Responses to Referee #1

We thank you for your time in reviewing our manuscript and for providing these insightful and constructive comments. Your valuable feedback is crucial for enhancing the quality and clarity of our work. We have carefully considered all the points raised and have revised the manuscript accordingly. Below, we provide a point-by-point response to your specific comments.

Comment 1:

How to generate the synthetic observations? How to determine the observation errors?

Our response:

Thank you for this question. The synthetic observation (visible reflectance) was generated using the Community Radiative Transfer Model (CRTM v2.1.3). The required atmospheric profiles and surface parameters were derived from the Nature Run. The essential cloud parameters (specifically, the cloud water path and effective particle radius) were calculated from the Nature Run's hydrometeor profiles using the formulas explicitly provided in Section 2.2.1 (Eqs. 1–6). We added the relevant description in the revised manuscript (Page 12, lines 256-260).

We modified the experimental design, and now the observation error is estimated by the following description:

The observation error of reflectance was estimated using the diagnostic method $(O-A)(O-B)^{\mathrm{T}}$proposed by Desroziers et al. (2005). Specifically, a pre-trial assimilation run was first conducted, in which the observation error for visible reflectance was set to the instrument noise level of 0.05. Based on the analysis fields from this pre-trial run, the reflectance observation error was estimated by computing the covariance between the observation-minus-background $(O-B)$ and observation-minus-analysis $(O-A)$ differences. The method yielded an estimated observation error of 0.1887, which is adopted in this study. We added the relevant description in the revised manuscript (Page 13, lines 284-289).

Comment 2:

Figure 3, can the simulated visible image be compared to a realistically observed one?

Our response:

We thank the referee for this valuable suggestion. We agree that comparing the simulated imagery with a real observation provides a more direct and convincing validation of the CRTM forward operator's performance under realistic conditions. Following your advice, we have updated Figure 3 by adding a panel for the real Himawari-8 AHI observed visible image at the corresponding time. Additionally, we have incorporated a corresponding analysis and discussion in the manuscript text (Page 8, lines 205-209).

Comment 3:

Are the ensemble members with different spin up time have the same target time?

Our response:

Thank you for this question. Yes, all ensemble members share the identical target time of 02:00 UTC on September 21, 2024, which serves as the initial time for the assimilation experiment (Page 9, line 229).

Comment 4:

I would assume that single-observation experiment assimilates one observation. But it seems like that single-observation experiment assimilates more than one observation? Moreover, 15-km horizontal localization seems too tight, and an observation error of 0.1 is really small? How about the observation-minus-background distributions? How to determine the observation error objectively?

Our response:

1. On the “single-observation” experiment design

The term "single-observation experiment" in this context refers to a series of independent single-observation experiments. As described in Section 2.3, 196 points were sampled, but for each point's analysis, only the observation at that specific location with a tight localization to prevent influence from other points (Page 13, line 281).

2.About the 15-km localization radius

The 15-km localization radius was specifically chosen for the single-observation experiments

to ensure that each observation, when assimilated at its target grid point, would not produce analysis increments that influenced the results at other sample points. Given that our sampling strategy selected points with a 75-km interval (as described in Section 2.3), this tight localization radius (significantly smaller than the sampling distance) guaranteed that each assimilation was spatially independent (Page 13, line 283).

3.The observation error and the O-B distributions

The observation error is no longer set to a static value of 0.1. In the revised experiments, the reflectance observation error is estimated objectively using the diagnostic covariance method proposed by Desroziers et al. (2005), i.e. $(O-A)(O-B)^{\mathrm{T}}$ (Page 13, line 285).

The obtained O-B distribution (shown in Fig. 1) has a mean of 0.023 and a standard deviation of 0.268. The Desroziers diagnostic yielded an optimal error estimate of 0.1887, which is the value adopted in all subsequent experiments. This approach provides a dynamically consistent, objective estimate of the observation error.

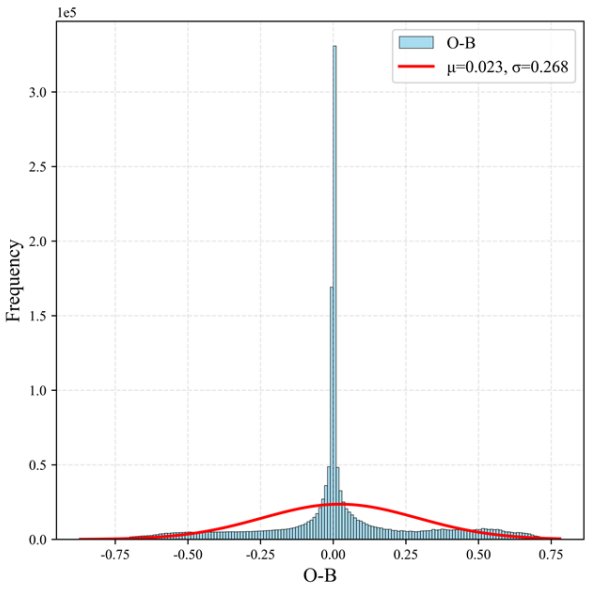

Figure 1: The PDF of observation-minus-background reflectance

4.How to determine the observation error objectively?

For the DA of VIS reflectance data, there are two main methods for establishing observation error.

The first is a static error model, which uses a fixed observation error based on instrument calibration data or empirical values. For example, Kugler et al. (2023) set the VIS reflectance observation error to 0.03 in convective-scale DA, aligning it with the instrument error. Schröttle et

al. (2020) used a fixed error of 0.2 in an idealized framework, while Scheck et al. (2020) also set a constant observation error of 0.2 for their convective-scale visible DA experiments.

The second method is a dynamic error model, whose core principle is to adaptively adjust error parameters based on the statistical relationships between observations, background fields, and analysis fields, allowing it to adapt to different scenarios. Currently, a commonly used dynamic error model for VIS DA is the Desroziers diagnostic optimization method (Desroziers et al., 2005). The observation error R is calculated using Equation (1):

$$R = E[d_a^0 \cdot (d_b^0)^T] \quad (1)$$

where $d_b^0 = O - B$, representing the observation-minus-background (O-B) deviation, and, $d_a^0 = O - A$, representing the observation-minus-analysis (O-A) deviation. $E[\cdot]$ denotes the statistical expectation, which is typically estimated using a large sample of data pairs.

Comment 5:

Cycling assimilation experiment uses the 40-km horizontal localization, which is still tight. Is the localization length scale optimally assimilated?

Our response:

Thank you for this question. After reviewing relevant literature, it was found that there is no universally optimal setting for the localization radius. In this study, following the approach of Kugler et al. (2023), a horizontal localization radius of 40 km (with a half-width of 20 km) was adopted. Their work demonstrated its effectiveness for assimilating visible satellite observations in a convective-scale OSSE with a 40-member ensemble. This choice aligns with the typical scale of convective structures and helps to mitigate sampling error while maintaining analysis stability. We acknowledge that sensitivity tests with varying radii could offer further insight, and we consider this a valuable direction for future, more extensive experimentation (Page 13, line 300).

Comment 6:

Figure 6, it is difficult to compare priors and posterior with ERA5 analyses, since ice water mixing ratio is not available in EAR5 analyses. It would be better to conduct an OSSE, in which the natural run is a simulation based on initial conditions and boundary conditions from EAR5 analyses.

Our response:

Thank you for this constructive suggestion. We fully agree that comparing ice water mixing ratio with ERA5 analyses is infeasible, and have revised our experimental design accordingly. As you recommended, the nature run in our OSSE is now a simulation driven by ERA5 reanalysis data for both initial and lateral boundary conditions. To ensure its independence from the ensemble forecasts, we configured it with a distinct 34-hour spin-up and a different set of physical parameterizations (e.g., WDM-5 microphysics, Tiedtke cumulus scheme). This setup, detailed in the revised manuscript (Page 10, lines 237-242).

Consequently, we can now perform an objective comparison of the prior and posterior ice water mixing ratios against this nature run itself, rather than against ERA5. We have updated the manuscript and replaced Figure 6 to directly demonstrate the analysis impact, which addresses the core issue you raised (Page 16).

Comment 7:

Lines 327-328, ‘its correction effect on cloud variables is significant’. How to examine the correction effect is ‘significant’? There are large corrections, but not sure the corrections are correct or not.

Our response:

Thank you for this comment. You are correct that the term “significant” should be reserved for results supported by statistical testing. In the original sentence, we used it informally to indicate the magnitude of the correction. To adhere to precise terminology, we have replaced “significant” with “substantial” in the revised manuscript (Page 16, lines 349), as the analysis shows a clear, sizable adjustment of the cloud variables toward the nature run.

Comment 8:

Figure 8, why STD can be negative? Are they STD or mean biases?

Our response:

Thank you for raising this point. The negative values in Figure 8 result from the use of a logarithmic scale (base 10) for the color bar, where values such as 10^{-3} (0.001) are displayed as -

3. All standard deviation values are positive. White areas in the plot indicate regions where the standard deviation exceeds the value set for the color scale (Page 19).

Comment 9:

Figure 10, is the red line for priors or posteriors of cycling assimilation? Since no DA is conducted for CTRL, why the CTRL errors do not continuously increase? Similarly, why the forecast errors of CTRL and EXP don't grow in free forecasts?

Our response:

Thank you for this question. The red line in Figure 10 (Page 21) represents the posterior errors from the cycled data assimilation experiment (EXP). It is important to note that in numerical weather prediction, errors do not increase monotonically indefinitely. Due to the concept of error saturation in nonlinear atmospheric models, after an initial phase of rapid error growth, the errors tend to fluctuate around a saturation level. For example, in Otkin et al (2010), Figure 5 shows that the error in the control experiment does not exhibit a clear monotonic increase but rather oscillates around a certain level. Similarly, Minamide et al. (Figures 11 and 12) demonstrate oscillating error behavior in the control (noOEI) experiment, which aligns with the pattern observed in our study.

Comment 10:

Figure 16, is it possible to perform a significance test on the scores, to examine whether the differences between two experiments are statistically significant?

Our response:

Thank you for this suggestion. We agree that statistical significance testing would be valuable. Therefore, we have expanded our experimental design: instead of conducting deterministic forecasts only for the first ensemble member, we now perform comparative analysis between 40 ensemble members and 40 CTRL experiments, and carry out significance tests on the precipitation scores. In the revised manuscript, we have updated Figures. 14 and 16 (now numbered as Figures. 13 and 15) and revised the related analysis and discussion accordingly (Page 24; Page 25).

Minor comments:

Comment 1:

‘ENSRF’ -> ‘EnSRF’

Corrected

Comment 2:

Line 251, typo: double ‘observation’

Corrected

Comment 3:

Caption of Figure 6, ‘(d) ice water liquid water mixing ratio’ -> ‘ice water mixing ratio’?

Corrected

Reference:

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. R. Meteorol. Soc., 131, 3385 – 3396, doi:10.1256/qj.05.108, 2005

Kugler, L., Anderson, J.L., and Weissmann, M.: Potential impact of all-sky assimilation of visible and infrared satellite observations compared with radar reflectivity for convective-scale numerical weather prediction, Q. J. Roy. Meteor. Soc., 149, 3623–3644, doi:10.1002/qj.4577, 2023.

Minamide, M. and Zhang, F. Q.: Adaptive Observation Error Inflation for Assimilating All-Sky Satellite Radiance. Mon. Weather Rev., 145, 1063 – 1081, doi:10.1175/MWR-D-16-0257.1, 2017.

Otkin, J. A.: Clear and cloudy sky infrared brightness temperature assimilation using an ensemble Kalman filter, J. Geophys. Res., 115, D19207, doi:10.1029/2009JD013759, 2010.

Scheck, L., Weissmann, M., and Bach, L.: Assimilating visible satellite images for convective-scale numerical weather prediction: A case-study, Q. J. Roy. Meteor. Soc., 146, 3165–3186. doi:10.1002/qj.3840, 2020.

Schröttle, J., Weissmann, M., Scheck, L., and Hutt, A.: Assimilating Visible and Infrared Radiances in Idealized Simulations of Deep Convection, Mon. Weather Rev., 148, 4357–4375, doi:10.1175/MWR-D-20-0002.1, 2020.