To reviewer 2:

Thanks so much for your general and detailed comments. We have fully revised this manuscript according to these comments. We are more than willing to conduct further revisions if additional requirements are given.

1. It is dangerous to evaluate the pollution control strategy by using only four pollution episodes. Too many parameters, especially the meteorological parameters, can influence the pollution level in one case, and would result in large uncertainties in the evaluation. A comparison based a long-period observation is needed. The current comparison between every two episodes, at least, not statistical significant.

R: This is a very good question. Thanks for pointing this out. Actually, the "2+26" strategy and regional air pollution alert with were contingent and implemented for severe pollution episodes. Therefore, the evaluation of short-term contingent local and regional emission-reduction measures were mainly conducted based on the analysis and simulation of $PM_{2.5}$ concentrations during short pollution episodes with different emission reduction measures. In this case, a large amount of studies (Jia et al., 2017; Cheng et al., 2017; Wang, et al., 2019; etc) were conducted simply based on one or two pollution episodes to evaluate the effects of different emission-reduction measures. On the other hand, as you pointed out, to evaluate long-term emission-reduction policies, instead of contingent emission-reduction measures, a long-term simulation should be conducted, which is another type of research based on other statistical methods (e.g. Chen et al., 2019).

Chen, Z., Chen, D., Kwan, M., Chen, B., Cheng, N., Gao, B., Zhuang, Y., Li, R., and Xu, B.: The control of anthropogenic emissions contributed to 80 % of the decrease in PM2.5 concentrations in Beijing from 2013 to 2017, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-1112, 2019.

Cheng, N., Zhang, D., Li, Y., Xie, X., Chen, Z., M, F., Gao, B.B., He, B.: Spatio-temporal variations of PM2.5 concentrations and the evaluation of emission reduction measures during two red air pollution alerts in Beijing, Scientific Reports, 7(1), 8220,2017.

Wang Q, Liu S, Li N, et al. Impacts of short-term mitigation measures on PM 2.5 and radiative effects: a case study at a regional background site near Beijing, China[J]. Atmospheric Chemistry and Physics, 2019, 19(3): 1881-1899.

Jia J, Cheng S, Liu L, et al. An Integrated WRF-CAMx Modeling Approach for Impact Analysis of Implementing the Emergency PM2. 5 Control Measures during Red Alerts in Beijing in December 2015. Aerosol and Air Quality Research, 2017, 17: 2491-2508.

2. In Fig. 2, it seemed to me the simulation result was too good. And the model can only underestimate PM2.5, but not overestimate, why? The author need to provide the comparison results of the chemical composition, but not only the mass concentration of PM2.5.

R：Thanks for pointing this out. Actually, the $PM_{2.5}$ simulation result was satisfactory, but not too good compared with similar studies. Maybe the plot figure caused this confusion. According to your comment, we also added the simulation of meteorological factors and chemical compositions and presented a comprehensive accuracy assessment table in the revised manuscript. Thanks again for this point. The WRF-CAMx model generally underestimate $PM_{2.5}$ concentrations, not every day (For some days, the observed $PM_{2.5}$ concentrations can be lower than the simulated values). But for a heavily polluted episode, the averaged simulated $PM_{2.5}$ mass concentrations were generally lower than observed $PM_{2.5}$ concentrations, which was revealed by relevant studies. The possible reason for the underestimation of $PM_{2.5}$ concentrations using WRF-CAMx model might be attributed to this: the emission inventories for running this model, including industry and other categories, could not fully reflect the actual emission scenarios. Firstly, not all emission-sources can be included in the emission inventories. Secondly, the contingent emission-reduction measures during pollution episodes may not be fully implemented by all factories. Therefore, the actually emitted precursors were more than model-predicted and thus the WRF may underestimate $PM_{2.5}$ concentrations.

Specific comments:

1. Remove “recent” in the title

R: Corrected

2. I would not recommend use ‘Orange air pollution alert’ in the title.

R: Corrected.

3. In Fig.3, this kind of direct comparison between two cases at different time did not make much sense.

R: Actually, the “2+26” regional emission-reduction strategy for improving air quality in

Beijing was recently proposed contingent policy and just implemented for twice. Therefore, to fully evaluate the effects of "2+26" strategy on $PM_{2.5}$ reduction, we selected two pollution episodes, one in March, 2013 with no emission-reduction measures and one in November, 2016 with local emission reduction measures to compare with the two pollution alerts with "2+26" emission-reduction measures, one in November, 2017 and one in March, 2018. Since the major meteorological conditions, initial $PM_{2.5}$ concentrations and the month between the pollution episodes in March, 2013 with no emission reduction measures and March, 2018 with "2+26" emission-reduction measures, and the pollution episodes in November 2016 with local emission-reduction measures and November 2017 with regional emission-reduction measures were generally similar. Therefore, comparing the corresponding pollution episodes were an effective approach for understanding the effects of local emission-reduction measures and regional emission-reduction measures on improving air quality in Beijing during pollution episodes. That is the reason we employed four pollution episodes to demonstrate the effects of "2+26" regional emission-reduction VS No emission-reduction, and "2+26" regional emission-reduction VS local emission-reduction.

4.In Fig. 4b, why there were such a high concentration of nitrite and chloride?

R: Thanks for pointing this out. Due to data recording errors, the NO_3^- in the previous manuscript was wrongly used as the nitrite. We corrected this and added additional OC and EC to the revised manuscript. The updated figure was listed as follows. Thanks again for pointing this out and we are very sorry for this confusion.

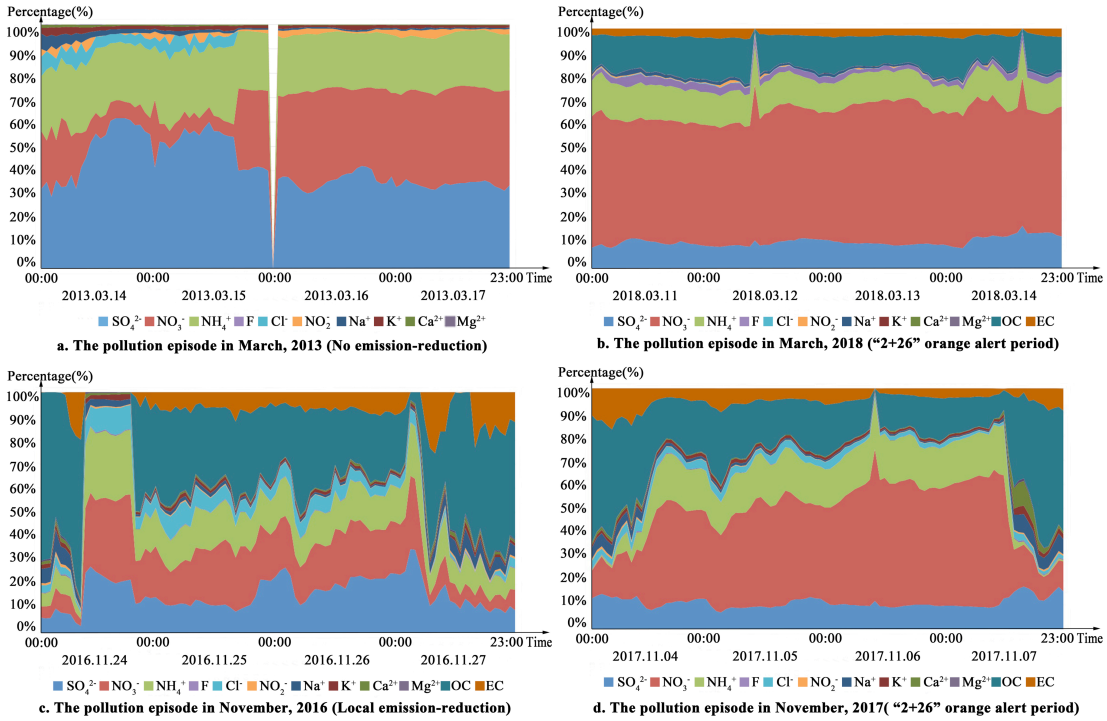

Fig 4. The variation of PM2.5 components in Beijing during four pollution episodes

5. In Fig. 5, compared to the previous pollution episodes, the contribution of coal combustion in March 2018 episode decreased a lot, but the November 2017 case did not, why?

R: As we know, $PM_{2.5}$ concentrations were highest in winter in Beijing, mainly due to the central heating (from November to March) required burning of coal materials. Since November, 2017, a large scale project "Coal to Gas" were implemented in the Beijing-Tianjin-Hebei region and a majority of coal ovens were replaced with equipment for gas burning in the "2+26" region, leading a notable decrease of the relative contribution of coal combustion to $PM_{2.5}$ concentrations. Based on the official assumption, the " coal to gas" project can lead to a 2 million-ton decrease in coal consumption in the Beijing-Tianjin-Hebei region.