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

## Summary and Recommendation – Reviewer #1

This work assesses the impact of model horizontal resolution on the simulation of local wave activity in the PRIMAVERA over the Northern hemisphere and Euro-Atlantic region. They classified each model to lower resolution (LR) and higher resolution (HR) versions, and then compared the ensemble mean of these two groups. They found no evident improvement of transient wave activity simulation for higher resolution. However, the conclusion based on the ensemble mean is questioned because the lower and higher bounds of model resolutions have a great heterogeneity (see the major comment). I recommend the authors to perform a major revision and reorganize the results by considering the comments listed below.

**Major comment:**

One of the goals of this work is to assess the impact of model resolution on the simulation of local wave activity. However, the classification of the model resolution onto HR and LR is questioned based on two reasons. On the one hand, the range of atmospheric and oceanic resolutions among different models are quite large. For example, the atmospheric resolution of LR in ECMWF-IFS is 50 km but the atmospheric resolution of HR in CNRM-CM6, EC-Earth3, HadGEM-GC31 is 50 km. On the other hand, the dynamics of LWA is dependent on spatial scales of atmospheric eddies. I thus suggest the authors to reorganize the results according to the range of atmospheric resolution: 250 km (Lower Resolution), 100 km (Standard Resolution), 50 km (Higher Resolution), 25 km (Highest Resolution) as in Scaife et al. 2019.

We acknowledge we have classified alla PRIMAVERA models in two broad categories: standard/Low resolution (LR) and High Resolution (HR), based only on the increase of the horizontal resolution of the HR compared to LR following Fabiano et al 2020 and Fabiano et al 2021. As both reviewers point out (see also comment made by the second reviewer) both the LR and HR sets contain simulation with an horizontal resolution which can vary significantly. The classification proposed by the reviewer would help to enlighten the representation of the LWA dynamics associated with atmospheric eddies with the horizontal resolution. In the revised version of the manuscript we will explore different ways of categorizing the models, possibly updating some of the figures (in particular Figure 7), and we will take into account the reviewer's concerns about the interpretation of the results.

**Minor comments:**

a. The authors are also encouraged to add results based on more models in this project if the data are available. This may be helpful to reduce the impact of single model on the ensemble mean.

In our work we included all the PRIMAVERA models for which daily data on some vertical pressure levels of the variables used to compute LWA (horizontal wind, temperature and geopotential height for the Mongtomery streamfunction ) are available. One model (AWI) was not included since such kind of data were not available.

b. In Figs. 5-6, 8-11, I suggest to plot the wave activity anomalies instead of total field to make consistency with the stream function anomalies.

We decided to include the full transient LWA field since we thought it would be of easier interpretation for the reader, given that LWA is not a widely used diagnostic yet. The idea is to identify the areas of high transient eddy activity with LWA and the associated circulation (cyclonic VS anticyclonic) with the Montgomery streamfunction anomalies. The stationary/climatological component has already been removed from transient LWA and it is somewhat consistent with the streamfunction anomalies, which are calculated with respect to a certain climatology. We agree that plotting transient LWA anomalies would help to identify

whether the error committed by the model against reanalysis is in the amplitude of the wave or a phase shift. We will produce such LWA anomalies maps and include them in the main body of the text or in the supplementary material.

In Fig 7, the spatial correlation should be correlation of the anomalous LWA between observation and models, since the high correlation in the current figure has contributions from the climatological pattern.

We computed the spatial correlation between models and observation using transient LWA, in which the stationary/climatological LWA has been removed. This is done to show how well the different models can simulate these patterns. We are not sure we completely understood the reviewer's comment in this respect.

c. The right bracket was missing before “and Montgomery”.

d. For Figs. 9-11, the captions can be changed to “As in Fig. 8 but for NAO-, SB, AR”.

e. Line 167: please delete the “DOI???”.

f. Line 270: please correct the phrase “in the four 4 WR”.

We thank the reviewer for spotting these mistakes/typos. We are going to correct them.

**References:**

Scaife, AA, Camp, J, Comer, R, *et al*. Does increased atmospheric resolution improve seasonal climate predictions? *Atmos Sci Lett*. 2019; 20:e922. https://doi.org/10.1002/asl.922

## Reviewer #2

This study by Ghinassi et al. aims to document the representation of midlatitude Rossby wave activity in state-of-the-art climate models of the PRIMAVERA project. The authors find that overall the models reasonably well represent the climatological mean wave activity (LWA). Increasing the resolution of the models generally improves the representation of the stationary LWA but not necessarily of the transient LWA. Further, it is found that an improvement of the models can only be observed when both the oceanic and atmospheric resolution is changed. For models where only the atmospheric resolution was changed a worsening of the models' ability to represent the LWA is detected.

The study is well written and the figures are clear. However, I have some major and in my view important comments which need to be addressed. Once these comments (which may change the interpretation) have been addressed I am happy to provide a more detailed review. Just to make sure, I am generally very positive about the work. Thus, I highly encourage the authors to submit a revised version of the manuscript to WCD.

**Most important comments:**

1) Statistical significance of results: An important part of the manuscript is the discussion of the differences between reanalysis and the PRIMAVERA simulations. In my view, the discussion lacks two important aspects. 1) More quantitative information concerning the biases would be very helpful. Accordingly, I suggest to revise e.g. Fig. 2 by showing the mean LWA as contours and the differences between the PRIMAVERA simulations and reanalyses in shading. This would clearly highlight the regions associated with the most pronounced biases. A calculation and discussion of the statistical significance of the differences is missing. Thus, I would like to ask the authors to provide some information on the statistical significance. For example, a bootstrap approach with replacement would be suitable to analyse the significance of the results.

We will provide some information on the statistical significance. A bootstrap approach with replacement is definitely a way to analyse the significance of the differences against interannual variability for the multimodel plots of Figure 2.

2) Choice of the isentropic level: I absolutely agree that the 320 K isentropic level is a suitable choice to investigate the midlatitude LWA during Northern Hemisphere winter. However, I am wondering how this level affects any interpretations concerning RWPs along the subtropical jet. To me it is quite unexpected that no signal of LWA activitiy is found along the subtropical jet which stretches from Northern Africa, across the Arabian Peninsula towards India during Northern Hemisphere winter. Therefore, I encourage the authors to either include an additional higher isentropic level in their analysis, or to at least comment on possible model biases in terms of LWA along the subtropical jet.

The point raised by the reviewer is certainly of interest, however in this first work we want to focus our attention on the extratropical LWA associated with the eddy driven jet. In general, we observed that at 320 K the wintertime LWA along the subtropical jet has a much weaker magnitude compared to LWA found along the eddy driven jet (as it can be observed in our Figure 1, but see also Figure 2 of Huang and Nakamura 2017 or Figure 1 panel a of Nakamura and Solomon 2011 part II). This is likely to be related with a weaker PV gradient associated with the subtropical jet compared to the PV gradient in the midlatitudes. Considering an additional higher isentropic level to analyse subtropical LWA we believe would add too much complexity to the present work and move the focus away from the mid-latitudes.

3) Classification into HR and LR: A major goal of the study is to investigate the impact of model resolution on the LWA. To this regard, the models are classified into LR and HR. However, in its current form the classification is questionable since LR and HR actually include model runs with the same atmospheric resolution. For example, the CNRM-CM6 with 50 km is classified as LR whereas the ECMWF-IFS with 50 km is classified as HR. Accordingly, I suggest to reorganize the classification so that each of them only contains models with a similar range of atmospheric resolution. In the same way, it would be intriguing to classify the simulation based on the ocean resolution (100 km vs 25 km). I would leave the final decision concerning this latter aspect to the authors.

As we discussed in the reply to reviewer 1, in the revised version of the manuscript we will explore different ways of categorizing the models, and we will take into account the reviewers' concerns about the interpretation of the results.

4) WR identification: The authors state that "to allow the comparison between different models and the observations we choose to work with the same reference reduced phase space for all simulations, defined by the 4 leading EOFs obtained from ERA5 reanalysis." Accordingly, the anomalies from the models are projected onto the reference space. Though I understand the reasoning behind, the reader is left wondering on how potential model biases affect the projection onto the reference space. Is any bias correction of M performed prior to the projection? This important information needs to be included and I suggest the authors to perform a bias correction prior to the regime identification.

For each model, the climatological mean field of M is removed before projecting on the reference space. So any mean state bias of the models does not affect the projection. As discussed in Sect. 3.1 of Fabiano et al. (2020), this step of the procedure is crucial for comparing results from different models/simulations. We will provide more detailed comments on this in the revised text.

**Minor comments:**

l. 40: Better write (e.g., wind, geopotential height, mean sea level pressure) instead of (wind, geopotential height, mean sea level pressure…)

l. 225: This shift is consistent with Quinting and Vitart (2019) who found the same behaviour in models of the S2S reforecast data base.

l. 248: The absence of a significant trend is consistent with Souders et al. 2014 who also investigated trends in RWP frequency, activity, and amplitude.

We thank the reviewer for the suggestions and we will include these citations, but probably in the conclusions where we comment on our results in the context of literature.

How did you estimate the significance of the trend?

In Figures 3 and 12 we plotted the transient LWA (coloured lines) plus or minus a standard deviation from the mean of each time series (shading). We did not perform a significance test to verify the presence/absence of trends but our conclusion was made analysing time series (with the shading/standard deviation representing the interannual variability) by eye. We will make this statement more rigorous in the revised version of the paper (e.g. performing a significance test).

l. 287: The relation between transient RWPs and blocking found in this study is consistent with the results of Altenhoff et al. (2008) and Quinting and Vitart (2019).

Thank you for suggesting these studies.

l. 365: Again, how is the significance of the trend determined? And can you actually quantify the magnitude of the trend?

Here we used the word “significant” as “substantial”, but no significance test has been performed. We will perform a significance test and add some quantitative comments in the revised version.

**References:**

Souders, M. B., Colle, B. A., & Chang, E. K. M. (2014). The Climatology and Characteristics of Rossby Wave Packets Using a Feature-Based Tracking Technique, Monthly Weather Review, 142(10), 3528-3548.

Altenhoff, A. M., Martius, O., Croci-Maspoli, M., Schwierz, C. & Davies, H. C. (2008) Linkage of atmospheric blocks and synoptic-scale Rossby waves: a climatological analysis, Tellus A: Dynamic Meteorology and Oceanography, 60:5, 1053-1063

**References:**

Huang, C. S. Y., and Nakamura, N. (2017), Local wave activity budgets of the wintertime Northern Hemisphere: Implication for the Pacific and Atlantic storm tracks, *Geophys. Res. Lett.*, 44, 5673– 5682, doi:10.1002/2017GL073760.

Nakamura, N., & Solomon, A. (2011). Finite-Amplitude Wave Activity and Mean Flow Adjustments in the Atmospheric General Circulation. Part II: Analysis in the Isentropic Coordinate, Journal of the Atmospheric Sciences, 68(11), 2783-2799, DOI: https://doi.org/10.1175/2011JAS3685.1