Comments from Review 2:

This paper described distinct pollution levels in two consecutive months of winter 2015 in North China. and argued that such a feature is regulated by meteorological patterns connected to the El Nino and Arctic Oscillations (AO). The study uses observations of one single super El Nino year to raise the hypothesis on ENSO/AO influence on haze distribution. The study then used WRF model simulations of three other super El Nino years to test those hypotheses. In terms of mechanism, this paper claims that the combination effect of El Nino and AO can influence the intensity of EAWM and thus result in an anomalous PM2.5 levels. I found the overall presentation straightforward and the topic is within the scope of ACP. However, the following concerns shall be addressed before its publication.

Response: We thank the reviewer for the comprehensive evaluation and comments to help to improve the manuscript. Please see the detailed responses to your comments below.

1. This paper states the importance of boundary layer height several times without mentioning the schemes and calculation used in the model. Please clarify.

Response: The PBL scheme of Mellor-Yamada-Janjic was used and added in the revised manuscript.

2. The simulation domain was not specified at all. It says “. . .with physics options same as those discussed in Gao et al. (2017). . .” . However, in Gao et al. (2017), the simulation was conducted in U.S. but this study is done for East Asia. Please specify the basic domain information.

Response: The domain information was added in the revised manuscript (the third paragraph of section 2), which is also shown below.

“The domain covers majority of East Asia (shown latter; i.e., Fig. 5b), with spatial resolution of 36 km by 36 km. The pressure of the model top is 50 hPa, with lambert conformal conic projection centered at 34°N, 110°E. A total of 34 layers were used, with the top of the first layer at about 40 meters.”

3. For biomass burning, the paper states “. . .biomass burning emissions include open burning of agricultural residue, calculated based on crop yields, fraction of biomass burned in the open field. . .” . But the paper did not specify the inventory used in this case. So which inventory did you use? FINN, GFED or any other self-developed source specified for this region?

Response: In this study, we used self-developed the emission inventory, developed by

Tsinghua University. The details have been added in the revised manuscript (last paragraph of section 2), which is also shown below:

"The emissions from open burning of agricultural residue have been included in the anthropogenic emission inventory developed by Tsinghua University. They were calculated based on crop yields, the ratio of residue to crop, the fraction of biomass burned in the open field, and emission factors (Wang and Zhang, 2008;Zhao et al., 2013;Zhao et al., 2018)."

4.The length of spin-up time is questionable. In this paper, the spin-up time for each simulation is only one week. Kumar et al. (2013) shows that it takes WRF-Chem about 10 days to spin-up for free atmosphere and 20 days to spin-up for surface level.

Response: Regarding the spin-up period, we have done a few tests previously, and found no significant differences once the spinning-off period reaches a week or longer. In some other studies, even fewer spinning-off period was used, i.e., Im et al. (2010) used 3-day and Sartelet et al. (2007) used one-day as spin-up period. Therefore, the one-week spin-up time should be sufficient.

Im, U., Markakis, K., Unal, A., Kindap, T., Poupkou, A., Incecik, S., Yenigun, O., Melas, D., Theodosi, C., and Mihalopoulos, N.: Study of a winter PM episode in Istanbul using the high resolution WRF/CMAQ modeling system, Atmos. Environ., 44, 3085-3094, https://doi.org/10.1016/j.atmosenv.2010.05.036, 2010.

Sartelet, K. N, H. Hayami and B. Sportisse. Dominant aerosol processes during high-pollution episodes over Greater Tokyo. Journal of Geophysical Research: Atmospheres, 2007, 112(D14).

5.Fig. 3, 4 and 5 all stated that "Stippled areas indicating exceedance of 90th confidence interval." But what kind of statistical test was applied here? Also what are the samples?

Response: The 90th confidence interval was calculated for each grid separately based on a two tailed t-distribution method using the thirty-year (1987-2016) data. The sample size is 30. For example, for December 2015 (Fig. 3) in terms of geopotential height (GHT) at 500 hPa, 90th confidence interval was calculated using GHT from December 1987-2016. Then, the value in December 2015 for each grid was compared to the 90th confidence interval. This has been reflected in the caption of Figure 3 in the revised manuscript.

6.The last paragraph on future research is also a bit puzzling, running SST forced experiment can be helpful for ENSO, but AO is difficult to be forced by SST.

Response:

The reviewer is correct that SST is directly related to ENSO. However, as was discussed in Geng et al. (2017) as well as our manuscript, we found that during the super El Niño events including 1982/83, 1997/98, and 2015/16, in particular during the peak of El Niño period, it was accompanied by a rapid sub-seasonal North Atlantic Oscillation (NAO)/Arctic Oscillation (AO) phase reversal from a positive to negative phase. The ENSO and AO is interconnected in this regard. Therefore, we propose to conduct a SST experiment to further elucidate the mechanism and the subsequent impact on haze formation.

Geng, X., Zhang, W., Stuecker, M. F., and Jin, F.-F.: Strong sub-seasonal wintertime cooling over East Asia and Northern Europe associated with super El Niño events, Sci. Rep., 7, 3770, 10.1038/s41598-017-03977-2, 2017.

1. Please consider to replace the sequential color schemes with divergent color schemes when showing the anomalies (e.g., Fig.1 and Fig.7). Fig 3 is a better example.

Response: The color schemes have been changed based on the reviewer's suggestions.

2. When drawing the boundaries of NCP in Fig.1, please be more rigorous. The box area is not entirely NCP. It also contains part of Bohai Sea and Inner Mongolia Plateau. Although this may not affect your final results but can cause misleading when saying this is NCP.

Response: The box does not include Inner Mongolia (Fig. 1 of the revised manuscript), and we added the relevant descriptions, i.e., the exclusion of Bohai area.

3. What is the unit of the wind vector in Fig. 3? m/s?

Response: The unit of m/s has been added in the caption of Fig. 3.

4. The captions of this paper need to be clearer. There are lots of figures with sub panels (e.g., a,b etc.) but they are not specifically mentioned in the caption. This way of description can be very confusing.

Response: All captions have been checked and made clearer by adding more specific descriptions.

5. Line 204-205: "the emissions of SO_2 in January is usually higher than December primarily due to a higher power demand" this statement needs a reference.

Response: We have rephrased the sentence and SO_2 emission was inferred by the concentration. The revised descriptions (last paragraph of section 3.2) are shown below:

Moreover, the anthropogenic emissions in January could be comparable to or higher than that in December, i.e., in January 2016, higher SO_2 concentration, implicative of SO_2 emissions, was found than December 2015 based on observed data (http://www.pm25.in; not shown).

6. Fig. 6b shows very low PM2.5 concentrations for major cities other than China and India. Why is this the case? How about cities like Tokyo, Osaka, Seoul and Bangkok etc.?

Response: Based on the data available to us, we added the evaluation of 13 stations in Japan, South Korea and Thailand. The observational data is from EANET (Acid Deposition Monitoring Network in East Asia, http://www.eanet.asia/product/index.html). The concentrations in these cities are lower than that from the major cities in China. Please note over Cheju and Kanghwa in South Korea, only the data in January 2016 is available. From the scatter plots shown below, we can tell that the model performs well (with low mean bias and error) among these stations. The descriptions as well as the figure and table have been added to the part 2 of the supporting information.

Table S1. Station information of EANET sites

	Japan										South Korea		Thailand
Stations	Rishiri	Ochiishi	Tappi	Sado-seki	Happo	Ijira	Oki	Banryu	Yusuhara	Hedo	Cheju	Kanghwa	Bankok
Latitude	45.12	43.20	41.25	38.25	36.68	35.57	36.28	34.67	32.73	26.78	33.52	37.74	13.75
Longitude	141.23	145.52	141.35	138.40	137.80	136.70	133.18	131.70	132.98	128.23	126.52	126.49	100.50

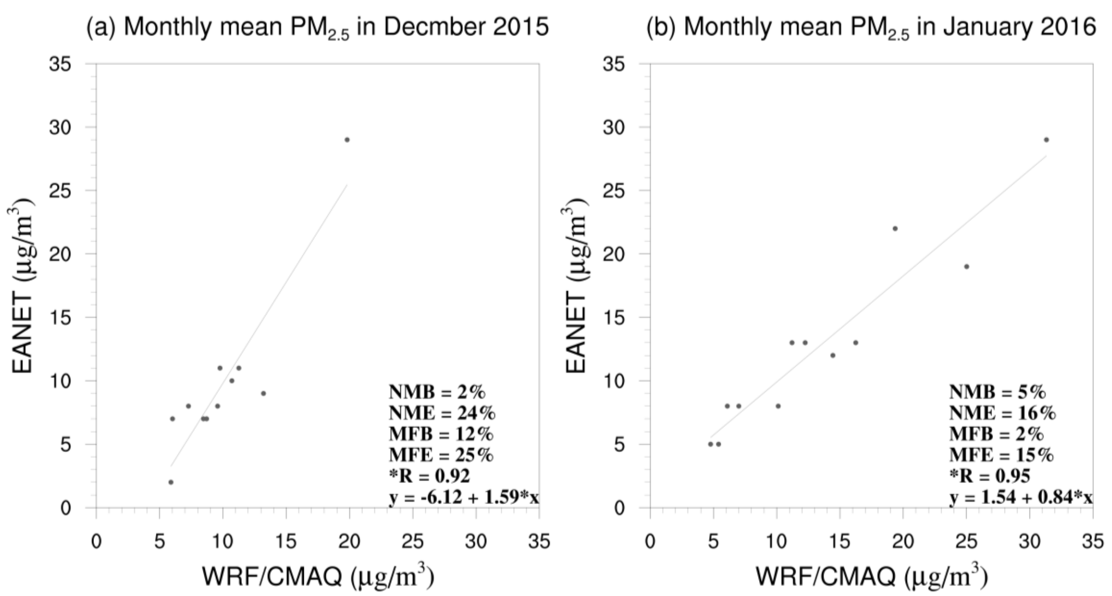

Figure S9 Evaluation of monthly $PM_{2.5}$ in CMAQ based on selected EANET observational data: (NMB: Normalized Mean Bias; NME: Normalized Mean Error; MFB: Mean Fractional Bias; MFE: Mean Fractional Error; R: correlation coefficient). The statistical significance of the linear correlation coefficient was performed and *R implies statistical significance at the 95% confidence level.

7. In Table 1, can you explain the large bias of 201512 WD10? It is almost 20 degrees and should not be treat as negligible.

Response: slightly larger bias (19.25°) of wind direction at 10-m (WD10) in December 2015 is partly attributable to the model comparison with observed values close to 0° or 360°, which may yield large bias albeit the small differences in reality (i.e., 10° in model vs. 350° in observation). This has been added in section 4.3 in the revised manuscript.

8. Line 62. “Formation” should be distribution. A relevant reference is Chen et al., (2018).

Response: “Formation” has been changed to “distribution”.