Response to Referee 1:

Review of Impact of Multiple Radar Wind Profilers Data Assimilation on Convective Scale Short-Term Rainfall Forecasts: OSSE Studies over the Beijing-Tianjin-Hebei region

General Comments:

This manuscript examines the impact of assimilating observations from various Radar Wind Profiler (RWP) networks on NWP forecasts of heavy rainfall in the Beijing–Tianjin–Hebei (BTH) region of China. These observation impact experiments are performed using an identical-twin observing system simulation experiment (OSSE) with only RWPs being assimilated. Results indicate that adding more RWP sites improves forecasts of heavy rainfall, with the configuration adding RWPs to both the ridge and foothills of the Taihang Mountains resulting in the largest benefits. Sensitivity tests indicate that the RWP impact is noticeably reduced when the detection altitude is reduced and when the vertical resolution is reduced. Interestingly, increasing the detection altitude past 8 km does not greatly improve forecasts. This information can be used in the future to help inform the further development of the RWP network in China.

Response: The authors appreciate the referee's insightful comments and constructive suggestions, which helped us significantly improve the quality of this manuscript. For clarity purpose, here we have listed the reviewers' comments in black font, followed by our response in blue font.

Overall, the study provides a useful step forward in better understanding observation impacts on mesoscale precipitation events in China, and I thank the authors for putting in the time to perform this study and write up the results. I do have a few of general suggestions, though, that I think will help improve the manuscript:

1. It seems like the OSSE framework used here is an identical twin (i.e., the forecast system and truth simulation use the same model) and omits observation types other than RWPs. While this is not inherently a problem, it would be very helpful if the authors discussed the limitations of such a setup and used these limitations to qualify their results. For example, not including other conventional observations (e.g., radiosondes, surface stations) might inflate the impact of RWPs. Additional information about the limitation of such a setup can be found in Hoffman and Atlas (2016).

   Response: Good point! Per your kind suggestion, we made the following modification in this revision:

   1) Lines 203-206: "This study utilizes an OSSE framework with an identical twin

setup, where the same numerical model is used for both the truth simulation and the forecast system. As noted by Hoffman and Atlas (2016), OSSEs with identical twin setups can lead to overly optimistic assessments of data impacts. Therefore, the results should be interpreted within the constraint."

2) Lines 252-256: "As our focus is to assess the impacts of assimilating wind observations from various RWP network layouts on convective-scale analysis and short-term severe weather prediction, only synthetic RWP data are assimilated in this study, excluding conventional observations such as radiosondes, surface stations, and satellite observations. This exclusion simplifies the analysis by isolating the impact of RWPs but may inflate their relative importance (Hoffman and Atlas, 2016)."

3) Lines 719-729: "As the fraternal twin scheme is used in this study, it does not account for model-related errors that occur in real-world applications. Consequently, the results might overestimate the actual benefits of RWP assimilation in operational systems. Furthermore, this study focuses exclusively on assimilating RWP data, without incorporating conventional observations or satellite data. While this approach simplifies the analysis by isolating the impact of RWPs, it may inflate their relative importance. Future research directions include: (1) Expanding the study to other precipitation types and high-impact convective events under diverse weather scenarios. (2) Evaluating the impact of RWP networks by assimilating RWPs together with more diverse observation types and incorporating non-identical twin setups to enhance realism and provide broader operational insights. (3) Investigating the benefits of assimilating real observational data on convective scale NWP once proposed RWP networks become available."

2. Some of the claims in section 4 do not appear to be fully supported by the figures. Can the authors revisit this section and double-check the claims made here? Some of the claims are also rather vague (e.g., on lines 433–436, it is not clear which metrics are being examined when claiming that one run is better than another run). Here are some specific examples of claims not fully supported by the figures (this list is non-exhaustive):

   Response: Thanks for pointing out this. The claims (line numbers have been updated to 479-486) have been reworded as "For the 20- and 30-dBZ thresholds, it is evident that FH_RD produces the highest ETS, POD, SR, and CSI scores during the 0-3 h forecast period, the improvement in BIAS values was minimal (Fig. 12a-d). However, for 40 dBZ, the RD experiment achives slightly higher ETS, POD, SR, and CSI scores than FH_RD does at most forecast lead times (Fig. 12e and f). It is also worth noting that, for 20- and 30-dBZ thresholds, FH produces higher ETS , POD, and CSI scores than RD does before the 2-h forecast lead time, while RD exhibits better forecast skill thereafter (Fig. 12a-d).". The authors have been reading through section 4 carefully and we hope most of the claims have been corrected.

   1) Lines 319–321: It looks like the RD and FH_RD experiments are also worse than CTL in the 550–500 hPa layer for u and that the RD and FH experiments are worse than CTL in the 700–550 hPa layer for v. So maybe it would be

better to simply say that "DA experiments assimilating ridge and foothill RWPs generally outperform CTL"?

Response: This has been corrected in lines 361-362: "It is seen that the DA experiments assimilating ridge and foothill RWPs generally outperform CTL". Thanks.

2) Line 499: There are some points on Fig. 15 that fall outside of the bias range listed here. E.g., 1-h, 30-dBZ CREF forecast for 20230712; 6-h, 40-dBZ CREF forecast for 20230721; 1-hr, 40-dBZ CREF forecast for 20230712.

   Response: This has been corrected in lines 557-559: "Except for 28 June 2023, the BIAS values of FH_RD fall within a reasonable range of 0.8–1.7 for reflectivity precipitation, indicating overall good forecast performance". Thanks.

3) Line 525: Based on Fig. 16b, it looks like all DA experiments actually increase the bias to make it farther from unity, not closer to unity.

   Response: Yes, this has been corrected in lines 587-590: "The BIAS values of the assimilation experiments are higher than that of the NoDA experiment (close to unity) at the analysis time, and then decreases slightly in the 1-6 h forecasts. However, the BIAS of NoDA increase consistently during 1–6 hours, making it farther from unity."

4) Line 556–558: The text states that POD initially increases, then quickly declines, but based on Fig. 17a, POD appears to increase during the entire 6-h forecast.

   Response: "POD and" has been removed in lines 622-623. Thanks.

3. The authors include some discussion as to why more RWP observations improve forecasts of heavy rainfall, which I think is really helpful. I think these discussions can be expanded, however, and made more clear. Some specific examples:

   1) Line 32: The phrase "enhances mesoscale dynamics" is vague.

      Response: This has been corrected to: "improves the representation of mesoscale atmospheric features".

   2) Lines 401–406: It is not clear what exactly is meant by "favorable environment for heavy rainfall". Is this the result of increased convergence?

      Response: Yes, the stronger southwesterly winds of the CTL experiment enhance moisture transport and convergence in the upstream mountains, leading to overestimated rainfall in those areas and underpredicted precipitation over Beijing. The corresponding text (lines 448-451) has been reworded.

3) Lines 553–556: It is not immediately clear to me why (b) results in RWP observations from the foothills having a more immediate impact on forecasts than RWP observations from the ridgeline.

   Response: For southwest-type rainfall events, the southwesterly wind propagates from upstream ridgeline stations to downstream foothill sites (Li et al., 2024). The foothill RWPs directly capture the moisture transport brought by the southwesterly flow and the evolution of convective systems. This has been clarified in lines 617-619.

   Li, N., Guo, J., Wu, M., Zhang, F., Guo, X., Sun, Y., Zhang, Z., Liang, H., and Chen, T.: Low-Level Jet and Its Effect on the Onset of Summertime Nocturnal Rainfall in Beijing, Geophysical Research Letters, 51, e2024GL110840, https://doi.org/10.1029/2024GL110840, 2024.

4) Line 578: Convergence zones are mentioned here, but nowhere else in the manuscript. If properly forecasting convergence zones is important to capturing heavy rainfall events in this area, it should be discussed in more detail somewhere in the manuscript.

   Response: Thank you for pointing out this. The expression has been corrected in lines 642-644: "In this research, observing system simulation experiments are performed to study the benefits of assimilating RWP observations for convective scale short-term heavy rainfall forecasts.".

In addition to these general comments, I also have several specific comments and technical corrections that are smaller in scope, which are listed below.

Specific Comments:

1. Line 42: Can the authors provide a range of vertical and temporal resolutions for RWPs here? I know this is included later, but I think it would be worthwhile to include it here to give the reader a sense of what is met by "high vertical and temporal resolution".

   Response: This is a very good suggestion. This has been added in lines 42-45: "Radar wind profilers (RWPs) are state-of-the-art meteorological observation instruments that provide wind profiles with vertical resolutions ranging from 60 to 240 meters and a temporal resolution of 6 minutes, enabling the detection of fine-scale atmospheric dynamic structures throughout the troposphere."

2. Lines 64–81: There is a lot of discussion here about the geography of the Beijing–Tianjin–Hebei region. I'd recommend including some references to the map in Fig.

3 so that the readers have a better sense of where exactly the mountains are relative to Beijing and what the density of the operational RWP network looks like.

Response: This is a very good suggestion. The references to Fig. 3 have been added in the corresponding text (lines 67, 71, and 79).

1) Line 124: A reference to Fig. 3 would also be helpful here. Does the domain of Fig. 3 perfectly match the WRF domain? If not, can the WRF domain be added to Fig. 3?

   Response: Yes, the domain of Fig. 3 is indeed the WRF simulation domain. The reference has been added (line 128).

3. Lines 96–99: Can a citation or two be added here to back up this claim?

   Response: This has been done in lines 101-103: "Meanwhile, the topography of the Taihang Mountains affects the distribution and intensity of the wind field, particularly during severe convective weather events (Li et al., 2024; Sheng et al., 2020)". Thanks for the suggestion.

4. Lines 156–158: Is there any DA in the truth run? My guess is that there is not (as is typically the case in OSSEs), but including the phrase "DA cycling" in this sentence makes it sound like the truth run might include DA. Can the sentence be reworded so that it is clear that the truth run does not include DA?

   Response: No, there is not any DA in the truth run. This has been reworded as "For each case, the model is initialized using the ERA5 data and integrated forward for 15 hours, with the boundary conditions also provided by the hourly ERA5 data." (lines 182-184) . Thanks for the suggestion.

5. Fig. 2: It is difficult to see the rain gauge observations in this figure, and it is also difficult (at least for me) to differentiate between the different dot sizes. Can the rain gauge data be plotted in a different manner to make it more clear? One idea: perhaps another row or column can be added to this figure that shows rain gauge observations as color-filled dots?

   Response: Thanks! This is a very good suggestion. Figure 2 has been updated to include a row of plots displaying rain gauge observations as color-filled dots, and we also reworded the corresponding text to reflect the change.

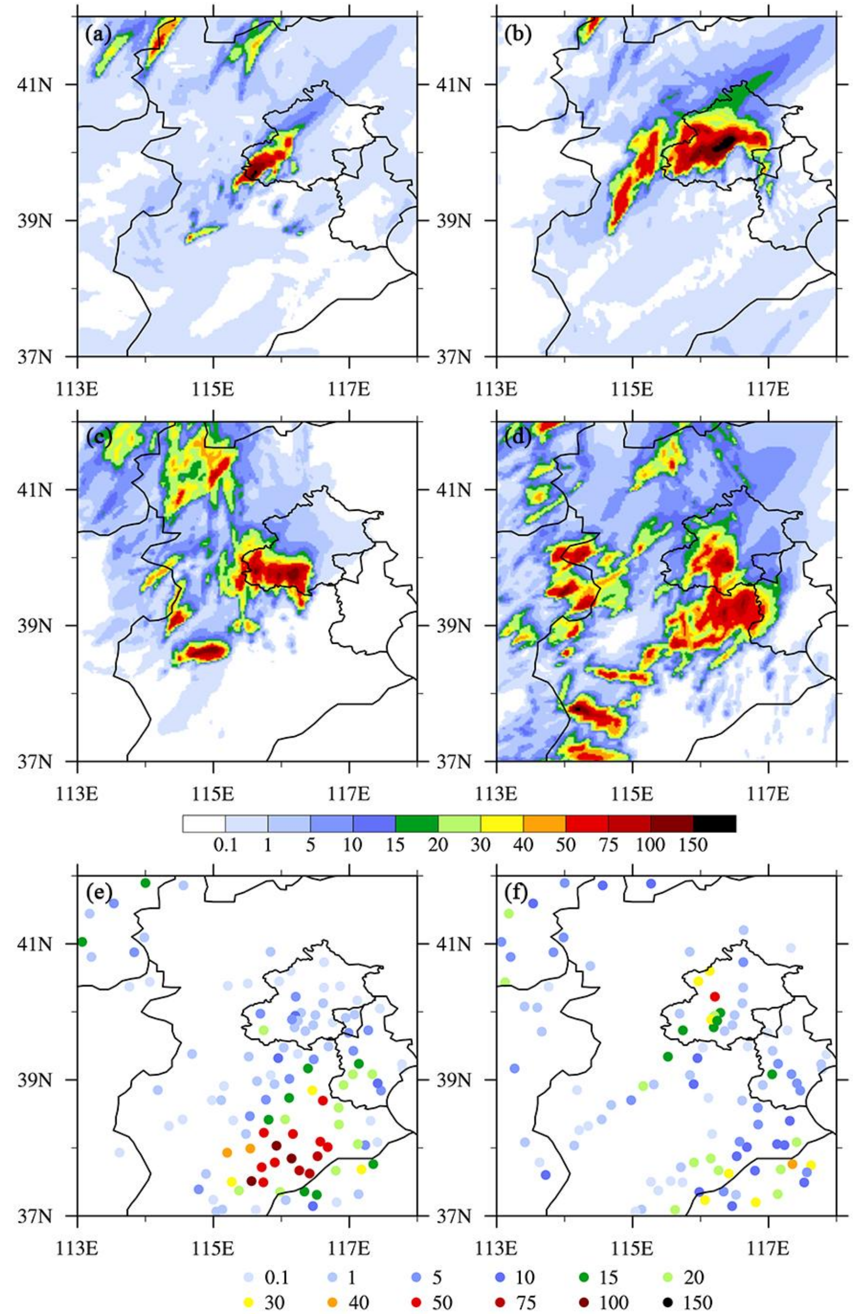

Figure 2. The 6-h accumulated precipitation (APCP) forecasts (mm, shaded) from 2100 UTC 20 July to 0300 UTC 21 July (left), and from 0300 UTC 21 July to 0900 UTC 21 July, 2023 (right) for (a)-(b) Truth, (c)-(d) NoDA experiments, and (e)-(f) the rain gauge measurements at national weather stations.

6. Lines 205–207: What exactly is the NOAA topographic dataset being used for here? Is this how RWP site elevations are determined?

Response: The dataset is the ETOPO1 Global Relief Model, a 1-arc-minute resolution topographic and bathymetric dataset provided by NOAA's National Centers for Environmental Information (NCEI). The dataset is widely used in geosciences, including oceanography, climate modeling, and environmental studies. Yes, we derive the RWP site elevations based on the topographic height in this dataset. We have added a citation of this dataset and made it clearly in lines 236-239: “The spatial locations for the foothill and ridge sites, with a total of 16 sites each, are determined based on the ETOPO1 Global Relief Model, a 1-arc-minute resolution topographic and bathymetric dataset provided by NOAA's National Centers for Environmental Information (Amante and Eakins, 2009).”.

7. Fig. 3: Can the Bohai Bay be made a different color to make it clear that this is a large body of water and not just flat land? Similar to terrain differences, differences in land cover between land and water can have a large impact on local meteorology.

Response: This is a very good suggestion. Figure 3 has been updated to fill the ocean in a light blue.

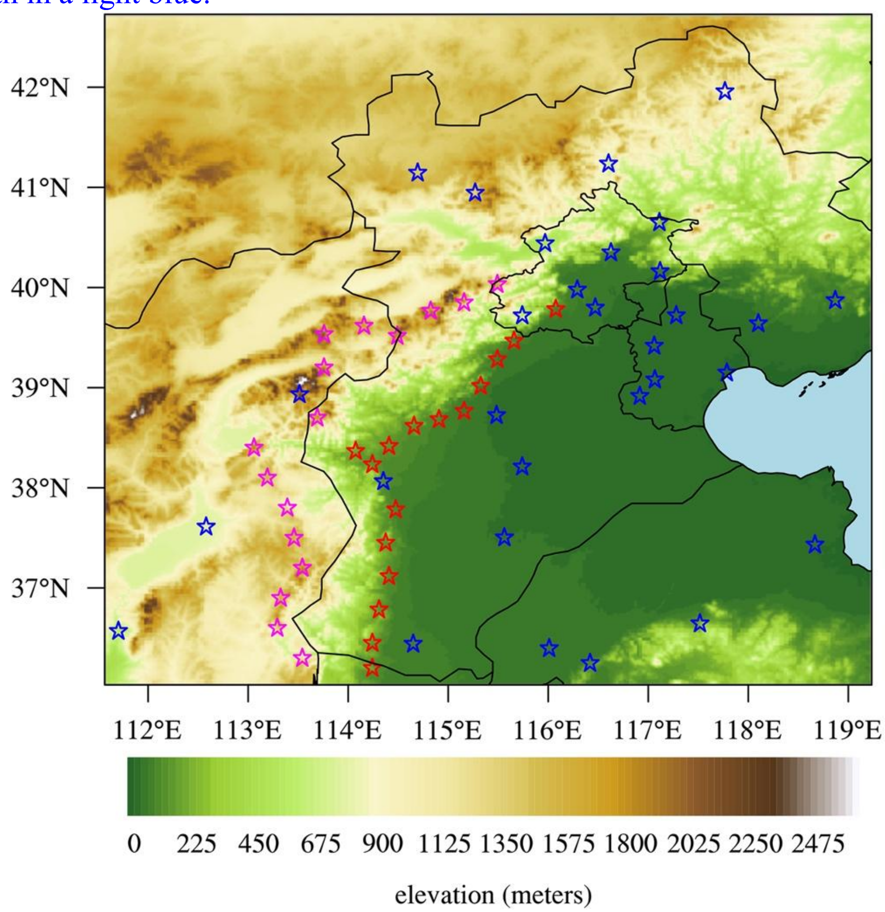

Figure 3. Spatial distribution of the operational RWP network (blue stars), and simulated RWP network along the foothill (red stars) and ridge (magenta stars) of the Taihang Mountains within the simulation domain. The terrain is represented by color shading, and

the ocean is shown in light blue.

8. Lines 224–225: There are certainly several similarities between the forecast system used here and WoFS (e.g., both use a very similar dynamical model), but the DA in WoFS seems very different. As mentioned in the manuscript, WoFS relies heavily on flow-dependent background error covariances derived from an ensemble (either in an EnKF or hybrid 3DEnVar), which is a notable difference from the pure 3DVAR approach used here.

   Response: Yes, we used a different assimilation method (3DVAR) compared to the WoFS system. Our approach primarily followed a cycling assimilation and forecasting workflow similar to that used in the WoFS real-time Spring Forecast Experiment (SFE) run, that is cycling DA for 9 hours at 15-min intervals. We have made it clearly in lines 263-271: "To mimic real-world operations, this OSSE study employs a DA and forecast cycle workflow similar to the Warn-on-Forecast System (WoFS) real-time Spring Forecast Experiment (SFE) runs, that is cycling DA for 9 hours at 15-min intervals (Heinselman et al., 2024; Hu et al., 2020; Jones et al., 2018) (Fig. 4). To minimize data contamination from precipitation, DA cycles are performed before widespread rainfall occurs in the simulation domain, as wind profile accuracy from RWPs can be degraded by falling hydrometeors (Zhang et al., 2017). The model initial and boundary conditions for all DA and forecast experiments are derived from the 12-h GFS forecasts. Unlike the SFE setupa 6-h free forecast in this study is launched every hour starting from the sixth hour of the analysis cycles, rather than from the first hour (Fig. 4).".

9. Lines 227–228: Can the authors confirm that the majority of RWP observations were collected in regions without precipitation? If not, does removing the RWP observations in regions of precipitation change the results?

   Response: The 3-h accumulated precipitation (APCP) forecasts of the truth simulation from the case on 21 July 2023 during the 9-h DA cycles overlapped by RWP site distributions are illustrated in Fig. r1 as an example. It is shown that in the first 3 hours of the DA cycles, none of the RWP sites is affected by precipitation (Fig. r1a). For 3-6 h, only one ridge station and two foothill stations are located in regions with weak precipitation (Fig. r1b). For 6-9 h, two ridge stations and three foothill stations are located in weak precipitation areas or at the boundaries of slightly stronger precipitation regions (Fig. r1c). In summary, the majority of RWP observations were collected in regions without precipitation, and two other cases showed similar situations. We believe that removing observations from a few sites during the final hours of the assimilation cycles may have a minimal impact on the assimilation results.

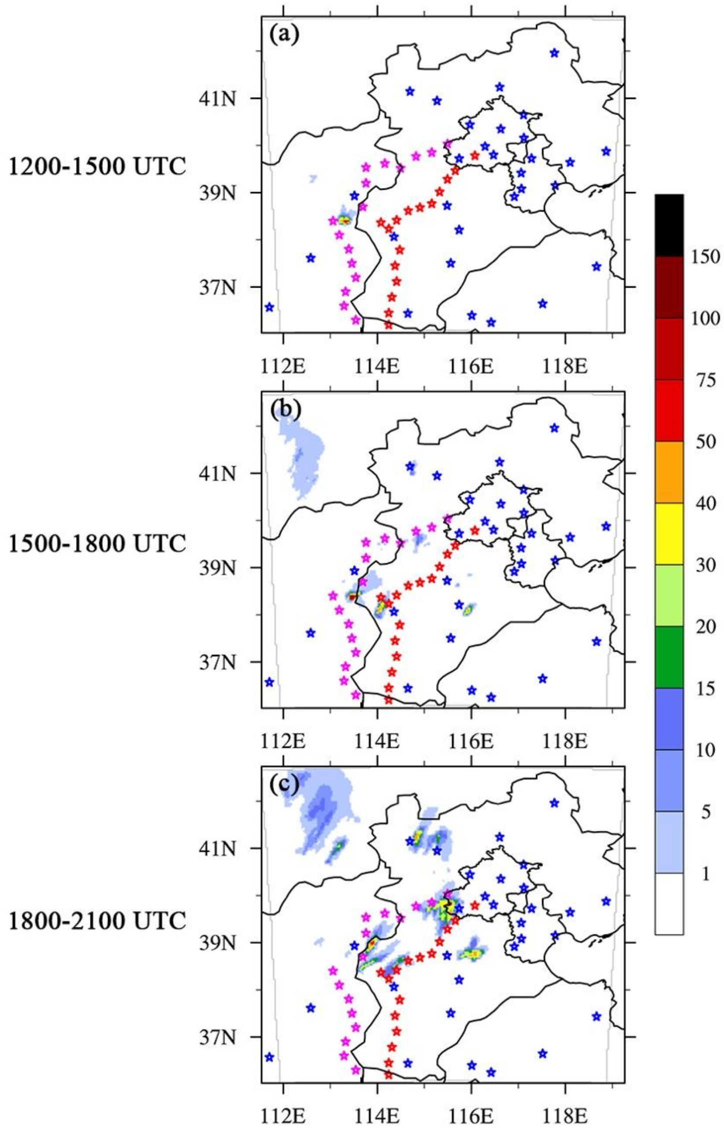

Figure r1. The 3-h accumulated precipitation (APCP) forecasts (mm, shaded) for the truth simulation (a) from 1200 UTC to 1500 UTC, (b) from 1500 UTC to 1800 UTC, and (c) from 1800 to 2100 UTC 20 July, 2023. Blue, red, and magenta starts represent the operational RWP network, the foothill RWP network, and the ridge RWP network, respectively.

10. Lines 229–230: Which GFS forecast hours are used for the initial and boundary conditions?

    Response: The 12-h GFS forecasts are used for the initial and boundary conditions. We have made it clearly in lines 268-269: “The model initial and boundary conditions for all DA and forecast experiments are derived from the 12-h GFS forecasts.”.

11. Lines 255–263: Based on this section, it sounds like there are no cross-variable correlations between RWP winds and other state variables (i.e., assimilating RWPs only updates the winds and not other variables like temperature and humidity). Can the authors confirm this? How will this impact the results from this study?

    Response: Thank you for pointing out this important aspect of the data assimilation process. The RWP observations primarily update the wind fields. However, the static background error covariance used in our assimilation system does include cross-variable correlations between wind and other state variables, such as temperature, humidity, and surface pressure. These correlations enable indirect updates to other state variables when wind observations are assimilated, although the correlations are predefined and static, which may limit their representativeness for certain dynamic atmospheric conditions. Despite the constraints, the primary focus of this study is to assess the influence of RWP observations on wind field analysis and their subsequent effect on precipitation forecasts. The indirect updates to temperature and humidity may also contribute to these improvements, although their impact is not explicitly evaluated in this study. We have acknowledged the potential benefits of employing advanced assimilation techniques, such as hybrid or ensemble-based methods, which can better represent flow-dependent cross-variable correlations and improve the assimilation of RWP data in the revised manuscript (lines 160-172, 729-736).

12. Line 256: Do the background wind errors increase or decrease with height? It is not clear from this sentence.

    Response: The background wind errors increase with height. This has been added in lines 295-297: "In all DA experiments, the background errors for zonal and meridional wind components are specified as 3–6 m/s, gradually increasing with altitude from the surface to 20 km above ground level (AGL).".

13. Line 286: Given that the model grid spacing is > 1 km, the term "convection-allowing" is more appropriate than "convection-resolving" because the convection is not fully resolved (e.g., Potvin and Flora 2015).

    Response: Corrected.

14. Lines 300–302: I am having trouble following the second part of this sentence. Maybe it can be reworded?

    Response: This has been reworded as "The superiority of FH_RD, RD, and FH over the CTL experiment persists during the subsequent 6-h free forecasts, highlighting the impact of wind profile observations gathered from ridge and foothill networks." (lines 341-343).

15. Line 377: This sentence is a bit confusing. Is the message that the intensity errors

are reduced in the FH experiment, but there is still a positive bias?

Response: This has been reworded as "The FH experiment produces stronger storms with a larger coverage area in Beijing compared to the CTL experiment, although the storm intensity remains slightly underestimated" (lines 419-421).

16. Lines 398–399: By "initial," are the authors referring to the 1-h APCP forecasts?

    Response: Yes, it is. We have made it clearly in line 442: "Although the areal coverage of rainfall in the 1-h forecast".

17. Lines 492–512: This section only compares the FH_RD experiments between the cases, but it seems like it might be more beneficial to compare each of these FH_RD experiments to their respective control runs. This comparison would provide a much clearer picture of how RWP observations improve NWP forecasts.

    Response: This is a very good suggestion. Figure 15 has been updated to include the forecast results from the control runs, and we also reworded the corresponding text to reflect the changes: "The results from the NoDA experiment are also shown to provide a clear picture of how RWP observations improve the short-term forecasts across different cases. For both the NoDA and FH_RD experiments, the forecast skills generally exhibit lower score metrics and more variability at higher thresholds. Overall, for different cases, the FH_RD experiment shows higher POD, CSI, and SAR values compared to the NoDA experiment, with more significant improvements observed in the first 3 hours. Most of the BIAS values for FH_RD are smaller than those for the NoDA experiment. Except for 28 June 2023, the BIAS values of FH_RD fall within a reasonable range of 0.8–1.7 for reflectivity precipitation, indicating overall good forecast performance. It is noted that some of the forecast scores do not decrease monotonically with increasing forecast lead time. For example, in the case 12 July 2023, smaller BIAS and FAR values are obtained for the 3- and 6-h reflectivity and precipitation forecasts, along with higher CSI. This occurs due to several factors: (a) initial scattered convection develops into a larger-scale west-east oriented system covering all of Beijing and central-northern Hebei at later times in this case, which models usually capture better; (b) errors in the timing and location of CI become less significant as convection evolves and forms clearer structures; and (c) for the free forecasts initialized from the first few hours, convection may not have started until the final forecast hour. CREF forecasts from FH_RD for the case 28 June 2023 show the best performance in terms of high POD, SR, and CSI. Meanwhile, persistent underprediction throughout the 1–6 h precipitation forecasts at all thresholds from this case can mostly be traced back to the difficulty in forecasting small-scale, short-lived, and relatively weak precipitation events. This phenomenon is more pronounced in the NoDA experiment, manifested by extremely low POD and CSI values." (lines 548-571).

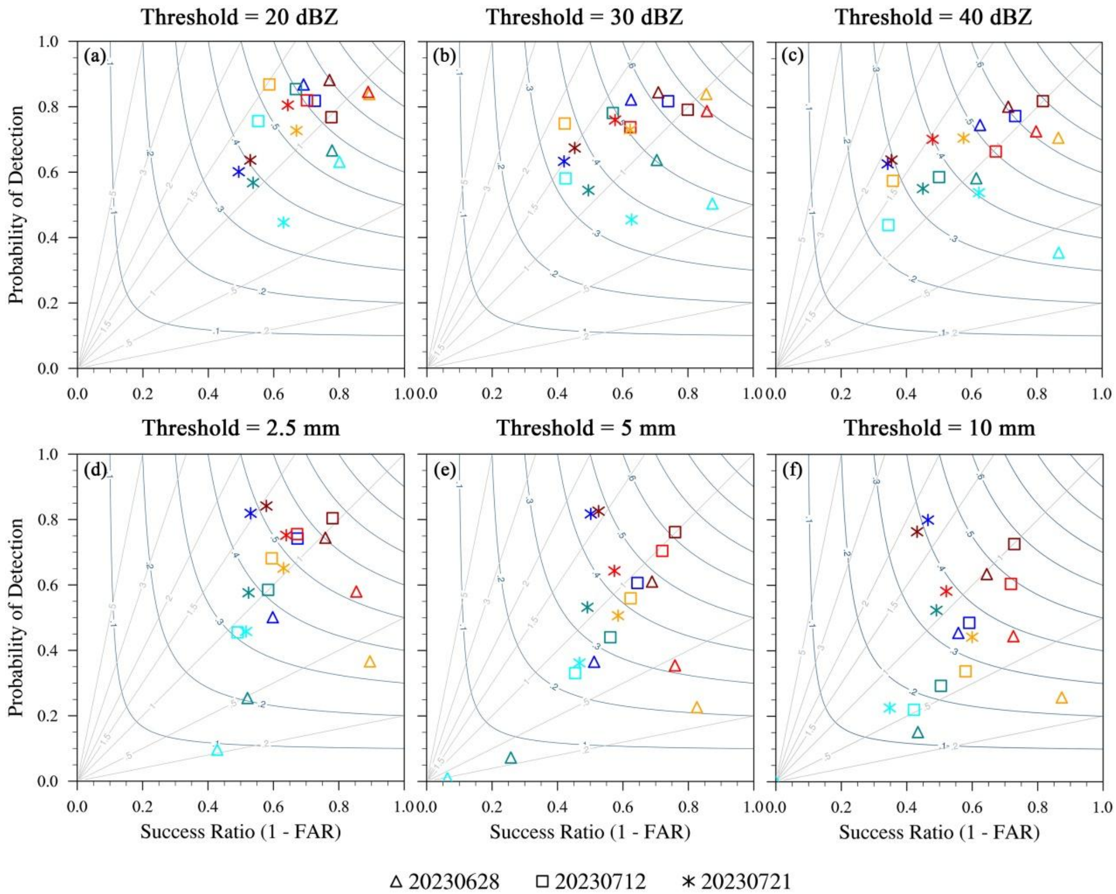

Figure 15. Performance diagram for 0-6 h CREF forecasts from the NoDA (cyan, dark cyan, and blue) and FH_RD (orange, red, and brown) experiments in each case at thresholds of (a) 20, (b) 30, and (c) 40 dBZ, respectively. (d–f) Same as in (a–c), but for 1-6 h HPRCP forecasts from each case at thresholds of (d) 2.5, (e) 5, and (f) 10 mm, respectively. Cyan and orange represent the analysis (1-h forecast for precipitation), dark cyan and red for 3-h forecasts, and blue and brown for 6-h forecasts. Results are shown for a neighborhood radius of 12-km.

18. Lines 525–526: How is statistical significance determined here? Based on Fig. 16, statistical significance can not be deduced because there appears to be overlap between the confidence intervals at all times. This does not necessarily mean that the difference in forecast statistics is not statistically significant (e.g., see 5.4.6 in this reference from Vanderbilt University: https://researchguides.library.vanderbilt.edu/c.php?g=69346&p=855555#), but we can not say for certain that the difference is significant without another type of statistical test. If another test has not been performed, I would recommend removing this mention of statistical significance.

Response: Thank you so much for your suggestion and for providing the reference materials. We have utilized all 6-h forecast samples from different initialization times across all cases and performed the calculations using the same method described in the references you provided. We have removed the mention of statistical significance in the revised manuscript (line 587).

19. Code and data availability: Can the authors also provide a link to the DA code used in this study? It would also be really helpful if the authors could provide the namelist files for WRF and the DA in a public-facing repository.

    Response: Thank you for your interest in the assimilation system used in this study. Regarding your request for the code, I regret to inform you that the system was developed collaboratively by our research team and involves proprietary components that cannot be made publicly available at this time. However, we are happy to share the namelist files for WRF and the DA (https://doi.org/10.5281/zenodo.14241597) and answer any specific questions you may have about the implementation. Accordingly, we have revised the Code and Data Availability section.

Technical Corrections:

1. Line 20: I'd recommend replacing "in rainfall forecasting" with "for rainfall forecasting".

   Response: Corrected as suggested.

2. Line 31: "benifiting" should be "benefiting".

   Response: Corrected as suggested.

3. Line 41: "state-of-art" should be "state-of-the-art".

   Response: Corrected as suggested.

4. Lines 55–57: This sentence sounds awkward, particularly the phrase "over 100 sites are deployed by 2020". Can it be reworded?

   Response: Corrected as suggested.

5. Line 59 (and elsewhere): Definite and indefinite articles (i.e., "the" and "a") are missing in several locations in sections 1–3. More information about when to use definite and indefinite articles can be found here: https://owl.purdue.edu/owl/general_writing/grammar/using_articles.html. Here are

some specific examples (italics indicate an added article; this list may not be exhaustive):

Response: Thanks for pointing out this. The authors have been reading through the manuscript carefully and we hope most of the articles have been corrected.

1) Line 59: “above national average” should be “above the national average”.

Response: Corrected as suggested.

2) Line 73: “in BTH region” should be “in the BTH region”.

Response: Corrected as suggested.

3) Line 76: “like Taihang mountains” should be “like the Taihang Mountains”.

Response: Corrected as suggested.

4) Line 80: “of RWP network” should be “of the RWP network”.

Response: Corrected as suggested.

5) Line 80: “in Taihang Mountains” should be “in the Taihang Mountains”.

Response: Corrected as suggested.

6) Line 99: “prior study” should be "a prior study”.

Response: Corrected as suggested.

7) Line 106: “over operational RWP network” should be “over the operational RWP network”.

Response: Corrected as suggested.

8) Line 114: “of NWP model” should be “of the NWP model”.

Response: Corrected as suggested.

9) Line 138: “enhance wind field analysis” should be “enhance the wind field analysis”.

Response: Corrected as suggested.

10) Line 139: “in PBL” should be “in the PBL”.

Response: Corrected as suggested.

11) Line 220: “distribution of operational RWP network” should be “distribution of the operational RWP network”.

Response: Corrected as suggested.

12) Line 221: “ridge of Taihang Mountains” should be “ridge of the Taihang Mountains”.

Response: Corrected as suggested.

13) Line 231: “hour of analysis” should be “hour of the analysis”.

Response: Corrected as suggested.

14) Line 237: “end of DA cycles” should be “end of the DA cycles”.

Response: Corrected as suggested.

15) Line 274: “using neighborhood radius” should be “using a neighborhood radius”.

Response: Corrected as suggested.

6. Line 64: “Heibei” should be “Hebei”.

   Response: Corrected as suggested.

7. Line 69: “plain” should be “plains”.

   Response: Corrected as suggested.

8. Line 72: Is there a reason why “large” is used instead of “synoptic” here?

   Response: The word “large” has been replaced with “synoptic”.

9. Lines 75–76: Add the word “in” between “network” and “concentrated”.

   Response: Corrected as suggested.

10. Line 78: “observation” should be “observations”.

    Response: Corrected as suggested.

11. Line 81: “forecast” should be “forecasts”.

    Response: Corrected as suggested.

12. Lines 89–92: Can this sentence be reworded? In particular, the phrases “few peer-reviewed published research” and “RWP network associated with complex terrain” (which also appears on line 93) are confusing. Perhaps the latter could be rewritten as “a RWP network in complex terrain”?

    Response: Reworded as suggested.

13. Line 106: “Does” should be “Do”.

    Response: Corrected as suggested.

14. Line 122: Remove the word “space” after “1.5-km”.

    Response: Corrected as suggested.

15. Line 136: Remove “and” and replace with a comma.

    Response: Corrected as suggested.

16. Line 136: “atmospheric motion vector” should be plural.

    Response: Corrected as suggested.

17. Line 141: Remove comma after “3DVAR”.

    Response: Corrected as suggested.

18. Line 143: Add “enhancements” after “RWP network”.

    Response: Corrected as suggested.

19. Line 168: Remove “of” between “southwest” and “Beijing”.

    Response: Corrected as suggested.

20. Lines 262–263: Please rewrite so that this sentence does not start with “And”.

    Response: Thanks for the suggestion. This has been corrected: “The horizontal correlation scale lengths are set to be 50 km in the first loop and 20 km in the second loop, while the corresponding vertical correlation lengths are 5 and 2 grid points, respectively.” (lines 301-303).

21. Line 269: “experiments” should be singular.

    Response: Corrected as suggested.

22. Line 298: “uncertiaties” should be “uncertainties”.

    Response: Corrected as suggested.

23. Line 314: It looks like the analyses are solid, not dashed.

    Response: Corrected as suggested.

24. Fig. 6: Note that the noDA experiment is black in the caption.

    Response: Corrected as suggested.

25. Line 335: Add this sentence to the caption of Fig. 7.

    Response: Added as suggested.

26. Line 410: A different word should be used in place of “estimates”. The rainfall from the truth run is the actual rainfall amount for the OSSE, not an estimate.

    Response: It has been replaced with “forecasts”

27. Line 413: "accumulate" should be “accumulated”.

    Response: Corrected as suggested.

28. Line 501: The transition “however” sounds out of place here.

    Response: This word has been removed.

29. Line 509: The transition “nevertheless” sounds out of place here.

    Response: “Nevertheless has been replaced with “Meanwhile”.

30. Line 664: “accsible” should be “accessible”.

    Response: Corrected as suggested.

References

Hoffman, R. N., and R. Atlas, 2016: Future Observing System Simulation Experiments. Bull. Amer. Meteor. Soc., 97, 1601–1616, https://doi.org/10.1175/BAMS-D-15-00200.1.

Potvin, C. K., and M. L. Flora, 2015: Sensitivity of Idealized Supercell Simulations to Horizontal Grid Spacing: Implications for Warn-on-Forecast. Mon. Wea. Rev., 143, 2998–3024, https://doi.org/10.1175/MWR-D-14-00416.1.

Response: Thanks for providing the references. They have been added in the revised manuscript.

Response to Referee 2:

Review of Impact of Multiple Radar Wind Profilers Data Assimilation on Convective Scale Short-Term Rainfall Forecasts: OSSE Studies over the Beijing-Tianjin-Hebei region

General Comments:

The study investigates the benefits of using radar wind profiler (RWP) observations for forecasting convective initiation (CI) in small-scale boundary layer convergence zones. The research employed the Weather Research and Forecasting (WRF) model along with the NSSL3DVAR data assimilation (DA) system. Synthetic RWP data, generated through Observing System Simulation Experiments (OSSE), were assimilated for three summer heavy rainfall events in the Beijing-Tianjin-Hebei region. The results indicated that assimilating RWP observations improved model initial conditions and enhanced short-term severe weather forecasts. Notably, improved forecasting outcomes were observed when combining operational, foothill, and ridge RWP data. Multiple sensitivity experiments were conducted to evaluate the impact of vertical resolution and maximum detection heights. The study identified optimal configurations for future real-time RWP data assimilation.

The research and results of the study are both interesting and significant. The manuscript is well-structured and clearly presented, offering a thorough analysis using the OSSE method. It is well-written and will be ready for publication after a few minor revisions and responses to my questions.

Response: The authors appreciate the referee's insightful comments and constructive suggestions, which helped us significantly improve the quality of this manuscript. For clarity purpose, here we have listed the reviewers' comments in black font, followed by our response in blue font.

Comments:

1. In lines 144-145, 'so we did not use the ensemble derived background error covariance, which is also incorporated in the variational framework'. What method is used to compute the background error statistics? What control variables are utilized to calculate the B-matrix? Is the covariance matrix a day or wet matrix. Write a few details about the calculation of the B-matrix.

   Response: Thanks! This is a very good suggestion. This has been added in lines 164-172: "The background error covariance matrix used in this study is constructed as the product of a diagonal matrix representing the standard deviations of background errors and a spatial recursive filter (Gao et al., 2004, 2013). The standard deviations for the pressure, potential temperature, relative humidity, zonal and meridional wind components are derived from the statistics of the Rapid Update Cycle (RUC, Benjamin et al., 2004) 3-hour forecasts over several years (Fierro et al., 2019b; Pan et al., 2021). The background error correlations are modeled by the recursive filter described by Purser et al. (2003a, b). The recursive filter can be applied in multiple passes (or outer loops), using different correlation length scales tailored to the scale of the weather systems represented by the assimilated observations."

2. Line 135, 'total precipitable water' are the products or sources of observations used for assimilating total precipitable water? The author needs to include appropriate references for each observation.

   Response: Thanks for pointing out this. The total precipitable water (TPW) is the satellite-retrieved product, such as GOES-16 TPW product, which contains water vapor information (Pan et al., 2018). The claims on lines 150-155 have been reworded as follows: "The NSSL3DVAR system assimilates multi-sensor high-resolution observations like radar radial velocity and reflectivity (Gao et al., 2013, 2016), sounding and surface data (Hu et al., 2021), and multiple satellite-retrieved products, such as cloud water path (Pan et al., 2021), total precipitable water (Jones et al., 2018; Pan et al., 2018), atmospheric motion vectors (Mallick and Jones, 2020; Zhao et al., 2021b, 2022), and Geostationary Lightning Mapper (GLM)-derived water vapor (Fierro et al., 2019a; Hu et al., 2020)."

3. Authors need to include additional information about the NSSL3DVAR assimilation system. This should encompass the minimization cost function, control variables, and the calculation of the B-matrix.

   Response: Thanks for pointing out this. Descriptions of cost function, control variables, and the calculation of the B-matrix have been added in the manuscript as follows:

   1) Lines 138-149: "In the NSSL3DVAR system, the analysis is derived by minimizing the cost function defined as the background term J_b and the observation term J_o plus the constraint term J_c:

$$J = J_b + J_o + J_c = \frac{1}{2}(\mathbf{x}-\mathbf{x}_\mathrm{b})^\mathrm{T}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_\mathrm{b}) + \frac{1}{2}(H(\mathbf{x})-\mathbf{y}_\mathbf{o})^\mathrm{T}\mathbf{R}^{-1}(H(\mathbf{x})-\mathbf{y}_\mathbf{o}) + J_c, \qquad (1)$$

   where $\mathbf{x}$ and $\mathbf{x}_\mathrm{b}$ are the analysis and background state vectors, respectively; H is the observation operator projecting analysis into the observational space; and $\mathbf{y}_\mathrm{o}$ is the observation vector. $\mathbf{B}$ is the background error covariance matrix, and $\mathbf{R}$ is the observation error covariance matrix. J_c represents weak constraints which include elastic mass continuity equation and diagnostic pressure equation constraints suitable for convective-scale data assimilation (Gao et al., 2004; Ge et al., 2012). Analysis variables include the three-dimensional wind fields, air pressure, potential temperature, water vapor mixing ratio, and the hydrometeors containing the mass mixing ratios for cloud water, rainwater, ice, snow, and graupel (Gao and Stensrud, 2012).

2) Lines 164-172: “The background error covariance matrix used in this study is constructed as the product of a diagonal matrix representing the standard deviations of background errors and a spatial recursive filter (Gao et al., 2004, 2013). The standard deviations for the pressure, potential temperature, relative humidity, zonal and meridional wind components are derived from the statistics of the Rapid Update Cycle (RUC, Benjamin et al., 2004) 3-hour forecasts over several years (Fierro et al., 2019b; Pan et al., 2021). The background error correlations are modeled by the recursive filter described by Purser et al. (2003a, b). The recursive filter can be applied in multiple passes (or outer loops), using different correlation length scales tailored to the scale of the weather systems represented by the assimilated observations.”

4. Line 132: Add a few references after the phrase water path, total precipitable water, and atmospheric motion vector. Jones, T. A., Wang, X., Skinner, P., Johnson, A., & Wang, Y. (2018). Assimilation of GOES-13 imager clear-sky water vapor (6.5 μm) radiances into a warn-on-forecast system. Monthly Weather Review, 146(4), 1077–1107. https://doi.org/10.1175/MWR-D-17-0280.1 Mallick, S., & Jones, T. A. (2020). Assimilation of GOES-16 satellite derived winds into the warn-on-forecast system. Atmospheric Research, 245, 105131. https://doi.org/10.1016/j.atmosres.2020.105131

Response: Thank you so much for your suggestion and for providing the references. They have been added in lines 152-154: “multiple satellite-retrieved products, such as cloud water path (Pan et al., 2021), total precipitable water (Jones et al., 2018; Pan et al., 2018), atmospheric motion vectors (Mallick and Jones, 2020; Zhao et al., 2021b, 2022)”.

5. Did the authors use any additional conventional observations or satellite observations in their assimilation system?

Response: No, we only assimilate RWP observations without incorporating additional conventional or satellite observations in this study. This exclusion simplifies the analysis by isolating the impact of RWPs but may inflate their relative importance. We will evaluate the impact of RWP networks by assimilating RWPs together with more diverse observation types to enhance realism and provide broader operational insights in future studies. The corresponding claims have been added in the manuscript as follows:

1) Lines 252-256: “As our focus is to assess the impacts of assimilating wind observations from various RWP network layouts on convective-scale analysis and short-term severe weather prediction, only synthetic RWP data are assimilated in this study, excluding conventional observations such as radiosondes, surface stations, and satellite observations. This exclusion simplifies the analysis by isolating the impact of RWPs but may inflate their relative importance (Hoffman and Atlas, 2016).”

2) Lines 721-728: “Furthermore, this study focuses exclusively on assimilating

RWP data, without incorporating conventional observations or satellite data. While this approach simplifies the analysis by isolating the impact of RWPs, it may inflate their relative importance. Future research directions include: (1) Expanding the study to other precipitation types and high-impact convective events under diverse weather scenarios. (2) Evaluating the impact of RWP networks by assimilating RWPs together with more diverse observation types and incorporating non-identical twin setups to enhance realism and provide broader operational insights."

6. Line 201, rewrite the line 'meridional wind components (u and v)' to 'meridional wind (u and v) components'.

   Response: Rephrased as suggested.

7. Line 217, 'The perturbations are assumed to be normally distributed Gaussian random errors with a mean of zero and a standard deviation of 2 m/s' , How do you calculate the observation error for each location?

   Response: Thank you for your question regarding the calculation of observation error. In this study, the observation errors at each location were modeled as normally distributed Gaussian random errors with a mean of zero and a standard deviation of 2 m/s. These errors were applied uniformly across all locations, assuming they reflect the typical uncertainties in RWP observations based on previous studies and operational experience (Hu et al., 2017; Zhang et al., 2016). The 2 m/s standard deviation represents a reasonable estimate of the observational error for RWP wind measurements. This value was chosen based on literatures on data quality assessment and data assimilation of wind profiles from the operational RWP network of China (Liu et al., 2020; Wang et al., 2020; Wang et al., 2023; Zhang et al., 2016; Zhang et al., 2017), ensuring consistency in evaluating the impact of different RWP network configurations.

8. In the summary section, write a few words on how incorporating flow-dependent background error covariances in data assimilation (DA) systems can enhance precipitation forecasting compared to the static background errors used in 3DVAR.

   Response: Thanks! This is a very good suggestion. This has been added in lines 729-736: "Moreover, future studies can address the limitations of static background errors in 3DVAR by incorporating flow-dependent background error covariances estimated from ensemble forecasts. As ensemble-based background error covariances can dynamically adapt to the evolving state of the atmosphere, the DA system will better represent the spatial and temporal variability of background errors, particularly in regions with complex topography or mesoscale features like convective systems. By leveraging flow-dependent background errors, the analysis can more accurately capture the initial atmospheric state, ultimately leading to more accurate precipitation predictions."

9. Based on OSSE results, what are the key considerations for designing and deploying optimal RWP networks in complex terrain regions?

Response: The OSSE results provide some insights into designing and deploying optimal RWP networks in complex terrain regions like the Taihang Mountains. Key considerations include:

a) **Station Placement in Key Terrain Areas**. The OSSE results consistently demonstrate that the FH_RD experiment, combining data from ridge, foothill, and operational wind profiler networks, delivers the most accurate short-term forecasts. Specifically, the assimilation of RWP data from ridge network significantly reduces wind errors in complex terrain, such as the Taihang Mountains upstream of Beijing. These regions are critical for convective initiation in Beijing and its surroundings. Therefore, RWP deployment in complex terrain should prioritize locations where terrain-driven meteorological processes are most pronounced, such as ridges, valleys, and foothills.

b) **Integrated Network Configuration**. The FH_RD experiment shows that combining ridge, foothill, and existing operational RWP networks delivers the most accurate short-term forecasts. This finding highlights the importance of integrating new RWP stations with existing observation resources to create a synergistic and complementary network, optimizing coverage and forecast accuracy.

c) **High Vertical Resolution and Detection Height**. Sensitivity experiments on vertical resolution and detection height further emphasize the importance of high vertical resolution and maximizing detection height in optimizing the RWP network for enhanced forecast accuracy. In complex terrain regions, where the lower atmosphere is strongly influenced by topography, accurate vertical wind profiles in the lower and mid-troposphere are essential for improving forecasts of convection and precipitation.

d) **Regional Adaptation to Terrain and Meteorological Characteristics.** Each complex terrain region has unique topographical and meteorological features. Tailored analyses are necessary to optimize station placement based on regional characteristics. For instance, in other mountainous regions, RWP networks might focus on capturing monsoon systems or localized storm development, requiring adjustments to station positioning accordingly.