This paper examines benefits of increasing resolutions in CMIP5 GCMs and EURO-CORDEX RCMs on extreme temperature, precipitation and wind indices. The paper examines a critical topic in the field of model development. In my opinion the paper serves to what was proposed. However, I have a couple of major comments about the methodology and a few of minor comments about the manuscript. I would strongly encourage the authors to add at least a qualitative discussion on the points mentioned below.

Major comments:

1) Authors have used empirically computed return periods. This method of computing return periods has a major limitation that it only considers the rank and not the actual magnitude of the data. Therefore, the largest return period computed here from 36 years of data cannot exceed 36 years. The method does not calculate return periods with sufficient accuracy in some cases such as a trend in the data or passage of a single storm of unusually very high intensity. In such cases the return periods will be underestimated. A proper way of estimating return periods is to fit a theoretical generalized extreme value (GEV) distribution to block maxima (annual maxima, here), and then compute return periods from using the parameters of the fitted distribution.
I would encourage authors to recognize this aspect of the limitation in their methodology and associated impact on return periods.

2) Authors have bias adjusted the data before computing return periods. This artificially reduces/ enhances the model maxima to appear closer to the observational estimates. In my opinion this hides the "true" model performance. For example, a comparison of Fig. 6 and S9 suggests that models perform poorer when evaluated without bias-correction. In my opinion an objective model estimation should not include bias-correction. Bias-correction should only be used after a model has been evaluated.

3) Authors have used "multimodel means". Multimodel means could strongly be affected by one or two outlier model. Instead, multimodel medians are more robust in a sense that they are rather insensitive to any outliers. I would encourage authors to discuss this limitation in their manuscript.

4) Also, it appears that authors have used all ensemble members from a model to compute "MM all" as in figures such as Fig. 1, 3 etc. This method of computing multimodel means gives more weight to a model with multiple ensemble members than a model with a single or a smaller number of ensemble members. This will also most likely result into model biases that are not representative of "common" model biases across different sets of GCMs.

Minor comments:

1) At several places (e.g., lines L201, 251, 265, 281 etc.) authors have used the term 'observation/s" for observational datasets (E-OBS, MESCAN etc.). These observational datasets are not "observations".
2) L206: What is OSTIA?
3) L80: "Precipitation extremes tend to get heavier and agree better with observations". When? With increasing resolution? If yes, please mention this.
4) L77-79: I do not understand this sentence completely. What do you mean by "for grid point models"? I suggest authors to use short and simple sentences instead of a long complex sentence (e.g., L477-480).
5) L318-319: The differences will be bigger when return periods are computed without bias-adjustment.
6) L323:324: This statement does not seem to be completely true. Cold biases in Scandinavian region increase considerably from CMIP5 to CORDEX 0.44. Also, the warm bias in the eastern part of Central Europe has decreased from CMIP5 to CORDEX. The "insensitivity" observed in Fig. 2 may be partly due to bias-adjustment.
7) Fig. 5 has color scales swapped between mean and bias figures. As a general comment, I would highly encourage authors to use different color schemes for representing totals and biases. Using once color scheme for both is very confusing while examining many figures.
8) L470-471: Return level plots are not distribution plots. Shape of the annual maximum can only be examined by estimating shape parameters of the fitted GEV distribution. Please correct this sentence.
9) L589-590: Are you referring to Fig. 6 here? If yes, please mention this for clarity.
10) L590-592: It appears that for Rx1day downscaling is more dominant over 3 out of 5 regions examined.
11) Table S1: Replace "donate" with "denote".
12) Table S1: "with those forming part of the "common subset" in **bold**. I think "bold" should be replaced by "colors".