Dear Reviewers:

We sincerely appreciate your valuable comments and suggestions, which have significantly improved the quality of our study. We have revised the manuscript accordingly and responded to all comments. Below are our point-by-point responses. The reviewers' comments are presented in black, our responses in blue, and the proposed changes to the manuscript in red.

Reply to Reviewer #1:

Thanks for addressing previous comments. The latest version has been greatly improved.

I have the following minor comments:

1. Line 145: "and is generally consistent with Chinese administrative divisions": It looks like this sentence is not needed and can be removed.

Response:

Thank you for your comment. This sentence has been removed.

2. Line 159: "34 (0.91%), 70 (1.42%), and 76 (0.72%) observations were rejected" -> why does a 6h window reject fewer observations with a larger percentage while a 1h window rejects more observations but with a smaller percentage?

Response:

Thank you for your comment. This study uses a 12-hour 3DVAR cycling assimilation framework, where the different experiments correspond to cycling intervals of 6 hours, 3 hours, and 1 hour for assimilating GMWR observations. In each assimilation cycle, only the MWR observations closest to the analysis time within ±10 minutes are selected. Therefore, the 1-hour cycling experiment includes a much larger number of available GMWR observations overall. As a result, although the total number of rejected observations increases, their proportion relative to all assimilated observations becomes smaller. For greater clarity, the corresponding revisions are as follows:

"Each experiment started at 12:00 UTC daily, followed by a 12 h cycling data assimilation period and a subsequent 24 h forecast.

Although the experiment with a 1 h assimilation interval rejects the largest number of observations, its rejection rate remains the lowest because it has the highest assimilation frequency and assimilates the largest volume of GMWR data."

3. Line 443-44: GMWR DA mainly improves forecasts 0~6h, so it is NOT that we cannot find improvements in the upper air, but we don't have radiosonde data to do the verification during corresponding period from 0~6h. One possible option to check the upper air forecast impact is to verify against aircraft data.

Response:

Thanks for your suggestion. We agree that GMWR DA mainly benefits 0–6 h forecasts, and the limited improvements indicated by the radiosonde-based evaluation should not be interpreted as "no improvement in the upper air." As an initial study, we primarily focus on implementing the direct assimilation of GMWR observations and demonstrating its potential; therefore, this work has some limitations. In the next step, within our planned satellite–ground combined assimilation framework, we will conduct a more comprehensive evaluation of the upper-air impacts using additional independent observations, including aircraft reports and radio occultation data. Nevertheless, the corresponding revisions in the radiosonde verification and discussion sections are as follows:

"Based on radiosonde-based verification at 12 and 24 h lead times, only limited improvements are evident after GMWR assimilation; however, this does not rule out larger impacts within 0–6 h, which cannot be robustly assessed here due to the limited temporal availability of radiosonde observations.

A more comprehensive evaluation of upper-air impacts could also be performed using additional independent observations, including aircraft reports and radio occultation data."

4. Line 450, Figure 12: Is this figure the average of forecasts at lead time of 12 and 24 hours? If yes, please be more descriptive in the figure caption.

Response:

Thank you for your question. Figure 12 shows the overall RMSE computed from all 12-h and 24-h forecast samples combined, rather than an average of individual forecast RMSEs. For greater clarity, the corresponding revisions are as follows:

"Figure 9. Verification of the initial conditions against radiosonde observations. RMSEs are computed from all samples over the ten-day assimilation experiment conducted from 13 to 22 October 2023.

Figure 12. Same as Fig. 9, but for forecasts at lead times of 12 and 24 h, with RMSEs computed from all forecast samples during the ten-day experiment."

5. Line 459: "relatively low FSS values observed around the 9 h forecast period": should it be "around the forecast hour 12" (based on Fig. 13)?

Response:

Corrected.

6. Fig. 13: Why is there a V shape in the FSS plots? Does it mean forecasts are dominated by another large scale feature?

Response:

Thank you for your comment. We re-checked the FSS calculation to clarify the origin of the V-shaped behavior. The V-shaped pattern is likely related to the diurnal variability of precipitation and a temporal phase mismatch between forecasts and observations (with the simulated precipitation occurring slightly earlier). Figure R1 shows the FSS together with the related diagnostic terms for the CNTL experiment. The FSS is defined as:

$$\mathrm{FSS} = 1 - \frac{FBS}{PFO + POB}$$

In simple terms, PFO and POB indicate how prevalent/widespread the threshold-exceeding precipitation signal is in the forecast and observed fields, respectively, while FBS measures how different the forecast and observed precipitation patterns are. Notably, POB reaches its minimum at the 12-h lead time, whereas PFO reaches its minimum at the 9-h lead time, indicating a lead-time (phase) mismatch in fractional coverage between forecasts and observations. Around the 12-h lead time, the overall event signal is relatively weak (small PFO+POB) but the forecast–observation difference is relatively large (large FBS), so the ratio FBS/(PFO+POB) becomes largest and the FSS reaches its minimum.

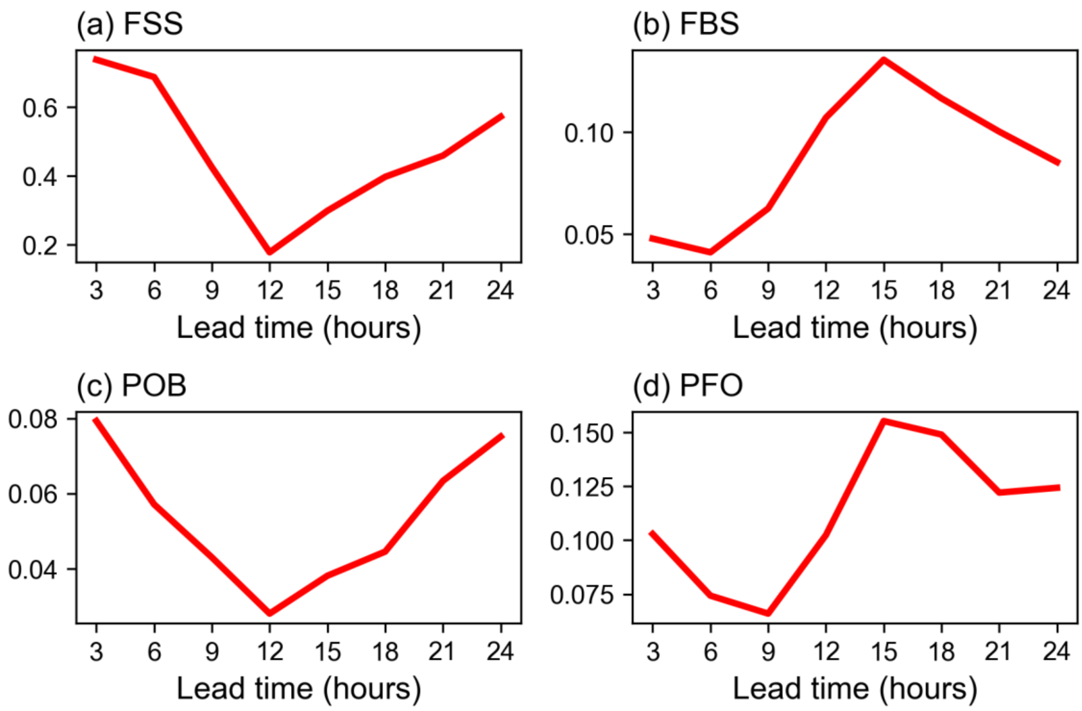

Figure R1. Time series of FSS, FBS, POB, and PFO from the CNTL experiment conducted from 13 to 22 October 2023. The scores were computed using 3-hour accumulated precipitation with a 3 mm threshold.

7. Line 469-470: The forecast verification against radiosonde does not agree with the conclusions here or that against surface observations.

Response:

Thank you for your comment. This sentence has been revised as follows:

“These findings are consistent with the above verification against surface station observations, suggesting that GMWR assimilation can improve forecasts and that higher-frequency assimilation leads to further enhancements.”

8. Line 503-509: This paragraph looks like some kind of duplicate with the paragraph in lines 482-487.

Response:

Thank you for pointing this out. Lines 503–509 partially overlapped with the results summarized in Lines 482–487. To avoid duplication, redundant content has been removed and the paragraph has been reorganized for clarity. The revised text is as follows:

“In the three-month O−B statistics, the STD in the K-band is larger than that in the V-band, consistent with Vural et al. (2024) and Cao et al. (2023). A notable positive O−B bias is observed at high-altitude stations over the Tibetan Plateau, which may be related to large-scale topographic effects. In this region, model simulations may contain errors, and RTTOV-gb coefficients may be inapplicable. The RTTOV-gb coefficients are based on global atmospheric profiles, which may differ significantly from the climatic conditions of plateau regions, potentially affecting simulation accuracy.”

9. "5 Conclusion and Discussion": The latest version added lots of good discussions. But the logic in section 5 can be further improved to make it easier to follow by readers. One possible revision is to do the discussion first and then conclude with the 3 main findings and the future plan about "doing satellite-based radiance and ground-based ones together".

Response:

Thank you for your helpful suggestion. Following your suggestion, we split Section 5 into two subsections—Section 5.1 (Discussion) and Section 5.2 (Conclusions and future work)—to improve the logic and readability. The revised Section 5 is as follows:

“

5 Discussion, conclusions, and future work

5.1 Discussion

This section discusses (i) O−B characteristics and their potential sources, (ii) the effectiveness and physical interpretability of the RF-based bias correction (BC), and (iii) the impacts of direct GMWR radiance assimilation on analyses and forecasts.

In the three-month O−B statistics, the STD in the K-band is larger than that in the V-band, consistent with Vural et al. (2024) and Cao et al. (2023). A notable positive O−B bias is observed at high-altitude stations over the Tibetan Plateau, which may be related to large-scale topographic effects. In this region, model simulations may contain errors, and RTTOV-gb coefficients may be inapplicable. The RTTOV-gb coefficients are based on global atmospheric profiles, which may differ significantly from the climatic conditions of plateau regions, potentially affecting simulation accuracy.

To mitigate these systematic O−B biases, a machine learning-based BC scheme using the RF technique was developed. The number and depth of trees are critical hyperparameters that must be predetermined. Training time increases approximately linearly with the number of trees, while performance exhibits a logarithmic-like saturation trend. In terms of tree depth, both training time and performance increase approximately logarithmically with depth. Thus, selecting a modest number (n_estimators) and depth (max_depth) of trees, such as 50 and 15, can balance efficiency and accuracy. Feature importance analysis for BC predictors revealed observed brightness temperature, atmospheric precipitable water, and surface pressure as key factors for correcting biases. The importance of brightness temperatures aligns with findings in satellite data bias correction (Liu et al., 2022; Zhang et al., 2023). Atmospheric precipitable water is essential for the K-band, a humidity-sensitive channel. Surface pressure plays a key role in temperature channels, thereby accounting for the positive bias observed in plateau regions. Although atmospheric thickness predictors contributed less overall, the 1,000–700 hPa thickness was relatively significant, likely due to GMWRs primarily sensing radiation from the lower atmosphere.

The machine learning-based BC scheme effectively mitigated the bimodal distribution and systematic errors in the O−B statistics. To assess its impact on the initial and forecast fields, a parallel experiment without bias correction was conducted, based on the 1-hour assimilation interval experiment (GMWR_1H). For the initial fields, as verified against radiosonde observations, the experiment without BC yielded only minor improvements in temperature and even degraded the water vapor field. As for the forecast fields, verification against surface station observations showed that the absence of BC led to a noticeable degradation in the forecast accuracy of 2 m temperature and relative humidity. These findings indicate that the machine learning-based BC scheme had a beneficial impact on both the initial conditions and the subsequent forecasts. Nevertheless, despite its demonstrated effectiveness, the scheme is subject to several limitations. Relying on offline O−B statistics, it implicitly assumes that all biases originate from the observations—an assumption that may not always hold and may, in some instances, mask model biases (Auligné et al., 2007; Eyre, 2016). These limitations motivate further improvements in future work (Sect. 5.2).

In this study, direct assimilation of GMWR radiances enhances both the initial conditions and the forecasts, showing potential for improving ABL and precipitation simulations. Although the assimilation of GMWR radiances yields slight improvements in the forecast wind fields (Fig. 11), it exerts an overall negative impact on the wind fields in the initial conditions (Fig. 9). This discrepancy stems from the use of distinct observational datasets: the initial conditions are verified against radiosonde observations, while the forecasts are evaluated using surface station data. It should be noted that assimilating GMWR improves the wind fields below 500 m AGL in the initial conditions. This improvement is consistent with the verification of the forecast, which demonstrates enhancements in the 10 m wind fields. Regarding the degradation of wind fields above 500 m AGL, the background error covariance may contribute to this negative impact. On the one hand, it determines the response of the wind fields to temperature and humidity adjustments made by RTTOV-gb. On the other hand, since RTTOV-gb's adjustments are primarily concentrated in the ABL, the response above the ABL may be propagated through the background error covariance.

5.2 Conclusions and future work

To investigate the impact of directly assimilating GMWRs in Southwest China, a GMWR assimilation module has been developed in WRFDA-4.5, where RTTOV-gb is used as the observation operator. Based on this module, a three-month sample dataset of O−B was collected to evaluate the bias and develop a BC model. Furthermore, 10-day assimilation experiments (Table 2) were conducted using this GMWR assimilation module and BC model to investigate the impact of direct GMWR assimilation and the effects of assimilation frequency. The main findings are as follows:

(1) Based on three months of hourly samples, noticeable O−B biases were observed, varying across sensors, channels, and geographical locations. The machine learning-based bias correction scheme, employing an RF model, effectively reduced these O−B systematic biases. After applying this BC model, both the bias and STD of the O−B were substantially reduced. Specifically, the bias and STD decreased by 0.83 K (97.1 %) and 1.63 K (64.6 %), respectively. For some channels, the original O−B distribution exhibited a bimodal pattern, which was transformed into a unimodal distribution after BC. The corrected O−B distributions exhibited Gaussian characteristics centered around zero.

(2) Assimilating GMWR enhances the accuracy of initial conditions, with higher assimilation frequencies amplifying the positive impact, particularly for temperature and humidity in the lower atmosphere. Evaluation against radiosonde observations shows that the temperature RMSE below 1 km AGL decreases by 3.67 % to 6.32 % as the assimilation frequency increases from 6 h to 1 h. For the water vapor mixing ratio, positive impacts extend up to 5 km AGL, with average RMSE improvements ranging from 1.98 % to 2.30 %. Verification against surface station observations further supports these findings, indicating that the RMSE for 2 m temperature decreases by up to 4.1 %, while the RMSE for 2 m relative humidity decreases by up to 1.3 % at the 1 h assimilation

frequency.

(3) The assimilation of GMWR observations leads to improvements in forecasts, and increasing assimilation frequencies have the potential to yield further improvements. In the first 6 hours of the forecast, the temperature RMSE decreases by 0.012 K, 0.014 K, and 0.019 K with 6 h, 3 h, and 1 h assimilation frequency, respectively. Similar trends are observed for relative humidity, the experiment with 1 h GMWR assimilation frequency shows the largest decrease in RMSE. GMWR assimilation also improves precipitation forecasts, with further enhancements seen as assimilation frequency increases. For 1 h GMWR assimilation, time-averaged FSS improvements reach 0.02 for both the 3 mm and 4 mm, 0.03 for 5 mm, and 0.04 for 6 mm thresholds.

Despite these encouraging results, this study has some limitations that motivate future work. Regarding bias correction, the offline scheme lacks anchoring observations, rendering the analysis fields more susceptible to model bias. Future efforts should consider bias correction strategies based on unbiased reference observations or adopt a constrained correction scheme, such as the constrained adaptive bias correction (Han and Bormann, 2016). The GMWR assimilation was implemented using 3DVAR, based on RTTOV-gb and WRFDA, and only static background-error covariances were employed in this study. The background error covariance matrix plays an important role in variational data assimilation, but this type of covariance is climatological, spatially homogeneous, and isotropic. This may limit the impact of GMWR assimilation, and flow-dependent error covariances should be considered in future work.

Moreover, only clear-sky GMWRs were assimilated in this study. Since precipitation processes are often accompanied by extensive cloud cover, few clear-sky GMWRs were available. To better explore the potential of GMWR assimilation, experiments were conducted during periods with abundant clear-sky GMWRs (e.g., a ten-day period in October 2023), which coincided with minimal heavy precipitation. Studies on satellite all-sky assimilation have shown that incorporating cloud- and precipitation-affected data improves forecasts (Ma et al., 2022; Xian et al., 2019), highlighting the need for future research on all-sky assimilation of GMWRs. Under such conditions, assimilation experiments could be conducted during a different or longer period, given that assimilated GMWR observations would be relatively more abundant. It is noted that GMWRs exhibit higher sensitivity and provide more valuable observations of the lower troposphere and planetary boundary layer compared to satellite-based microwave radiometers (Shi et al., 2023). Building on this study, future research could explore the joint direct assimilation of satellite-based and ground-based microwave radiometers. A more comprehensive evaluation of upper-air impacts could also be performed using additional independent observations, including aircraft reports and radio occultation data. By leveraging their complementary observational capabilities, this approach has the potential to further enhance the accuracy of atmospheric analysis and improve forecasting across multiple layers of the atmosphere."