

Response to Referee #2' comments

The authors appreciate the referee's insightful comments and constructive suggestions, which helped us significantly improve the quality of this manuscript. Below are itemized, point-by-point responses to the referee's comments. For clarity purpose, here we have listed the reviewers' comments in black plain font, followed by our response in blue plain font.

General Comments:

The authors investigate the potential value of low-altitude meteorological drone (MD) observations for convective-scale short-term forecasting. In this manuscript, an Observing System Simulation Experiment (OSSE) is conducted for a rainfall event over the Beijing–Tianjin–Hebei region, in which synthetic MD observations of temperature, specific humidity, and horizontal wind are generated from a free-running “truth” simulation. A comprehensive analysis is carried out through five sets of sensitivity experiments, examining the impacts of horizontal resolution, observation height, spatial distribution, joint assimilation of thermodynamic and wind variables, and observation errors. The results indicate that assimilating MD temperature and humidity observations improves both thermodynamic analyses and precipitation forecasts. Furthermore, denser MD networks and observations extending to higher altitudes provide additional forecast benefits, and the inclusion of wind observations yields further improvements.

This research is valuable, and the manuscript is generally well written. However, several issues require clarification and further consideration:

1. Vertical levels of MD observations: In the experiments, only three observation levels (120 m, 300 m, and 600 m) are considered. However, Lines 282–283 indicate that MD observations are typically conducted within the lowest 1–3 km of the atmosphere, while Lines 279–295 suggest that future MD development will focus on low-altitude airspace below 1 km. Does this imply that future MD deployments will be limited to the three selected levels? If not, what is the rationale for choosing only these levels? It would be helpful to justify this selection or consider including additional vertical levels to assess their impact..

Response: Thank you for this important comment. Meteorological drones are capable of providing vertically resolved observations through a substantially deeper portion of the lower atmosphere, commonly within the lowest 1–3 km depending on the platform and mission. The present study, however, focuses specifically on representative sampling levels within the low-altitude atmosphere below 1 km, which is of particular relevance to emerging low-altitude aviation and meteorological applications in China. The 120-, 300-, and 600-m levels are therefore selected as representative lower-, middle-, and upper-level sampling heights within this low-altitude range. In particular, the 120- and 300-m levels

correspond to operationally relevant low-altitude layers, while 600 m is included as an additional upper-level sampling height to extend the vertical sensitivity assessment over a broader portion of the PBL. These three levels are not intended to represent fixed or exclusive operating levels for future MD deployments, nor do they represent the full vertical profiling capability of MD platforms. Rather, they provide a controlled experimental framework for examining how the impact of MD data assimilation depends on observation height. We have revised Sect. 3.2 and the description of the observation-height sensitivity experiments to clarify this rationale and to distinguish the representative levels used in the present OSSE from the broader vertical sampling capability of real MD platforms. We also note in the Summary and conclusions that vertically continuous profiles and a broader range of observation heights should be investigated in future work. The corresponding revisions are as follows:

1) Lines 290-292: “These representative sampling heights provide a controlled experimental framework for quantifying the sensitivity of DA impact to observation height within the low-altitude atmosphere”.

2) Lines 295-302: “These levels are selected as representative lower-, middle-, and upper-level sampling heights within the low-altitude atmosphere to examine the sensitivity of the DA impact to vertical placement. Notably, 120, 300, and 600 m correspond to typical allowable flight altitudes for unmanned aircraft in China’s uncontrolled airspace. Previous studies have examined the assimilation of vertically profiling UAS observations (Jensen et al., 2021; Murdzek et al., 2026; Pinto et al., 2024), whereas the present study specifically considers observations at representative low-altitude operating levels”.

3) Lines 940-942: “Future work should therefore extend the present framework to multiple cases, assess the added value of vertically continuous MD profiles, ...”.

2. Observation density across levels: In the experiments examining different observation heights, are the numbers of observations at each level kept consistent? If not, variations in observation count may influence the results and should be discussed or controlled for.

Response: Thank you for raising this important point. In the three single-level observation-height experiments (MD_H5_Z120, MD_H5_Z300, and MD_H5_Z600), the same 5-km horizontal MD network is used, with identical horizontal observation locations, observation variables, and temporal sampling intervals. Therefore, these three experiments contain the same number of observations at each assimilation time, and observation height is the only difference among them. The differences among these single-level experiments can therefore be interpreted primarily as sensitivity to the vertical placement of the observations. By contrast, the baseline MD_H5 experiment assimilates observations at all three levels and therefore contains a larger total amount of observational information than the individual single-level experiments. Consequently, its advantage reflects the combined effects of vertically distributed sampling and increased observational information, rather than the effect of vertical distribution alone. The corresponding

revision is as follows:

1) Lines 367-370: “The three single-level experiments use identical horizontal sampling locations, observation variables, and temporal sampling intervals, and therefore contain the same number of observations at each assimilation time”.

3. Representativeness of results: The study focuses on a single heavy rainfall event. Including additional cases would strengthen the generality of the conclusions. If expanding the dataset is not feasible, the authors should be cautious in presenting broad conclusions. For example, how representative are the findings (e.g., Lines 808–811)? To what extent can they be generalized to other convective events with different synoptic conditions?

Response: This study is based on a single heavy-rainfall event, and its findings should therefore be interpreted within this specific setting. The principal objective of the present study is to examine, within a controlled OSSE framework, the relative sensitivity of convective-scale analyses and forecasts to different MD observing configurations, rather than to establish universally applicable or optimal network specifications. To address this concern, we have revised the Abstract, Introduction, Results and discussion, and Summary and conclusions to make the case-specific nature of the findings more explicit. Statements that could be interpreted as implying broader robustness or universal optimality have been moderated using wording such as “for the present event,” “for this case,” and “among the configurations examined here.” We also explicitly state that additional cases covering different synoptic and convective environments and geographical regions are required to assess the generality of the sensitivities identified in this study. In addition, we have added a spatial moving-block bootstrap to estimate the sampling uncertainty of the categorical verification metrics while accounting for spatial dependence in the verification fields. A 50-km block size and 2000 resamples are used, and the resulting 95% confidence intervals are added to Figs. 9, 11, and 14. The corresponding discussions have been revised to avoid overinterpreting relatively small differences among experiments when the confidence intervals overlap. The corresponding revisions are as follows:

1) Lines 41-42: “For the present event, the 5- and 10-km networks generally provide the largest forecast improvements among the tested configurations”.

2) Lines 409-414: “To quantify the sampling uncertainty of these categorical verification metrics while accounting for spatial dependence in the verification fields, a spatial moving-block bootstrap with a 50-km block size and 2000 resamples is applied (Hamill, 1999; Wilks, 1997). The resulting 95% confidence intervals quantify the uncertainty associated with spatial sampling within the present event and do not represent case-to-case sampling variability”.

3) Lines 600-604: “Among the tested configurations, the 5- and 10-km MD observation networks generally yield higher forecast skill than the coarser configurations across the evaluated lead times and precipitation thresholds for this case, suggesting that relatively dense MD observations are more effective for

improving convective-scale short-term precipitation forecasts”.

4) Lines 872-880: “For this case, the 5- and 10-km networks generally provide greater forecast benefits than the coarser configurations, whereas the 40-km network tends to yield smaller improvements, suggesting that its spacing is too coarse to adequately represent mesoscale and smaller-scale thermodynamic structures relevant to convective precipitation evolution. However, the 5-km spacing represents the densest configuration examined here rather than an optimal network spacing, and further network densification may yield diminishing benefits as the observation spacing approaches the characteristic spatial scales of the model and DA system”.

5) Lines 912-914: “The sensitivities identified in this OSSE provide useful guidance for the design and evaluation of future MD observing networks, while their generality requires further assessment across additional convective regimes”.

6) Lines 937-942: “Fourth, the conclusions are based on a heavy rainfall event over the Beijing-Tianjin-Hebei region. Additional cases covering different synoptic and convective environments and regions are needed to assess the robustness and generality of these findings. Future work should therefore extend the present framework to multiple cases, ...”.

4. Line 485 states that specific humidity in the analysis and forecast is examined. In Figure 6 and related figures, it is labeled as Q_v . Could the authors clarify whether the variable shown is specific humidity or water vapor mixing ratio?

Response: Yes, the variable denoted by Q_v represents the water vapor mixing ratio, rather than specific humidity. We have corrected the terminology throughout the manuscript.

5. In Lines 337–340, it is important to explicitly state that the data assimilation (DA) experiments are designed to evaluate the upper bound of the information content provided by MD observations, given that no other observational data are assimilated. This context is critical for interpreting the results. Additionally, it would be valuable to discuss how the impact of MD observations might change when assimilated alongside conventional and remote-sensing observation networks.

Response: Thank you for this important comment. We have expanded the discussion in the Summary and conclusions to clarify that, in an operational DA system already assimilating conventional and remote-sensing observations, part of the information provided by MD observations may overlap with existing observing systems, and their incremental benefit could therefore be smaller than that obtained in the present experiments. At the same time, high-frequency MD observations may provide complementary information within the data-sparse PBL. Evaluating the added value of MD observations when assimilated together with conventional, radar, satellite, and other dense observations is therefore identified as an important direction for future work. The corresponding revision is as follows:

1) Lines 916-928: “First, only synthetic MD observations are assimilated, while conventional observations (e. g., radiosonde and surface data) as well as radar and satellite observations are excluded. This experimental design facilitates a clearer attribution of forecast impacts to MD observations, while the forecast improvements reported here should therefore be interpreted as an estimate of the upper-bound impact of MD observations. In operational applications, part of the information provided by MD observations may overlap with existing observing systems, so their incremental benefit could be smaller than that obtained in the present experiments. Nevertheless, high-frequency MD observations may still provide complementary information within the data-sparse PBL. Future work should therefore evaluate the added value of MD observations when assimilated together with conventional, radar, satellite, and other dense observing networks”.

6. In figure 6 and others, is the statistics for Q_v ? As stated in Line 485, specific humidity of analysis and forecast is studied. However, in figure 6 and others, it is labeled Q_v . Just want to clarify is it specific humidity or water vapor mixing ratio.

Response: Yes, Q_v in Fig. 6 and the related figures denotes the water vapor mixing ratio. We have corrected the terminology throughout the manuscript so that Q_v is consistently referred to as water vapor mixing ratio rather than specific humidity.

7. Line 337-340, it is very important to point out that the DA experiments are designed to evaluate the upper bound of the information content provided by MD observations. Since no other observations are included in the experiments. It will be very interesting to see its impact on top of conventional and remote-sensing observation network.

Response: Thank you for pointing this out. As described in our response to Comment 5, we have revised the manuscript to explicitly clarify that, because no conventional, radar, or satellite observations are assimilated, the present experiments are intended to estimate the upper bound of the information content provided by MD observations, rather than their incremental value in an operational multisource DA system. We have also added discussion in the Summary and conclusions noting that the incremental benefit of MD observations may be smaller when other observing systems are already assimilated because of information overlap and redundancy. Evaluating the added value of MD observations on top of conventional and remote-sensing observation networks is now identified as an important direction for future work.