Response to Referee #1' 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:

This study uses simulated meteorological-UAV observations to investigate the impacts of assimilating meteorological UAV data on short-range numerical weather prediction. The manuscript systematically analyzes how assimilation performance is affected by different observation data types, observation locations, and prescribed observation errors. It further highlights the value of assimilating observations within the planetary boundary layer. This work is of great significance for improving numerical forecasts of high-impact weather, and also provides valuable references for designing meteorological UAV observation strategies. The research results demonstrate good scientific and application values. Nevertheless, it is confusing that the UAV observations are only performed at three fixed altitude levels. The authors are suggested to clarify in the data description whether this configuration arises from technical limitations of UAV platforms or is an empirical setup.

Response: Thank you for the positive assessment of our work and for raising this important point. The use of synthetic observations at 120, 300, and 600 m is not imposed by a technical limitation of meteorological UAV platforms, but is a deliberate experimental design related to the specific objective and application context of this study. Actually, meteorological UAV platforms can provide vertically resolved profiles with much finer vertical sampling than the three discrete levels considered here.

First, this study is motivated by the emerging development of low-altitude meteorological observing systems in China, with increasing interest in observations within the lowest several hundred meters above ground for low-altitude aviation and meteorological applications. One primary consideration for selecting these three heights (120, 300, and 600 m) is that they represent typical allowable flight altitudes for unmanned aircraft within China's uncontrolled airspace. The three heights adopted here were therefore selected as representative low-altitude operating levels rather than as an approximation of the full vertical sampling capability of a profiling UAV platform. Second, the purpose of the height-sensitivity experiments is to isolate the relative value of observations obtained at different representative low-altitude levels for convective-scale data assimilation and precipitation forecasting. In this respect, our experimental design differs from previous UAV studies that have focused on assimilating vertically profiling observations (e.g., Jensen et al., 2021; Murdzek et al., 2026; Pinto et al., 2024).

We have revised the manuscript to clarify this motivation and to distinguish the objective of the present height-sensitivity experiments from studies of vertically

continuous UAV profiling. We also explicitly identify the simplified vertical sampling as a limitation of the present study and note that future work should examine vertically continuous profiles with finer vertical spacing to quantify their additional value. The corresponding revisions are as follows:

1) 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".

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

Specific comments:

1. Line 197-198: "MD assimilation module is developed and incorporated into the DA system." Why is a separate assimilation module constructed for MD observations, rather than directly adopting the existing assimilation module for conventional observations? The authors are suggested to provide relevant explanations in the manuscript.

   Response: Thank you for this important question. The dedicated MD component serves primarily as an observation-processing interface that enables NSSL3DVAR to ingest MD observations and the associated information required by the DA system. Once processed, the MD observations are assimilated within the same 3DVAR framework as the other supported observations, using the same cost function, control variables, background-error covariance formulation, and minimization procedure. We have revised the corresponding description in Sect. 2 to make this distinction explicit. The corresponding revision is as follows:
   1) Lines 183-186: "To enable the assimilation of MD observations, a dedicated MD observation processing interface is implemented within NSSL3DVAR to handle the MD-specific observation input and associated information required by the DA system".

2. Line 216-217: "The chosen correlation scales are proportional to the observation spacing, reflecting the characteristic scales resolved by the MD network." The authors are suggested to discuss the relationship between the relevant scales and the MD network. It should be clarified whether the 5-km configuration represents the optimal setting for the MD network, so that readers can better understand the purpose of this setup.

Response: Thanks for this helpful comment. The correlation length scales discussed here are parameters of the recursive filter used to represent spatial correlations in the static background-error covariance. They control the spatial extent over which observational information is distributed through the analysis increments, rather than representing the effective spatial resolution of the MD observations themselves. In the present experiments, the horizontal correlation length scales are adjusted in proportion to the MD network spacing so that the spatial spreading of observational information remains broadly consistent with the sampling density of each network configuration. We have revised the manuscript to clarify this relationship and to avoid implying that the correlation scales represent spatial scales directly resolved by the MD network. We also clarify that the 5-km configuration is the densest network examined in this study. It is adopted as the baseline for the subsequent sensitivity experiments because it generally provides the largest analysis and forecast improvements among the tested horizontal-spacing configurations. However, the 5-km spacing should not be interpreted as an optimal network configuration. The present experiments are designed to examine the sensitivity to network spacing rather than to determine an optimal spacing. Further experiments with finer sampling and a broader range of configurations would be required for such an assessment. The corresponding revisions are as follows:

1) Lines 205-209: "The horizontal correlation length scales are adjusted in proportion to the MD network spacing so that the spatial spreading of observational information remains broadly consistent with the sampling density of each network configuration. These correlation scales characterize the spatial spreading of analysis increments through the static background error covariance".

2) Lines 361-364: "Among these experiments, MD_H5, representing the densest network examined in this study, is selected as the baseline reference for subsequent sensitivity experiments".

3) Lines 877-880: "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".

3. Line 262-267:"These deficiencies indicate considerable … highly sensitive to thermodynamic and kinematic conditions in the PBL". This paragraph is difficult to follow. Why does the inconsistency in the spatial distribution of precipitation indicate that precipitation is related to the thermodynamic and dynamic characteristics of the planetary boundary layer? The authors are suggested to elaborate on this point.

Response: The precipitation discrepancies between the NoDA forecast and the truth simulation primarily demonstrate that the background forecast contains appreciable errors and therefore provides a meaningful baseline for evaluating the impact of additional observations. The relevance of PBL observations is instead motivated by the well-established sensitivity of convective initiation and evolution to low-level thermodynamic and kinematic conditions (Crook, 1996; Weckwerth and Parsons,

2006). We have therefore revised this paragraph to separate these two arguments and to clarify why low-altitude MD observations are relevant to the present convective event. The corresponding revision is as follows:

1) Lines 254-264: "These deficiencies indicate considerable precipitation forecast errors in the NoDA experiment and therefore provide a meaningful baseline for evaluating the impact of assimilating additional observations. Low-altitude observations are particularly relevant to convective forecasting because convective initiation and evolution are strongly influenced by thermodynamic and kinematic conditions within the PBL. Low-level temperature and moisture regulate atmospheric buoyancy and convective instability, while low-level winds affect moisture transport, convergence, and convective organization (Crook, 1996; Weckwerth and Parsons, 2006). The present event therefore provides a suitable testbed for assessing the potential contribution of MD observations to convective-scale analyses and short-term precipitation forecasts".

4. Line 306: "Gaussian random errors with a mean of zero are added" It is suggested to clarify which scenario the Gaussian-distributed errors correspond to. For instance, some instruments tend to exhibit larger errors in high-temperature regions and smaller errors in low-temperature regions. Could such characteristics violate the assumption of Gaussian distribution?

Response: Yes, for some real instruments, the magnitude of observational error may vary with the atmospheric state or observing environment. This does not necessarily invalidate the Gaussian assumption itself, because the conditional error distribution under a given environmental state may still be approximately Gaussian; rather, it indicates that the error variance may not be constant across all conditions. In OSSE studies, synthetic observations are commonly generated by adding zero-mean Gaussian random perturbations to the model truth, with the prescribed standard deviations representing the expected magnitude of observational uncertainty. This OSSE study adopts the commonly used state-independent Gaussian-error formulation in order to provide a controlled framework for assessing the relative impacts of different MD observing configurations. In addition, the observation-error sensitivity experiments explicitly examine how changes in the prescribed error magnitude affect the analyses and forecasts. We have revised the manuscript to clarify the interpretation and scope of the Gaussian-error assumption as follows:

1) Lines 315-317: "In this formulation, the error variance is assumed to be independent of the atmospheric state; possible state-dependent variations in real instrument errors are not explicitly considered".

2) Lines 931-936: "Third, the present experiments adopt state-independent Gaussian random observation errors with spatially uniform standard deviations. In real applications, observation error characteristics may also vary with atmospheric conditions, terrain, and observing environments. Future work could further

examine state-, environment-, and region-dependent error specifications when sufficiently detailed error characteristics of operational MD observations become available".

5. Line 416-418: "especially near the Yanshan and Taihang Mountains, while weak cold biases are found over northeastern Hebei, northern Shandong, and the adjacent coastal waters" The authors are suggested to provide appropriate physical interpretations. The error characteristics show similar patterns to the terrain height variation. Is such error mainly caused by the model's inability to accurately resolve mountain locations, which consequently yields poor simulations of weather variations induced by high-elevation terrain?

Response: In this OSSE study, the truth and forecast experiments use the same WRF domain, horizontal grid, and static terrain representation. Therefore, the larger temperature errors near the Yanshan and Taihang Mountains are unlikely to be primarily caused by differences in the geographical representation or location of the mountains between the truth and forecast simulations. A more likely explanation is that the lower-tropospheric thermal structure over complex terrain is particularly sensitive to differences in the initial and lateral boundary conditions. Terrain-induced circulations, strong elevation-dependent thermal gradients, and boundary-layer processes can amplify these differences, resulting in larger temperature departures over mountainous regions. We have added this interpretation to the revised manuscript. At the same time, we avoid attributing the error to a single physical mechanism because a detailed diagnosis of the terrain-related error sources is beyond the scope of the present study. The corresponding revision is as follows:

1) Lines 432-440: "This indicates that the background run does not adequately represent the mesoscale thermal structure, particularly over complex terrain. As the truth and forecast simulations use the same model grid and static terrain representation, these terrain-related error patterns are unlikely to arise from differences in the geographical representation of the mountains themselves. Instead, the complex terrain may amplify differences in the lower-tropospheric thermal structure arising from the different initial and lateral boundary conditions through terrain-induced circulations, elevation-dependent thermal gradients, and boundary-layer processes".

6. Line 427-429: "which suggests that the densest MD network most effectively constrains the lower-tropospheric thermal analysis." I acknowledge that it is impractical to conduct an unlimited number of experiments with diverse resolutions. Nevertheless, the authors are suggested to appropriately discuss the matching relationship between observation resolution and model resolution. Higher observation resolution does not necessarily yield better performance. In addition, it is worth exploring whether there exists an upper limit for the improvement achievable via resolution enhancement.

Response: Yes, a denser observation network does not necessarily lead to continuously increasing analysis or forecast benefit. In this study, the forecast model has a horizontal grid spacing of 1 km, whereas the MD networks examined have horizontal sampling spacings of 40, 20, 10, and 5 km. The 5-km configuration therefore represents the densest network tested in this study rather than an optimal network spacing. The relationship between observation spacing and model resolution is not determined by the model grid spacing alone. Although a 1-km model grid can represent smaller-scale structures than the MD networks considered here, the effective information contributed by increasingly dense observations also depends on the model's effective resolution, the spatial correlation scales used in the background-error covariance, and the representativeness and spatial redundancy of neighboring observations. As the observation spacing becomes sufficiently small relative to these characteristic scales, neighboring observations are expected to provide increasingly redundant information, and the incremental benefit of further increasing the network density may therefore diminish. The present experiments do not include MD spacings finer than 5 km and thus cannot determine the spacing at which such saturation occurs. We have revised the manuscript to clarify that the 5-km configuration is the densest network examined rather than an optimal configuration, and have added a discussion of the relationship between observation spacing, model resolution, and the potential diminishing returns associated with further network densification. The corresponding revisions are as follows:

1) Lines 604-612: "However, this does not imply that continued network densification below 5 km would necessarily yield further improvement. Although the model grid spacing is 1 km, the benefit of increasing observation density also depends on the model's effective resolution, background-error correlation scales, and observational representativeness. As the observation spacing becomes sufficiently small relative to these characteristic scales, neighboring observations may provide increasingly redundant information, leading to diminishing incremental benefits. Since spacings finer than 5 km are not examined here, the corresponding saturation scale cannot be determined from the present experiments".

2) Lines 580-584: "Collectively, these score metrics support the visual impression from Fig. 7 and 8 that, among the sampling spacings examined here, the 5- and 10-km MD networks more effectively predict the precipitation evolution and yield more accurate quantitative precipitation forecasts than the coarser configurations".

3) Lines 877-880: "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".

7. Line 610-611: "This suggests that restricting the MD network to the plain area degrades the temperature analysis relative to the full-network configuration," The term "degrade" is not recommended, as it may mislead readers into thinking that adding observations over plain regions would degrade model performance. As the

authors mentioned, temperature errors differ between plain and plateau areas. However, identical observation errors are prescribed across all regions in this study. It is suggested that region-dependent observation errors be considered, which can be listed as one of the directions for future exploration.

Response: The term “degrades” has been replaced with more explicit comparative wording in the revised manuscript. In the present controlled OSSE, identical observation-error standard deviations are prescribed across the domain so that the effect of network deployment can be isolated without simultaneously introducing spatial variations in observation quality. This assumption facilitates a direct comparison between the plain-only and full-network configurations. The region-dependent observation error specifications have been added as one of the directions for future exploration. The corresponding revisions are as follows:

1) Lines 647-650: “This suggests that restricting the MD network to the plain area results in larger temperature analysis errors than the full-network configuration, likely because observations are excluded from the mountainous region where the background temperature errors are relatively large (Fig. 4b)”.

2) Lines 667-669: “Similarly, the plain-only configuration yields larger humidity RMSEs than the full-network experiment across much of the lower troposphere”.

3) Lines 931-936: “Third, the present experiments adopt state-independent Gaussian random observation errors with spatially uniform standard deviations. In real applications, observation error characteristics may also vary with atmospheric conditions, terrain, and observing environments. Future work could further examine state-, environment-, and region-dependent error specifications when sufficiently detailed error characteristics of operational MD observations become available”.