Dear referee #1:

Thank you very much for your thoughtful and insightful comments on this study. Your suggestions are very helpful to improve the quality of our manuscript. Based on your comments, we have revised the manuscript accordingly, including improving the clarity of the figures and providing the quantified results of the relevant statistical metrics. In addition, we have provided a sufficient introduction of model configuration and used datasets, which has been moved to supplementary information considering the length. Below, we provide a point-by-point response to your comments. The texts in blue are the comments, those in black are our response. The referenced line numbers correspond to the revised manuscript with changes marked.

Major comments:

1. 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?

Reply: Thank you very much. During this ozone pollution in the GBA, high ozone concentrations were observed in the land region within the black square (Fig. R1a). Therefore, we selected simulated results from land-based grids to discuss the formation of the high ozone episode (Fig. R1b). In Fig. 4 of the original manuscript, the averaged time series were calculated using all simulated ozone concentrations from land grids within the black square. The water grids were not included in this study.

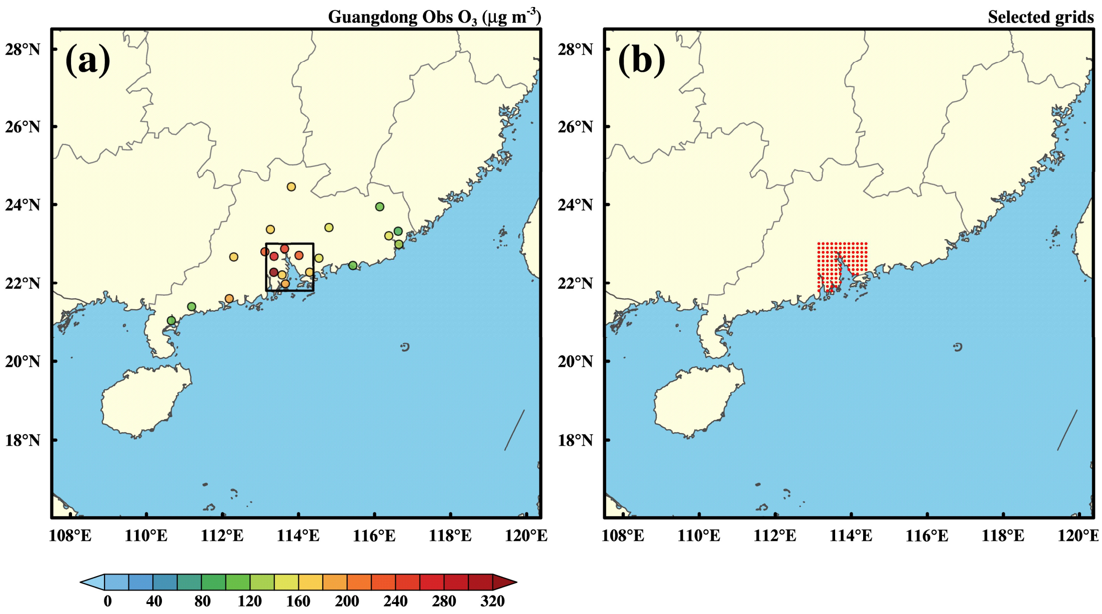

Figure R1. The observed ozone concentrations in Guangdong Province at 14:00 local time on July 31 (a); The selected grids used to discuss the formation of this high ozone episode in this study (b).

The mean time series of ozone_good and ozone_bad indeed began to diverge in the evening of July 30. Notably, ozone_good exhibited a more significant decrease during the nighttime and a stronger increase in the following morning compared to ozone_bad. To validate the simulations, we added the mean time series of the observed ozone (ozone_obs), calculated by averaging measured ozone from stations within the black square, into the Fig. R2. The ozone_obs similarly showed significant decrease in the evening of July 30 and a subsequent increase in the next morning. This pattern aligns more closely with ozone_good, indicating that the good EMs more accurately reproduced the diurnal cycle of ozone compared to the bad EMs.

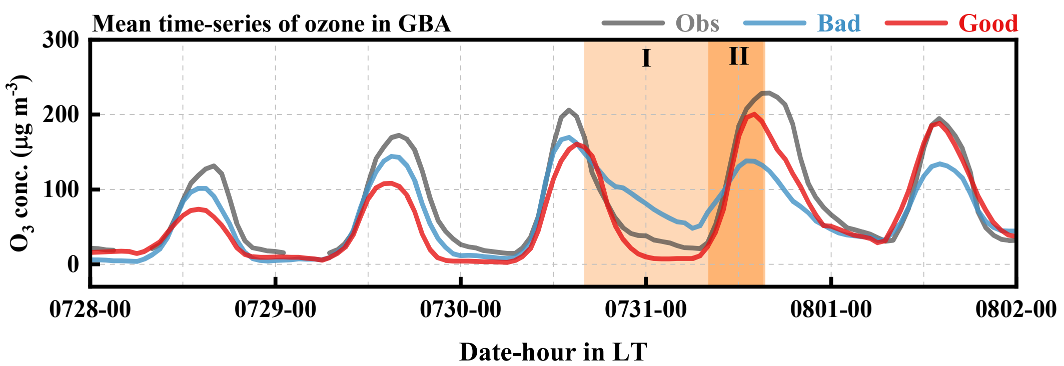

Table R2. Mean time series of ozone in GBA from July 28 to August 02 in local time (LT). The observations, bad EMs and good EMs are denoted by grey, blue, and red lines, respectively.

2. 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?

Reply: Thank you very much. In this study, we implemented the ensemble initialization method to conduct ensemble simulation (Meng and Zhang, 2008a, b). The initial and boundary condition files (wrfinput_d0* and wrfbdy_d01) from the base simulation were perturbed using the "cv3" background error covariance option in the WRF-3DVAR package. This procedure generated 30 new initial/boundary condition sets by randomly perturbing the meteorological variables: horizontal wind components, potential temperature, and water vapor mixing ratio. The perturbation magnitudes were constrained within one standard deviation of each variable (Zhu et al., 2016; Xiao et al., 2023).

These 30 perturbed condition sets were subsequently used to conduct the numerical simulation, producing 30 ensemble members (EMs). For the numerical simulation, we used identical reanalysis dataset and model configurations. In addition, nudging options were not applied. Divergences among EMs arose from the perturbations of initial/boundary conditions and the nonlinear error growth during numerical integration.

This ensemble simulation method has been utilized in data assimilation and weather analysis studies, and the findings have been widely validated and accepted by scientific community. In our study, we applied this method to improve the model performance ozone simulation during the extremely high ozone episode in GBA preceding the landfall of typhoon NIDA. Among the 30 EMs, several EMs performed better both in ozone concentrations and meteorological variables compared to the base simulation, which suggested that this method can effectively fulfill the demands of this study.

Minor comments:

1. 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.

Reply: Thank you for your comment. We completely agree with you. In a modeling study, we believe that introducing the model configuration is crucial. We also welcome other scientists to replicate the experiment using our simulation method. We have provided an introduction to the model configuration for this study, which includes the domain setting, spatial resolution, parameter selection, and input datasets. Considering the length of the introduction, we have moved it to the supplementary information. Please refer to Text S1, Table S1, and Table S2 in the supplementary information for the relevant details.

2. 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.

Reply: Thanks for this comment. The values of R, MNB and the Index of each EM are listed in Table R1. The R values range from 0.79 to 0.92, which suggested that all EMs could well reproduce the variation patterns of surface ozone in the GBA. However, the MNB values range from -46.99% to -10.92%. According to the recommended threshold values of MNB (±15%) provided by EPA (2005, 2007), only EM17 could meet the criterion, while the other EMs did not.

There is significant discrepancy in the validation result between R and MNB. To draw a common conclusion of the model performance in ozone, we introduced an index that quantitatively evaluates the performance of each EM. This index considered the effects of both R and MNB, which could evaluate the variation pattern and the high concentrations of ozone simultaneously. Based on the index, the top three EMs were EM05, EM16, and EM17, while the bottom three were EM20, EM23, and EM28.

Table R1. The R, MNB and Index of each EM on ozone in the GBA

EMs	R (×100)	MNB (%)	I_{dis}
EM01	86.23	-30.97	33.90

EM02	88.44	-27.40	29.74
EM03	90.67	-32.58	33.89
EM04	83.25	-24.80	29.93
EM05	90.79	-21.42	23.32
EM06	89.75	-35.62	37.07
EM07	86.60	-34.00	36.77
EM08	82.77	-33.74	37.88
EM09	88.52	-31.62	33.64
EM10	85.64	-29.89	33.16
EM11	88.48	-32.46	34.44
EM12	82.06	-36.74	40.89
EM13	92.30	-23.47	24.70
EM14	90.01	-35.37	36.76
EM15	85.06	-28.46	32.14
EM16	91.03	-18.35	20.43
EM17	92.08	-10.92	13.49
EM18	87.63	-30.95	33.33
EM19	83.59	-31.34	35.37
EM20	81.18	-37.64	42.08
EM21	84.54	-30.11	33.85
EM22	89.51	-29.16	30.99
EM23	85.15	-46.99	49.28
EM24	86.04	-26.11	29.60
EM25	88.55	-32.95	34.89
EM26	87.68	-29.10	31.60

EM27	87.89	-29.39	31.79
EM28	81.99	-40.32	44.16
EM29	79.68	-31.42	37.41
EM30	89.28	-28.72	30.66

Table R1 has been moved to the supplementary information, please check it there. In addition, the R and MNB values of each EM are also listed in the revised manuscript (Table 2). We have also added the relevant discussion on R and MNB in the revised manuscript. Please check the details at the lines 164-175.

Typos:

1. Line 75: The authors mention " black square" but they don't mention Figure 1. Modify to (Figure 1, black square)

Reply: Thank you very much. The figure 1 has been modified in accordance with the comment from referee #3. The black square is now in Fig. 1b. Thus, we have modified the "black square" to "black square in Fig .1b". Please check the details at line 75 in the revised manuscript.

2. Line 133: Fig.2 to substitute into Fig.2 (b)

Reply: Thank you for this comment. In the original manuscript, figure 2b did not cover all 17 geophysical source regions, whereas figure 2a did. In addition, figure 1 in the revised manuscript now presents the information about the source regions. Therefore, the "Fig. 2" has been modified to "Fig. 1a". Please check the details at line 144 in the revised manuscript.

References

Meng, Z. and Zhang, F.: Tests of an Ensemble Kalman Filter for Mesoscale and Regional-Scale Data Assimilation. Part IV: Comparison with 3DVAR in a Month-Long Experiment, Mon Weather Rev, 136, 3671-3682, 10.1175/2008mwr2270.1, 2008a.

Meng, Z. and Zhang, F.: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study, Mon Weather Rev, 136, 522-540, 10.1175/2007mwr2106.1, 2008b.

Xiao, H., Liu, X. T., Li, H., Yue, Q., Feng, L., and Qu, J.: Extent of aerosol effect on the precipitation of squall lines: A case study in South China, Atmos Res, 292, 106886, 10.1016/j.atmosres.2023.106886, 2023.

Zhu, L., Wan, Q., Shen, X., Meng, Z., Zhang, F., Weng, Y., Sippel, J., Gao, Y., Zhang, Y., and Yue, J.: Prediction and Predictability of High-Impact Western Pacific Landfalling Tropical Cyclone Vicente (2012) through Convection-Permitting Ensemble Assimilation of Doppler Radar Velocity, Mon Weather Rev, 144, 21-43, 10.1175/Mwr-D-14-00403.1, 2016.