We would like to thank the two reviewers for their comments and suggestions to our revised manuscript. We now answer their comments, before resubmitting a newly revised version of the manuscript. The original reviewers' comments are in black, while our replies are in blue.

Reviewer #1:

The authors have answered most of my questions and the manuscript has been improved significantly. However, the author did not adequately address my minor concern b. When analyzing the four weather regimes, it is better to plot the composite anomalies of transient LWA instead of total field of transient LWA because the former will help to focus on the LWA anomalies in the key region related to each regime, and thus the spatial pattern will not be affected by other regions (e.g., the high correlation over North Pacific). I thus recommend the authors to perform a minor revision by considering the comments listed below:

a) In Figs. 5-6, 8-11, I still suggest to show the composite maps of the transient LWA anomaly instead of the total transient LWA. Here I want to clarify that the composited transient LWA anomaly is calculated as the composites of total transient LWA in each weather regime minus the climatology of transient LWA in all days. It is important to note that the climatology of transient LWA is different from the stationary LWA (Huang and Nakamura 2017).

b) Similarly, in Fig. 7, it is better to show the spatial correlation of anomalous transient LWA between model and observation. It is interesting to see whether the spatial correlation of the anomalous transient LWA between model and reanalysis will become higher under higher model resolution.

We would like to thank the reviewer for suggesting this alternative analysis considering the anomalous LWA instead of the full field. We have done the suggested analysis producing analogues of Fig. 6,7 and Figs. 8-11 in terms of anomalous transient LWA. However we decided to keep these plots and their discussion as supplementary material to our manuscript. The main reason behind this is that the LWA theory in the primitive equations in isentropic coordinates in combination with Weather Regimes to our knowledge has not been applied yet to a climatological analysis involving reanalysis and several models. Therefore, in this first article we prefer to show the full LWA field (and not the anomalies) for both reanalysis and PRIMAVERA as reference for future work.
Another reason is related with the fact that from the full transient LWA field the regions of weak zonal wind can be deduced for each regime due to the non acceleration theorem, and this would not be straightforward when considering LWA anomalies. We refer to the supplementary material, where the analysis of WR in terms of anomalous LWA is presented and discussed at line 388-390:
"In addition to the analysis of WR in terms of transient LWA we repeated our approach but considering the transient LWA anomaly (i.e. the transient LWA in each WR minus the climatology of transient LWA for DJF) to exclude the model biases in the mean state. The results are presented in the supplementary material."

and in the conclusions at lines 492-93:

"The same analysis but in terms of anomalous transient LWA, which can be found in the supplementary material, also confirmed the results discussed above."

Reviewer #2:

This manuscript by Ghinassi et al. assesses the representation of Rossby wave activity in PRIMAVERA simulations and its sensitivity to model resolution. The authors addressed my previous comments and the manuscript has greatly improved. I still identified some minor issues which should be addressed prior to publication. In particular, the discussion of Figures 8-11 is difficult to follow since references to the individual figure panels are not provided. Such an addition would help the reader tremendously. After these comments have been addressed I recommend the study to be published in WCD.

Minor comments:

l. 3: In its classical physical definition the amplitude of a wave can be quantified. Is that what "strength" is referring to? Please consider to write "amplitude" instead and use this terminology consistently. Further, I think that "in terms of Rossby wave activity" is redundant since it is already being mentioned that you are using a diagnostic based on finite amplitude local wave activity. In my opinion, the sentence could end with "... of Rossby waves."

The reviewer is right, the use of strength can create confusion so we changed it to amplitude to have a consistent terminology for all the manuscript.

l. 43: Is it on purpose that the authors are using the term "Rossby wave train" here?

Yes, we used “Rossby wave train” to denote a longitudinally extended series of ridges and troughs containing both contributions from planetary, quasi stationary waves and transient wave packets.

l. 48: "thus" is somewhat redundant since the authors already state at the beginning of the sentence "this implies".

We eliminated it, thanks for the suggestion.

l. 69: A critical reader may wonder why your diagnostic is more robust than previous diagnostics. Providing a brief explanation concerning this aspect would certainly strengthen your argument. It could be sufficient to state that evidence for your statement is provided in the following (l. 72ff).

We added “This contrasting results motivate us to use a robust diagnostic based on Finite Amplitude Local Wave Activity (LWA), which is able to objectively identify Rossby waves, as we will discuss in the following paragraph.” at lines 68-70.

l. 92: I think this sentence is a bit confusing. Are you really assessing how well the large scale circulation over Europe and the North Atlantic is represented in observations? What would then be your reference you are comparing to. Please revise the sentence if necessary.

We clarified this point and rephrased to “The aim of this work is to assess how well the large scale circulation over Europe and the North Atlantic is represented in state of the art, high resolution, global climate models, using observations (reanalysis) as reference.” at lines 92-93.

l. 143: Please insert space between m and s^{-1}. E.g., in Latex through m\,s^{-1}.

Thank you for spotting this, we corrected it.

l. 154: This explanation is very helpful. Please consider to provide this earlier in the manuscript, e.g., when first introducing vertical gradients (e.g., l. 135) since I assume that they are heavily affected by the coarse vertical resolution.

We thank the reviewer’s suggestion, however we prefer to keep the two statements separated since the sentence at line 135 refers to the theoretical aspects of LWA, while the discussion at lines 154-158 refers to our methodology and the availability of data.

l. 185: Quite often weather regimes are defined as circulation patterns that persist for several consecutive days. Are you using any minimum persistence criterion in your work or is it possible that weather regimes sometimes persist for a single day only? Please explain.

Thanks for the comment. We do not adopt here a persistence criterion for defining the regimes, so all days in the timeseries are considered in the clustering and regimes can persist even for a single day. This is in line with the analysis in Fabiano et al. (2020).

l. 197: Are you meaning the northeastern instead of the northwestern Pacific? Actually, the storm track is quite active over the northwestern Pacific at least in terms of cyclone activity.

Yes, we meant northeastern Pacific, since we are referring to the regions in which the meridional PV gradient appears weaker.

l. 201: This sentence is slightly confusing. I would rather state that "The local LWA clearly maximizes at the downstream end of the storm tracks". Compared to the classical definition of storm tracks, e.g., Chang et al. 2002, the maxima in LWA are much further east than the maxima in cyclone activity or geopotential height standard deviation.

We changed our sentence according to the reviewer's suggestion.

l. 254: This is an important result! Based on this I was wondering if the authors would like to reconsider the title of their paper. In its current form the title gives the impression of a pure verification study. However, seeing these results there is an important message about the model resolution to adequately represent Rossby waves. Perhaps asking "what resolution is needed to better represent Rossby waves in climate models?" could increase the visibility of the study. At least from the hemispheric perspective there is a clear result. For Europe, however, things seem to be much more unclear. I would leave the final decision about the title of the study to the authors.

We thank the reviewer for this nice suggestion, however we prefer to keep the title as it is.

l. 324: This statement is confusing. Figure 4b clearly shows the benefit of the HR. Am I missing something important here? Please clarify.

The reviewer is right, our previous statement was confusing. We referred to the fact that in Figure 2 (and 13) the impact of HR is not very clear over the EAT sector. We have changed the sentence at lines 323-325 with:

"As anticipated in Section~3, in our comparison we focus only on transient LWA associated with RWPs, since the benefit of a higher resolution are less evident in the LWA distribution of the multimodel mean (compare Fig.2 (c) and (d))."

l. 329: How is this threshold of 0.5 chosen? In weather forecasts, a pattern correlation (anomaly correlation coefficient) of less than 0.6 is considered to be useless. In this study it is the pattern correlation of the climatological mean. Would one not expect to see considerably higher correlations?

We are not aware of other studies which computed pattern correlations using LWA as a variable therefore we chose this value pretty much arbitrarily.

Such lower values of pattern correlation compared to anomaly correlation coefficient computed using geopotential height may be due to the fact that LWA is a highly derived quantity and its field it's not as smooth as geopotential height. We are applying here a very challenging diagnostic for models, since this requires both that the regimes are well represented (as in Fig. 3 of Fabiano et al. 2020), and that additionally the LWA composites for each regime are well represented. This is why we did not expect very high correlations here. In addition our correlation is a spatial correlation of a climatological mean which does not involve any time dependence.
With that said, there are many models that have much better performances (above 0.8), so the choice of the 0.5 threshold has been probably too conservative and we changed it to 0.6.

We modified our sentence at lines 328-331 in:

"It can be seen how the majority of the PRIMAVERA models represents the transient LWA pattern in a satisfactory way, with values of pattern correlation larger than 0.6 (apart from the CMCC model for the AR regime and NAO+ in the HR), with some of the models having a correlation larger than 0.8 (the best model in this sense is EC-Earth which has a correlation coefficient larger than 0.8 for all four WRs).”

l. 344: Given the many panels in each Figure, please label these (a, b, c etc.). This would help the reader enormously.

We labelled the panels of Figs.8-11 and now we refer to the labelling in the discussion of WRs in PRIMAVERA at lines 347-384. Thanks for this suggestion.

l. 366: Remove double "the".

Thank you for spotting this, we removed it.

l. 405 and elsewhere: Why are the authors switching to past tense? Please check carefully for consistent tense in this last section.

We removed the inconsistencies in the tenses using the present tense. Thank you.

References:
Chang, E. K. M., Lee, S., & Swanson, K. L. (2002). Storm Track Dynamics,
Journal of Climate, 15(16), 2163-2183.

References:

Fabiano, F., Christensen, H.M., Strommen, K. *et al.* Euro-Atlantic weather Regimes in the PRIMAVERA coupled climate simulations: impact of resolution and mean state biases on model performance. *Clim Dyn* **54,** 5031–5048 (2020). https://doi.org/10.1007/s00382-020-05271-w