Responses to Referee #2

We thank you for the careful review and constructive comments, which have helped us improve the manuscript. We have revised the paper accordingly and provide point-by-point responses below. All changes in the revised manuscript are highlighted.

Major comments:

1) Introduction and conclusions

Comment 1:

Section 1 (Introduction): The paper would benefit substantially from stating explicit research questions or objectives at the end of the introduction. This would help readers understand what to expect and what progress is being claimed. For example: What is the impact of assimilating visible reflectance/radiances in the GSI-EnKF-CRTM configuration? Please also briefly indicate the approach/methods used to address these questions (OSSE design, cycling strategy, evaluation metrics).

Our response:

Thank you for your valuable suggestion. We have revised the introduction by explicitly stating the research questions and briefly outlining the experimental approach at the end of Section 1 (Page 3, lines 97-110). The added text clarifies that the study aims to evaluate (1) the correction of cloud fields (especially CWP), (2) indirect effects on non-cloud variables, and (3) improvements in precipitation forecasts. It also summarizes the OSSE design, including single-observation and hourly cycled assimilation experiments, and the key metrics used for evaluation (MB, MAE, RMSE, ETS, FAR). This addition provides a clearer roadmap for the readers.

Comment 2:

Section 4 (Conclusions): Section 4 requires substantial revision to address the open issues raised in this review. Please avoid speculation and ensure the conclusions directly answer the research questions/goals proposed in the introduction. The conclusions would also benefit from a clearer discussion of implications of the findings, including (if appropriate) limiting steps toward operational application (computational cost, bias handling, etc.).

Our response:

Thank you for your constructive feedback regarding the conclusions section. In response, we have thoroughly revised Section 4 to provide a more direct, evidence-based conclusions that addresses the points raised in your review.

The revised conclusions are now explicitly structured around the three research questions posed in the Introduction. Each subsection directly answers one question: (1) how VIS reflectance DA corrects cloud fields, (2) its limited direct but potential indirect impact on non-cloud variables, and (3) its positive effect on precipitation forecasts. We have removed speculative statements and ensured all conclusions are firmly grounded in the presented results (Page 26).

As requested, we have also added a clearer discussion of the implications and operational considerations. The final paragraph now explicitly notes the high computational cost of multi-physics ensembles, the need for bias-correction when using real observations, and the importance of testing across diverse weather regimes, which are key limiting steps toward any future operational application (Page 26, lines 547-549).

Comment 3:

Line 492: The meaning of "intrinsically" is unclear in context; please rephrase.

Our response:

We appreciate your insightful comment on the ambiguity of "intrinsically". To accurately reflect the scientific definition of CWP and eliminate ambiguity, we revised the sentence to: "Since CWP is defined as the vertical integral of the sum of the liquid water mixing ratio (Ql) and the ice water mixing ratio (Qi), the effective adjustment of the CWP vertical distribution indicates a consequent optimization of Ql and Qi" (Page 26, lines 524-526)

Comment 4:

Lines 493–494: To me, it's unclear why the reported sign changes occur and what the implications are. Please add an explanation and link explicitly to the DA update mechanism and the observation operator sensitivity.

Our response:

Thank you for your valuable comment. In response, we have added a clear explanation in the revised manuscript (Page 26, lines 519-522) that directly addresses why the innovation δ_{REF} and the CWP analysis increment δ_{CWP} exhibit consistent sign changes. The added text explains that a positive δ_{REF} indicates the first guess underestimates the cloud field, and through the EnSRF update, the system corrects this by increasing the analyzed cloud hydrometeor content, leading to δ_{CWP} >0 and vice versa for a negative innovation.

2) VIS DA results and their implications

Comment 1:

Section 2.1 / Lines 145–156: I disagree with the wording that cloud-related control variables "must" be included. VIS observations can be assimilated to update only conventional variables (e.g., temperature/humidity) via ensemble covariances or adjoint sensitivities; cloud/hydrometeor control variables are optional, though they may increase impact substantially. Please revise the text accordingly.

Our response:

Thank you for raising this important point. We agree that assimilating VIS observations to update only conventional variables via ensemble covariances is also a viable approach. In revising the manuscript (Page 5, lines 151-153), we have deleted the absolute term 'must' to clarify that our extension of the control variables to include cloud hydrometeors is a deliberate choice for this study. This design allows the VIS reflectance DA to directly and efficiently act on cloud-related variables, thereby better achieving our research objective of quantifying and correcting cloud-field errors.

Comment 2:

Line 327 "VIS reflectance DA has a limited effect on non-cloud variables": Based on the presented results, I am not convinced that VIS DA has a limited or negligible effect on non-cloud variables. Explanation of why the stated behavior occurs is missing. Discuss whether this is consistent with prior studies. This should be explicitly revisited in the conclusions/discussion and placed in the context of existing literature on VIS assimilation.

Our response:

Thank you for this important comment regarding the impact of VIS DA on non-cloud variables. We have revised the manuscript to address the three key points you raised.

First, in the discussion (Section 3.1), we have now explicitly explained why the direct impact is limited. We state that the VIS reflectance is primarily sensitive to cloud optical properties, leading to a fundamental difference in the VIS observation operator's sensitivity to these two classes of state variables. This results in a negligible Kalman gain for non-cloud variables, preventing their direct effective constraint. (Page 16, lines 349-352)

Second, we have placed this finding in the context of existing literature. Immediately following the physical explanation, we now cite prior VIS assimilation studies (Scheck et al., 2020; Zhou et al., 2022) that report a similarly impact, confirming that our result is consistent with the established understanding in the field. (Page 16, lines 356-358)

Third, as requested, this point has been explicitly revisited in the revised Conclusions (Section 4). The finding is summarized and directly linked to the physical mechanism and prior studies with the statement: "This finding aligns with prior VIS DA studies (Scheck et al., 2020; Zhou et al., 2022)." (Page 26, line 535)

Comment 3:

Line 332 "for non-cloud variables approximates zero ($\boldsymbol{H}$n≈0).": This statement appears to contradict Figure 6, where temperature and moisture do show some adjustments. Please reconcile the text with the evidence shown in the figure (and clarify whether the plotted profiles represent a member or the ensemble mean).

Our response:

Thank you for this important comment. While the observation operator ***H*** for non-cloud variables is indeed close to zero (***H***n≈0), the ensemble-based assimilation system may still produce negligible adjustments in these variables due to sampling errors arising from the limited ensemble size (40 members). These sampling errors can introduce spurious background error covariances between cloud and non-cloud variables, leading to very minor and spatially isolated increments in temperature and moisture. However, as shown in the updated Figure 6, there are no adjustments for temperature and moisture.

Regarding the figure, the vertical profiles in Fig. 6 are plotted for the ensemble mean, as indicated in the figure caption. (Page 16, line 361)

Comment 4:

Lines 379–390 and Figure 9: Figure 9 suggests notable systematic differences between the nature run and the model climatology/distributions. This raises an important DA question: to what extent is the DA correcting systematic biases versus random/displacement errors? Classical DA assumes unbiased errors; if systematic differences prevail, they can affect innovations and lead to suboptimal updates. Please discuss this explicitly and summarize the implications.

Our response:

We thank the reviewer for raising this critical point. In the revised manuscript, we have updated the experimental setup: the nature run is now generated by running WRF driven by ERA5 reanalysis data (Page 10, lines 237-242). As shown in the updated Figure 9, the systematic differences between the nature run and the model distributions are substantially reduced compared to the previous version. To further quantify this, we computed the observation-minus-background (O-B) statistics from our cycling experiments. As shown in the new Figure 1, the O-B distribution has a mean (μ) of 0.023 and a standard deviation (σ) of 0.268, and closely follows a Gaussian distribution. This confirms that the innovations entering the DA system are approximately unbiased, satisfying the classical DA assumption.

Importantly, despite the absence of large systematic biases between the nature run and the model, the VIS reflectance DA continues to produce clear improvements in cloud analyses, particularly for thin clouds (reflectance <0.2, CWP <0.1 $kg \cdot m^{-2}$), and consistently reduces CWP errors throughout the cycling window. The conclusions remain consistent with our previous findings. This demonstrates that the effectiveness of VIS DA is not limited to correcting systematic biases; it also successfully reduces random/displacement errors and improves cloud field structure even when the background and nature run are already in close agreement (Page 20).

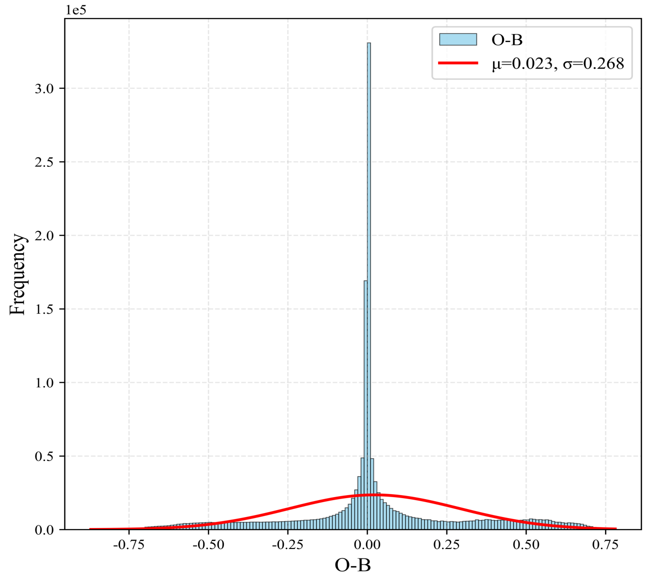

Figure 1: The PDF of O - B reflectance

Comment 5:

Lines 388–389 (thin-cloud conclusion): The conclusion that VIS DA mainly corrects thin clouds may depend on evaluating the ensemble mean. Please clarify whether this conclusion holds at the member level. How would the conclusion change if you analyzed each member's increments or departures? Relatedly, how might results differ if the OSSE did not exhibit strong systematic differences between the nature run and the ensemble?

Our response:

Thank you for these insightful questions regarding our thin-cloud correction conclusion.

We have clarified in the manuscript that the PDF analysis presented in Figure 9 is based on a single ensemble member, confirming that the identified effect is evident at the member level and is not an artifact of ensemble averaging (Page 19, line 410).

To further address your concern about the influence of systematic differences between the nature run and the ensemble, we have revised the OSSE setup by generating the nature run with

ERA5-driven WRF simulations, which substantially reduces the systematic biases compared to the original configuration. As shown in the new Figure 1, the observation-minus-background (O-B) distribution has a mean of only 0.023 and closely follows a Gaussian distribution, indicating that systematic differences are now negligible. Despite this change, the VIS reflectance DA continues to produce clear improvements, particularly for thin clouds (reflectance <0.2, CWP <0.1 kg·m^{-2}), and consistently reduces CWP errors throughout the cycling window. This demonstrates that the effectiveness of VIS DA is not limited to correcting large systematic biases; it also successfully reduces random errors and improves cloud analyses even when the background and nature run are already in close agreement (Page 20).

Comment 6:

Line 395 onward: As noted above, this section would benefit from an explicit evaluation/discussion of the extent to which improvements arise from correcting biases (systematic) versus random errors (e.g., displacement or wrong CWP). Please consider adding suitable diagnostics/metrics, and at a minimum discuss this limitation.

Our response:

We thank the reviewer for this insightful comment. In response, we have added RMSE (Root Mean Square Error) as an additional diagnostic metric to complement the MAE analysis, as RMSE is sensitive to large deviations and provides insight into the spread of errors, while MAE reflects the average error magnitude, together they help infer whether improvements arise from reducing systematic biases or random errors. The updated results show that at 08:00 UTC, EXP achieved maximum instantaneous reductions of 6.8% in MAE and 17.8% in RMSE, with mean reductions of 2.5% in MAE and 2.2% in RMSE over the entire period. The larger relative improvement in RMSE at the peak suggests that the assimilation effectively mitigates larger deviations or outliers, potentially associated with random errors such as cloud displacement or inaccurate CWP. However, we acknowledge that a complete separation of systematic bias correction versus random error reduction remains challenging with the current diagnostic framework, and we have now explicitly discussed this limitation in the revised manuscript (Page 21, lines 426-428).

Comment 7:

Lines 398–399: The manuscript primarily shows impacts during the assimilation window. Please clarify and show (or discuss) how long-lasting the impact is with the forecast lead time. Have forecasts been verified in addition to the analysis states?

Our response:

Thank you for this important question regarding the forecast impact longevity. In the revised manuscript, we have addressed this by presenting the temporal evolution of CWP MAE and RMSE throughout the entire experimental period, which includes both the assimilation window (02:00–08:00 UTC) and the subsequent forecast period (08:00–02:00 UTC next day). This is shown in the updated Figure 10 (Page 21).

As illustrated in Figure 10, the positive impact of VIS reflectance DA on CWP forecasts persists throughout the 18-hour forecast period, with EXP consistently outperforming CTRL. Notably, the maximum error reduction occurs at 13:00 UTC on 21 September, with MAE and RMSE. These results demonstrate that the assimilation benefits are not confined to the analysis states but extend well into the forecast, indicating robust and lasting improvements in cloud field predictions.

Comment 8:

Line 427: If the statement is correct and supported by results, it raises the question of the persistence of the imposed cloud increments. If temperature/humidity are not adjusted consistently toward saturation, a cloud increment introduced by VIS DA could quickly evaporate after forecast initialization. Please discuss model balance/consistency and the expected persistence of cloud increments.

Our response:

We thank the reviewer for this important question regarding the persistence of cloud increments and the physical consistency of the analysis. In the revised manuscript, we have updated Figure 11 (formerly Figure 12) to show the mean bias (MB) profiles at 04:00 UTC, which is two hours into the forecast initialized from the 02:00 UTC analysis. As shown in the figure, compared to the CTRL, the vertically averaged MB (dashed lines) for U, V, and T is closer to zero in the analysis and first guess, while the impact on Q remains neutral. These results demonstrate that, although the direct adjustment of non-cloud variables during assimilation is negligible, the improved cloud fields introduced by VIS DA exert a weakpositive influence on temperature and wind during the early forecast stage.

Regarding model balance and consistency: the EnSRF algorithm updates cloud variables using flow-dependent background error covariances, which implicitly account for multivariate correlations. As a result, the adjusted cloud fields are statistically consistent with the temperature and humidity fields, minimizing initial imbalances. Subsequent model integration further adjusts these fields through physical parameterizations, ensuring that the cloud increments remain physically plausible and do not rapidly evaporate.

We have also removed the ambiguous statement at the original Line 427 that may have caused confusion, and we have revised the corresponding discussion to reflect these findings (Page 21, lines 438-452). We thank the reviewer for prompting this clarification, which has strengthened the interpretation of our results.

3) OSSE setup and synthetic observation generation

Comment 1:

Section 2.2.3: It is not sufficiently clear how the synthetic observations were generated. Please add a concise but complete description: what fields from the nature run (ERA5) were used, how CRTM was driven, whether observation noise was added, and how representativeness was handled.

Our response:

We thank the reviewer for raising this point. In response, we have added a concise description in Section 2.2.3 to clarify the generation of synthetic observations (Page 12, Lines 256-260). The added text explains that the atmospheric variable profiles, surface parameters, and hydrometeor contents were extracted from the nature run, interpolated to the model grid via cubic spline interpolation to minimize representativeness errors arising from spatial mismatch, and then used to compute cloud water path and effective particle radii via Eqs. (1)-(6) before driving the CRTM forward operator. It also explicitly states that no observation noise was added to the synthetic reflectance values.

Comment 2:

Line 260 (observation error choice): The observation error specification is critical for any OSSE. The discussion is currently too limited. Please expand: how does the chosen observation error compare to (i) ensemble spread/background error and (ii) typical departure statistics? Did you consider estimating errors using O–B/O–A statistics? Were sensitivity tests performed (even limited)? A spread–error comparison would be very informative and would clarify the relative weighting between background and observations.

Our response:

Thank you for raising this important point. We agree that the specification of observation errors is critical, and we have expanded the discussion accordingly.

In the revised experiments, the observation error for visible reflectance is no longer set to an arbitrary static value (e.g., 0.1). Instead, we objectively estimated it using the diagnostic covariance method proposed by Desroziers et al. (2005), i.e., $(O-A)(O-B)^{\mathrm{T}}$. This method provides a dynamically consistent estimate by leveraging the innovation (O-B) and analysis residual (O-A) statistics (Page 13, lines 284-289). The resulting O-B distribution (typical departure statistics, shown in Fig. 1) has a mean of 0.023 and a standard deviation of 0.268. The Desroziers diagnostic yielded an optimal observation error estimate of 0.1887, which was adopted in all subsequent experiments. This value is physically reasonable, as it falls between the mean O-B (indicative of bias) and the standard deviation of O-B (indicative of random error). We did not perform extensive sensitivity tests on observation errors, as the Desroziers method provides a theoretically sound and objective estimate.

We also compared this estimated error to the ensemble spread (background error) in observation space. The ensemble spread was approximately 0.149, which is slightly smaller than the estimated observation error (0.189). This indicates that the background uncertainty is on average somewhat lower than the observation uncertainty, yielding a balanced weighting with a slight tendency to trust the background more in the assimilation system.

Comment 3:

Figure 9 (comparability/resolution): How comparable are ERA5/nature run "truth"/synthetic observations to model/ensemble outputs in terms of effective resolution? Would it be appropriate to super-ob to a common resolution for fair comparison? Please comment and/or justify the chosen verification approach.

Our response:

We thank the reviewer for raising this important point regarding resolution comparability. To address this concern, the verification reference used in the revised manuscript has been updated (Page 10, lines 237-242) Instead of directly interpolating ERA5 data onto the model grid, we now generate the ruth" by running WRF driven by ERA5 (hereafter referred to as ERA5-WRF). This approach ensures that the verification reference and the model forecasts (both EXP and CTRL) share identical grids, terrain following coordinates, thereby eliminating discrepancies in effective resolution. Consequently, the point-to-point comparison is already equitable, and super-obbing to a common resolution is not necessary. Therefore, the chosen verification approach is adequate to support the conclusions regarding the relative benefit of data assimilation.

4) Use of "significant" and "scientific significance."

Comment 1:

Line 252 (and other occurrences): "significant" is not appropriate unless a statistical test is performed. Please replace with alternatives such as "substantial," "notable," or "pronounced," depending on meaning.

Our response:

Thank you for your thorough review and valuable suggestion regarding the use of the term "significant." We fully agree that its use should be reserved for contexts where statistical tests have been performed. In accordance with your recommendation, we have systematically revised all instances where "significant" was used descriptively (rather than statistically) throughout the manuscript, replacing it with more appropriate terms such as "substantial," "notable," "pronounced," or "considerable," depending on the intended meaning.

We have retained the term "significant" in Section 3.2.3 (Page 23, "Impact on precipitation"), as well as in the related conclusions and the Abstract, on the grounds that the results that are supported by statistical testing. We believe the revision enhance the precision and academic rigor of the manuscript. Thank you again for your insightful and constructive feedback.

Comment 2:

Line 306: The phrase "scientific significance" is not appropriate here; no statistical evidence is provided. Please reword and frame conclusions as case-study findings.

Our response:

We appreciate your comment regarding the use of "scientific significance". In response, we have revised the sentence to frame the finding. The revised text (Page 15, lines 330-331) now reads: "This observational finding provides strong empirical support for the key conclusion discussed in Sect. 2.1 regarding experimental setup and observational sensitivity." This adjustment ensures the conclusion is presented as a case-study finding rather than a claim of scientific significance. Thank you for your valuable suggestion.

Comment 3:

Lines 328 and 375: Same issue, please revise terminology accordingly.

Our response:

Thank you for pointing out the specific instances in Lines 328 and 375. We confirm that these occurrences have been addressed as part of our systematic revision in response to Comment 1. Specifically, "significant" in Line 328 has been replaced with "substantial," and in Line 375 it has been changed to "notable." Both modifications are included in the comprehensive list provided in our response to Comment 1 above.

Minor comments:

Comment 1:

Lines 86–94: This paragraph lacks a clear lead sentence; as a reader, it is unclear why it is introduced here. Consider adding a first sentence explaining why perturbations/multi-physics are used and then reorganizing (potentially moving lines 95–102 earlier).

Our response:

Thank you for this suggestion. We have revised the paragraph by adding a clear lead sentence at the beginning (Page 3, lines 87-90). This sentence now states that the performance of ensemble-based DA depends on the ensemble spread, explains its role in representing forecast uncertainty and controlling state adjustments, and thereby naturally introduces the discussion of the two primary methods for generating this spread. This restructuring has improved the logical flow of this paper.

Comment 2:

Figure 2: Consider using consistent y-axis limits (0–1) in both panels to better highlight solar/viewing-angle impacts on reflectance. It may also help to relate this figure to similar sensitivity illustrations in the literature (e.g., see Geiss et al. 2021 / https://doi.org/10.5194/acp-21-12273-2021)

Our response:

Thank you for your valuable suggestion regarding Figure 2 (Page 8). Following your recommendation, we have revised the figure by applying a consistent y-axis limit of 0–1 to both panels, which better highlights the impact of solar/viewing geometry on reflectance. These adjustments enhance the clarity and comparability of the figure. We appreciate your constructive feedback.

Comment 3:

Section 2.2.1: Please discuss computational cost: how expensive is CRTM VIS simulation, and is it fast enough for operational application (or what developments are needed)?

Our response:

We thank the reviewer for raising the important point regarding computational cost. In response, we have added a discussion in Section 2.2.1 (Page 8, lines 210-214) that quantitatively assesses the runtime of the CRTM forward operator. The assessment confirms that while the computational cost of the current serial implementation is non-negligible and would require optimization for operational application, it is fully acceptable for the scope and objectives of this OSSE-based feasibility study.

Comment 4:

Figure 3 (and other figures): Please use colorblind-friendly colormaps. Avoid using the same colormap within a multi-panel figure when panels show different quantities; this can be confusing. For Figure 3(a), maximum reflectances appear below ~0.6 for a tropical cyclone; please explain whether this is expected given geometry/cloud properties or indicates a limitation in the simulation.

Our response:

We sincerely thank the reviewer for their valuable comments. In response, we have carefully revised all figures throughout the manuscript to ensure they are colorblind-friendly. Specifically, Figures 3, 4, 6, 7, 8, 9, 12, 13, and 15 now employ colorblind-friendly colormaps. In multi-panel figures that display different physical quantities, we have used distinct colormaps for each quantity to avoid any potential confusion.

Regarding the maximum reflectance of below ~0.6 in the tropical cyclone region (Fig. 3b), we have added an observed satellite image (Fig. 3a) for the corresponding time, which shows that the maximum observed reflectance is similarly around 0.6, confirming that our simulation is consistent with the actual observation. At the same time, this value is physically expected given the observing geometry. The simulation corresponds to 00:00 UTC, when the solar zenith angle over the typhoon area is large (55°–65°). As demonstrated in the sensitivity analysis in Fig. 2b, such a large solar zenith angle substantially reduces the top-of-atmosphere reflectance even for optically thick clouds. Figure 2b shows that for a solar zenith angle of 60°, the maximum reflectance for various cloud phases and particle sizes remains around 0.5. Hence, the simulated reflectance is consistent with the theoretical expectation and reflects the actual conditions under oblique solar illumination, rather than a limitation in simulation (Page 9).

Comment 5:

Table 1 / ensemble: Since the ensemble/OSSE was constructed specifically for this study, it would be important to demonstrate that the ensemble is not severely under- or over-dispersive. Please consider adding diagnostics such as:

- spread vs. error over time (for temperature and/or reflectance/CWP),
- rank histograms (or similar calibration diagnostics),
- or at minimum discuss whether dispersion was evaluated.

Our response:

We thank the reviewer for this valuable suggestion regarding ensemble dispersion diagnostics. In the revised manuscript, the analysis of the ensemble spread evolution addresses this point. Specifically, Figure 8 now presents the standard deviation (i.e., spread) of the ensemble for both the observed variable (reflectance) and the state variable (CWP) at the DA times (02, 04, 06, 08 UTC).

As discussed in the text (Page 19, line397), regions with larger spread correspond well with areas where the analysis increments are more pronounced (Figure 7), which is consistent with the theoretical expectation of the ensemble Kalman filter. Furthermore, the spread remains nearly constant over the assimilation cycles, indicating that the ensemble does not exhibit excessive collapse or explosive growth.

Comment 6:

Section 2.2.3 (H operator): Please clarify whether is treated as a linear or nonlinear operator in the EnSRF context. Since the VIS operator is nonlinear, explain whether any linearization is assumed and how this affects the results.

Our response:

Thank you for your insightful comment. In ensemble Kalman filter (EnKF) and its variants (e.g., EnSRF, LETKF, EAKF), the observation operator is indeed linearized when computing the Kalman gain. Specifically, these methods approximate the nonlinear observation operator by a linear approximation to update the state. This linearization introduces analysis errors because the visible observation operator is strongly nonlinear, especially in cloudy and precipitating regions. Consequently, the linear approximation cannot accurately capture the true nonlinear mapping between visible observations and model state variables, leading to suboptimal updates and potential degradation in the analysis of cloud-related fields (Page 12).

Comment 7:

Line 257: Please verify the assumption that single-observation experiments are independent. Depending on the definition of localization radius (e.g., Gaspari–Cohn half-width vs cutoff), the cutoff distance can exceed the nominal radius.

Our response:

Thank you for raising this point. In our single-observation experiments, observations were selected at every fifth grid point, corresponding to a horizontal spacing of 75 km. The localization radius was set to 15 km. The effective influence range of this Gaspari–Cohn function is typically twice the radius (i.e., about 30 km), meaning the cutoff distance is approximately 30 km. Since the observation spacing (75 km) is considerably larger than this cutoff distance, the assimilation influence regions of individual observations do not overlap. Therefore, these single-observation experiments can be considered mutually independent (Page 13, line 283).

Comment 8:

Lines 263–278: Please rewrite for clarity; the motivation is understandable after careful reading, but the current phrasing is difficult to follow.

Our response:

Thank you for your suggestion to improve the clarity of the experimental design description. We have rewritten the paragraph concerning the cycled DA time window (Page 13, lines 292-299) to present the rationale in a more logical and direct manner. The revised text now clearly states the overall daytime window available for VIS reflectance (00:00–08:00 UTC), explains the decision to use a refined assimilation period (02:00–08:00 UTC). We believe this restructuring makes the experimental setup and its justification clearer for the reader.

Comment 9:

Figure 6 (analysis diagnostics): Consider whether it would be helpful to show mean and standard deviation of first-guess and analysis departures vs height over all samples to demonstrate whether both biases and errors are reduced by VIS DA. Also clarify in the caption whether the figure shows a single random profile, or whether it is an ensemble mean or a selected member. Please also comment on the representativeness of ERA5 vs model profiles.

Our response:

Thank you for your valuable comments on Figure 6 (analysis diagnostics). We have carefully considered your suggestions and provide the following detailed responses:

Regarding the suggestion to show mean and standard deviation of first-guess and analysis departures versus height over all samples, we agree that such statistics would provide a comprehensive view of bias and error reduction. However, given that this study conducts single-assimilation experiments, we consider that analyzing a representative random profile is sufficient to demonstrate the diagnostic capability of the VIS DA system.

As revised in the figure caption, we have explicitly stated that the profile in Figure 6 represents the ensemble mean of the assimilation experiment.

Following your comment, we have optimized the experimental design of the Nature Run: instead of directly using ERA5 reanalysis data, we now drive the WRF model with ERA5 to generate the Nature Run (detailed parameter settings are provided in Page 10, lines 237-242 of the revised manuscript). We acknowledge that ERA5, as a reanalysis product, may have limitations in fully representing the true atmospheric state. However, it is important to note that this study adopts an OSSE framework, where the Nature Run serves as the "theoretical truth" to quantify the relative improvement effect of VIS DA. The primary objective of our analysis is therefore to evaluate the relative improvement achieved by the VIS DA system, comparing the analysis against the first guess within this closed experimental setting, rather than to perform an absolute validation against real-world observations.

Comment 10:

Figure 7: Use a diverging colormap for increments centered at zero (white at zero, red/blue for positive/negative). Consider consistent vmin/vmax across the increment panels to facilitate time-to-time comparison.

Our response:

Thank you for your suggestion regarding the visualization of increments in Figure 7. Following your recommendation, we have updated the CWP increment panels (the third column, a3, b3, c3, d3). Specifically, we have (1) applied a diverging colormap with white centered at zero, red for positive increments, and blue for negative increments; and (2) implemented consistent, symmetric limits across all four increment subplots to enable direct visual comparison of the adjustment magnitudes throughout the assimilation cycles (Page 18).

Comment 11:

Line 372: The phrase “mismatched spatial patterns” is unclear. Please explain what is meant and why it is attributed to nonlinear model physics rather than, for example, sampling error or representativeness differences.

Our response:

Thank you for your valuable comment regarding the unclear attribution of the “mismatched spatial patterns.” In direct response, we have revised the corresponding paragraph (Page 19, lines 397-403) to provide a clearer explanation. The text now explicitly states that the observed anomaly, where significant increments are occasionally absent in regions of high CWP STD, occurs because a large spread in the state variable does not guarantee a proportionally large spread in the simulated observation. We have expanded the attribution to include three key factors: (1) the highly nonlinear relationship between CWP and reflectance (which is primary and depends on cloud microphysics and viewing geometry); (2) the finite sampling error to 40-member ensemble; (3) the representativeness limitation of using CWP to characterize reflectance signal. This revision addresses your point by moving from a single, vague cause to a multi-factor explanation that acknowledges the roles of nonlinear physics, sampling, and representativeness.

Comment 12:

Figure 8: Use consistent color scaling across panels (b1–b4) to allow visual comparison.

Our response:

Thank you for your suggestion regarding Figure 8. We have updated the reflectance standard deviation panels (b1–b4) by applying a consistent color scale across all four subplots. This allows for a direct visual comparison of the spatial distribution and magnitude of ensemble spread in simulated reflectance across the different DA times (Page 19).

Comment 13:

Figure 10 (optional): For future studies, consider spatial/ensemble verification (e.g., Brier Score, Fractions Skill Score) to better interpret probabilistic/spatial precipitation performance.

Our response:

We thank the reviewer for this constructive suggestion. We agree that metrics like the Brier Score and Fractions Skill Score would provide a more comprehensive evaluation of probabilistic and spatial precipitation performance. As the current study focuses on developing and initially validating the GSI-EnKF-CRTM-Vis technique's impact on cloud and precipitation variables, our chosen verification framework aligns with this primary goal. We acknowledge the value of the suggested advanced verification and will explicitly include it as a direction for future work.

Comment 14:

Lines 420–421: The conclusion of negligible impact on non-cloud variables may depend on using mean profiles. Consider whether evaluating member-wise departures (first guess vs analysis) and then averaging could reveal a clearer signal.

Our response:

We thank you for this thoughtful suggestion. In our current diagnostics for Figure 11(Page 22), we have already implemented the recommended approach: we first compute member-wise departures (analysis minus first guess) for each of the 40 ensemble members, and then average these departures to obtain the mean bias profiles shown in the figure.

It is worth noting that for linear quantities such as bias, this member-wise then average calculation is mathematically equivalent to first averaging the analysis and first guess fields separately and then computing their difference (i.e., mean(analysis - first guess) = mean(analysis) - mean(first guess)). Therefore, our results are robust regardless of the order of operations. The nearly identical vertical profiles between analysis and first guess across all non-cloud variables confirm that assimilating VIS reflectance has a negligible direct impact on these fields, consistent with single observation experiments.

Comment 15:

Lines 433–434: Please clarify how this conclusion is supported by the presented results.

Our response:

Thank you for your insightful comment. We have carefully reconsidered the explanation and revised the corresponding text to more clearly link the conclusion to the presented results. We have revised the experimental setup : instead of using ERA5 directly as the nature run in an OSSE framework. we now use ERA5 to drive WRF and generate simulated observations. We removed the earlier discussion about OSSE and model differences. This made the final sentence (' Future studies should validate simulated non-cloud variables against real observations. This will better reveal how VIS reflectance DA actually affects these variables.') abrupt, so we have deleted it as well. The conclusion is now more concise and directly supported by the results presented.

Comment 16:

Figure 12: Panels are difficult to read at the current size. Consider focusing on a subset (e.g., one time) and providing larger panels. It could also help to show analysis/first-guess departures (both mean and standard deviation) rather than only mean bias.

Our response:

We thank the reviewer for this helpful suggestion. In response, we have revised Figure 12 (now numbered as figure 11) to focus on a single time (04:00 UTC) and enlarged the panels to improve readability. Additionally, following the reviewer's recommendation, we now present both the mean bias (MB) and the standard deviation (STD) , rather than showing only the mean bias across four times. The updated figure provides a clearer and more complete assessment of the assimilation impact on non-cloud variables. Corresponding modifications have been made to the figure caption and the related discussion in the text (Page 21, lines 438-452).

Comment 17:

Figures 13 and 15: Clarify whether panels (c) show the ensemble mean or a selected member. Ensure consistent color scaling or colorbars between Figures 13 and 15. Also consider changing the lower boundary when displaying light precipitation. For me, the threshold of (<10/30 mm) for non-white colors is not optimal, as white incorrectly suggests "no precipitation." Better use, e.g., 1mm for both figures.

Our response:

Thank you for your valuable suggestions regarding Figures 13 and 15 (now numbered as figures 12 and 14). In response, we have made the following revisions:

First, on Page 23, line 464 of the manuscript, we have added the explicit statement "Based on the 40-member ensemble mean" to clarify that the precipitation fields shown in panels (b) and (c) represent the ensemble-mean analysis.

Second, the lower display threshold for precipitation has been raised from 0 mm to 1 mm so that the white background no longer incorrectly suggests "no precipitation."

Third, since Figure 12 shows a 7-hour accumulation (during the assimilation window) while Figure 14 shows a 24-hour accumulation, we retained slightly different value ranges to preserve the spatial detail characteristic of each period. A completely uniform color scale would obscure important details in the short-term accumulated precipitation. All figures in the manuscript now employ colorblind-friendly palettes.

Comment 18:

Figure 14: As mentioned above, spatial verification metrics would be valuable in future work.

Our response:

We thank the reviewer for highlighting this point. We agree that spatial verification metrics, such as the skill score mentioned in relation to Figure 14 (now numbered as figure 13), would offer a more nuanced assessment of spatial precipitation patterns. In this study, ETS and FAR served as our primary metrics for evaluating precipitation forecast performance, which met the core objectives of the current work. We will consider incorporating spatial verification approaches in future, more comprehensive evaluations.

Language/Formatting comments:

Comment 1:

Line 251: Remove the word “observation” (duplicate/awkward phrasing).

Corrected

Comment 2:

Figure 8 caption: “of of” → “of”.

Corrected