

Review of WCD manuscript egusphere-2025-4969 “Storm-Resolving Models Advance Atmospheric Blocking Simulations and Climate Change” by Edgar Dolores-Tesillos et al. (2025)

Summary

The study illustrates how storm-resolving models represent European blocking using historical simulations of ICON and IFS-FESOM in comparison with eight CMIP6 models and ERA5. Despite regionally and seasonally-confined improvements, coupled storm-resolving model simulations reveal that typically known biases related to European blocking with respect to frequency, duration and spatial extent are not considerably reduced relative to CMIP models. Instead, an IFS AMIP Experiment forced with realistic SSTs indicates substantial improvement. Thus the study highlights the importance of realistically simulated SSTs for European blocking and emphasizes that a grid refinement towards the storm-resolving scale does not solve the European blocking bias problem. Another part of the study explains a negative trend of European blocking in specific regions, whereas the strong bias does not allow for reliable and accurate interpretations.

Assessment

The study provides interesting insights into the performance of storm-resolving models in simulating European blocking and associated trends. The study addresses different contributing factors, however, the study lacks a well-structured and consistent presentation of the results. Overall, there is a large number of major concerns that limits the study to be published in its current form. Given the scientific relevance of the study, I recommend a **major revision** of the manuscript by addressing the raised concerns before the manuscript can be considered to be accepted for publication in the journal of Weather & Climate Dynamics.

General comments

- 1.) The title can be misleading. The word ‘Simulations’ can be understood as climate model simulations, but this is not meant here. Either use the singular form ‘Simulation’ or the word ‘representation’ here. Also, the word ‘advance’ should not be used as the study concludes that the increased horizontal resolution in storm-resolving models does not lead to overall reduced blocking biases. Further, the phrasing ‘Climate Change Insights’ sounds very general and can eventually be tied to a specific result or removed due to reasons mentioned in the third comment.
- 2.) There is a general lack of motivation to several steps being performed in the study. For instance, the motivation about using storm-resolving models is too inadequate and requires further clarification of their advantages in the *Introduction* section. Further, several paragraphs need a re-structuring and enhanced consistency to guide the reader through the results more carefully.
- 3.) I consider the subsection 4.3 about the insights into European blocking trend too separate and too incomplete and would recommend removing this section from the study. Particularly, the fact that the European blocking biases largely remain in storm-resolving models complicate the assessment about their trends and therefore need a more careful consideration, which might be part of a follow-up study.

4.) The discussion section includes paragraphs, in which results are mentioned without being discussed in the context of related studies. Thus, the discussion appears too repetitious to me and needs to be more focused on how the results are related to the current research state.

5.) In general, the results of the included figures are not adequately and concisely presented. For some figures, certain subplots are not addressed in the results, even though they contain relevant features which would have been worth to be mentioned (see minor comments below). Further, there is an inconsistent and not convincing presentation as for Fig. 11. The figure is part of the subsection 4.2 explaining the blocking simulation during summertime, however I would expect such figure with a similar framework presented in the subsection 4.1 for wintertime as well.

6.) There are many incorrect or misplaced figure references. I highlight a couple of them in the detailed point-by-point responses below. Also, for the sake of a better readability, the authors should move figure references towards the end of a sentence.

Additional minor comments are provided in the line-by-line comments below.

Minor comments

Line 7-17: The paragraph appears somewhat unstructured due to jumps between AMIP and CMIP results. It would be more straightforward for the reader, if the paragraph is structured with the CMIP results first and followed by the AMIP results.

Line 8: Why do you choose 8 models from the CMIP6 ensemble? The authors should provide a reason why these models are chosen from the larger CMIP6 ensemble.

Line 15-17: How does this result relate to ERA5?

Line 21: The readability would be improved, if the paragraph is separated into two paragraphs here, i.e., separate the results related to the SSP3-7.0 forcing and the summary & implications from each other.

Line 38-40 and following paragraphs: The authors identify several factors contributing to the misrepresentation of blocking in climate models, namely (1) horizontal resolution, (2) storm-track biases, and (3) parameterisations of moist processes. However, the subsequent paragraphs introduce additional aspects (jet waveguide and SST representation) that are not included in the initial list. As also mentioned in major comment 2, I recommend revising this section for consistency by either following the original order of the listed processes or expanding the list to include all relevant factors. Providing a complete, numbered list upfront and then discussing each item in the same order would substantially improve clarity.

Line 74 - 77: The third and fourth research questions are very similar, and both refer to the final section on blocking representation under climate change. Since this section is intended as an addition with *first insights* rather than a main part of the paper, I

recommend raising only one research question for this section, if any, and combining the two existing questions into a single question.

Line 89: I recommend removing the phrasing ‘the nextGEMS models’, as it may lead to misunderstandings. Projects like nextGEMS do not own the models. They rather contribute to model development, conduct experiments, or use the models for research purposes.

Line 92: Why do you refer them to be storm-resolving Earth system models? The authors need to provide a justification. For instance, by stating the typical grid-spacing threshold below which a model is considered storm-resolving.

Line 92-93: This statement is too vague. The authors should clarify why the results are considered promising. Do the cited studies compare storm-resolving models with typical CMIP models and demonstrate an outperformance? The paper would benefit from a more detailed motivation for using storm-resolving climate models. This explanation could also be included in the introduction, which currently lacks a clear rationale for assessing storm-resolving models.

Line 93: ‘...for the **representation/simulation of** climate extremes...’

Line 94: ‘A key distinction between **both** models’

Line 116: As mentioned above. Why do you use the subset of eight models from the CMIP6 ensemble.

Line 139: Is there a physical explanation or geographical direction associated with the three gradients that could be given here? I would assume they correspond to northern, southern, and equatorial gradients. Explicitly stating this would make the blocking index easier to understand.

Line 159: Why do the summer and winter seasons are the ‘main seasons’. The authors may consider to replace this phrase by ‘winter and summer seasons’.

Line 175: For the comparison between IFS and CMIP, the authors have to refer to Fig. 1a,b,d.

Line 179-180: The authors should take care to describe the locations precisely. According to Fig. 1, I would rephrase the sentence to ‘...underestimation of blocking frequency over **the North Atlantic** and eastward displacement of blocking maxima **towards Eurasia.**’

Line 180: I would expect a reference to Fig. 1c rather than Fig. 1d.

Line 186: There is no compensation of biases over the North Atlantic and Eastern Europe (Fig. 1e). This should be stated here as well.

Line 189: Why does the blocking frequency biases result from the size and duration? To my understanding, the blocking frequency might only be influenced by size or duration if a blocked event is too small or too short-lived to be classified as a blocked event. Thus, the relationship emerges due to the constraints of the blocking indices rather than due to

a physical relationship/dependence. The authors should clarify the dependences of the blocking characteristics to each other as well as the influence of the methodology in order to avoid confusion.

Line 194: The sentence only describes the North Atlantic. Thereby, a reference to Fig. 2b is sufficient.

Line 195: ‘...but capture **mean and** the 95th percentile...’

Line 197: ‘...slightly shorter events (Fig. 2d,e).’

Line 194-200: This paragraph is intended to explain the models’ blocking duration, as introduced at the outset. However, the subsequent sentences conflate duration, frequency, and size. I recommend guiding the reader systematically through these three diagnostics and then, in a separate paragraph, explaining the biases by drawing on the contributions of the blocking characteristics.

Figure 1, caption: Subplot labels are not correct. Please replace d), e) and g) by c), d) and e).

Line 216: The models are not multidecadal. The authors have to be more concise and may replace this part by ‘...in the multidecadal **simulations/experiments of both** storm-resolving models.’

Line 237: Do you refer to Table 4?

Line 248: ‘Negative SST **bias**...’

Line 270: ‘...during **wintertime**.’

Line 271: ‘**Over** the **North** Atlantic,...’

Line 291: The areas on which the averages are performed, do not really fit to the terminology of ‘North Atlantic’ and ‘North Pacific’. I would suggest to adjust and reduce the size of the areas to the ocean basins to avoid misunderstandings with the terminology.

Line 294/295+ 303/304: To my understanding, the conclusions about the performance of the ICON simulation does not agree in these two sentences. While the authors emphasize that ICON has a small blocking frequency bias (RMSE=0.33) (line 294/295), they describe an underestimation of the number of blocking events by referring back to the Atlantic frequency bias. Please clarify if the presentation of the results are correct. If so, I would recommend to revise the paragraphs with a more structured presentation of the results.

Line 297: ‘...more closely...’ If the authors refer with these findings to Fig. 6a,b, I would not conclude that IFS AMIP more closely matches ERA5. There are still significant differences with a similar magnitude compared to IFS hist over the North Pacific. The extent is only somewhat smaller in IFS AMIP. I would recommend a more concise presentation of the differences between the models with regard to blocking characteristics.

Line: 301: I would assume the authors only refer to IFS hist and not to both IFS simulations here. 'In the North Atlantic, **IFS hist produces ...**'

Line 316: 'The AMIP simulation shows larger blocks **in general, i.e. the entire distribution is shifted towards an overestimation,**...' I would generally suggest to more concisely describe the distribution differences and highlight that not only the mean or median is different, but also the entire distribution is shifted relative to ERA5.

Line 320-321: ICON hist and not IFS AMIP has the lowest RMSE for blocking frequency in the North Atlantic according to Table 5. Also, what is meant by the term 'balanced performance'? Do we observe a compensation of biases between different blocking characteristics leading to an overall low RMSE? Please clarify the contribution of all three blocking characteristics.

Line 325: ICON does not simulate smaller blocks in the Pacific according to Fig. 7f.

Line 326-327: ICON-hist and not IFS AMIP has the highest block count over the North Pacific according to Fig. 7d.

Line 332: ',.... than the coarse-resolution **CMIP** models.'

Fig. 7d: Where is the line for the mean of IFS amip?

Figure 7, caption: '...and their properties'. Providing the explicit list of properties/ characteristics with subplot references would be helpful here, e.g., count (a,d), duration (b,e), size (c,f)

Line 335: '...midlatitudes.' The authors need to provide a reference to a figure of the their study or cite a previous study.

Line 339-340: IFS hist shows a positive U bias here as well (Fig. 8a).

Line 346: ',...more intense jets...' Do you refer to multiple jets here? Subtropical and polar jet? Please clarify.

Line 347/348: To my understanding, CMIP indicates a slight equatorward shift rather than a broader jet over Eurasia.

Line 348/349: The bias in CMIP is generally lower than in IFS AMIP, however, the RMSE for IFS amip is generally lower than for CMIP. Does this inconsistency emerges due to a stronger cancellation of biases with opposite signs in IFS amip and thus the limitation of the large size of areas, upon which the averaged RMSE values are calculated on?

Line 353: The overestimation of blocking largely occurs over North America and thus downstream of the poleward jet shift. Thus, it would not contradict the common inverse relationship between jet intensity and blocking occurrence.

Line 369: 'negative SST **bias**'

Line 375: 'warm SST **bias**'

Line 376: ‚Such positive **bias**’

Line 380: ‚the impact of this **cold** bias’

Line 386: ‚...when biases are strong...’ - Why is the influence of SST anomalies on blocking only present when biases are strong? Bias differences in SST and blocking just indicate that a reduced SST bias might be related to a reduced blocking bias. Also what is meant by ‘spatially structured’? I would recommend to remove the part after the comma.

Figure 9, caption: (c) instead of (d)

Line 292: Why is the acronym (IFS hist) introduced here again?

Line 396: ICON simulations feature a weaker storm track compared to IFS AMIP or ERA5?

Line 403: What is meant by ‘meridionally spread’? This needs to be better expressed, for instance: ‚the jet is not exclusively tied to the mid-latitudes, and rather present on a wider latitude range’ ...do you mean something like that?

Line 406 - 416 + Fig. 11: There are a couple of shortcomings related to Fig. 11 and its description, as mentioned in major comment 5. Why does the figure illustrates only the relationship of blocking and storm tracks? Why is the relationship in this framework only shown for summer and not for winter in the previous section 4.1? The results are not really convincing as the relationship is relatively weak. Is this again a consequence of averaging over too large domain (smoothing and/or cancellation of signals)? I would recommend to revise the figure accordingly or consider removing it or move it to SI.

Figure 9+10: There is no hatching visible, eben though differences are quite large regionally. Please double-check if there is really no significance.

Figure 11, caption: What is meant by ‘selected simulations’?

Line 420/421+ line 443 +448: The authors regularly point out that substantial biases must be taken into account when assessing blocking trends. As mentioned in the major comment 3, I would thus recommend to remove section 4.3 from the study.

Line 431: The suggested equatorward shift is not confirmed by a negative pattern directly northward of the positive pattern. The negative pattern appears over Eurasia (Fig. 12a).

Line 435: Fig. 13d clearly shows the typical North Atlantic warming hole. So the reduced blocking does not align with a North Atlantic warming pattern.

Line 439-442: The results on the number of events (count) should be presented here as well (Fig. 12c,f,i,l).

Line 443-455: Even though there is a small reference to it earlier in the text, this paragraphs lacks a more detailed discussion on the results related to the SST trend (Fig. 13c,d).

Line 489-507 + Line 527-536: These paragraphs are an extension of the results rather than a discussion of your results in the context of other studies. These paragraphs need to focus more on how the results relate to current research and should be shortened for clarity, as mentioned major comment 4.

Line 515: ‘These remaining biases could be due to the **mis**representation of moist processes...’

Line 541-553: None of the raised bullet points include a reference to blocking under climate change. Instead, the most important future research branch motivated by the present study is the improvement of the representation of blocking in multi-decadal historical climate simulations to facilitate a more reliable estimate on how blocking will change regionally and seasonally under climate change conditions.

Line 571 + 574/575: This key finding indicate that storm-resolving models still have blocking biases and do not advance blocking simulation. As mentioned in the major comment 1, I would recommend revising the title of the study.

Line 622: Even though the authors mention the assistance of Large-Language Models (LLM) such as ChatGPT, I would suggest being careful to include longer sentence structures. Sentences in which an insertion is included separated by long dashes (em-dashes) is a typical output feature by ChatGPT.