Dear referee,

Thank you for your professional suggestions and guidance on our manuscript which are very important to help us to improve the quality of our paper. We have carefully studied all the comments and have made corresponding revisions. Here are our responses and revision details. The Reviewer's comments are in black and our responses are in blue. All line numbers correspond to the revised version with markers.

Sincerely,

Siting Li and co-authors

General Comments:

The article by Siting et al. is about the implementation of an Ensemble Optimal Interpolation (EnOI) method in the numerical chemistry & weather prediction system GRAPES_Meso5.1 / CUACE and the application of the method to try to improve $PM_{2.5}$ and visibility forecasts of pollution episodes in Eastern China. The authors strive to calibrate the parameters of the EnOI, namely the "localization length scale", which is the spatial range of the assimilation, the ensemble size, and the time at which assimilation should be carried out. They also investigate the impact of assimilating $PM_{2.5}$ observations on the simulated $PM_{2.5}$ concentration field. The mean error (ME) and the root mean square error (RMSE) on the initial $PM_{2.5}$ concentration field is reduced when assimilating the data. According to the authors, the forecasts of the $PM_{2.5}$ concentration and visibility fields are also improved throughout the lead time window of 24 hours, especially when the assimilation time is 1200 UTC because the discrepancy between simulation and observation is larger compared to 0000 UTC. Again, according to the authors, the visibility forecasts by assimilating $PM_{2.5}$ are further improved for light pollution episodes in comparison with heavy pollution episodes, which are more affected by humidity. Thus, for extreme low visibility during severe haze pollution, the authors recommend to assimilate both $PM_{2.5}$ and humidity observations.

The work carried out by Siting et al. essentially consists of optimal interpolation where the model error covariance matrix is evaluated by an ensemble approach with the members of the ensemble being previous timeframes of the $PM_{2.5}$ concentration field. Beyond the fact that the $PM_{2.5}$ field which assimilates data at a given time and at the close following times actually presents better statistics compared to the observations, the results obtained do not seem very convincing to me. However, this work constitutes an approach which deserves to be studied and is therefore worthy of publication. The article could be improved by taking into account the following remarks.

Response: Thank you very much for taking the time to read our paper. Sincerely thanks for the positive general evaluation and the detailed and professional comments and this is valuable for the paper improving, all the comments have been carefully addressed. Enclosed below are our point-to-point responses to these comments. The major revisions to this manuscript are:

- rewrite the introduction
- conduct more experiments with only 50% observations to be assimilated by EnOI and evaluate the impact of DA against the non-assimilated sites in sensitivity experiments as well as cyclic DA experiments. The accuracy of the forecast or analysis through DA method is measured by model-observation agreement mainly using RMSE and CORR as metrics and the improvement by EnOI is represented by reduction of RMSE, MB or enhancement of CORR. These results are presented

in Table 3,4,5 and Figure 5,7 in revised manuscript.

- In Section 3.4.1 impact on $PM_{2.5}$ forecast, we add comparison between our results and previous paper. The results in our study show that using EnOI to assimilate ground-based $PM_{2.5}$ observations for the model chemical initial field can reduce over 9.8% of RMSE for 24h forecast in average. Park et al. (2022) implemented an ensemble Kalman filter (EnKF) in the Community Multiscale Air Quality model (CMAQ model v5.1) for data assimilation of ground-level $PM_{2.5}$. They found using EnKF with 40 ensemble number can reduce 9% of RMSE for 24h forecast. Comparing their results with ours, we can find that, while EnOI is sub-optimal, it can give improvement of forecast that are comparable to those of the EnKF. Moreover, the computational cost of EnOI is typically about N times less than that of EnKF. Therefore, we suggest that EnOI may provide a practical and cost-effective alternative to the EnKF for the applications where computational cost is a main limiting factor, especially for real-time operational forecast.

L55 to 61 – Instead of listing a very large number of references, authors should briefly indicate what they contain.
Response: Based on your suggestion, we have made modifications to the original text accordingly and briefly elaborated on the cited literature to make our paper more readable. Lines 73 to 94 in the revised manuscript

L95 – Develop the acronym "EnOI" at least in the title of the section.
Response: Thanks for the suggestion, we have dropped the use of abbreviations at the headings. The title of Section 2.1 has been revised to "The ensemble optimal interpolation"**.** Line 149 in the revised manuscript

L102 – I don't understand the point of the double notation "x" and "psi". If there is a good reason, the authors should give it, otherwise the notations should be simplified.
Response: Thank you for the suggestion. " x" represents the ensembles obtained by using the Ensemble Kalman filter (EnKF) method (ensemble forecasting after perturbing the model uncertainty). " psi" represents the statistical sample of the long-time integration of the model in the ensemble optimal interpolation (EnOI) method. As you mentioned, to avoid too many notations, we have unified the two notations. Equations (1) to (8) in the text are a brief introduction to the EnKF method, while EnOI is introduced starting from Equation (9). Lines 170 to 204 in the revised manuscript

L110 - In fact, the ensemble is used to estimate the error covariance matrix of the model. The authors could say it simply and directly.
Response: Yes, you're right. We have made the appropriate adjustments as you suggested. Lines 196 to 197 in the revised manuscript

L112 – The relationship between "psi_a", "psi_f" and "psi_i" should be given.
Response: "psi_f" represents the background field of the model, which is the original chemical initial field of the model we need to assimilate, and "psi_a" represents the analysis field, which is the chemical initial field after assimilation. "psi_i" is the hour-by-hour forecast sample of the model, which is used to estimate the background error covariance of the model. Line 173, line 174 and line 179 in the revised manuscript.

L113 - What the acronym "AF" corresponds to is not indicated (even if we understand well). The many acronyms used throughout the text should be made explicit and, in my opinion, fewer abbreviations should be used.

Response: Thanks for your suggestion. Because of the length of this article, using abbreviations for infrequent words would only make it more difficult to read this article. Based on your suggestion, we have eliminated the abbreviations for words that occur infrequently.

L114 – As the scalar "alpha" is used to weight the model and the observations, one would expect to see "1 – alpha" in formula (6). Are the authors sure about this formula?

Response: yes, we sure about this formula. According to Evense(2003) and Oke(2008) , $\alpha \in (0,1]$ is the parameter giving different weights for the forecast and measurement error covariances, which can be adjusted to the size of the covariance for a specific application. The variance of a stationary ensemble over a long period usually overestimates the instantaneous variability; therefore $\alpha<1$. We have inserted the above as an explanation in the manuscript. Lines 199 to 203 in the revised manuscript

(References: Evensen, G. The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynamics, 53, 343-367, 10.1007/s10236-003-0036-9, 2003.

Oke, P., Brassington, G., Griffin, D., and Schiller, A.: Ocean Data Assimilation: a case for ensemble optimal interpolation, Australian Meteorological and Oceanographic Journal, 59, 10.22499/2.5901.008, 2010.)

L118 – Although the use of a length scale or spatial range in data assimilation is understood, this concept is poorly introduced. The expression "to avoid all observations…" is not clear at all and needs to be rephrased.

Response: Thanks to your suggestion, we have added a new Section 2.2 in the text dedicated to localization. Lines 206 to 219 in the revised manuscript.

L118 – What does the "localization scheme" look like? The authors should give the formula of it.

Response: Thank you very much for your suggestion, we have given the formula in Fig.1 and lines 209 to 219 in the revised manuscript.

L125 – We understand later in the reading of the article how the ensemble of PM2.5 concentration fields is constructed. It would be better if the authors explained it in this section of the article.

Response: We explain it in Section 2.2. Lines 225 to 230 in the revised manuscript.

L129 – The sentence "Compared with the traditional EnOI..." can only be understood after reading the rest of the article. This sentence should be explained at this point in the article.

Response: We have added a corresponding explanation in the article. Line 196 in the revised version describes the traditional EnOI ensemble sample acquisition, Lines 225 to 230 in the revised manuscript describes how the ensemble samples are obtained in this paper.

L134 – The horizontal and vertical dimensions and resolutions of the simulation domains used by the GRAPES_Meso5.1 and CUACE models should be indicated.

Response: GRAPES_Meso5.1/CUACE is an online coupled model, they have the same resolution and the simulation area of the model is adjustable. In Section 2.5, we describe the horizontal and vertical dimensions and resolution of the simulation domain used by the GRAPES_Meso5.1/CUACE model. Lines 285 to 287 in the revised manuscript.

L166 – What do the authors call a "warm" restart? Is that really the right term?

Response: The term "warm restart" refer to using initial condition with the previoud forecast. There are several papers mentioned this term but it is not very common phase so we have deleted it in our paper and rewrite the description of experimental setup. Lines 289 to 290 in the revised manuscript.

References:

Xu, Z., F., Hao M., Zhu L., J.: On the Research and Development of GRAPES-RAPES, Meteorological Monthly, 39(4):466-477, 2013, in Chinese.

Feng, J., M. Chen, Y. J. Li, and J. Q. Zhong: An implementation of full cycle strategy using dynamic blending for rapid refresh short-range weather forecasting in China. Adv. Atmos. Sci., 38(6), 943−956, https://doi.org/10.1007/s00376-021-0316-7, 2021.

L183 – Finally it is said that "the N hourly model forecasts before the assimilation moment were used as the ensemble samples to approximate" the model error covariance matrix. It could have been mentioned before.

Response: We have realized that it would be very unfriendly for the reader to understand this study without some previous mention of how the assimilation ensemble was obtained. As you suggested, we have clarified it in the Section 2.2. Lines 226 to 230 in the revised manuscript.

L207 – If we look at figure 4, it seems very difficult to me to see what is the optimal "localization length scale" especially since the metrics on the correlation (CORR) and the errors (RMSE, MB and ME) are close.

Response: Thanks for pointing out the problem. Initially, we intended to use the scatter plot to directly compare the difference between 'without data assimilation' and 'data assimilation'. However, as you mentioned, the difference between 'assimilation' and 'without assimilation' is obvious, but not between the assimilation experiment with different localization length-scale. After considering your suggestion, we decided to show the differences between the statistics directly in table 3. Line 835 in the revised manuscript.

L248 – This is the same problem as for my previous remark. The number of members in the ensemble seems to me chosen in a practical, if not "ad hoc", way. It should be studied whether other pollution episodes lead to the same choice of parameters "ensemble size" and "localization length scale".

Response: Thanks for your pertinent advice. As you mentioned, providing only one pollution episode for sensitivity experiments would reduce the credibility of our study results. We have added the sensitivity test for the whole month of December 2016, in which three different levels of pollution process occurred in China, as evidence for the sensitivity test. Also, we will replace Table 3 and Table 4, with the results for one month. Section 2.4 Lines 269 to 273 in the revised manuscript. Section 2.5 Line 294 in the revised manuscript

L265 – Thanks to the authors for expanding the acronyms to facilitate reading.

Response: We apologize for any unnecessary acronyms in the article that may have caused the reader trouble.

L271 – What is a "sheet-like" concentration distribution? Is it really the correct term?

Response: Thank you for this question, the word "sheet like" is the wrong expression. We have modified it accordingly. Lines 444 to 445 in the revised manuscript.

L287 – In Figure 8, there are very significant differences between the forecasts both with and without data assimilation and the observations, in particular on December 19 and December 23. Not only the amplitude, but mainly the dynamics of the concentrations are very different. Do the authors have an explanation on this point?

Response: We did not draw the figure properly, which caused your misunderstanding. After assimilating the observations, the original initial field of the model changes, but the internal dynamical processes of the model do not change in any way. Since the model is very sensitive to the initial conditions, a change in the initial conditions will inevitably cause a change in the trend of the subsequent simulation. As assimilation reduces the $PM_{2.5}$ initial field error, the model forecasts move closer to observations, leading to changes in amplitude. However, it is difficult to change the initial field alone to have significant improvement in the subsequent $PM_{2.5}$ forecasts for all 24 hours. After a certain period, the assimilation effect gradually disappears.

In order to visually describe the results of the daily assimilation of this process, we replaced the figure with the results of the whole process assimilated every 12 hours as Figure 9. As can be visualized in the figure, the assimilation does not change the dynamical background of the model.

L291 – The final part of the sentence "… and relatively consistent in (20)" seems to be lacking. Please, correct.

Response: Thanks for the correction. We removed the original part of Section 3.4.1 and rewrote it in the revised manuscript. Line 470 in the revised manuscript was deleted.

L305 – Figure 9 does not seem to me to show convincingly that the assimilation at 1200 UTC (DA12) is better than the assimilation at 0000 UTC (DA00). Have the authors looked at, from a more fundamental point of view, why this should be the case?

Response: This is a mistake in our drawing. We replotted it to new Figure 8 (Line 915 in the revised manuscript) and marked the average relative RMSE of DA00 and DA12. From the new figure, we can see that the assimilation effect of 1200 UTC is better than that of 0000 UTC. The difference in assimilation efficiency due to assimilation at different moments is mainly caused by the error between forecast and observation. In terms of the analytical equation of EnOI：

$$\psi^a = \psi^f + \mathrm{K}(d - H\psi^f)$$

$(d - H\psi^f)$ represents the error between forecast and observation, K is the gain matrix and is related to the ensemble size and the selection of the localization length-scale. if there are no bugs in the implementation, the poorer the prior compared to the assimilated observations, the more spectacular the shift towards the observations after assimilation.

L311 – Is it a general property that assimilation at 1200 UTC would be better than assimilation at 0000 UTC? The authors should be more careful about their assertion.

Response: Thanks for pointing out our problems. Our conclusion is indeed hasty and not rigorous enough. The conclusion that the assimilation at 1200 UTC is better than that at 0000 UTC is mainly derived from the average situation of this process. We calculated that the average RMSE improvement for "Tolal" and "NC" assimilation at 0000 UTC was 12.3 and 9.8 μg m^{-3}, respectively, while that at 1200 UTC was 14.4 and 14.0 μg m^{-3}, respectively, so that the average improvement at 1200 was more significant than that at 0000 UTC.

However, in this paper we only analyzed one pollution episode and did not give evidence for the conclusions we gave. As you said, it cannot be simply assumed that the assimilation effect at 1200 UTC is better than that at 0000 UTC from the results we have given. We add the relative RMSE averages in Figure 8 and qualify the conclusions. Lines 523 to 539 in the revised manuscript.

L323 – There is no linear relationship between visibility and PM2.5 concentration. Thus, it is not surprising that the result of assimilation improves the result on visibility for light pollution episodes, whereas this improvement does not exist or is insignificant for heavy pollution episodes.

Response: yes, the relationship between visibility and $PM_{2.5}$ is very complicated, and improvements in visibility are influenced by a variety of factors. Visibility is more correlated with $PM_{2.5}$ concentration when the pollution is lighter. Visibility receives a greater influence from relative humidity during severe pollutants, and the model simulates large errors in relative humidity during this time, so only assimilating $PM_{2.5}$ does not solve all the problems of visibility. This is the next step of our upcoming work for the assimilation of meteorological initial field moisture for atmospheric chemistry models.

L334 – Also in Figure 11 (as reported for Figure 8), there are large discrepancies between the PM2.5 concentration predictions with or without data assimilation and the observations. Would the authors have an explanation on not only the amplitudes, but above all the dynamics which are very different?

Response: The 12-hour time series of $PM_{2.5}$ and visibility for four stations are shown in Figure 11 (only the results for the first 12 hours are provided because the first 12 hours are more obvious). The initial field of 0000UTC was assimilated, and at 0100 UTC, the model forecast rapidly approached the observation, and subsequently the assimilation test gradually approached the observation test as the forecast time increased. However, during this period, the assimilation experiment trends are not completely consistent with the control experiment, which we believe is mainly influenced by two reasons:

1, within 12 hours, the assimilation effect does not completely disappear, so the assimilation test and the control test do not completely converge;

2, for a single site is susceptible to advective diffusion, and after assimilating the ground-based observations, the incremental $PM_{2.5}$ analysis upwind gradually diffuses with the wind field to affect the concentrations at other grid points, leading to differences in the change trends. However, this effect of diffusion does not persist consistently, and eventually the assimilation disappears, and the two experimental trends converge.