The present article aims to investigate the physical and chemical drivers of extreme levels of Ozone in the Great Bay area of South China during the typhoon approach. The article shows the results of an intercomparison work of numerical simulations and the application of a source apportionment method to analyse the causes of extreme events and ozone high levels.

The content of the present manuscript is of interest to the scientific community both for the occurrence, more frequency of extreme weather events and their link with air pollution, and for the evaluation of the performance of CTM such as WRF-Chem in reproducing these events. From this point of view, the manuscript has certain relevance in the field and for the journal. On the other hand, I found the manuscript need more work to improve the clarity of the facts exposed and a better choice of images to show.
One of the starting questions reading the manuscript, is relative to the configuration of the model, the spatial resolution, and the inputs used for the weather and emissions. These elements should be mentioned in the text and in a table to let other scientists replicate the experiment in the future.

Secondly, the validation of the model performance should be more evidently quantitative. I appreciate the tentative to make the evaluation easy to understand by keywords (good/bad) and colours (red/blue) but it would be also good to know the real numbers behind the model performance. How is the best bias? How's the worst? These are numbers that in the most important intercomparison exercises are provided.
Said that I found interesting and intuitive the use of the Index linking MNB and R to show the model performance and I believe it's a kind of combined metric that would find replication in other works.
Finally, the images chosen for the manuscript are too small and too many. I suggest selecting those that need to stay in the main text and making them bigger in order to allow the reader to have a clearer way to examine them. Alternatively, I suggest combining the information provided in some multi-panel images in a small number.

Major comments:

- Section 3.2.1: How do the authors calculate the average time series that they show in Figure 4? Are they averaging the values in the whole GBA? Are they accounting only for the land part or also the water? Considering that the two models start to diverge in a night-day cycle could be possible to see also how the observations perform in that cycle?

- Section 3.2.2: Lines 250 - 255: this change in the wind direction between two simulations that should use the same re-analysis data as input of meteorology needs more clarification. Assuming the spin-up between the simulations is the same, the input data are the same and all the other parameters are the same, how do the authors justify this difference? Do the authors use any kind of nudging option in WRF to constrain the model outputs to the initial re-analysis fields?

- **Minor comments:**
  the authors should provide a more detailed description of the model configuration adopted for the simulations. In particular, there is information relative to spatial resolution, initial and boundary conditions used to feed the model, emission inventories for the chemistry and input of meteorology for WRF that should be mentioned and motivated in the methodology section to make the experiment replicable.

- Even if we appreciate that the authors are trying to make the validation easier to read it would be advisable to generate a range of confidence in model performance on the basis of the values of MNB, R and Index.

Typos:
Line 75: The authors mention " black square" but they don't mention Figure 1. Modify to (Figure 1, black square)
Line 133: Fig.2 to substitute into Fig.2 (b)