<u>Response to Reviewer #2:</u>

The manuscript entitled "An approach to refining the ground meteorological observation stations for improving PM2.5 forecasts in Beijing-Tianjin-Hebei region" introduced an approach to refine the ground stations by identifying the sensitive areas for targeted observations. The study is highly related to the studies of predictability, target observation and data assimilation. And it provides a scientific guidance on optimizing the ground stations. I believe the approach is not only useful for air quality forecasts, but can also be used to the forecasts of extreme weather events. Nevertheless, there is a gap between publication and the manuscript in current version. I hope the following comments will help authors improve the manuscript.

Response: We thank your appreciations.

<u>Specific comments:</u>

1. Line 42. There are a great many publications addressing the meteorological conditions on PM2.5 variations, but the authors only cite one, which is not enough. More references are needed here.

Response: We thank the reviewer's suggestions. We will add the references on Line 42 (Lou et al., 2019; Chen et al., 2020).

2. Line 68. "assimilating more observations may not necessarily lead to much higher forecast benefits." References are needed here.

Response: We will add the references here (Li et al., 2010; Liu et al., 2021).

3. Line 75. How are the worse forecast skills possible when the sensitivities are low? Please provide a detailed explanation here.

Response: We will add a detailed explanation in the revised manuscript. Theoretically, if the observations in the area where the forecast is not sensitive to the initial errors are assimilated, the forecast skills might be slightly improved or neutral. However, in realistic forecasts, the imperfect assimilation procedure or the unresolved scales and processes in the model may induce additional errors and lead to the worse forecasts when the observations in the area where the forecast is not sensitive to the initial errors are assimilated (Janjic et al., 2018). For example, in Yu et al. (2012), removing the initial error in the area that is not the most sensitive area could worsen the prediction results of ENSO. That emphasized the importance of identifying the most sensitive area and suggests that additional observations should be assimilated more carefully in this sense.

4. Line 195-202. The descriptions are insufficient and confuse me. Please add more details and make it clear.

Response: Sorry for confusing the reviewer. We will rewrite the paragraph and make it clearer.

"The spectral projected gradient 2 (SPG2) method is used to solve the optimization problem in Eq. (3). It is noted that the SPG2 algorithm is generally designed to solve the minimum value of nonlinear function (cost function) with an initial constraint condition, and the gradient of cost function with respect to the initial perturbation represents the descending direction of searching for the minimum of the cost function. Therefore, in this study, we have to rewrite the cost function Eq.(3) as $J'(\delta x_0^*) = \min_{\delta x_0{}^T C_1 \delta x_0 \leq \beta} - [M(x_0 + \delta x_0) - M(x_0)]^T C_2 [M(x_0 + \delta x_0) - M(x_0)]$ and the WRF adjoint model is used to compute the gradient of the cost function. Specially, to calculate the CNOP, a first guess initial perturbation is projected into the constraint condition ($\delta \boldsymbol{x}_0^{(0)}$) and superimposed on the initial state ($\boldsymbol{x}_0$) of the WRF model. After the forward integration of WRF, the value of cost function, -[$M\left(\boldsymbol{x}_0 + \delta \boldsymbol{x}_0^{(0)}\right)$- $M(\boldsymbol{x}_0)$], can be obtained. Then, with the adjoint model of WRF, the gradient of the cost function with respect to the initial perturbation ($g(\delta \boldsymbol{x}_0^{(0)})$) is calculated. Ideally, the gradient presents the fastest descending direction of the cost function. However, in realistic numerical experiments, the gradient presents the fast-descending direction but not necessarily the fastest. So we need many more times of integrations. After iteratively forward and backward integrations of the WRF model governed by SPG2 algorithm, the initial perturbation is optimized and updated until the convergence condition is satisfied. Here, the convergence condition is $\left\| P\left(\delta \boldsymbol{x}_0^{(p)} - \boldsymbol{g}(\delta \boldsymbol{x}_0^{(p)})\right) - \delta \boldsymbol{x}_0^{(p)} \right\|_2 \leq \varepsilon_1$, where ε_1 is an extremely small positive number, $P\left(\delta x_0^{(p)}\right)$ projects the initial perturbation to the constraint condition. Finally, the CNOP ($\delta \boldsymbol{x}_0^{(p)}$) which presents the initial perturbation that causes the largest forecast errors using the SPG2 method can be obtained.

5. Line 313. Please clarify that the real "meteorological" observations are not in public archive, because in section 3.1, the authors have compared the simulations with the observed PM2.5 concentrations.

Response: We will clarify that the real meteorological observations are not available in public archive in the revised manuscript. The sentence will be corrected into "Since the real meteorological observations are not in public archive, the "additional observations" are correspondingly taken from the initial field of the truth run (i.e. the ERA5 data) and called as "simulated observations" according to the OSSEs.".

6. Line 322. Is the CNOP-type initial error that what has been described in section 2.3? It is suggested to add a detailed description on what variables the CNOP-type errors have contained here.

Response: Yes, the CNOP-type initial error is what has been described in section

2.3. We will add a detailed description of CNOP-type error here. The sentence will be corrected into "the CNOP-type initial errors which includes wind, temperature and water vapor mixing ratio components at the ground level are calculated for each of the 48 PM2.5 forecasts with the application of WRF and its adjoint model by using the SPG2 solver (see section 2).

7. Line 349. "the area with larger values of TME can be regarded as the sensitive areas". It is ambiguous. Is there a threshold for the definition of sensitive area or just determined subjectively?

Response: The TME is applied to measure the comprehensive sensitivity of PM2.5 forecast uncertainties on initial meteorological perturbations. When we identify the essential observational network, we take the 3% as the threshold to determine the sensitive area. Then a total of 424 sensitive grid points is obtained. We select the 3% as the threshold here because the number 424 of sensitive grid points is close to the number of 481 of the meteorological stations within and surrounding the BTH region.

8. Line 453. "the essential stations can indeed provide additional observations that help increase the skill of the PM2.5 forecasts, in comparison to other constructed stations but not in the sensitive grids". The authors did not do any comparison experiments to show the improvements are higher than assimilating the station observations which are not in the sensitive grids. How can they get such conclusions?

Response: We will correct the sentence into "It is clear that the essential stations can indeed provide additional observations that help increase the skill of the PM2.5 forecast in the BTH much significantly".

9. As shown in Figure 7 (a1, a2), assimilating the observations will lead to worse forecasts since the AEv and AEM are negative. It is hard to understand. Why will assimilating the observations will lead to worse forecasts?

Response: The negative PM2.5 forecast skills occurred at the AF initialized at 20:00 on Nov 18th 2016. For the AF initialized at 20:00 at Nov 18th 2016, the PM2.5 concentration in the truth run increases from 139.5 $\mu g\, m^{-3}$ to 151.5 $\mu g\, m^{-3}$ averaged over the BTH region; while the control run forecasts the PM2.5 concentration of 159.6 $\mu g\, m^{-3}$ averaged over the BTH region, 20.1 $\mu g\, m^{-3}$ higher than the PM2.5 concentration in the truth run. When all the constructed station observations are assimilated, the PM2.5 concentration averaged over the BTH region is 153.11 $\mu g\, m^{-3}$ at the forecast time, much closer to the truth run. However, the improvements in the BTH region are uneven (Figure 2), and the number of grids showing negative improvements overweigh those showing positive improvements, which results in a negative AEv and AE_M.

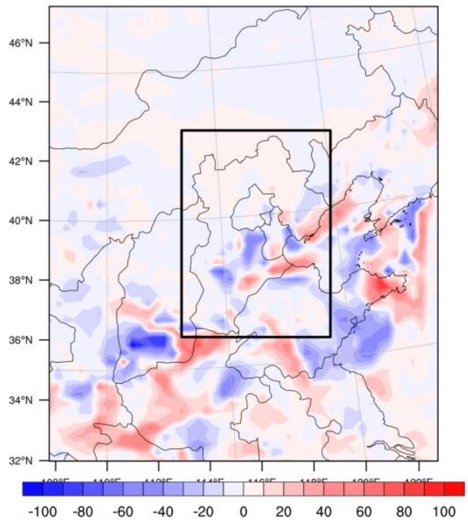

Figure 2 The improvements of PM2.5 forecast skills when all the constructed station observations are assimilated.

10. Line 703. "It is clear that assimilating the fewer observations can lead to higher PM2.5 forecast skills". It is inaccurate. It is suggested to rephrase it more carefully.

Response: We will correct the sentence into "It is clear that assimilating the fewer sensitive observations may lead to higher PM2.5 forecast skill".

11. It is suggested to mention the limitation of the study in section6 that the results are based on OSSEs. If the real observations are available, how the refined station observations help improve the air quality forecasts deserve deeper studies.

Response: We will add the limitations in the revised manuscript. Due to the unavailable of the meteorological observations from the Monitor center, we have to assimilate the simulated observations (the ERA5 data) to the control run to show the effectiveness of the cost-effective observation network. If the real meteorological observations are available, how the real observations from the refined station network can help improve the PM2.5 forecasts in the control forecast against the observations still needs further studies.

References:

Janjić, T, Bormann, N, Bocquet, M, Carton, JA, Cohn, SE, Dance, SL, Losa, SN, Nichols, NK, Potthast, R, Waller, JA, Weston, P., 2018. On the representation error in data assimilation, Q J R Meteorol Soc. 144: 1257– 1278.

Li, X., Zhu, J., Xiao, Y., and Wang, R., 2010. A Model-Based Observation-Thinning Scheme for the Assimilation of High-Resolution SST in the Shelf and Coastal Seas around China, Journal of Atmospheric and Oceanic Technology, 27(6), 1044-1058.

Liu C., Zhang, S., Gao, Y., Wang, Y. and coauthors, 2021. Optimal estimation of initial concentrations and emission sources with 4D-Var for air pollution prediction in a 2D transport model. Science of the Total Environment, 773, 145580.

Lou, M., Guo, J., Wang, L., Xu, H., Chen, D., Miao, Y., Lv, Y., Li, Y., Guo, X., Ma, S., Li, J., 2019. On the relationship between aerosol and boundary layer height in summer in China under different thermodynamic conditions. Earth Space Sci. 6 (5), 887–901.

Yu, Y., Mu, M., Duan, W., Gong, T., 2012. Contribution of the location and spatial pattern of initial error to uncertainties in El Niño predictions. Journal of Geophysical Research, 117, C06018.