Orographic resolution driving the improvements associated with horizontal resolution increase in the Northern Hemisphere winter mid-latitudes by Paolo Davini, Federico Fabiano, and Irina Sandu

The study presents a systematic analysis of the effects of the small-scale orography on the large-scale circulation and in particular, the blocking phenomena. Due to coarse model resolution, climate models do not resolve the small-scale orographic features in sufficient detail. These features can impact the large-scale eddy forcing of the atmosphere with consequences for frequency and evolution of atmospheric blocking events resolved by these models. Using a set of carefully designed climate model integrations, the study demonstrates that increasing the orographic resolution in models, more so than increasing the atmospheric resolution, leads to a reduction of biases in model blocking frequency, jet latitude and eddy forcing (as compared to ERA5). Ultimately, the study concludes that properly tuning physical parameterizations in high resolution climate runs is key to eliminating potential energy imbalances caused due to changing resolution, and to fully reaping the benefits accompanying increased model resolution.

The experimental setup is scientifically sound, and the figures complement the discussion well. I do find the conclusions derived, at places, to be potentially misleading and slightly exaggerated, not being fully consistent with what the figures suggest. I address some of the concerns below. However, overall, the study could form an important contribution to the existing body of literature, and can be of value to HighResMIP and other climate model intercomparison projects. I recommend the manuscript for publication **pending major revisions**, after the following concerns have been addressed.

1. **Impacts on Blocking:** The T511 and T799 runs are integrated without the TOFD and SSO scheme and with other parameterizations tuned at a T255 resolution only. The parameterizations are not fully accurate (and not even accurately tuned in some instances) and in most cases respond to the climatology generated by the model. The climatology generated by the T511 and T799 are notably different (Figure 1 and 2). Given this, how do the authors isolate/remove the significant systematic errors that can be introduced by the untuned parameterizations? Moreover, it would be great if the authors can discuss how linear the response is between simply changing the atmospheric resolution and changing both the atmospheric resolution and orographic resolution in concert.

2. **Scaling:** The changes due to changing resolution in Figure 4(b,c) and Figure 4(e,f) are much weaker when compared to the T255-ctrl biases in Figure 4(a) and Figure 4(d) respectively. If computationally possible, can the authors also illustrate the biases for the tuned T255-ctrl-param run as well (and if they are very different despite similar TOA budget)?

Moreover, given the current magnitude of the biases, some of the conclusions in Section 4 could be misleading. For instance, the changes due to increase in atmospheric resolution vs orogoraphic resolution appear comparable to me (and not strikingly different as claimed). Moreover, the magnitudes resulting from a 3-4 fold increase in horizontal resolution seem to be small, and even qualitatively, both types of resolution increases lead to a reduction in biases over Scandinavia and parts of Azores. I suggest the authors to please re-consider (mellow) such strong claims based on the experimental setup and, appropriately rewrite the findings.

Finally, can the authors provide some insight into how these biases are expected to scale with horizontal resolution? The planetary and synoptic scales are very well resolved at T255 resolution as well. Given this, is there a critical resolution, by which eddy feedback of the small-scale onto the large-scale, and its impact on blocking, is completely resolved? I appreciate that this could be an open question with no clear answer as of now, but given that the relative reduction in biases even after a 3x resolution increase is moderate at best, to what extent do the conclusions from this study confirm the impact of small-scales in influencing large-scale blocking?

3. **Model Tuning:** The final sentence of the abstract stresses that the findings point to the importance of tuning. This conclusions will be more readily acceptable if the authors could, if computationally possible, provide a comparison of their findings with a tuned version of T799. The parameterizations employed in the model are not scale aware and the improvements in the T511-ctrl and T799-ctrl run are not significant. Thus, to be able to confidently claim that model tuning is imperative to the blocking frequency, it would help to show the bias reduction in a tuned T799 run.

   In addition, can the authors comment on how the parameterizations in the T255-ctrl-param were tuned? For instance, what constraints were used, and was it tuned to produce realistic winds in the Northern hemisphere only or to optimize the TOA radiative budget?

4. **Mechanism:** Based on my understanding, a mechanism explaining how small-scale changes due to a more resolved orography project onto the planetary-scale does not clearly stand out from the study. The biases in transient forcing are in-phase with improvements in baroclinic energy conversion over the Pacific and North America, but not over the Tibetan Plateau. Also, the improvements in barotropic energy conversion over the Atlantic are more robust than actual impovements in blocking frequency.

   Another concern I have is that the changes in eddy forcing with changes in resolution should be correlated with the TOA imbalance induced by resolution change (in cloud cover, precipitation etc). Moreover, the changes in the stationary wave forcing (as mentioned around L270) are similar when either the atmospheric resolution is

increased or the orographic resolution in increased (Figure 2(b,c)). So, it is not totally clear to me how it is the changes around the orography that project onto the Atlantic jet stream, far away from the orography. I appreciate that it is not the focus of the study, but it would be great if the authors can provide some explanation for this for the general reader.

Other Comments

1. L15 : which 'has' resulted
2. L21 : Please provide an appropriate citation, if possible
3. L30 : Please elaborate more on the first sentence - providing reasons - as to why the effective resolution is lower. For example - hyperdiffusion damping things grid-scale dynamics etc.
4. L43 : well-known
5. L52 : 'Thus, ' many questions related to ... remain open.
6. L58 : can remove the parenthesis to improve readability.
7. L60 : (this paragraph) suggests that previous studies have already established the role of a more resolved orography. Thus, by simply reading the paragraph, it is unclear how this study improves upon the findings of Kanehama et al. (2019). Please elaborate some key limitations of the previous study and how those are addressed in this new study providing more details than just 'in a more quantitative manner'.
8. L71 : Section 3 'and Section 4' include the analysis ... variability 'respectively'.
9. L81 : uppermost one $\rightarrow$ model top
10. Section 2.1 : Please provide the vertical grid resolution in the troposphere.
11. L86 : is a one-year spinup sufficient? If I understand correctly, comprehensive model runs with a seasonal cycle need to be spun-up longer when focusing on climatological statistics.
12. L109 : Why is a three-member ensemble employed only for the T255 resolution integrations?
13. L117 : I am not sure if I understand "protocol" in this context. How about "project" perhaps.
14. L127 : Please remove 'are presented ... paragraphs.'
15. L131 : whose 'accurate' simulation still 'present a challenge' for ...

16. L140 : (paragraph) Please mention in some reasonable detail how sensitive the results are to the definition of blocking itself. The changes in blocking frequency over mainland Europe are not very significant - can those changes be sensitive to the definition?

17. L151 : investigating

18. L171 : Similar to the

19. L179 : lower middle troposphere? please be more specific by providing the rough pressure level you are refering to (500 hPa? 750 hPa?)

20. L177 : reference pressure

21. L153 : standard dev wrt the zonal mean? Please be more specific.

22. From a reabability point of view, I think it would be better to swap section 2.2 and 2.3 and place the (current) section 2.3 right after the model setup. This will ensure all the model runs specific details are discussed together.

23. L206 : There is no noticeable shift in the jet height. Please add the old and new tropopause height to support the claim.

24. L210 : larger 'orographic' wave activity, I assume?

25. L215 : How is the streamfunction defined? Is it simply the geopotential?

26. L245 : Please mention the key reasons why models have such a modality bias.

27. L251 : While I agree that in a quantitative sense T799 is closest to ERA5, it also develops a fourth peak around 50°N and has the northmost peak around 57°N as opposed to ERA5 and T511 which has similar structure in terms of modality and gradients. Can you please comment more on this, why the distribution does not converge qualitatively with increases in atmospheric resolution?

28. L257 : please remove 'obviously'

29. L261 : I do not fully agree with the conclusion. Especially over mainland Europe, increase in orography resolution worsens the bias. In addition, over the UK and Scandinavian regions, changes due to atm res. are comparable (and in cases larger) than the changes due to oro. res. Therefore, the statement is inarguably true only for the lower latitude Azores Atlantic basin.

30. L268 : Again, I do not fully agree with the conclusion. The changes in blocking duration are definitely comparable to the bias between T255 and ERA5, along with worsening of the bias over mainland Europe. While I support the conclusion over the Antantic, concluding the same for mainland Europe is not so stratightforward.

31. L270 : The stationary patterns by changes in atmos. res and orog. res. are affected similarly (Figure 2b,c). Thus I think the final statement of the paragraph should be framed more carefully.

32. L275 : Is the bias significantly stronger than for the tuned T255 run?

33. Figure 5 : Is the large red band over Eurasia in the bias (Figure 5a) explained exclusively by the Alps? (similar to the bias over western North America)

34. L283 : ... orographic resolution. The changes to the ... I suggest a paragraph break here.

35. L286 : same terms (not shown)?

36. L295 : This is not totally clear to me, can you please explain why?

37. L317 : significantly for almost

38. L357 : I think it decreases by 1% at T511 and by 2% at T799 instead. Isn't it?

39. L396 : ... might be considered responsible for the absence ...

40. Figure 1(b) : please mention the maximum temperature difference associated with stratospheric cooling. The colorbar saturates early. It would also be great if the authors can mention in the Figure 4's caption the maximum value of the biases in Figure 4a as the colorbar saturates early for lower values. More generally, please increase the font size in all the figures, and wherever the colorbar saturates, provide the maximum value of the statistic.

41. How does the Jet Latitude Index (JLI) compare for the different models integrations over the Pacific? This is part of my broader concern, regarding how the wind changes due to flow blocking around the Rocky Mountains and Tibetan Plateau relate to changes in the Atlantic JLI.