Response to reviewers of the manuscript

"*A 3D-Var Assimilation Scheme for Vertical Velocity with the CMA-MESO v5.0*"

H. Li, Y. Yang, J. Sun, Y. Jiang, R. Gan, and Q. Xie

for Geoscientific Model Development

Response to Reviewer #1

Reviewer #1

In this study, a vertical velocity (w) assimilation scheme is developed in the 3DVar method. In this w assimilation scheme, the Richardson equation is employed as the operator for w, enabling the 3DVar system to update dynamic and mass variables from w assimilation. The w assimilation scheme looks reasonable to me. However, I have concerns about the experimental configurations for both the single case study and batch experiments in the manuscript. Therefore, I recommend a major revision.

Response: We gratefully thank the reviewer for your time spent making the constructive remarks and useful suggestions, which has significantly raised the quality of the manuscript and has allowed us to improve the manuscript. Each suggestion has been accurately incorporated and addressed in the revised manuscript. Please see our detailed response below in blue.

Major comments:

1. The authors highlight the use of the Richardson equation as the operator for *w* to reduce imbalance, but there is a lack of corresponding discussion and figures to support this assertion. I strongly recommend the authors incorporate related discussions and provide figures that illustrate the effectiveness of employing the Richardson equation for *w* in reducing imbalance.

Response: We appreciate the reviewer's valuable suggestion. The authors here expressed that using the Richardson equation as an observation operator for assimilating w to reduce imbalance is in comparison to the direct assimilation of w by expanding it as a control variable (the observation operator is simplified as an interpolation operator from model space to observation space). Unfortunately, w is not treated as a control variable in the current 3D-Var system of CMA-MESO. Therefore, the authors regret that, at present, we are unable to conduct further experimental validation and can only rely on the algorithm's design to explain this issue. Using w as an extended momentum control variable may introduce excessive noise, as w only has an impact on itself after assimilation, unless a physical constraint is incorporated through a physical transformation in the background error or a physical weak constraint is added to the cost function. To avoid this problem, Chen et al. (2020) assimilated the pseudo-horizontal wind convergence derived from the w fields using the mass continuity equation. In our study, the adiabatic Richardson equation is employed as the observation operator for w. This observation

operator combines the continuity equation, adiabatic equations, and hydrostatic relation. As a result, the update of other variables is physically constrained when w observations are assimilated.

Nevertheless, the authors attempted to analyze the correlation between w and other variables (including the zonal and meridional winds u and v, temperature t, and water vapor mixing ratio q_v) based on 120 forecast samples. Fig. R1 shows the correlation coefficients between the time series of variable w and the other model variables at each point on the 700 hPa level. It can be seen that the correlation between variable w and the other variables is complex. In Fig. R2, the correlation coefficients between variable w at the point (32.5°N, 112.51°E) and the other model variables at different points show that there are some spurious correlations exist at distant distances.

Considering the non-hydrostatic characteristics of meso-microscale convective systems, the authors plan to conduct non-hydrostatic assimilation experiments for w in which w is expanded as a control variable in our upcoming work. This will involve re-estimating the background error covariance and allow for further investigate the impact of assimilating w in this manner on the update of other model variables.

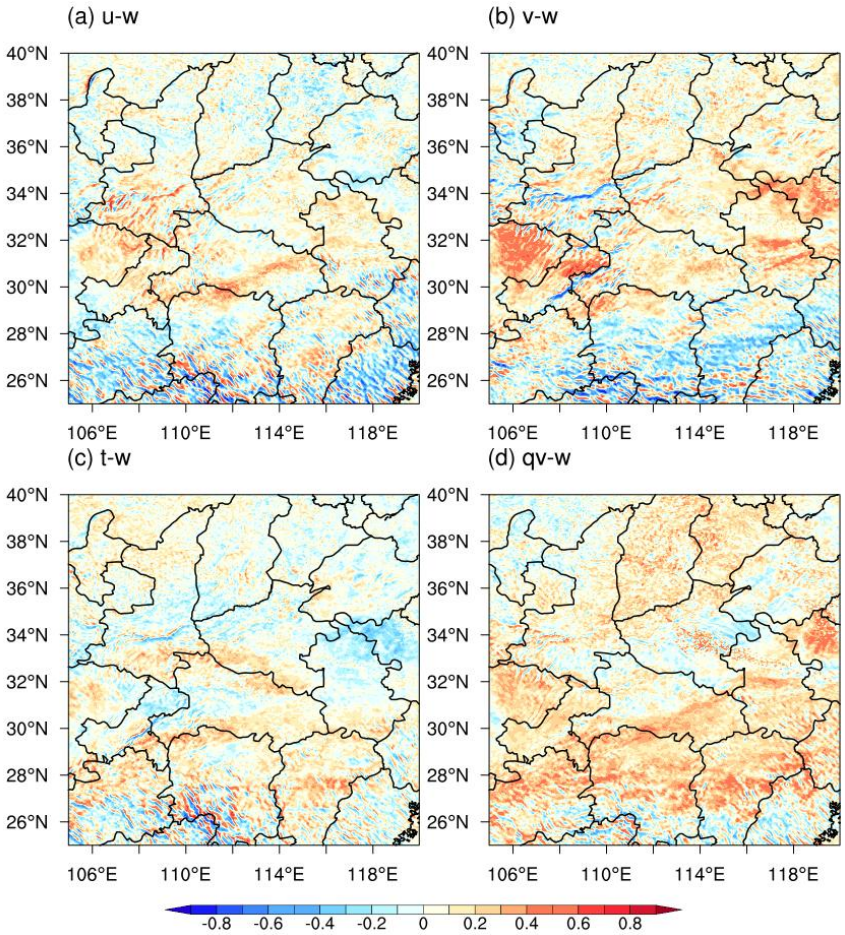

Figure R1 The spatial distribution of correlation coefficients between variable w and (a) zonal wind u, (b) meridional wind v, (c) temperature t, and (d) water vapour

mixing ratio q_v at 700 hPa. These statistics are based on 120 forecast samples.

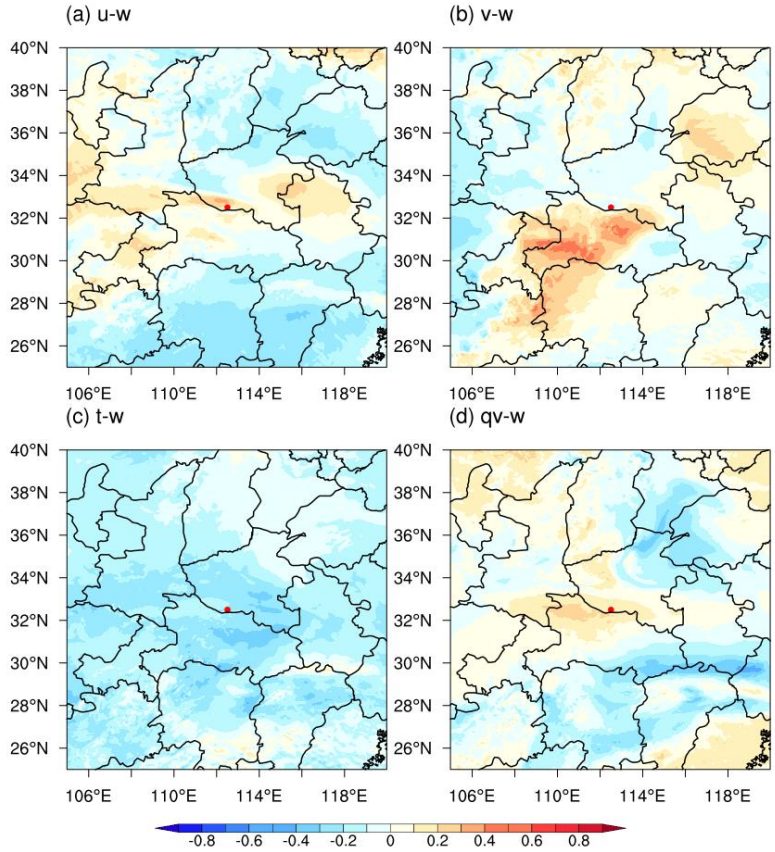

Figure R2 The spatial distribution of correlation coefficients between the time series of variable w at the point (32.5°N, 112.51°E) and the time series of (a) zonal wind u, (b) meridional wind v, (c) temperature t, and (d) water vapour mixing ratio q_v at other model points on the 700 hPa level.

2. In the single observation test, only horizontal increments are presented, while the *w* observation captures vertical motions. It would be more informative to display vertical cross-sections of the increments, revealing the vertical spread of assimilating *w* single observations.

Response: Thank you for your suggestion. Following the reviewer's suggestion, the vertical cross-sections of the analysis increment have been included as Fig. 2 in the revised manuscript. Additionally, the following statement has been added (Lines 148–152 in the revised manuscript): "From the vertical cross-section of the analysis increment for each variable (Fig. 2), it can be seen that the increase in specific humidity is primarily concentrated in the lower layer below the observation location, while the increases in the other three variables are distributed throughout the entire layer. Regarding the increase in horizontal wind u, below the single point observation, there is a convergence of u wind that extends to the ground. Above the single point observation, there is a divergence of u wind that extends to approximately 150 hPa".

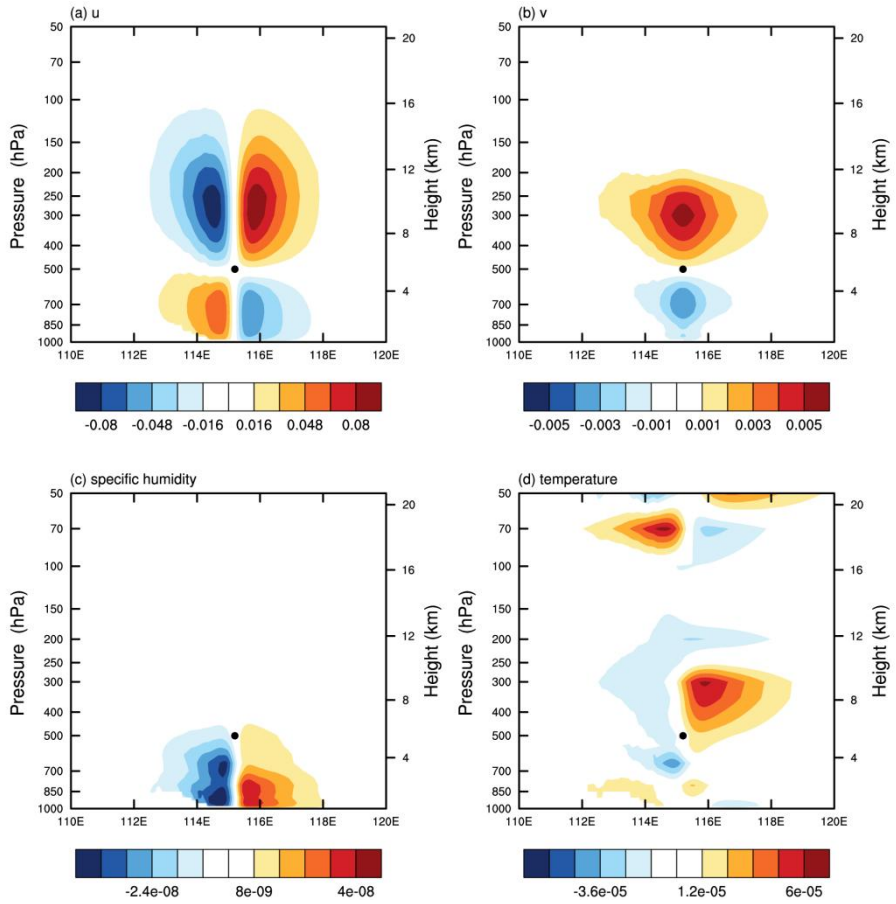

Figure 2 of the revised manuscript. The analysis increments of (a) zonal wind u, (b) meridional wind v (unit: m s^{-1}), (c) specific humidity (unit: kg kg^{-1}), and (d) temperature (unit: K) in a vertical cross-section at 38.0 °N at 1500 UTC on July 4, 2020, for the single observation experiment. The solid black dot in the figure represents the location of the single observation.

> I am particularly interested in understanding the vertical range of assimilating w. Does it extend from the bottom level to the top level? Are there any constraints limiting the impacts of w assimilation?

Response: Yes, the assimilation of w extends from the bottom level to the top level. The current design does not impose any constraints on the impacts of w assimilation in the vertical direction, although it is necessary to set limits, particularly for unreasonable increments at higher levels of the model. However, a specific method for rationalizing the vertical constraint has not been devised yet, except for a crude approach that restricts impacts below a certain model layer. Exploring more reasonable ways to limit the impact of w assimilation in the vertical direction will be part of the author's future work.

> Additionally, based on Fig. 1, the range of significant increments appears to be approximately 8 degrees. Have you considered tuning the decorrelation scales of BEC for *w* assimilation in real case experiments? I believe adjusting these

scales could be crucial for constraining the impact radii of *w* observations in convective-scale data assimilation.

Response: Yes, regarding the optimal decorrelation scales for static background error covariances, the authors conducted several sensitivity experiments for the real case study. Specifically, the length scales of 50 km are employed for horizontal wind components based on the study conducted by Xiao et al. (2020), which utilized the length scales of 48 km for horizontal wind components. However, from Fig. R3, there is a slight difference in the forecast composite radar reflectivity between the control (CTRL) and DA-W (w pseudo-observations are cyclically assimilated with an interval of 1 h from 1300 to 1500 UTC on July 4, 2020) experiments. In the study, a relatively large scale factor of 400 km was used for several reasons. First, the derived w pseudo-observations are mostly small values (shown in Fig. R4 (c)–(c3)), with the majority (90.4%) falling within the range of 0.0-0.6 m/s. Second, the observation operator for w, the adiabatic Richardson equation, combines the continuity equation, adiabatic equations, and the hydrostatic relation, making it more suitable for large-scale systems.

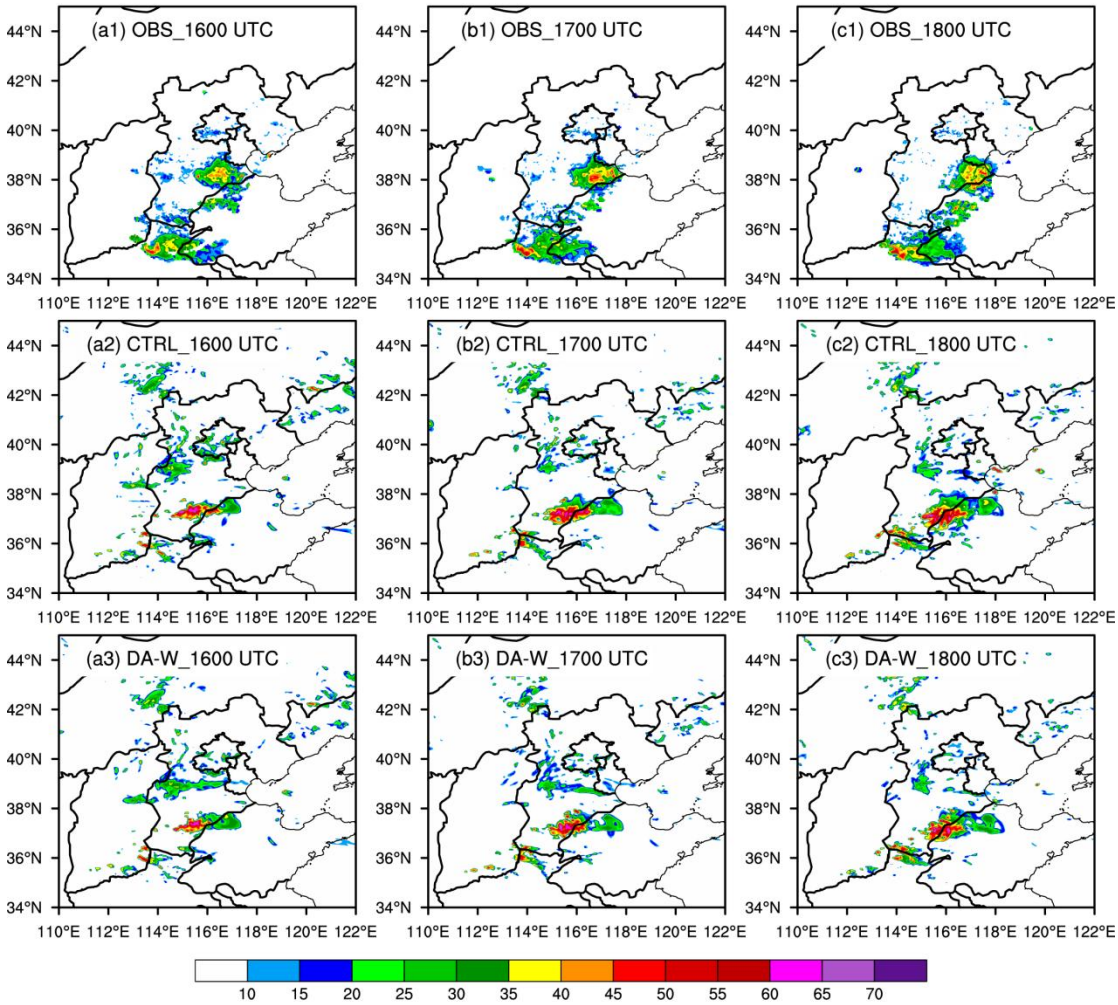

Figure R3 The observed (OBS; (a1)–(c1)) composite radar reflectivity (units: dBZ) and the forecast composite radar reflectivity of the (a2)–(c2) CTRL and (a3)–(c3)

DA-W experiments at (a1)–(a3) 1600 UTC, (b1)–(b3) 1700 UTC, and (c1)–(c3) 1800 UTC on 4 July, 2020.

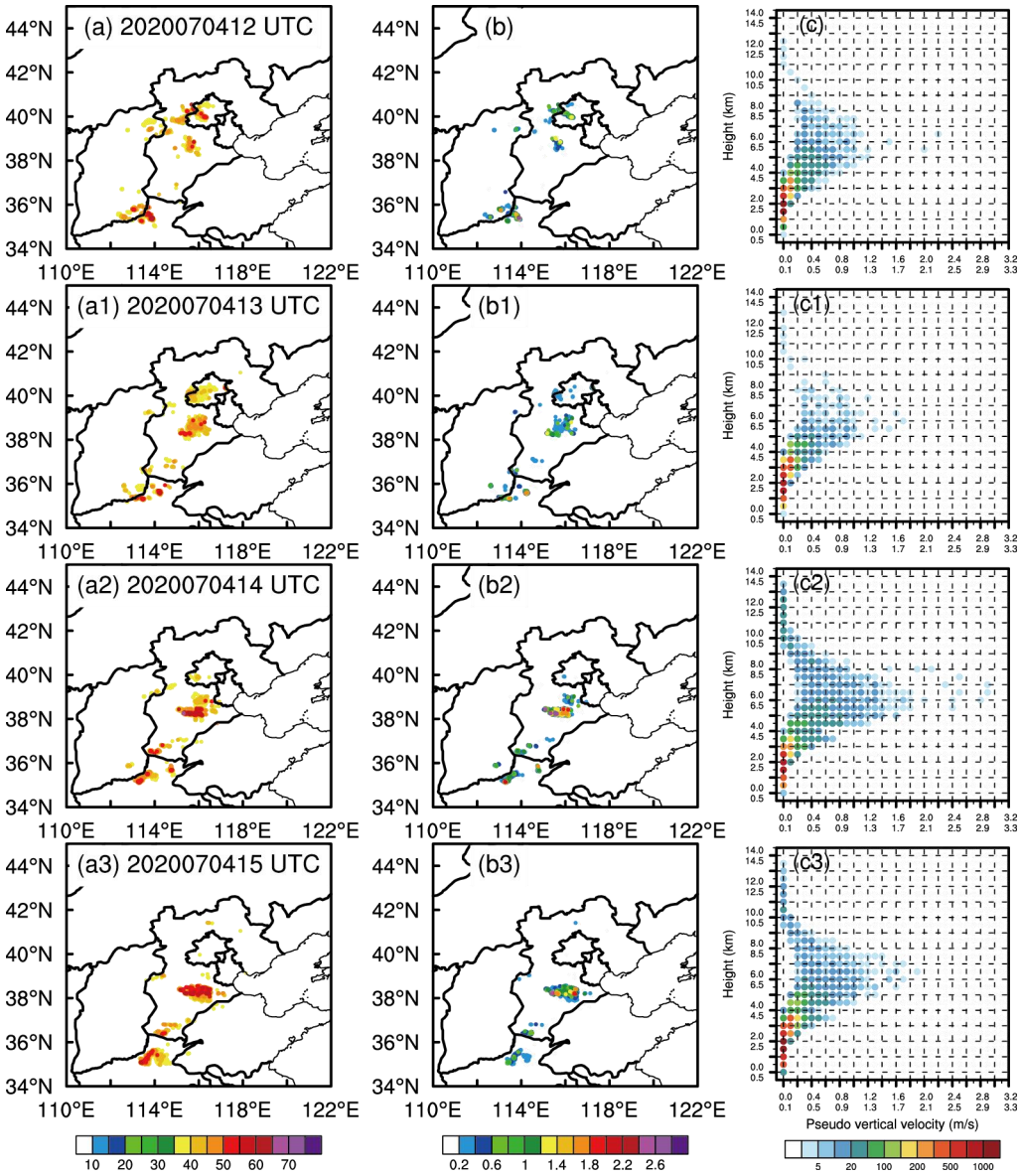

Figure R4 The observed composite radar reflectivity greater than 35 dBZ ((a)–(a3); units: dBZ), the derived maximum w pseudo-observations ((b)–(b3); units: m s^{-1}), and the frequency distribution of w pseudo-observations ((c)–(c3)) at (a)–(c) 1200 UTC, (a1)–(c1) 1300 UTC, (a2)–(c2) 1400 UTC, and (a3)–(c3) 1500 UTC on July 4 2020.

3. The assimilation configurations in both the heavy rainfall case study and batch experiments appear unconventional. In the case study, CNTL performs no assimilation, while DA-W assimilates *w* hourly for only three hours. This leads to DA-W being initialized four times additionally, potentially introducing significant impacts. It's challenging to attribute these impacts solely to the assimilation of *w*. In the batch experiments, both cases seem to involve single-time assimilation each day, likely contributing to minimal

impacts from *w* assimilation. I recommend the authors rerun both batch and single-case experiments, considering an increased number of assimilation cycles per day. For consistency, the single case experiments could from one day of the batch experiments.

Response: Thank you for your valuable comments. Based on your suggestion, we have conducted a new set of batch experiments. In Figure 3 (b) of the revised manuscript, both the CTRL and DA-W experiments were initialized at 0000 UTC daily from July 1 to July 10, 2020, and run until 1200 UTC each day. The first 3 hours were considered as a "spin-up" period. In the CTRL experiments, observations from aircraft measurements, radiosondes, and other sources (for a comprehensive list, refer to Fig. 3 (b)) were assimilated from 0300 to 0600 UTC with a 1-hour assimilation interval (radial velocity observations are available at each analysis time, while other data sources are only available at 0300 and 0600 UTC). The CTRL-1CY experiment indicates assimilation at 0300 UTC only, while the CTRL-2CY experiment represents assimilation at 0300 and 0400 UTC, and so on (the number preceding the experiment name "CY" represents the assimilation iterations). The DA-W experiments are similar to the CTRL experiments, but include the assimilation of *w* pseudo-observations (*w* pseudo-observations are available at each analysis time).

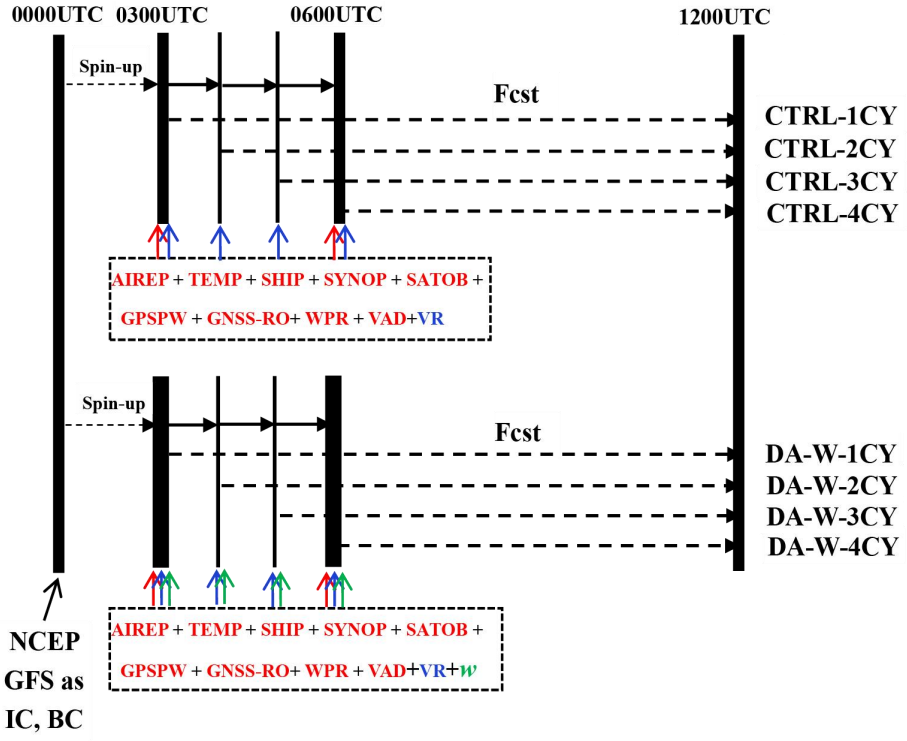

Figure 3 (b) of the revised manuscript. Illustration of the numerical experimental scheme for the CTRL and DA-W batch experiments. Both experiments utilize NCEP GFS data as the initial condition (IC) and boundary condition (BC). The abbreviation "Fcst" represents forecast. The assimilated data comprises conventional observations from aircraft measurements (AIREP), radiosondes (TEMP), ships (SHIP), and ground stations (SYNOP). In addition, cloud-track-wind (SATOB), precipitable water derived from the Global Positioning System (GPSPW), refractivity radio-occultation data

from the Global Navigation Satellite System (GNSS-RO), wind profiler radar (WPR), velocity-azimuth display (VAD) wind, and the radar radial velocity (VR) are assimilated. The pseudo-w data is also assimilated for the DA-W experiments.

Figs. 4-6 in the revised manuscript present the 10-day averaged forecast skills for hourly accumulated precipitation from 0600 UTC to 1200 UTC. For the threshold of 1 mm h^{-1}, it is not always the case that the ETS score improves as the number of assimilation times increases for both CTRL and DA-W experiments. However, with an increase in the scoring threshold, especially for 20 mm h^{-1}, a higher score is generally achieved with more assimilation times, indicating a positive impact of multiple assimilation on the forecast. However, with an increase in the scoring threshold, especially for 20 mm h^{-1}, a higher score is generally achieved with more assimilation times, indicating a positive impact of multiple assimilation on the forecast. When comparing the ETS scores of the CTRL and DA-W experiments with the same assimilation times, it can be seen that the DA-W experiment has a neutral or slightly worse effect on the forecast at threshold of 1 mm h^{-1} compared to the CTRL experiment. However, at thresholds of 5 and 20 mm h^{-1}, the DA-W experiment achieves higher scores than the CTRL experiment in most situations, regardless of multiple or single assimilation. Moreover, the experiment with 3 assimilation times (denoted by experiment names ending with "3CY") demonstrate the most significant improvements compared to experiments with other assimilation times.

The FSS scores provide clearer results: for experiments with the same assimilation times in CTRL and DA-W (e.g., DA-W-2CY compared to CTRL-2CY experiment), the DA-W experiment consistently achieves better scores, indicating that the assimilation of w has a better adjustment effect on the forecast of precipitation location. From the BIAS scores, the DA-W experiments have a neutral impact on the forecast compared to the CTRL experiments. In the first 3-hour forecast, the DA-W experiment generally performs worse than the CTRL experiment (with the same assimilation times) for each threshold value, primarily due to producing more false alarms. However, in the latter 3-hour forecast, the DA-W experiment demonstrates better scores compared to the CTRL experiment.

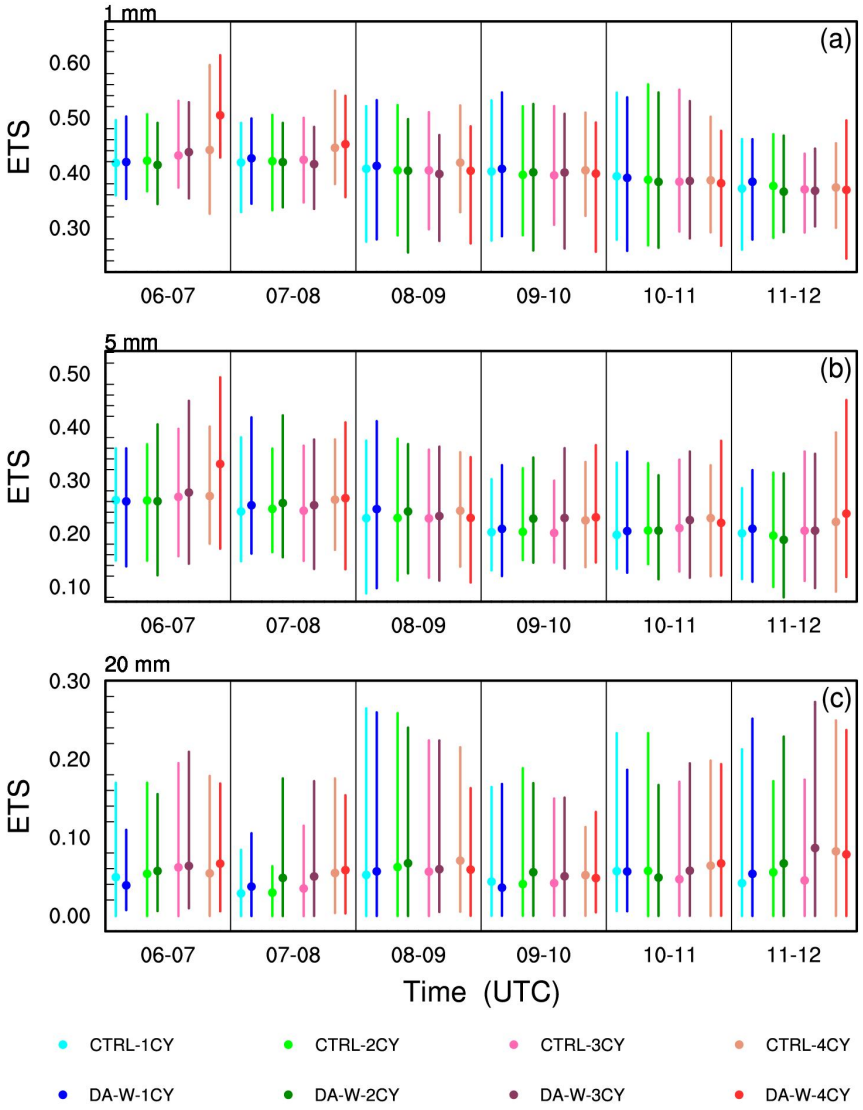

Figure 4 of the revised manuscript. The 10-d (July 1 to July 10, 2020) averaged equitable threat score (ETS; solid dots) of the predicted hourly accumulated precipitation from 0600-1200 UTC of the CTRL and DA-W experiments for thresholds of (a) 1 mm h^{-1}, (b) 5 mm h^{-1}, and (c) 20 mm h^{-1}. The top (bottom) of the line that passes through the solid dot corresponds to the maximum (minimum) ETS value for those 10 days.

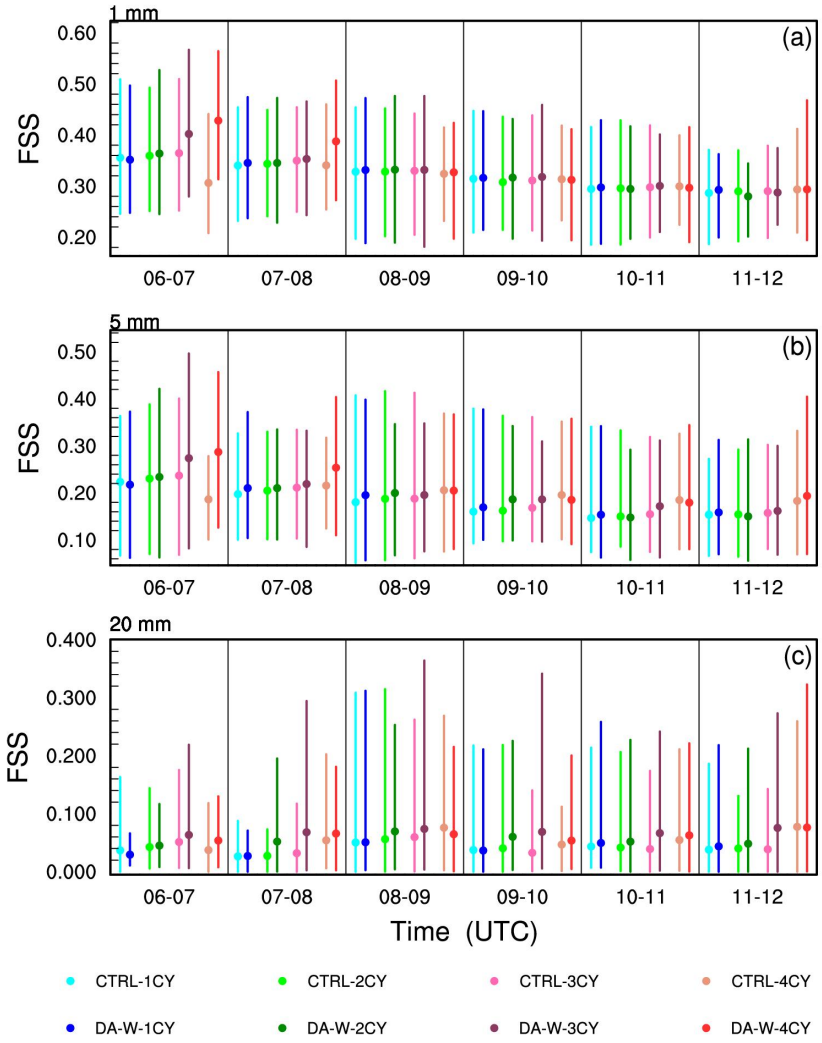

Figure 5 of the revised manuscript. Same as Fig. 4 but for the neighborhood-based fractions skill score (FSS).

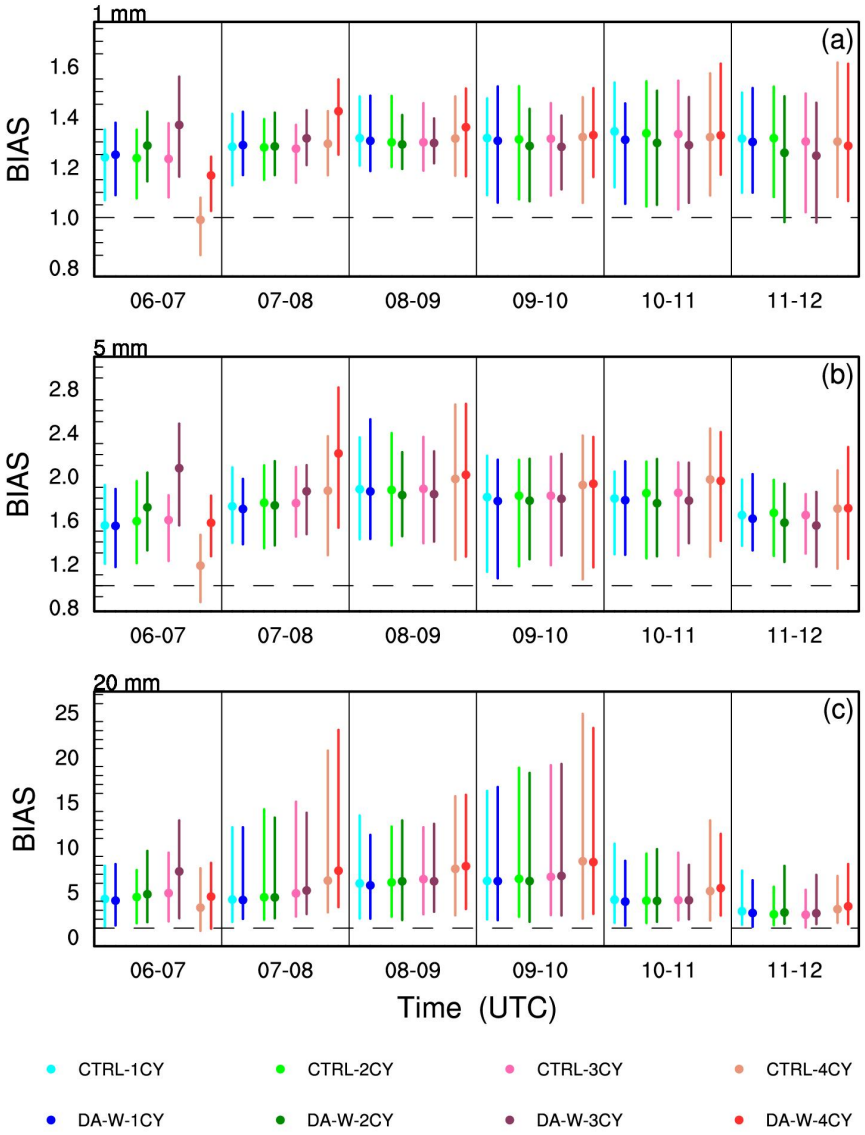

Figure 6 of the revised manuscript. Same as Fig. 4 but for the bias score (BIAS). The black dashed line represents BIAS value equals 1.

Following the reviewer's suggestion, we have replaced the individual test with a case in the batch experiments, specifically the case initialized on July 9, 2020. Fig. 7 presents the ETS, FSS, and BIAS scores for different thresholds. In both the CTRL and DA-W experiments, increasing the assimilation times does not necessarily result in higher ETS scores, particularly for the 1 mm h^{-1} threshold. However, when comparing the CTRL and DA-W experiments with the same assimilation times, the DA-W experiment consistently achieves higher scores. The FSS scores indicate that, except for the period from 0600 to 0700 UTC, the hourly accumulated precipitation exhibits higher scores with more assimilation times, and the DA-W experiment consistently outperforms the CTRL experiment. Regarding the BIAS scores, the DA-W experiment has a neutral effect on the forecast compared to the CTRL experiment.

Fig. 8 displays the 6-hour accumulated precipitation of the CTRL-4CY and DA-W-4CY experiments, with the majority of precipitation occurring in Jiangxi

Province. The heavy precipitation center exhibits a maximum 6-hour accumulated precipitation exceeding 100 mm. The CTRL-4CY experiment successfully captures the forecast location of this heavy rainfall area, although the overall precipitation intensity is low. In contrast, the DA-W-4CY experiment performs better in forecasting the intensity of heavy precipitation.

In Fig. 8 (a), line A-B represents the observed main precipitation belt. Fig. 9 shows the sections along line A-B for the CTRL-4CY and DA-W-4CY experiments at 0700 UTC on July 9, 2020. The DA-W-4CY experiment effectively enhances the w values across the entire model layers, resulting in a negative water vapor increment below the model's 850 hPa compared to the CTRL-4CY experiment. Simultaneously, positive increments of water vapor are observed in the middle and upper layers of the model.

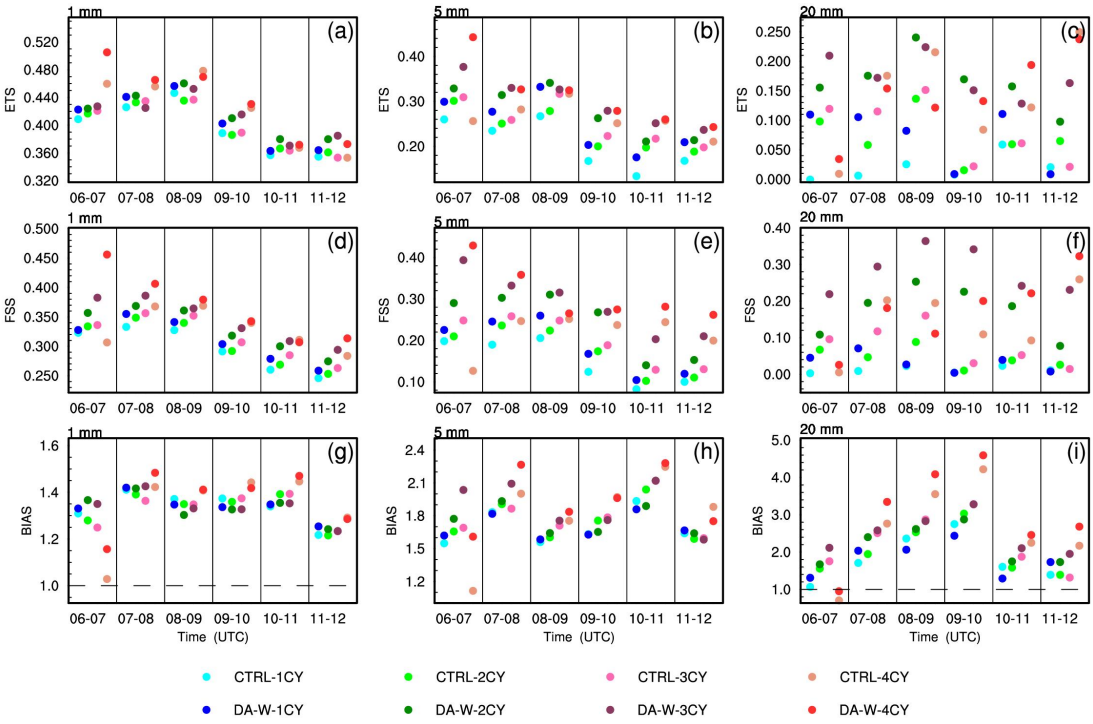

Figure 7 of the revised manuscript. The equitable threat score (ETS; (a)–(c)), the neighborhood-based fractions skill score (FSS; (d)–(f)), and the bias score (BIAS; (g)–(i)) for the predicted hourly accumulated precipitation of the CTRL and DA-W experiments (the black dashed lines in (g) and (i) represent BIAS value equals 1). The analysis focuses on thresholds of 1 mm h^{-1}, 5 mm h^{-1}, and 20 mm h^{-1} for the case initialized at 0000 UTC on July 9, 2020.

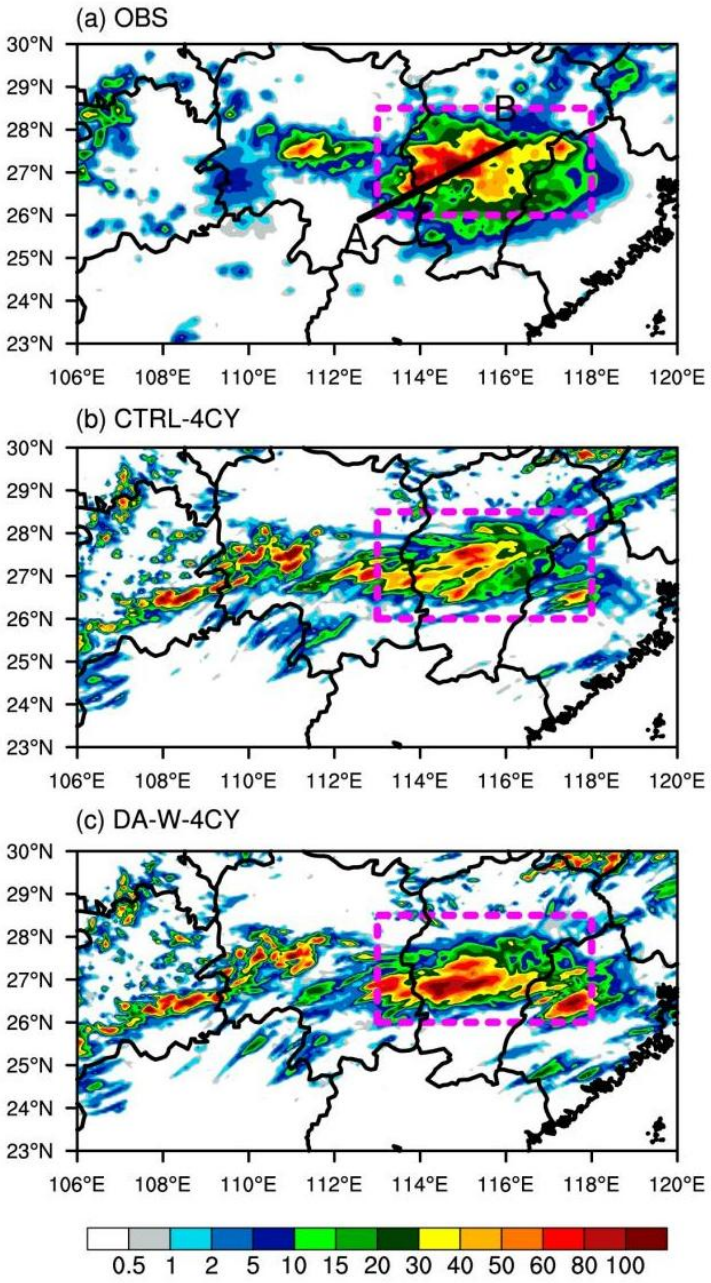

Figure 8 of the revised manuscript. The 6-hour (0600-1200 UTC) accumulated precipitation (units: mm) on July 9, 2020 for (a) observations (OBS), (b) CTRL-4CY, and (c) DA-W-4CY experiments. The areas enclosed by dotted purple lines indicate regions with observed strong rainfall.

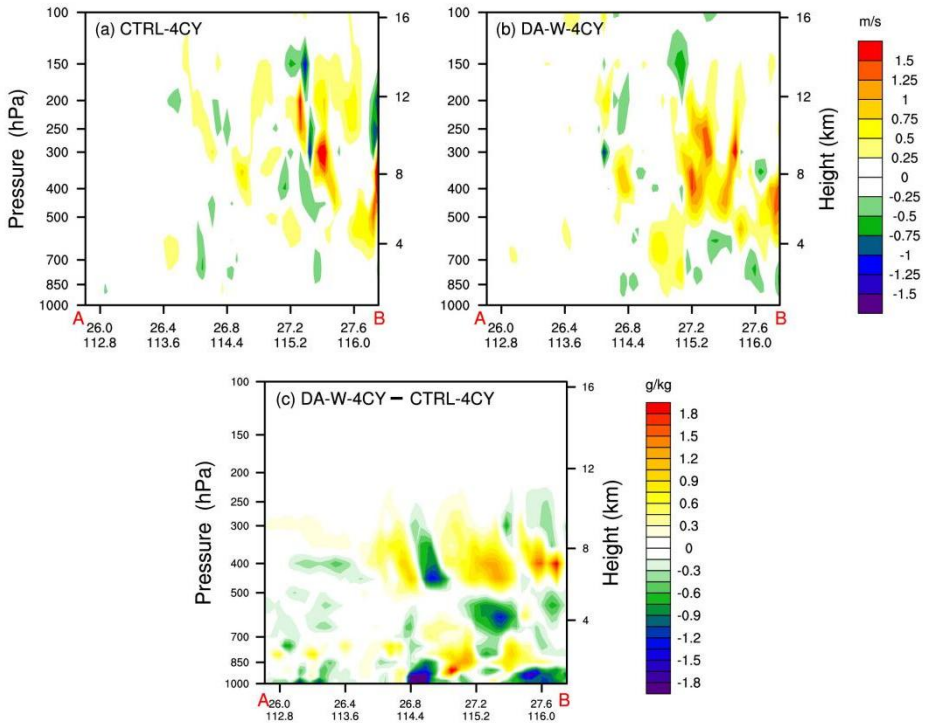

Figure 9 of the revised manuscript. Cross sections of the w (units: m s^{-1}) at 0700 UTC on July 9, 2020, along line A–B in Fig. 8 (a) for the (a) CTRL-4CY and (b) DA-W-4CY experiments. (c) represents the difference in water vapor between the CTRL-4CY and DA-W-4CY experiments (units: g kg^{-1}).

Minor comments:

Line 42: Change "vertical velocity" to "w". Check it throughout the manuscript.

Response: Thank you for your suggestion. We have revised it as the reviewer suggested throughout the manuscript.

Line 54: Change "computing" to "computational".

Response: Thank you for your suggestion. It has been revised as suggested (Line 57 in the revised manuscript).

Line 54: The physical constraint is implicitly considered in the ensemble BEC in the EnKF-based methods.

Response: Yes, we agree. Thank you for your suggestion, and we apologize for our less rigorous statement. The statement "2) lack of strict physical constraints except for the four-dimensional variational method" has been revised to "2) lack of strict physical constraints" (Line 58 in the revised manuscript).

Line 58: the observation operator of w is not the problem. The issue is the effectiveness of assimilating w in the forecasts.

Response: Thank you for your careful consideration. Following the reviewer's suggestion, we have revised the sentence as "Within the 3D-Var framework,

assimilating w faces numerous challenges, the most significant of which is the development of an effective assimilation method that produces a reasonable positive impact on forecasts" to improve its clarity and appropriateness (Lines 62–63 in the revised manuscript).

Line 86: "wind" should be "winds".

Response: Thank you for your suggestion. It has been revised as suggested (Line 86 in the revised manuscript).

Line 91: It is not accuracy. H links the model state variables to the observed variables, not the control variables.

Response: We apologize for this mistake. The "control variables" has been revised to "the model state variables" in the revised manuscript (Line 94).

Line 96-97: This sentence is not accuracy enough. The observation operator combines the dynamic and mass fields, but it cannot adjust these variables. Maybe it could be modified to "enabling the 3DVar method to adjust ….".

Response: Thank you for your suggestion. Yes, our previous statement was not precise enough. According to the reviewer's suggestion, we have revised the sentence to "The Richardson equation combines the continuity equation, adiabatic equations, and hydrostatic relation, which enables the 3D-Var method to adjust the dynamic and mass fields simultaneously and result in a more balanced analysis field." (Lines 99–100 in the revised manuscript).

Line 100: add "height" after "top".

Response: Thank you for your suggestion. Added as suggested (Line 103 in the revised manuscript).

Line 116-119: Make it simple. Maybe the (2) and (3) can be combined.

Response: Thank you for your suggestion. The statement does cause tediousness here, and has been revised to "the tangent linear of the observation operator and its adjoint for the w term are included to calculate the cost function and its gradient values" as the reviewer suggested (Line 119–120 in the revised manuscript).

Line 122: It is common to use bold H to represent the tangent linear observation operator and with the subscript T to represent its adjoint operator.

Response: Thank you for your suggestion. The symbols H' and H^T have been revised to $\mathbf{H}$ and $\mathbf{H}^T$, respectively, to represent the tangent linear observation operator and its adjoint operator throughout the entire revised manuscript.

Eq. (9): The cost function J in Eq. (1) corresponds to analysis state x, while in Eq. (9) it is for the control variable. Modify it to make it consistent.

Response: Following the review's suggestion, the cost function J in Eq. (1) has been revised to $J(c_v)$ to maintain consistency with Eq. (9) (Eq. (10) in the revised manuscript).

Fig. 3: How is the bias score calculated? Is it the frequency bias? For me, the values of bias are too large, indicating a significant overestimation of both experiments.

Response: We appreciate the reviewer's careful attention, and we apologize for any confusion caused. The BIAS score, which represents the frequency bias, is calculated using the following equation:

$$\text{BIAS} = \frac{\text{hits+false alarms}}{\text{hits+misses}}, \quad \text{(R1)}$$

where the indicator "hits" refers the event forecast to occur and did occur; "false alarms" refers to the event forecast to occur, but did not occur; "misses" refers to the event forecast not to occur, but did occur. The specific calculation for each indicator was performed using the neighborhood-based method described in reference Clark et al. (2010). "false alarms" are assigned when an event is forecast at a grid point and not observed within a neighborhood framework of the forecast. However, The author mistakenly treated "false alarms" as when an event is not observed at a grid point and forecast within a neighborhood framework of the forecast. We sincerely apologize for this error, which has now been corrected throughout the entire revised manuscript. Fig. R5 presents the BIAS scores for the case study mentioned in the original manuscript, and the values have become reasonable.

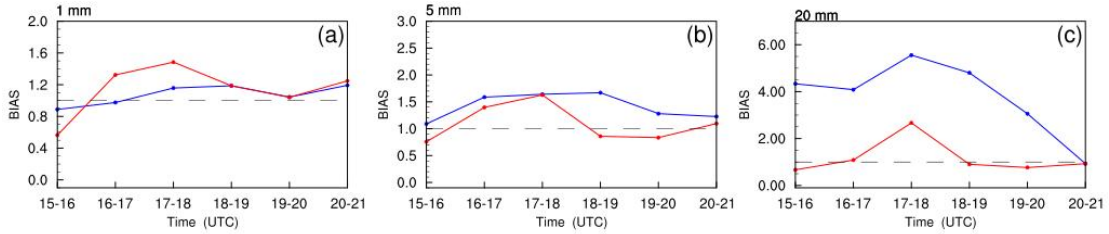

Figure R5 The bias score (BIAS; (a)–(c)) of the predicted hourly accumulated precipitation of the CTRL (the blue solid line) and DA-W (the red solid line) experiments for thresholds of 1 mm h^{-1}, 5 mm h^{-1}, and 20 mm h^{-1} for the heavy rainfall case on 4 July 2020.

Fig. 4: The flowchart is a little bit confusing to me. Is it continuous cycling or partial cycling? Were the assimilations done only at 06Z every day? The description of the flowchart is not clear enough to me.

Response: We apologize for this confusing. It is not a continuous cycling, as assimilation is only performed at 0600 UTC every day. However, we have revised the batch experiments, and the corresponding flowchart has also been updated as Fig. 3 (b) in the revised manuscript.

Fig. 5 and 6: The authors may enhance clarity by presenting the average forecast skills of ETS, FSS, and BIAS over a ten-day cycling period instead of displaying

results for each case. By utilizing ten days' samples, the inclusion of error bars in the figures can provide a more comprehensive representation of variability and uncertainty.

Response: Thank you for your suggestion. We have rerun the batch experiments, and the revised manuscript includes the 10-day (July 1 to July 10, 2020) averaged ETS, FSS, and BIAS scores as suggested by the reviewer (Figs. 4-6). Additionally, error bars have been included in the figures to provide a more comprehensive representation of variability and uncertainty.

We thank the reviewer again for taking the time to review our manuscript.

References

Clark, A. J., Gallus, W. A. Jr., Weisman, M. L.: Neighborhood-based verifcation of precipitation forecasts from convection-allowing NCAR WRF Model simulations and the operational NAM, Wea. Forecasting, 25:1495–1509, doi: 2010WAF2222404.1, 2010.

Chen, Z., Sun, J., Qie, X., Zhang, Y., Ying, Z., Xiao, X., and Cao, D.: A method to update model kinematic states by assimilating satellite-observed total lightning data to improve convective analysis and forecasting, J. Geophys. Res. Atmos., 125(22), 1–26, doi:10.1029/2020jd033330, 2020.

Xiao, X., Sun, J., Qie, X., Ying, Z., Ji, L., Chen, M., and Zhang, L.: Lightning data assimilation scheme in a 4DVAR system and its impact on very short-term convective forecasting, Mon. Weather Rev., 149(2), 353–373, doi:10.1175/mwr-d-19-0396.1, 2021.

Response to Reviewer #2

Reviewer #2

The article presents a 3DVAR assimilation scheme for w, which appears reasonable and has a positive impact based on results from a heavy-rain event and a 10-day batch experiment.

The authors sincerely appreciate the valuable time invested by the reviewer in reviewing our manuscript. We have addressed your queries in the subsequent responses.

However, there are some doubts:

1. From the batch experiment results, it was found that w assimilation has a small impact on the forecast. However, in the case study where w was assimilated in a cyclical manner, it led to significant improvements in the forecast. It is suggested that the authors also set multiple w assimilations in the batch experiment section.

Response: Thank you very much for your comments and professional advice. Based on your suggestion, we have conducted a new set of batch experiments to explore multiple assimilation iterations for vertical velocity (w). In Figure 3 (b) of the revised manuscript, both the CTRL and DA-W experiments were initialized at 0000 UTC daily from July 1 to July 10, 2020, and run until 1200 UTC each day. The first 3 hours were considered as a "spin-up" period. In the CTRL experiments, observations from aircraft measurements, radiosondes, and other sources (for a comprehensive list, refer to Fig. 3 (b)) were assimilated from 0300 to 0600 UTC with a 1-hour assimilation interval (radial velocity observations are available at each analysis time, while other data sources are only available at 0300 and 0600 UTC). The CTRL-1CY experiment indicates assimilation at 0300 UTC only, while the CTRL-2CY experiment represents assimilation at 0300 and 0400 UTC, and so on (the number preceding the experiment name "CY" represents the assimilation iterations). The DA-W experiments are similar to the CTRL experiments, but include the assimilation of w pseudo-observations (w pseudo-observations are available at each analysis time).

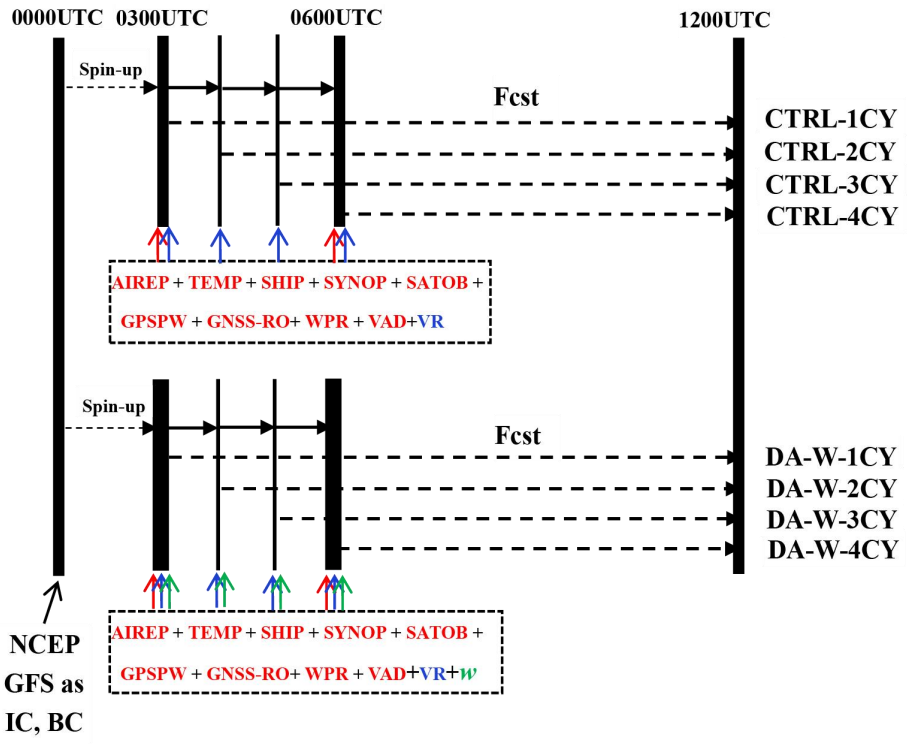

Figure 3 (b) of the revised manuscript. Illustration of the numerical experimental scheme for the CTRL and DA-W batch experiments. Both experiments utilize NCEP GFS data as the initial condition (IC) and boundary condition (BC). The abbreviation "Fcst" represents forecast. The assimilated data comprises conventional observations from aircraft measurements (AIREP), radiosondes (TEMP), ships (SHIP), and ground stations (SYNOP). In addition, cloud-track-wind (SATOB), precipitable water derived from the Global Positioning System (GPSPW), refractivity radio-occultation data from the Global Navigation Satellite System (GNSS-RO), wind profiler radar (WPR), velocity-azimuth display (VAD) wind, and the radar radial velocity (VR) are assimilated. The pseudo-w data is also assimilated for the DA-W experiments.

Figs. 4-6 in the revised manuscript present the 10-day averaged forecast skills for hourly accumulated precipitation from 0600 UTC to 1200 UTC. For the threshold of 1 mm h^{-1}, it is not always the case that the ETS score improves as the number of assimilation times increases for both CTRL and DA-W experiments. However, with an increase in the scoring threshold, especially for 20 mm h^{-1}, a higher score is generally achieved with more assimilation times, indicating a positive impact of multiple assimilation on the forecast. However, with an increase in the scoring threshold, especially for 20 mm h^{-1}, a higher score is generally achieved with more assimilation times, indicating a positive impact of multiple assimilation on the forecast. When comparing the ETS scores of the CTRL and DA-W experiments with the same assimilation times, it can be seen that the DA-W experiment has a neutral or slightly worse effect on the forecast at threshold of 1 mm h^{-1} compared to the CTRL experiment. However, at thresholds of 5 and 20 mm h^{-1}, the DA-W experiment achieves higher scores than the CTRL experiment in most situations, regardless of multiple or single assimilation. Moreover, the experiment with 3 assimilation times (denoted by experiment names ending with "3CY") demonstrate the most significant

improvements compared to experiments with other assimilation times.

The FSS scores provide clearer results: for experiments with the same assimilation times in CTRL and DA-W, the DA-W experiment consistently achieves better scores, indicating that the assimilation of w has a better adjustment effect on the forecast of precipitation location. From the BIAS scores, the DA-W experiments have a neutral impact on the forecast compared to the CTRL experiments. In the first 3-hour forecast, the DA-W experiment generally performs worse than the CTRL experiment (with the same assimilation times) for each threshold value, primarily due to producing more false alarms. However, in the latter 3-hour forecast, the DA-W experiment demonstrates better scores compared to the CTRL experiment.

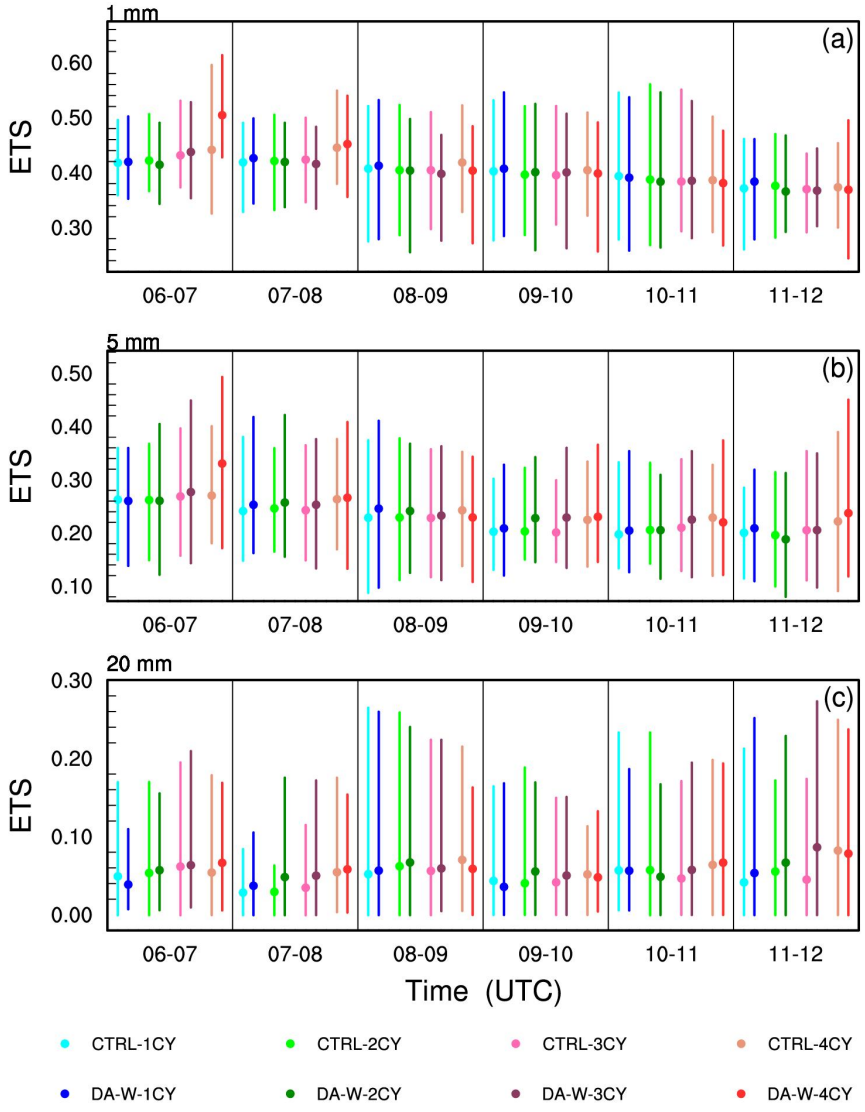

Figure 4 of the revised manuscript. The 10-d (July 1 to July 10, 2020) averaged equitable threat score (ETS; solid dots) of the predicted hourly accumulated precipitation from 0600-1200 UTC of the CTRL and DA-W experiments for thresholds of (a) 1 mm h^{-1}, (b) 5 mm h^{-1}, and (c) 20 mm h^{-1}. The top (bottom) of the line that passes through the solid dot corresponds to the maximum (minimum) ETS value for those 10 days.

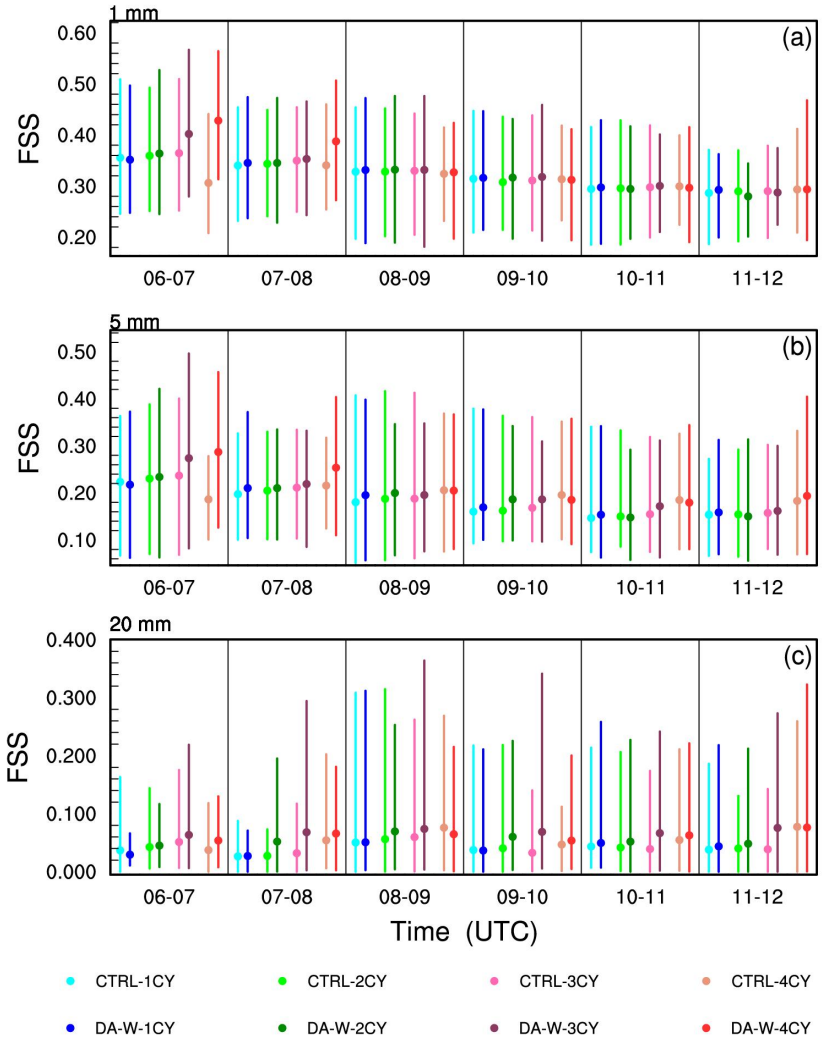

Figure 5 of the revised manuscript. Same as Fig. 4 but for the neighborhood-based fractions skill score (FSS).

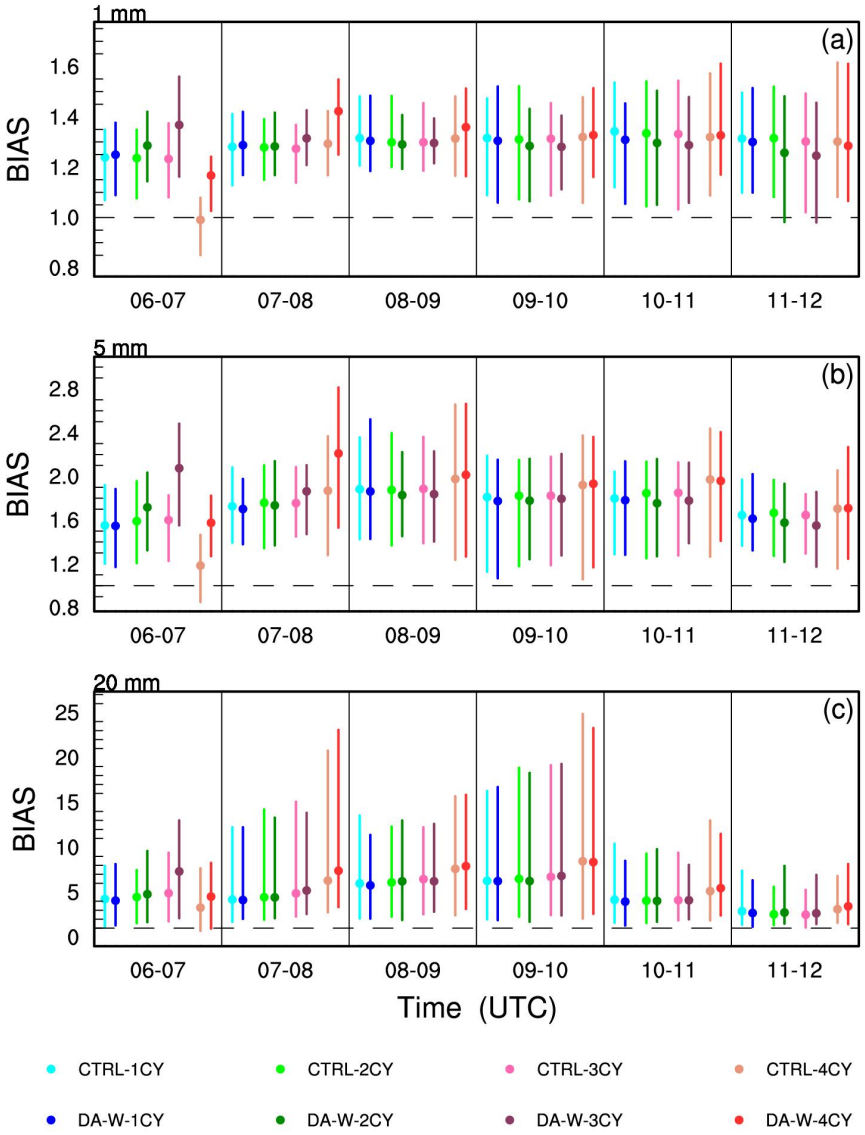

Figure 6 of the revised manuscript. Same as Fig. 4 but for the bias score (BIAS). The black dashed line represents BIAS value equals 1.

2. As mentioned by the author, some studies assimilate w from total lightning data. In this article, however, the w observations are derived from radar reflectivity data. It is unclear how the authors obtained at the approximate magnitude of the w values using this method.

Response: Yes, in this study, the pseudo-w observations are derived from radar reflectivity data. The authors conducted statistical analysis of the pseudo-w observations. From the frequency distribution of the pseudo-w observations in different value and height bins (Fig. R6 (a)–(d)), it is observed that the majority (90.4%) of pseudo-w observations have small values, ranging from 0.0 to 0.6 m s^{-1}.

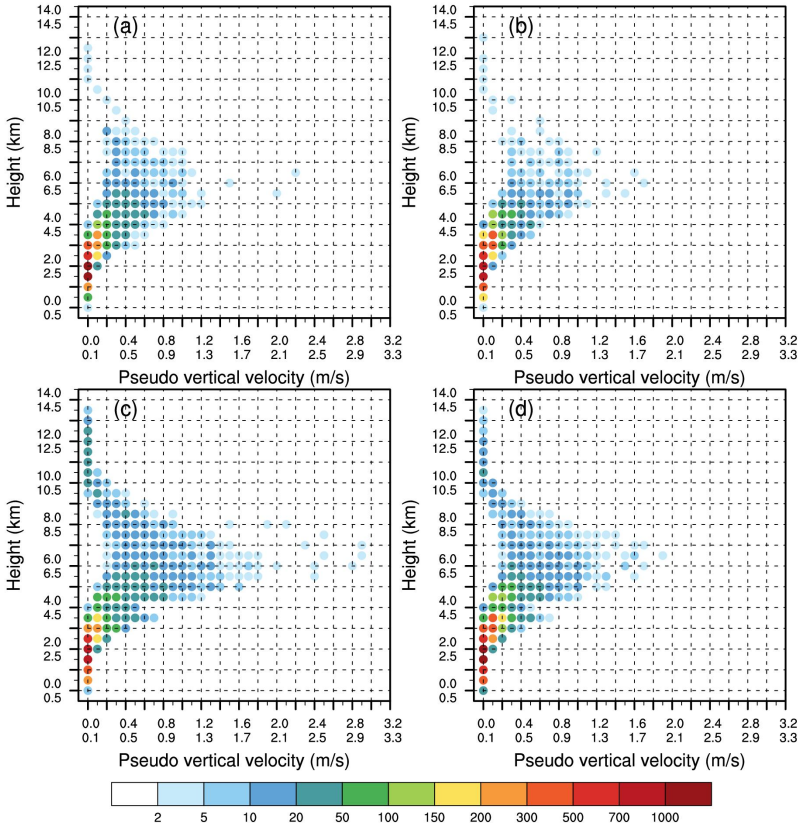

Figure R6 The frequency distribution of pseudo-w observations at (a) 1200 UTC, (b) 1300 UTC, (c) 1400 UTC, and (d) 1500 UTC on July 4, 2020.

In addition, were the authors able to compare the radar reflectivity-derived w with the w of the model background field to determine any differences in magnitude between the two values? If there is a significant difference, it may be necessary to remove the larger w values during the assimilation process.

Response: We appreciate the valuable suggestion from the reviewer. Fig. R7 presents the mean w profiles (blue solid lines) calculated from the background at different analysis times for the case studied in our manuscript (please refer to the caption of Fig. R7 for specific area boundaries. The statistics are calculated based on grid points where the composite radar reflectivity exceeds 35 dBZ). The shape of the profile shows good agreement with the results of Yuter and Houze (1995). In the convection area, the mean w increases from low values at lower levels to a peak value occurring at middle to upper levels (approximately 6-8 km), and then decreases at higher altitudes. The maximum mean w of the background is about 0.5 m s^{-1} at most analysis times, except for 1500 UTC, which is almost 1.0 m s^{-1}. These values are close in magnitude to the pseudo-w observation values, of which 90.4% are smaller than 0.6 m s^{-1}.

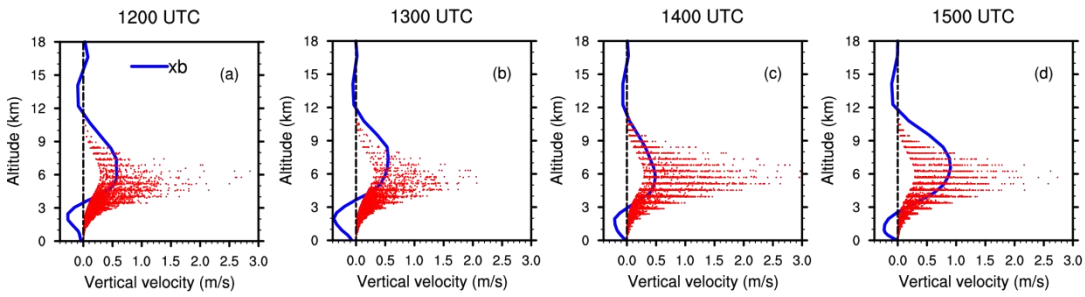

Fig. R7 The distribution of pseudo-w observations (red dots) retrieved from radar reflectivity data at different analysis times on July 4, 2020. The blue solid lines represent the averaged w profiles calculated from the background field (xb), and the averaging is performed over the area of (35–44°N, 113–119°E).

Some minor revisions are as follows:

Page 1:

Line 21: Change "the result indicates" to "the results indicate".

Response: Thank you for your suggestion. This sentence has been removed in the revised manuscript.

Line 22: The statement "leading to improved equitable threat score (frequency skill score) for the first 1 h (3 h) precipitation forecasts" may cause confusion, please describe it in detail.

Response: Thank you for your suggestion. Since we reran the experiments, this sentence has been removed in the revised manuscript.

Line 23: Change "assimilated" to "assimilation".

Response: Thank you for your suggestion. This sentence has been removed in the revised manuscript.

Page 2:

Line 36: Change "allows they" to "allows them".

Response: Thank you for your suggestion. Revised as suggested (Line 40 in the revised manuscript).

Line 41-42: Delete "of vertical velocity".

Response: Thank you for your suggestion. Deleted as suggested (Line 45 in the revised manuscript).

Line 48: Add "s" to the word "field".

Response: Thank you for your suggestion. It has been revised as suggested (Line 52 in the revised manuscript).

Page 3:

Line 75: Delete "real".

Response: Thank you for your suggestion. This sentence has been removed in the

revised manuscript.

Lines 89-91: This sentence is quite difficult to understand. I suggest that it be described simply and clearly.
Response: We apologize for any confusion caused. We have revised the related text to "The observation operator H is used to derive the equivalent of the observations from model state variables" in order to improve clarity. (Line 94 in the revised manuscript).

Line 92: Delete "to assimilate w observation directly".
Response: Thank you for your suggestion. Deleted as suggested (Line 95 in the revised manuscript).

Line 96: Add "s" to the word "adjust".
Response: Thank you for your suggestion. The word "adjust" has been revised to "enables" (Line 99 in the revised manuscript).

Page 14：
Line 279: Change "wish" to "wishes".
Response: Thank you for your suggestion. It has been revised as suggested (Line 315 in the revised manuscript).

We would like to thank the reviewer for taking the time to review our manuscript.

References

Yuter, S. E. and Houze, R. A.: Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus: Part II. Frequency distribution of vertical velocity, reflectivity, and the differential reflectivity, Mon. Weather Rev., 123, 1941–1963, doi:10.1175/1520-0493(1995)123<1941:TDKAME>2.0.CO;2, 1995.