200 hPa. ELIAS2.0 predicts conditional probabilities of WCB inflow, ascent, and outflow using a sigmoid activation function in the final layer of a U-Net convolutional neural network. WCB inflow is defined as WCB air masses being located below 800 hPa on the basis of trajectory data in the original training data set. The ascent stage, which typically occurs with a time lag of 24 hours after the inflow, represents WCB air masses between 800 and 400 hPa. WCB outflow comprises all WCB air masses above 400 hPa and occurs typically with a time lag after the ascent stage. WCB masks in ERA5 are derived every 6 hours (at 0, 6, 12, and 18 UTC), with each grid cell assigned a binary value: 1 for the presence of a WCB and 0 for its absence". Concerning the reliability of ELIAS2.0, we now extended the discussion of potential limitations in Section 5.

L160: here you mention "WCB masks" . This terminology might not be clear to every reader, I would clarify.

We will clarify the term mask in lines 160 to 161 as follows:
WCB masks in ERA5 are derived every 6 hours (at 0, 6, 12, and 18 UTC), with each grid cell assigned a

binary value: 1 for the presence of a WCB and 0 for its absence.

L180: …blocking frequency by more than 80%. ...this value does not correspond to the colour scale which is shown in Fig.1, please adapt.

We apologise for the confusion, here, we refer an underestimation of 80% relative to ERA5. We have added it to the text.

L182: … Figs. 1i and 2e. $\Longrightarrow$ which figures are you referring to?

We thank you for pointing this out. The reference to figure 2e was wrong, we have removed it.

L190 V: This paragraph does not read very well, please rewrite (for example: Although the biases are calculated for each climate model, we present the multi-model CMIP6 mean bias for concision. Fig. 2a,b show the mean and mode of Z for ERA5. A closer spacing of Z contours can be observed when considering the mode rather than the mean value)

We have followed your advice and adapted the paragraph taking your example.

L194: Using the mode results in sharper gradients in the mid-latitudes as Swanson (2001) $\Longrightarrow$ Which has also been described in Swanson (2001).

We have modified the line 194 as suggested.

L205: … Fig. 2d? $\Longrightarrow$ which figure do you mean here?

We apologise for the typo. We replaced “Fig. 2d” with “Fig. 2c”.

Fig. 3d: the black and green lines could be thicker, and I would use solid and not dashed.

We think that using dashed lines is more intuitive since they are negative values. However, we agreed that using solid lines would help to distinguish the contours, thus we have increased the thickness and used solid lines.

L228: …divergence in Figure 3d $\Longrightarrow$ difference of divergence between CMIP6 and ERA5?

Thanks for noticing this. We have replaced “their divergence in Figure 3d” with “the difference of $\nabla \cdot \mathbf{E}$ between CMIP6 and ERA5 in Figure 3d”

L237: …looking at the biases $\Longrightarrow$ …looking at the differences?

We have replaced “biases” with “differences”

L276: ...tend to produce less frequent WCB outflows

We have modified the sentence in line 274 “ the models tend to produce less WCB outflow” with “the models tend to produce less frequent WCB outflows”

L275: …of WCB inflow frequencies.

We have added the term “frequencies”

L281: …bias over eastern Europe $\Longrightarrow$ …bias over western Europe?

We have replaced “eastern” with “western”

L288: …ascending trajectories by the WCB activity are crucial $\Longrightarrow$ …ascending WCB trajectories are crucial

We have replaced the sentence “ascending trajectories by the WCB activity are crucial” with “ascending WCB trajectories are crucial”

L312: . . . analysing WCB outflow frequencies.

We have added the term “frequencies”.

L313,314: i) misrepresenting diabatic processes $\Longrightarrow$ maybe more precisely “diabatic heating and the associated cross-isentropic flow”.

We thank you for the suggestion and have adapted the sentence.

Fig. 5d and L328-331: Can you explain in more detail how the eddy heat flux is linked to the WCB inflow, what is the role of this flux for WCBs? And why is its maximum to the north of the main WCB inflow regions?

WCB inflows typically exhibit a pronounced meridional component, suggesting a positive but relatively small $v'T'$ signal. The maximum $v'T'$ is observed over and slightly north of the mean baroclinic zone, where temperature perturbations (T') remain substantial, and meridional wind perturbations (v') are still significant. The WCB inflow primarily occurs within the warm sector of a cyclone and is therefore situated south of the main baroclinic zone.

L369: yellow region in Fig.6 $\Longrightarrow$ there is no yellow region, please adapt.

We apologise for the typo. We have replaced “yellow” with “red”

Fig.6: I very much like schematic summaries. However here you might also reconsider the causalities I mentioned above when describing the figure in the text and figure caption and formulate the conclusion a little bit more careful.

We agreed that the interaction among the different processes must be addressed with more caution. Thus, we have rephrased the the figure description in the text and the Figure 6 caption.
In the text line 367, we added:
“Note that the blocking may produce some biases in the background flow and dry processes, as described in the discussion. Thus, we refer to linkages rather than causalities, and the following chain of processes should be taken with caution.”
And Figure 6 caption:
“Note that blocking may bias the background flow and dry processes and the process chain should be interpreted cautiously.”

Reviewer 4

General comments:

This study investigates the representation of atmospheric blocking in eight CMIP6 models and relates negative biases in the blocking frequency in the Euro-Atlantic region to biases of the background flow, dry dynamics and moist diabatic processes. They show that a too zonal and equatorward shifted background flow and an underestimation of the WCB outflow in the central and eastern North Atlantic contribute to the underestimation of the blocking frequency. The results are very interesting, the method is appropriate, the paper is well written, and the storyline is clear. Most of the below comments can be addressed with some additional explanations and rewording.

Major comments:

1. You argue that the too zonal and too far equatorward flow explains the negative blocking bias, but couldn't it be the other way around, i.e., that the negative blocking bias leads to a too zonal and too far equatorward flow?

   Thank you for highlighting this point. It is a concern raised by other reviewers as well. We address this in more detail in the discussion section as follows:
   "Establishing the direction of causal relationships is challenging due to the complex feedbacks between the various synoptic features analysed in this study. However, we argue that a causal connection exists between the mean state and the blocking bias. Specifically, the mean state is not only influenced by atmospheric blocks but also by factors such as the representation of topography, land-sea distribution, and tropical convection. Several studies show that coarse-resolution models struggle to accurately capture these elements (e.g., Scaife et al., 2011; Pithan et al., 2016; Kleiner et al., 2021). Carefully designed sensitivity studies would be needed to quantify the causal relationship between diabatic processes and Euro-Atlantic blocking biases. A first step could be numerical experiments testing the overall sensitivity of the large-scale flow to WCB activity and its associated latent heat release. A rather subtle approach would be experiments with and without stochastic perturbation schemes in the region of WCBs. For example, Dawson and Palmer (2015) and Christensen et al. (2015) show that stochastic perturbations in general improve the representation of blocking in the Atlantic-European region. However, we still lack an understanding of exactly how diabatic processes contribute to this improvement."

2. I would find it helpful if you could discuss in more detail the limitations of using a machine learning method to identify WCBs in climate models. Which additional biases could arise from this method? Maybe you could also discuss in some more detail the differences between the WCBs identified with ELIAS2.0 and those identified by Joos et al. 2023 based on trajectory calculations. Furthermore, in Section 3.4 it would be helpful if you could add more details on how the ELIAS2.0 WCB identification method works.

   Many thanks for this suggestion which is in line with the comments of the other reviewers. Following your suggestion, we now provide additional details on the ELIAS2.0 WCB identification method in Section 3.4 of the manuscript. To keep the manuscript short nevertheless, the interested reader is referred to the original paper by Quinting and Grams (2022) for more details. Concerning the additional biases that could arise, we now include more details in the last paragraph of Section 5. For example, as shown by Joos et al. 2023, the ascent rate of WCBs in climate models is slower than in reanalysis data. Accordingly, also the WCB outflow height after 48 hours of ascent is lower. This lower outflow height, however, can not be diagnosed with ELIAS2.0 and rather appears as an underestimation of WCB outflow frequency. In order to also identify the outflow height, a separate machine learning model would need to be trained which is beyond the scope of this study. Further limitations are given in Section 5.

Minor comments, wording and typos:

3. Lines 58-59: “however, this link is found when applying anomaly block indices” – I have difficulties to follow here, can you explain what you mean?

   We apologise for the lack of clarity. We have added a small description as follows:
   “There are two commonly used types of indices: absolute and anomaly indices (e.g., Barriopedro et al., 2010). Absolute indices rely on absolute fields, such as flow gradients, to identify the circulation associated with blocking events. In contrast, anomaly indices detect deviations in variables like geopotential height (Z) or potential vorticity (PV) from a given climatological baseline.”

4. Lines 70-71: I would move the reference to Schemm, 2023 to the end of the previous sentence.

   We have moved the reference ti the end of previous sentence.

5. Line 73: Shouldn’t it be downstream “of”?

   We have replaced “to” with “of”.

6. Section 2: Could you briefly explain how Palmer et al. (2023) quantify the ability of the models to represent blocking frequency? Without having read this paper, I find it confusing that in Table 1 the models termed unsatisfactory are regarded as being adequate. Also, could you motivate a bit better why you excluded CMIP6 models with inadequate blocking representation from your analysis? In which regard are they inadequate? It would be interesting to know whether they exhibit the same biases as the ones you are investigating, did you look into this a bit?

   We agreed that it needs more detail. We have rephrase lines 98-99 as follows:
   “Palmer et al. (2023) used a combination of quantitative metrics—such as RMSE, bias, variance, and correlation—and qualitative assessments, including the examination of circulation wind patterns. Based on these evaluations, the models have been categorized into four classifications: inadequate, unsatisfactory, satisfactory, and not available. We analyse only models with adequate performance, excluding the inadequate models, as suggested by Palmer et al. (2023).”
   Furthermore, we have remove the fourth column in Table 1 to avoid confusions. Finally, we agreed that it would be interesting to look to more models but it is beyond the scope of the present study.

7. What do the abbreviations in the third column of Table 1 (Member) stand for?

   This refers that for each model are different members, where each run varies in realization, initial conditions, physics, forcing or a combination. For instance: realization index=2, initialization index=1, physics index=3, and forcing index=222, the member label is “r2i1p3f222”. We haved added this explanation in Table 1 as follows: “The member notation indicates variations in realization, initial conditions, physics, forcing, or a combination of these factors. For example, if the realization index is 2, the initialization index is 1, the physics index is 3, and the forcing index is 222, the corresponding member label is denoted as ”r2i1p3f222””

8. Line 110 and references: Marco Rohrer et al. should be Rohrer et al.

   We thank you for noticing this. We have corrected the reference.

9. Line 125: “featuring a Rossby Wave breaking” – I don’t understand what you mean here.

   We apologise for the lack of clarity. We meant that the flow reversal index captures blocking associated with Rossby wave breaking, typically during the mature stage. This is different from the anomaly indices, which can capture blocking in earlier stages and with other configurations. We will rephrase this sentence to:
   “This method captures blocks in their mature stage associated with Rossby Wave Breaking (RWB) (e.g. Pinheiro et al., 2019).”

10. Line 156-157: “the” is missing before Madden-Julian Oscillation and “a” before WCB.

    We have adapted the line 156-157.

11. Line 168: delete “the” in front of ERA5

    We have deleted “the” in front of ERA5.

12. Caption Fig. 2: I assume that for the zonal wind the mean is shown and not the mode, but it is not specified in the caption.

    We have adapted the caption as follows:
    “a,b) Climatological a) mean and b) mode of Z at 500 hPa for ERA5. The mean of zonal wind at 500 hPa is shown as a purple contour (15 $m\,s^{-1}$). c,d,e) CMIP6 multi-model mean biases of c) mean $|\frac{dZ}{dy}|$, d) mode of $|\frac{dZ}{dy}|$, and e) mean of zonal wind at 500 hPa against ERA5. Black contours depict the ERA5 $|\frac{dZ}{dy}|$ climatology in c) and d) (5, 10, 20, 30, 40 $m\,km^{-1}$ intervals). Stippling denotes regions where more than 80% (6 models) of the ensemble members indicate a bias of the same sign.”

13. Line 200: “north Atlantic” should be “North Atlantic”. And I would always write “North Atlantic” and not just “Atlantic” (e.g., lines 203, 216, 265, etc.).

    We have followed your advice, we have only used “North Atlantic”.

14. Figs 2c,d and lines 204-205 and Figs. 2: How can it be interpreted that the highest dZ/dy of the mode and of the mean over western Europe are located in different regions?

    As we discussed, the use of the mode allows a better identification of the background flow. Thus, we expect that the influence of background flow is more crucial at higher latitudes.

15. Line 205: “South Europe” should be “Southern Europe”

    We have replaced “South Europe” with “Southern Europe”.

16. Line 225: I assume you mean Fig. 2?

    We thank you for noticing this, yes, we meant Figure 2. We have corrected it.

17. Fig. 4: e,d in the third line should be e,f

    We have replaced “e,d” with “e,f”.

18. Line 270: “see green contour farther to the black ...” This part of the sentence is difficult to understand, consider rephrasing.

    We have rephrased as follows:
    “see that the green contour is farther from the black contour on the polar side compared to the equatorial side.”

19. Line 281: According to Fig. S1f in Joos et al. 2023 there aren’t any strong biases in the WCB outflow over eastern Europe, but rather over western and central Europe and the eastern North Atlantic.

    We apologise for the typo. We have replaced “eastern” with “western”.

20. Line 288: “ascending trajectories by the WCB activity” – strange wording, I would just write “ascending WCB trajectories”

    We have replaced “ascending trajectories by the WCB activity” with “ascending WCB trajectories”.

21. Why do you discuss Fig. 5 at the end of Section 5 and not already in Section 4.3? I would find it easier to follow if you showed these results before the discussion in Section 5.

    We believe this fits better in the final discussion, where we synthesize the overall findings, whereas Section 4.3 focuses specifically on biases in the WCBs.

22. Fig. 5b: Why is the eddy heat flux maximum located north of the WCB inflow maximum?

    WCB inflows typically exhibit a pronounced meridional component, suggesting a positive but relatively small $v'T'$ signal. The maximum $v'T'$ is observed over and slightly north of the mean baroclinic zone, where temperature perturbations (T') remain substantial, and meridional wind perturbations (v') are still significant. The WCB inflow primarily occurs within the warm sector of a cyclone and is therefore situated south of the main baroclinic zone.

23. Line 313: misrepresenting – better: misrepresented

    We have replaced “misrepresenting” with “misrepresented”

24. Line 369: which yellow region do you mean?

    We apologise for the confusion. We refer to the red region. We have corrected it.

References

Attinger, R., Spreitzer, E., Boettcher, M., Wernli, H., and Joos, H. (2021). Systematic assessment of the diabatic processes that modify low-level potential vorticity in extratropical cyclones. *Weather and Climate Dynamics*, 2(4):1073–1091.

Barriopedro, D., García-Herrera, R., and Trigo, R. M. (2010). Application of blocking diagnosis methods to general circulation models. part i: a novel detection scheme. *Climate Dynamics*, 35(7-8):1373–1391.

Brunner, L. and Steiner, A. K. (2017). A global perspective on atmospheric blocking using gps radio occultation – one decade of observations. *Atmospheric Measurement Techniques*, 10(12):4727–4745.

Christensen, H. M., Moroz, I. M., and Palmer, T. N. (2015). Simulating weather regimes: impact of stochastic and perturbed parameter schemes in a simple atmospheric model. *Climate Dynamics*, 44(7):2195–2214.

Davini, P., Corti, S., D’Andrea, F., Rivière, G., and Von Hardenberg, J. (2017). Improved Winter European Atmospheric Blocking Frequencies in High-Resolution Global Climate Simulations. *Journal of Advances in Modeling Earth Systems*, 9(7):2615–2634.

Dawson, A. and Palmer, T. N. (2015). Simulating weather regimes: impact of model resolution and stochastic parameterization. *Climate Dynamics*, 44(7):2177–2193.

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al. (2011). The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

Dolores-Tesillos, E., Teubler, F., and Pfahl, S. (2022). Future changes in north atlantic winter cyclones in cesm-le – part 1: Cyclone intensity, potential vorticity anomalies, and horizontal wind speed. *Weather and Climate Dynamics*, 3(2):429–448.

Greeves, C., Pope, V., Stratton, R., and Martin, G. (2007). Representation of northern hemisphere winter storm tracks in climate models. *Climate Dynamics*, 28:683–702.

Harvey, B., Cook, P., Shaffrey, L., and Schiemann, R. (2020). The response of the northern hemisphere storm tracks and jet streams to climate change in the cmip3, cmip5, and cmip6 climate models. *Journal of Geophysical Research: Atmospheres*, 125(23):e2020JD032701.

Hoskins, B. J. and Hodges, K. I. (2002). New perspectives on the northern hemisphere winter storm tracks. *Journal of the Atmospheric Sciences*, 59(6):1041–1061.

Joos, H., Sprenger, M., Binder, H., Beyerle, U., and Wernli, H. (2023). Warm conveyor belts in present-day and future climate simulations – part 1: Climatology and impacts. *Weather and Climate Dynamics*, 4(1):133–155.

Kleiner, N., Chan, P. W., Wang, L., Ma, D., and Kuang, Z. (2021). Effects of climate model mean-state bias on blocking underestimation. *Geophysical Research Letters*, 48(13):e2021GL094129. e2021GL094129 2021GL094129.

Palmer, T. E., McSweeney, C. F., Booth, B. B. B., Priestley, M. D. K., Davini, P., Brunner, L., Borchert, L., and Menary, M. B. (2023). Performance-based sub-selection of cmip6 models for impact assessments in europe. *Earth System Dynamics*, 14(2):457–483.

Pinheiro, M. C., Ullrich, P. A., and Grotjahn, R. (2019). Atmospheric blocking and intercomparison of objective detection methods: flow field characteristics. *Climate dynamics*, 53:4189–4216.

Pithan, F., Shepherd, T. G., Zappa, G., and Sandu, I. (2016). Climate model biases in jet streams, blocking and storm tracks resulting from missing orographic drag. *Geophysical Research Letters*, 43(13):7231–7240.

Quinting, J. F. and Grams, C. M. (2022). Eulerian identification of ascending airstreams (elias 2.0) in numerical weather prediction and climate models – part 1: Development of deep learning model. *Geoscientific Model Development*, 15(2):715–730.

Quinting, J. F., Grams, C. M., Chang, E. K.-M., Pfahl, S., and Wernli, H. (2024). Warm conveyor belt activity over the pacific: modulation by the madden–julian oscillation and impact on tropical–extratropical teleconnections. *Weather and Climate Dynamics*, 5(1):65–85.

Scaife, A. A., Copsey, D., Gordon, C., Harris, C., Hinton, T., Keeley, S., O'Neill, A., Roberts, M., and Williams, K. (2011). Improved atlantic winter blocking in a climate model. *Geophysical Research Letters*, 38(23).

Schemm, S. (2023). Toward eliminating the decades-old "too zonal and too equatorward" storm-track bias in climate models. *Journal of Advances in Modeling Earth Systems*, 15(2):e2022MS003482.

Wandel, J., Büeler, D., Knippertz, P., Quinting, J. F., and Grams, C. M. (2024). Why moist dynamic processes matter for the sub-seasonal prediction of atmospheric blocking over europe. *Journal of Geophysical Research: Atmospheres*, 129(8):e2023JD039791. e2023JD039791 2023JD039791.

Wandel, J., Quinting, J. F., and Grams, C. M. (2021). Toward a systematic evaluation of warm conveyor belts in numerical weather prediction and climate models. part ii: Verification of operational reforecasts. *Journal of the Atmospheric Sciences*, 78(12):3965–3982.

Woollings, T., Barriopedro, D., Methven, J., Son, S.-W., Martius, O., Harvey, B., Sillmann, J., Lupo, A. R., and Seneviratne, S. (2018). Blocking and its response to climate change. *Current Climate Change Reports*, 4(3):287–300.

Zappa, G., Shaffrey, L. C., Hodges, K. I., Sansom, P. G., and Stephenson, D. B. (2013). A multimodel assessment of future projections of north atlantic and european extratropical cyclones in the cmip5 climate models. *Journal of Climate*, 26(16):5846–5862.