

Response to Community' 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.

Review of Impact of Low-altitude Meteorological Drone Data Assimilation on Convective-Scale Short-Term Rainfall Forecasts: An Observing System Simulation Study

General Comments:

This manuscript presents an OSSE to evaluate the potential impact of assimilating low-altitude meteorological drone (MD) observations on convective-scale analyses and short-term precipitation forecasts over the Beijing-Tianjin-Hebei region. The authors generate synthetic MD observations of temperature, specific humidity, and horizontal wind from a truth run and assimilate them into the WRF model using the NSSL 3DVAR system. Sensitivity experiments are performed to assess the influences of horizontal resolution, observation height, spatial distribution, combined thermodynamic-wind assimilation, and prescribed observation errors. The results indicate that denser and vertically distributed MD networks improve both thermodynamic analyses and QPFs, and that wind observations provide added value. The study addresses a timely topic relevant to emerging low-altitude meteorological observation networks and provides useful guidance for future network design.

The manuscript is generally well written and organized, the experimental design is logical, and the figures are mostly clear. However, several important methodological limitations and scientific questions need to be addressed before the paper can be considered for publication. My detailed comments are provided below.

Major Comments:

1. The entire analysis is based on one heavy rainfall event. While the authors note that additional cases are needed, this limitation is critical for drawing general conclusions about MD network design. I recommend that the authors either (a) extend the study to include multiple convective events with different synoptic pattern or (b) at minimum, include a discussion on how the representativeness of this case may limit the transferability of the results. Without this, the claim that “the 5- and 10-km networks provide the most robust and consistent forecast benefits” is not statistically supported.

Response: Thank you for this important comment. The principal objective of this 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 optimal network specifications. To address this concern, we have revised the Abstract, Introduction, Results, and Summary and Conclusions to clarify that the quantitative network-design sensitivities obtained here are specific to the present event. Statements implying broader robustness or universal optimality have been moderated, and the need for evaluation across additional convective regimes is now explicitly discussed. 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 600-602: “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, ...”.

3) Lines 872-876: “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”.

2. Real meteorological drones typically provide continuous vertical profiles with high vertical resolution, yet the synthetic observations here are restricted to only three discrete levels (120, 300, and 600 m AGL). The authors acknowledge that drone profiling can have vertical resolutions of tens of meters, but the experiments do not exploit this capability. This choice may lead to an underestimation of the potential benefit of MD data, especially for constraining the vertical thermodynamic structure within the boundary layer. The authors should either discuss why only three levels were chosen or include an additional sensitivity experiment that assimilates a more complete vertical profile (e.g., 50-m spacing up to 1 km) to quantify the added value of higher vertical resolution.

Response: Thank you for pointing this out. Yes, real meteorological drones can provide observations with much finer vertical sampling than the three discrete levels considered in this study. The use of 120-, 300-, and 600-m observations was, however, a deliberate choice related to the specific objective and application context of this work. First, this study is motivated by the emerging development of low-altitude meteorological observing systems in China. Current low-altitude aviation planning and meteorological service development place particular emphasis on the atmosphere within several hundred meters above ground, with the 120- and 300-m levels representing operationally relevant low-altitude layers. The 600-m level is included as an additional upper-level sampling height to extend the vertical sensitivity assessment within the PBL. 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 MD 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 sense, our experimental design differs from previous UAS 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 UAS 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 much finer vertical spacing to quantify this additional benefit. 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, ...”.
3. The quantitative verification metrics (ETS, POD, FAR, etc.) show differences among experiments, but no measure of statistical significance is provided. Given the single-case nature of the study, it is unclear whether the differences are robust or within noise. The authors should consider using bootstrap resampling or another method to estimate confidence intervals for the skill scores, or at least discuss the uncertainty associated with the verification.

Response: Thank you for this valuable suggestion. In the revised manuscript, we have introduced a spatial moving-block bootstrap to estimate the sampling uncertainty of the categorical precipitation verification metrics while accounting for spatial dependence in the verification fields. A 50-km block size and 2000 bootstrap resamples are used. The same spatial blocks are sampled for all experiments in each bootstrap realization to ensure consistent spatial sampling across the experiments. A brief description of this bootstrap procedure and the interpretation of the resulting confidence intervals has also been added to the opening paragraph of Section 4 (Results and discussion). The resulting 95% confidence intervals have now been added to the ETS time series in Figs. 9 and 11 and to the ETS, BIAS, POD, and FAR scores in Fig. 14. The corresponding figure captions and discussions have also been revised to explicitly describe these confidence intervals and to avoid overinterpreting relatively small differences among experiments when the intervals overlap. We further clarify in the revised manuscript that these confidence intervals

quantify spatial sampling uncertainty within the present event and do not represent case-to-case sampling variability. The corresponding revisions are as follows:

1) 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”.

2) Figures 9, 11, and 14: The confidence intervals are now presented in the corresponding verification figures, with additional explanations in the figure captions and Results and discussion.

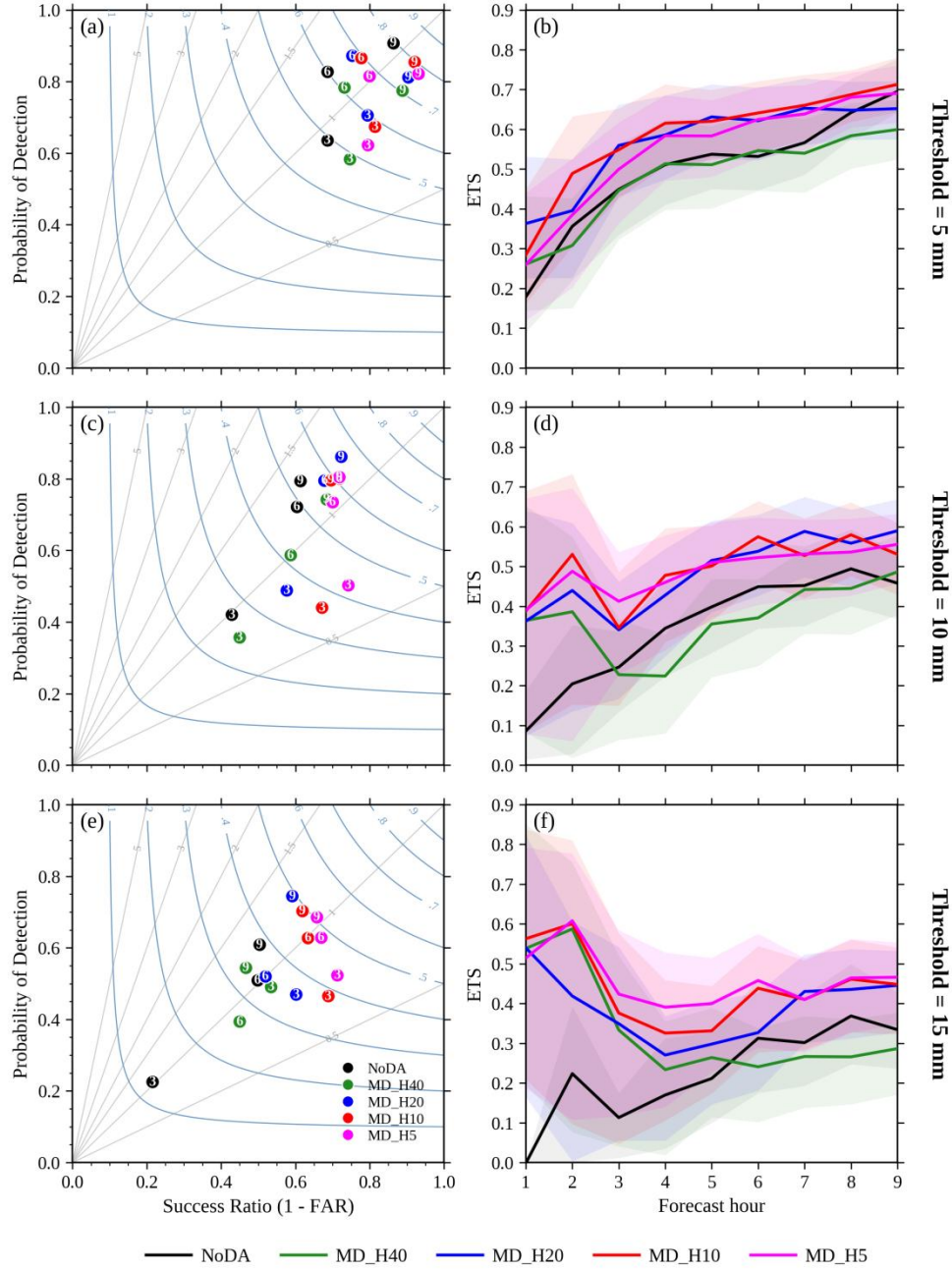


Figure 9. Impact of horizontal spacing on score metrics of precipitation forecast.

Performance diagrams (left) and equitable threat score (ETS; right) for the 1-9 h hourly precipitation forecasts from NoDA (black), MD_H40 (green), MD_H20 (blue), MD_H10 (red), and MD_H5 (magenta) at thresholds of (a, b) 5 mm, (c, d) 10 mm, and (e, f) 15 mm. Numbers in the performance diagrams denote the 3-, 6-, and 9-h forecast lead times. Shading in the ETS panels indicates the 95% confidence intervals derived from 2000 spatial block bootstrap resamples. All scores are calculated using a 12-km neighborhood

radius.

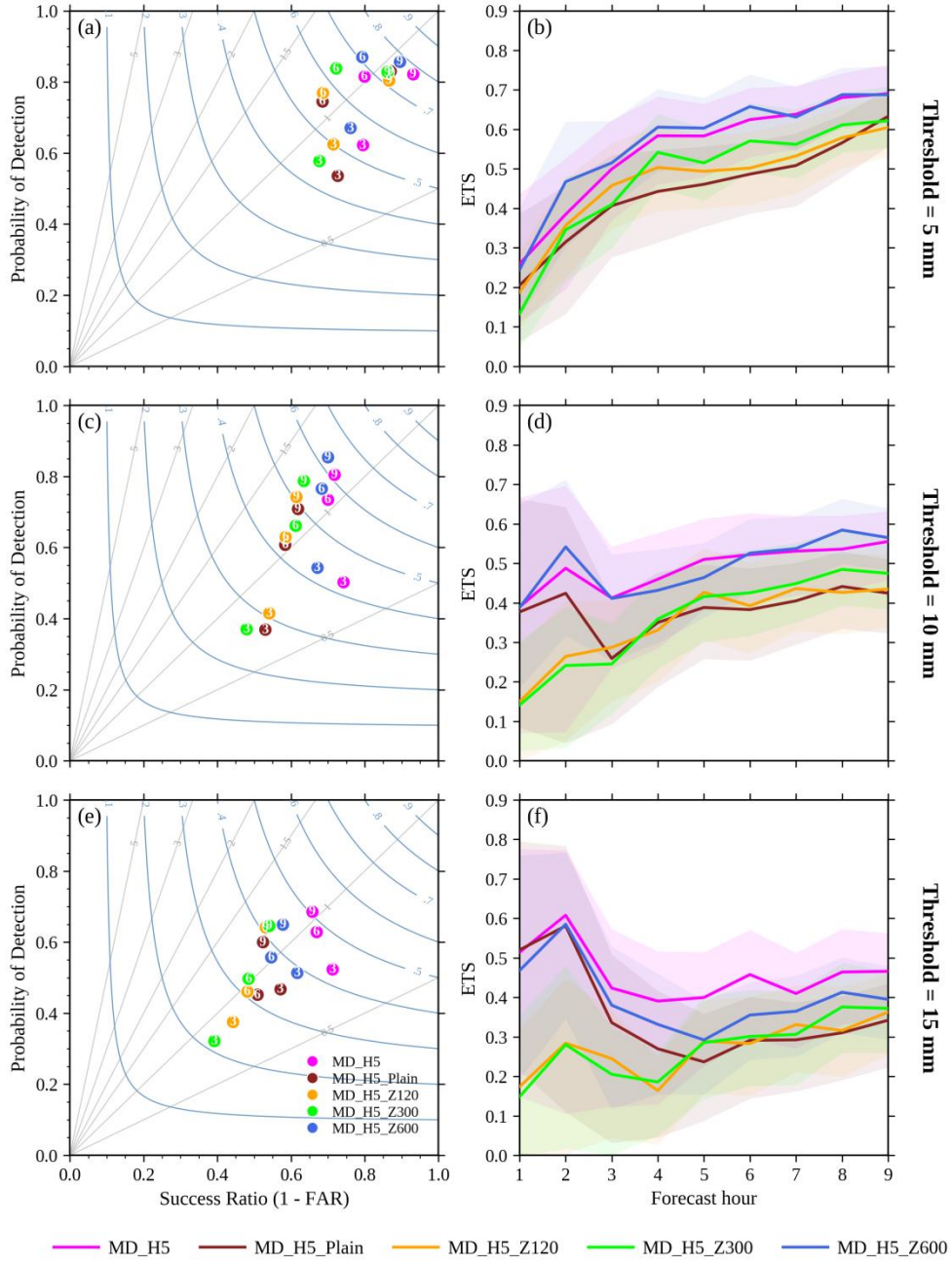


Figure 11. Impact of observation height and spatial distribution on precipitation forecasts.

Same as Fig. 9, but for the MD_H5 (magenta), MD_H5_Plain (brown), MD_H5_Z120 (orange), MD_H5_Z300 (green), and MD_H5_Z600 (blue) experiments.

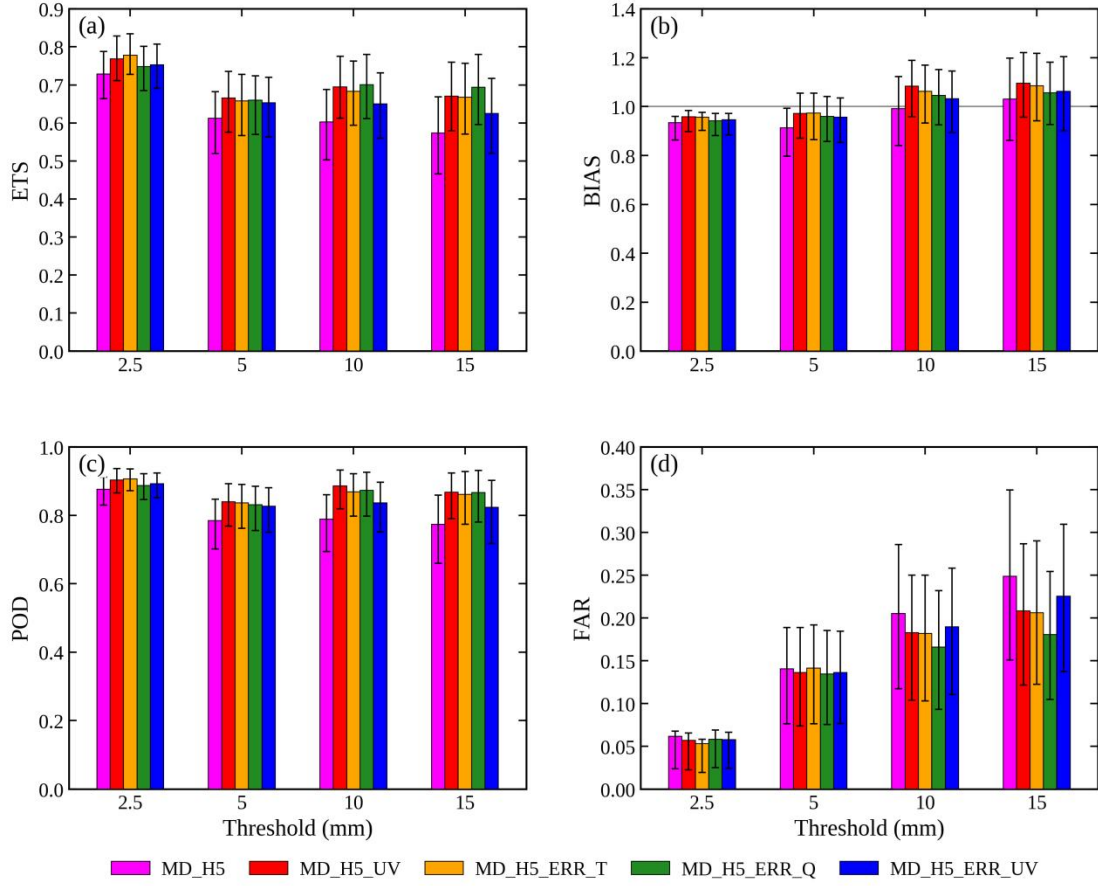


Figure 14. Impact of joint thermodynamic–wind assimilation and observation error on precipitation forecasts.

(a) The equitable threat score (ETS) , (b) bias (BIAS), (c) probability of detection (POD), and (d) false alarm ratio (FAR) of 6-hourly accumulated precipitation forecasts from the MD_H5 (magenta), MD_H5_UV (red), MD_H5_ERR_T (orange), MD_H5_ERR_Q (green), and MD_H5_ERR_UV (blue) experiments at 2.5-, 5-, 10-, and 15-mm thresholds, respectively. Error bars indicate the 95% confidence intervals derived from 2000 spatial block bootstrap resamples. All scores are calculated using a 12-km neighborhood radius.

Minor Comments:

1. The introduction is somewhat too long and could be tightened, especially the description of Chinese MD development programs, which may be of less interest to international readers.

Response: We have substantially shortened this part of the Introduction by retaining only the information directly relevant to the motivation of the present study, including recent meteorological drone applications, the emerging development of low-altitude observing systems, and the remaining research gap in MD network design and data assimilation. Detailed descriptions of individual programs and application scenarios have been removed. The corresponding revisions are as follows:

1) Lines 95-105: “In China, MD applications have expanded in recent years, including typhoon-oriented upper-air sounding, coordinated offshore observations, and multi-aircraft field campaigns (Wen et al., 2025). More recently, increasing attention has been given to the development of low-altitude meteorological observing systems, in which MDs are expected to support fixed-point profiling, route-based observations, and targeted mobile observations, primarily within the lowest 1 km of the atmosphere. Several pilot programs have also begun exploring MD-based low-altitude observations for aviation and boundary-layer applications. Despite these developments, peer-reviewed studies evaluating the potential value of low-altitude MD observations remain limited, particularly with respect to observing network design and convective-scale data assimilation”.

2. In Section 4.2, the observation height sensitivity experiments use only single-level observations. The conclusion that multi-level data are better is intuitive, but the manuscript would benefit from a more detailed analysis of why the 600-m level performs better than 120 m or 300 m alone. For instance, are the larger improvements due to better constraining the low-level jet or moisture flux?

Response: Thank you for pointing this out. As only temperature and humidity observations are assimilated in these experiments, the improved precipitation forecast performance of the 600-m observations should primarily be attributed to their influence on the thermodynamic structure rather than to a direct constraint on the low-level wind field. Observations at 600 m sample a higher portion of the PBL than those at 120 m and therefore provide direct information on temperature and moisture conditions above the immediate near-surface layer that are relevant to convective instability and moisture availability. In addition, vertical background-error correlations in the assimilation system and model integration between successive assimilation cycles allow the information introduced at this level to influence a broader depth of the lower atmosphere. This interpretation is supported by the vertical RMSE profiles in Fig. 10, which show that the differences among the observation-height experiments are concentrated mainly in the lower troposphere and that the level of strongest thermodynamic constraint shifts upward

as the observation height increases. The 600-m observations therefore provide relatively stronger constraints on temperature and moisture above the near-surface layer, whereas the influence of the 120-m observations is more concentrated close to the surface. The discussion in Section 4.2 has been expanded to provide a clearer interpretation of the height dependence. The corresponding revisions are as follows:

1) Lines 659-669: “These height-dependent differences may help explain the subsequent differences in precipitation forecast skill. Compared with observations at 120 m, observations at 300 and 600 m provide direct constraints on temperature and moisture farther above the immediate near-surface layer, while vertical background-error correlations and model integration between assimilation cycles allow this information to influence a broader portion of the lower atmosphere. Such differences in the vertical thermodynamic constraint may contribute to the generally greater precipitation forecast benefits obtained from the higher single-level observations. Similarly, the plain-only configuration yields larger humidity RMSEs than the full-network experiment across much of the lower troposphere”.

3. In Line 575–577, “Among the tested configurations, the 5- and 10-km MD observation networks show the most consistent benefits”. When “consistent” is used in a single case, it is better to rephrase.

Response: This has been corrected in lines 600-602: “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”. Similar wording elsewhere in the manuscript has also been revised where necessary.

4. In the references part, some references are incomplete or inconsistently formatted (e.g., “Benjamin et al., 2004a” is cited but only one 2004 Benjamin entry appears in the reference list). Please verify all citations.

Response: Thanks. The in-text citations have been carefully cross-checked against the reference list and incomplete or inconsistent entries have been corrected. In particular, the previous citation “Benjamin et al. (2004a)” has been corrected to “Benjamin et al. (2004)”. We also corrected several incomplete bibliographic entries, DOI formatting errors, and the year suffixes of the Zenodo references, and standardized the reference formatting throughout the manuscript. In addition, several citations in the Introduction were revised to ensure that they more directly support the corresponding statements.

5. In the Data availability part, the Zenodo links are provided, but it would be helpful to also include the scripts used to generate the synthetic observations and compute verification metrics, as these are essential for reproducibility.

Response: We have archived the key Fortran subroutines used to generate the synthetic MD observations and calculate the precipitation verification metrics in the

Zenodo repository. The Code and Data Availability section has been revised accordingly: “The namelist files for WRF and the assimilation system, together with the key Fortran subroutines used to generate the synthetic MD observations and calculate the verification metrics, are accessible online (<https://doi.org/10.5281/zenodo.22297689>, Zhao, 2026a)”.