#Reviewer 5

General comments:

This is an interesting paper, addressing an important topic, and appropriate for publication in ACP. At times the description of the dynamics is hard to follow, and so I suggest some attention be given to making the arguments simple and clear.

The paper mainly falls down, I think, in presenting the meteorology of the convective system that is alleged to be responsible for bringing the ozone down to the surface. This may simply be because the authors are meteorologists by training, and forget that most ACP readers are not. Please explain more! A prime example is Figure 11, which purports to demonstrate the downward transport: “We performed cross-section analyses of the bow-echo MCS (Fig. 11b-c), and the results clearly show a rearward pathway through which the stratospheric ozone-rich airmass was transported to the surface by the rear inflows descending from stratiform clouds to the leading convective line.” Maybe they clearly show that to the authors, but unfortunately not to this reader. Is the salient point that the “tracers” are now mostly below 3.6 km? Or that some of the wind vectors are pointing down? Is the reader supposed to be able to see “the rear inflows descending from stratiform clouds to the leading convective line”? Remember that many of your readers won't know where the "leading convective line" is found.

Thank you for your points. We have tried out best to add more related information in the manuscript and modified the expressions. The point-to-point responses are listed below.

Minor points:

The presentation and grammar need some good editing. It is sometimes difficult to understand what the authors are trying to say.

Sorry for this, we have tried out best to improve the language of this paper. Many thanks to the reviewers for pointing out the grammar errors.

Abstract: This is a bit long, and some of it reads like an introduction to the paper, rather than a brief summary of new results. At the least, the last sentences of each paragraph (lines 20-21 and 30-32) should be moved or deleted.

Thank you, we have modified the sentences and shortened the abstract.

Lines 66, 449, 450: "…wrapped around the anvil". An "anvil" is a block of iron that a blacksmith hammers upon. "Rearward anvil" and "forward anvil" are meteorologist's slang. Most readers will know that an anvil-shaped cloud is often associated with a thunderstorm, but no more. Please be clear about what you are describing, and why.

The anvils form and extend in both directions because the updrafts of thunderstorm divide into two branches at upper levels. Readers are referred to Houze (2004) for more details about the anvil clouds and MCSs.

Houze Jr., R. A.: Mesoscale convective systems, Rev. Geophys., 42, RG4003, https://doi.org/10.1029/2004RG000150, 2004.

Lines 111-116: The instrumentation should be identified, and/or the uncertainty and detection limits cited.

The ground-based air pollutant data from CNEMC have been widely used in researches concerning air quality and air pollution in China. Readers are referred to Lu et al. (2018), Li et al. (2019, 2020), Han et al. (2020), Wang and Zhang (2020) for more details about the observation network.

Lu, X., Hong, J., Zhang, L., Cooper, O. R., Schultz, M. G., Xu, X., Wang, T., Gao, M., Zhao, Y., and Zhang, Y.: Severe surface ozone pollution in China: A global perspective, Environ. Sci. Technol. Lett., 5, 487–494, https://doi.org/10.1021/acs.estlett.8b00366, 2018.

Li, K., Jacob, D. J., Liao, H., Shen, L., Zhang, Q., and Bates, K. H.: Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China, P. Natl. Acad. Sci. USA, 116, 422, https://doi.org/10.1073/pnas.1812168116, 2019.

Li, K., Jacob, D. J., Shen, L., Lu, X., De Smedt, I., and Liao, H.: Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences, Atmos. Chem. Phys., 20, 11423–11433, https://doi.org/10.5194/acp-20-11423-2020, 2020.

Han, H., Liu, J., Shu, L., Wang, T., and Yuan, H.: Local and synoptic meteorological influences on daily variability in summertime surface ozone in eastern China, Atmos. Chem. Phys., 20, 203–222, https://doi.org/10.5194/acp-20-203-2020, 2020.

Wang, X. and Zhang, R.: Effects of atmospheric circulations on the interannual variation in PM2.5 concentrations over the Beijing–Tianjin–Hebei region in 2013–2018, Atmos. Chem. Phys., 20, 7667–7682, https://doi.org/10.5194/acp-20-7667-2020, 2020.

Lines 189-190: I think this is saying that FLEXPART-WRF used the 3-km resolution output of WRF-ARW, but it isn't really clear.

Sorry for this. We have modified the sentence in the manuscript.

Lines 208-201: "It is a common practice to use 25th percentile of ozone concentration distributions as a background value (e.g., Parrington et al., 2013), which yields an even severer ozone enhancement in the surface." I think the authors are trying to suggest that the ozone amount is more significant because it is all transported from elsewhere, and so could be measured against some "background" value (arbitrarily defined). This is a dubious comparison that will only serve to confuse the reader. Delete.

Thank you for your suggestion, this sentence has been deleted.

Line 226: After "…not reduced in Qingdao and Weihai" I suggest adding "…which were outside of the path of influence of the MCS, as noted in the preceding paragraph."

Thank you. Added in the manuscript.

Line 230: I suggest referring to Figure 1 here.

Thank you for this suggestion. Added in the manuscript.

Line 281: Bohai Bay is not indicated in Figure 1.

Sorry for this. We changed it to "Bohai Sea" in the manuscript.

Lines 287-288: Is that the blue areas? In other words, does a positive vertical velocity in Pa imply downward motion? This is not clear.

Yes, the vertical velocity here uses a unit of Pa s^{-1}, and hence the positive values (in blue colors) represent the downward air motions. We added more description in the caption of Fig. 5

Line 307, Figure 6: Why use dewpoint depression? This metric will be unfamiliar to those without meteorological training (most ACP readers!).

Sorry for this. The temperature and dewpoint are frequently used in the radiosonde observations, and the differences between them (dewpoint depression) directly imply the saturation of airmass. For this reason, we compared the dewpoint depressions to track the descent of stratospheric dry air.

Line 344: The term "bow-echo MCS" is used here without definition. The description appears later, beginning on line 362. Please move it ahead of this.

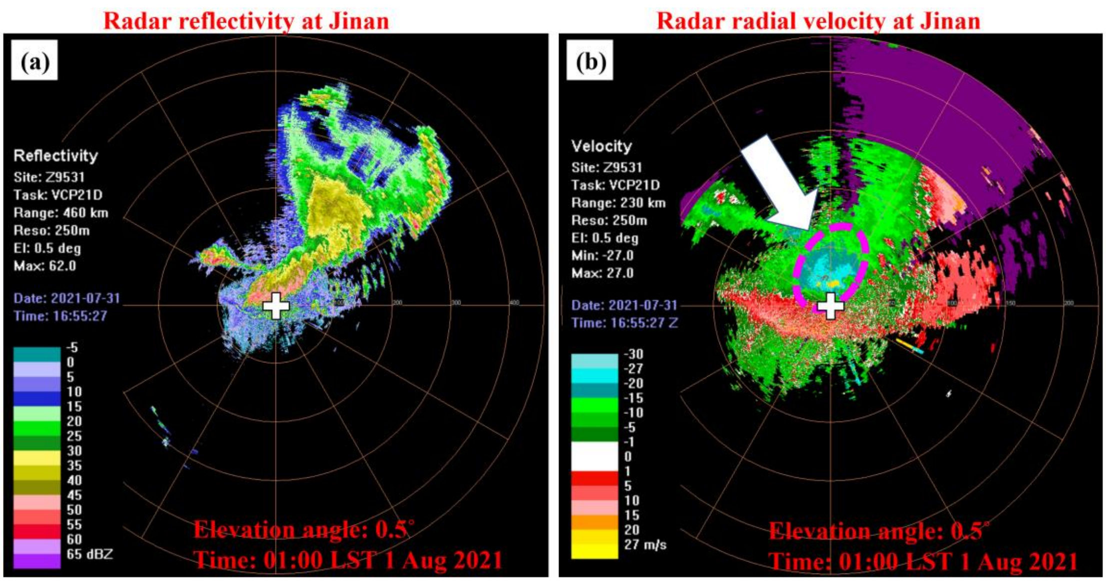
Thank you for this valuable suggestion. The description of bow-echo MCSs was moved to Line 356-360.

Line 363: "produce"? Perhaps "are associated with" would be better. The radar echoes don't cause the winds.

Thank you. Modified in the manuscript.

Line 367: Perhaps they should be shown? I find the evidence of descent unconvincing at present, and this is an important part of the paper. Also, a few lines below, you claim that "…strong radar reflectivities were confined below 6 km altitude (480 hPa, -9 ℃) suggesting limited vertical extension of convective storms." It would be helpful to see those data.

The figure below shows the radar reflectivity and radial velocity observed by Jinan Radar station (white cross symbol) at 01:00 LST on 1 August 2021 with a radar elevation angle of 0.5°. The storm had moved close to the radar station (figure a). In the radial velocity filed (figure b), the negative values represent that the airmasses move toward the radar station (white cross symbol here) and the positive values represent the airmasses move away form the radar. A region associated with large negative radial velocities exceeding 20 m/s (the magenta dashed circle) appeared in the north of radar station due to the development of descending rear inflows of bow echoes.



“the strong radar reflectivities were confined below 6 km altitude (480 hPa, -9 °C) suggesting limited vertical extension of convective storms.”. The supporting data and figure is shown in Fig. S8 in the supplementary file.

Lines 388-394: This plot and description give me no useful information with which to evaluate the model performance. What exactly is being simulated? What observations are being compared? Does a POD of 0.8 mean we have 80% perfect agreement, or 80% chance of seeing something similar within 20 km? What does the SR of 20-80% mean, and what is a frequency bias (FR)?

In contrast, I do get some information from comparing Figures 8 and 9. Perhaps instead of S9 you could simply describe the agreement between these figures. It looks to me like WRF is simulating a system of similar size and strength in pretty much the same place.

Yes, we can qualitatively evaluate the model performances by comparing Figures 8 and Figure 9. The WRF model does a good job capturing the evolution of the MCSs. Additionally, we intended to quantitatively compare the model simulation and observations in the form of performance diagram, which is frequently used in high-resolution weather models. The simulated radar reflectivity is compared with the observed one with the following performance metrics. Success ratio (SR), which equals 1-Far (False Alarm Rate), POD (probability of detection), CSI (critical success index) and FB (frequency bias). According to the contingency table (Table S1), there are hits, misses, false alarms, and correct negatives.

Table Contingency table for observations and simulations

		Observations	
		Yes	No
Simulations	Yes	Hits	False alarms
	No	Misses	Correct negatives

And hence we calculated the following metrics with a 20-km searching radius.

POD = $\frac{\text{Hits}}{\text{Hits+Misses}}$, a ratio that ranges from 0 to 1 and represents the ratio of correctly forecasted objects to the total number of observed objects.

FAR = $\frac{\text{False alarms}}{\text{False alarms +Correct negatives}}$, a ratio ranges from 0 to 1 and represent the ratio of overproduced forecasts to the total number of objects forecasted, which also ranges from 0 to 1. Thus SR (= 1-FAR) represent the ratio of correctly forecasted objects to the total number of objects forecasted.

CSI = $\frac{\text{Hits}}{\text{Hits+False alarms +Misses}}$, a ratio of correctly forecasted objects to the total number of observed and forecasted objects

FB = $\frac{\text{Hits+False alarms}}{\text{Hits+Misses}}$, a ratio of the positive forecasts (both true and false) to the number of observed objects.

Lines 442-448: This description and Figure 11 are quite confusing to me, as noted above. I'm not at all sure what the lines labelled "tracers" represent. Are they contours of particle counts? At 3.6 km?

The magenta contour lines labelled "tracers" represent the number of tracer particles at each altitude. We have modified the caption in Fig. 11.

Figures 2, 3 & 4: I find the times on the X-axis hard to read.

The times on the X-axis are expressed as month/day/hour, and we have added more description in the figure caption.