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:

Overall, this paper is clearly written, including detailed descriptions of the experimental design. The analysis and presentation of the results are also generally clear. However, I have some concerns regarding the bias correction method and the way the experiments are compared. I would suggest that the authors address these points before the paper can be considered for publication.

Major comments:

1. The paper includes a focus on bias correction, but its impact on the analysis and forecast fields is not discussed in depth. While the improvements in GMWR diagnostics in observation space (e.g., Figs. 6–8) are somewhat expected, the more crucial aspect is how the bias correction affects the model space. It would strengthen the paper to include a comparison between assimilation experiments with and without bias correction for GMWR.

Response:

Thank you for your valuable suggestion. Based on the 1-hour assimilation interval experiment (GMWR_1H) already presented in the manuscript, an additional experiment (GMWR_1H_noBC) without bias correction (BC) was conducted. As shown in Figure R1, the initial conditions were verified against radiosonde observations. The results indicate that the experiment with BC reduced the RMSE of temperature and water vapor in the lower atmosphere. In contrast, the experiment without BC yielded only minor improvements in temperature RMSE and even increased the RMSE for water vapor. Specifically, for temperature below 1 km, the experiment with BC reduced the RMSE by

6.32%, whereas the experiment without BC achieved only a 0.49% reduction. For water vapor below 5 km, the experiment with BC reduced the RMSE by 1.98%, while the experiment without BC resulted in an 8.47% increase.

Similarly, as shown in Figure R2, the forecast fields were verified against surface station observations. The results show that the experiment with BC reduced the RMSE of 2m temperature and relative humidity at a lead time within 12 hours. In contrast, the experiment without BC degraded the forecast accuracy of 2 m temperature and relative humidity. Specifically, the time-averaged temperature RMSE decreased by 0.019 K with BC, whereas it increased by 0.001 K without BC. For relative humidity, the time-averaged RMSE decreased by 0.102 with BC, while it increased by 0.079 without BC.

These findings suggest that omitting BC weakens the effectiveness of assimilation and may even introduce negative impacts on both the initial and forecast fields. The above results have been covered in the discussion section, as shown below.

"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."

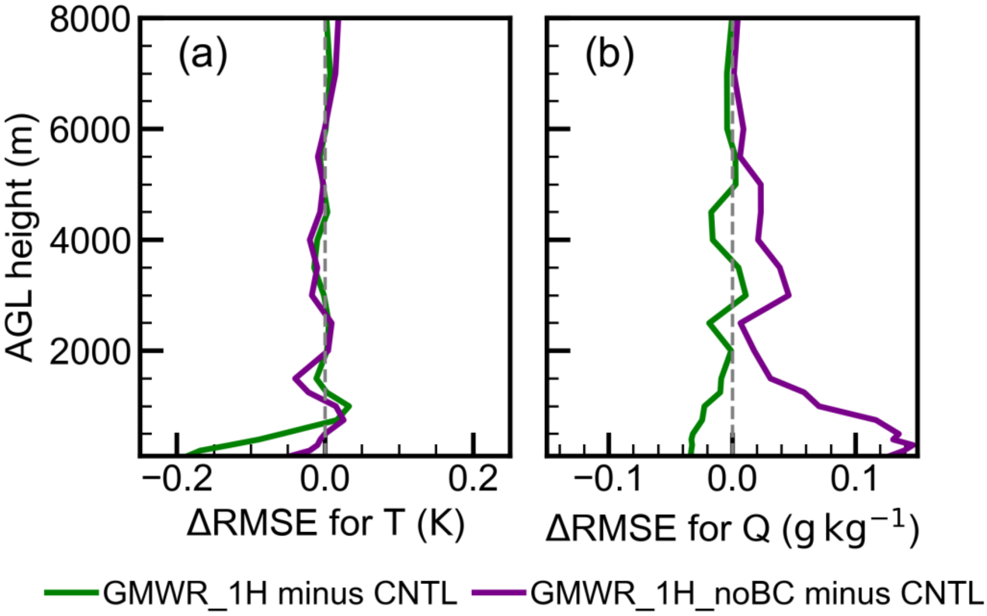

Figure R1. Verification of the initial conditions against radiosonde observations, based on the ten-day assimilation experiment conducted from 13 to 22 October 2023. Differences in root mean square error (RMSE) for (a) temperature and (b) water vapor mixing ratio between experiments.

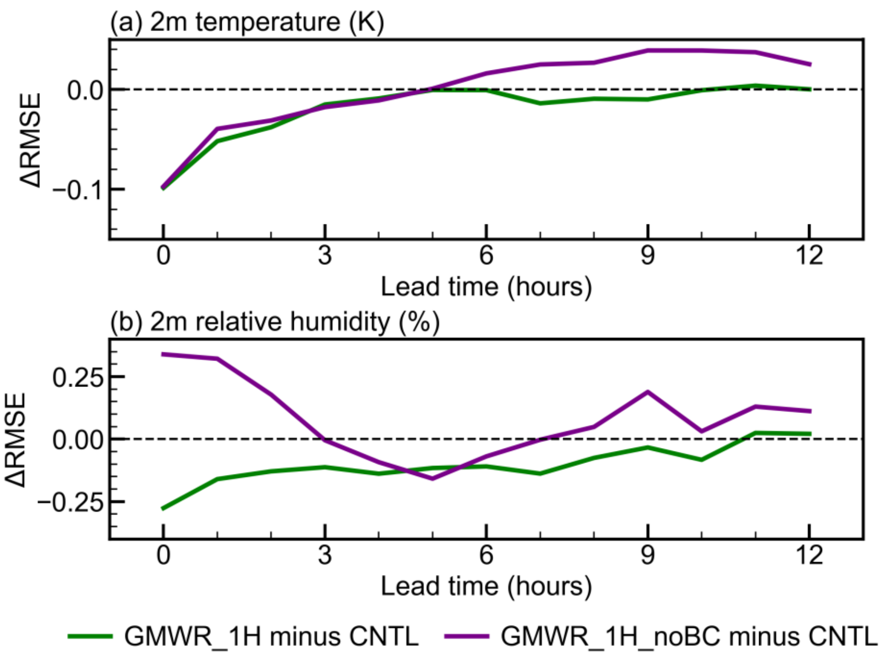

Figure R2. Verification of the forecast against surface station observations, based on the ten-day assimilation experiment conducted from 13 to 22 October 2023. Differences in root mean square error (RMSE) for (a) temperature and (b) relative humidity between experiments.

2. The bias correction approach based on offline O–B statistics essentially assumes that all biases originate from the observations. However, this

assumption may not always hold, and such correction could potentially mask model bias (e.g., Aulígné et al., 2007; Eyre, 2016), especially when more complex predictors or bias-prediction schemes are used to make the correction more expressive. Therefore, such offline bias correction that simply brings the O–B mean close to zero does not necessarily indicate a successful correction. It may be a sign of success, but could also be a result of compensating for model bias, rather than removing observation bias. Although in practice it remains difficult to fully separate the sources of O–B bias, I believe it is important for the paper to acknowledge this fundamental limitation of the current offline bias correction method based on O-B statistics.

Aulígné, T., McNally, A. P., & Dee, D. P. (2007). Adaptive bias correction for satellite data in a numerical weather prediction system. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 133(624), 631-642.

Eyre, J. R. (2016). Observation bias correction schemes in data assimilation systems: A theoretical study of some of their properties. Quarterly Journal of the Royal Meteorological Society, 142(699), 2284-2291.

Response:

We sincerely appreciate your valuable suggestion. Following Zhang et al. (2023, 2024), a nonlinear bias correction scheme based on O–B statistics was applied in this study. The assimilation experiment incorporating this bias correction demonstrated improvements in both the initial and forecast fields. However, as you rightly pointed out, the background field inherently contains biases, and the use of an offline correction method may potentially mask these model bias. In response to your suggestion, we conducted comparative experiments with and without this bias correction scheme. Although the results show that the bias correction had a positive impact on both the initial conditions and forecasts, it is admitted that this approach has some limitations in differentiating between model- and observation-related components of systematic biases.

This limitation has now been clearly stated in the discussion follows:

“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). Moreover, 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)."

Zhang, X., D. Xu, X. Li, and F. Shen, 2023: Nonlinear Bias Correction of the FY-4A AGRI Infrared Radiance Data Based on the Random Forest. *Remote Sensing*, **15**, 1809, https://doi.org/10.3390/rs15071809.

Zhang, X., D. Xu, F. Shen, and J. Min, 2024: Impacts of Offline Nonlinear Bias Correction Schemes Using the Machine Learning Technology on the All-Sky Assimilation of Cloud-Affected Infrared Radiances. *Journal of Advances in Modeling Earth Systems*, **16**, e2024MS004281, https://doi.org/10.1029/2024MS004281.

3. Regarding the wind analysis and forecast, the results appear somewhat inconsistent depending on the diagnostic used. For example, Fig. 9 shows a degradation, while Figs. 10–11 indicate marginal improvements. However, the paper does not seem to acknowledge or discuss these discrepancies across different diagnostics. A brief discussion of these differences would help clarify the interpretation of the results.

Response:

Thanks for your comments. The discussion of these differences has been added to the discussion as follows:

"Although the assimilation of GMWR radiance 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 field below 500 m AGL in the initial conditions. This improvement is consistent with the verification of the forecast, which demonstrate enhancements in the 2-meter wind fields. Regarding the degradation for wind field above 500m 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."

Minor Comments:

1. Figure 3(b)(d): It would be useful to show the same scatter plot after bias correction to examine whether the two distinct clusters merge.

Response:

Thank you for your suggestion. The two distinct clusters can result in a bimodal distribution in the O−B probability density function (PDF). For example, a cluster shifted to the right of the diagonal in the scatter plot leads to a peak on the positive x-axis of the O−B PDF. In the "Bias Correction" section, the bias correction model was tested. As shown in Figure 6b, d, the O−B PDF transitions from a bimodal to a unimodal distribution, indicating that the two clusters are effectively merged. Figure 6 is used to illustrate this issue rather than adding an additional scatter plot, as the manuscript already includes 17 figures. The corresponding revisions we have implemented are as follows:

"The bimodal feature in the O–B PDF corresponds to the two distinct clusters observed in the scatter plots shown in Fig. 3. Specifically, when one cluster is concentrated along the diagonal and the other shifts to the right, a peak forms on the positive x-axis of the O−B PDF, resulting in a bimodal distribution. From the O–B PDF (Fig. 6), both instruments exhibit a positive bias with a unimodal distribution in the K-band. In contrast, the V-band displays a bimodal distribution: the second peak appears on the right for HATPRO and on the left for MP3000A. These results are consistent with the scatter plots shown in Fig. 3. After BC, the O–B PDF changes from a bimodal to a unimodal distribution, indicating that the two clusters have been effectively merged.

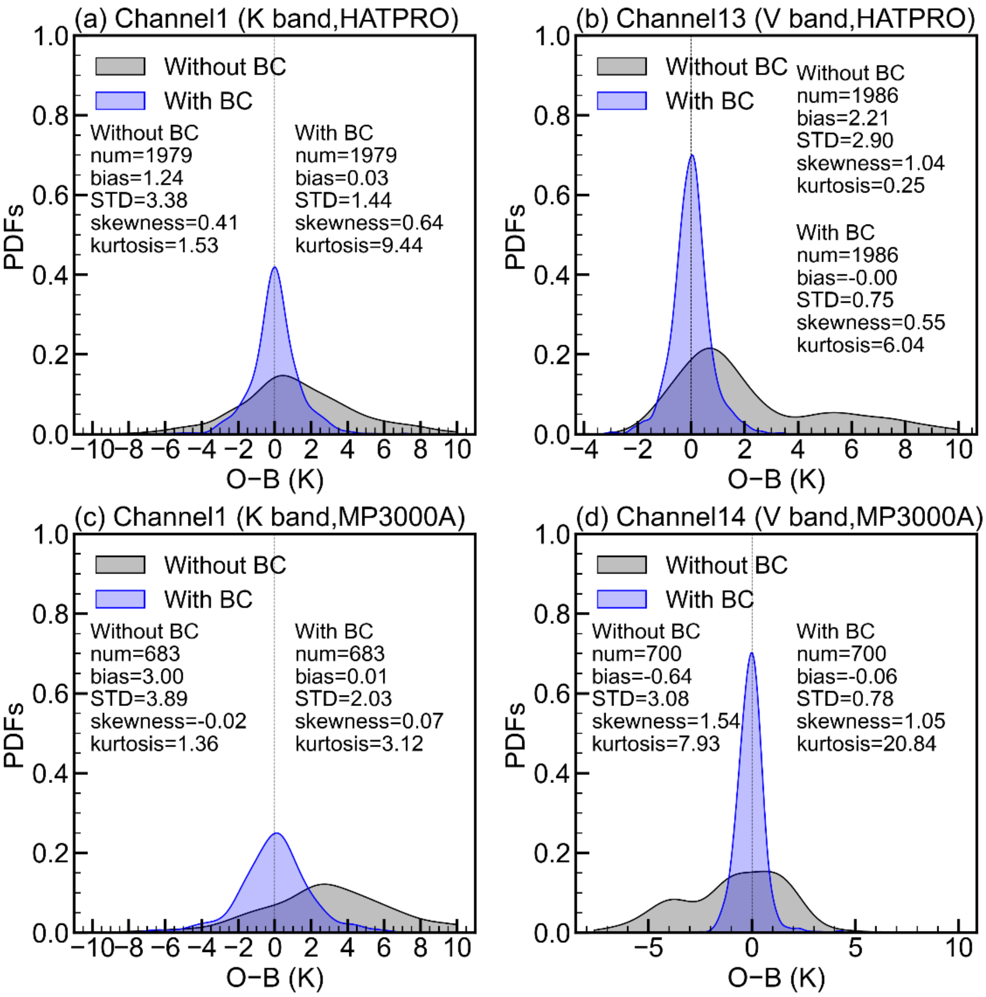

Figure 6: Probability density functions (PDFs) of the O−B distributions, based on a test set randomly selected from 30% of the three-month sample dataset collected from August to October 2023. The top and bottom rows correspond to the HATPRO and MP3000A sensors, respectively, while the left and right columns represent the K-band and V-band. Each panel displays the number of samples (num), the mean (bias), standard deviation (STD), skewness, and kurtosis of the distributions."

2. L233-235: Existing approaches can also address nonlinear relationship between the physical variables, e.g., skin temperature (TS), and the bias since the selection of predictors can be completely general. E.g., consider predictors $p1 = TS$, $p2 = (TS)^2$, $p3 = (TS)^3$, $p4 = (TS)^4 \ldots$

Response:

Thank you for your comment. Most offline bias correction methods assess the relationship between biases and predictors using multivariate linear regression. Although incorporating nonlinear operators into physical variables can enhance their ability to handle nonlinear relationships, machine learning approaches provide a more efficient means of capturing such complexities. The corresponding revisions we have implemented are as follows:

"O–B bias is commonly represented using multiple linear regression with several predictors. Compared to linear estimates, nonlinear approaches show improved performance in reducing systematic biases (Zhang et al., 2023, 2024). Following these works, this study employed a machine learning–based bias correction scheme using the Random Forest (RF) technique (Breiman, 2001)."

3. L326-327: Since Figure 8 does not show the CTRL results, it is difficult to determine whether the assimilation of GMWR really improves the fit (even though such improvement is expected)

Response:

Following your suggestion, we have included the results of the CNTL experiment. The corresponding revisions in the manuscript are as follows:

"O–A statistics were also computed based on the initial fields from the CNTL experiment. The O–A bias in CNTL is slightly larger than that in the GMWR assimilation experiments, with a more noticeable difference in the V band. Regarding the STD, the CNTL experiment shows higher values across all channels, with the largest difference approaching 1 K. These results suggest that assimilating GMWR data improves the consistency between the initial fields and the observed brightness temperatures.

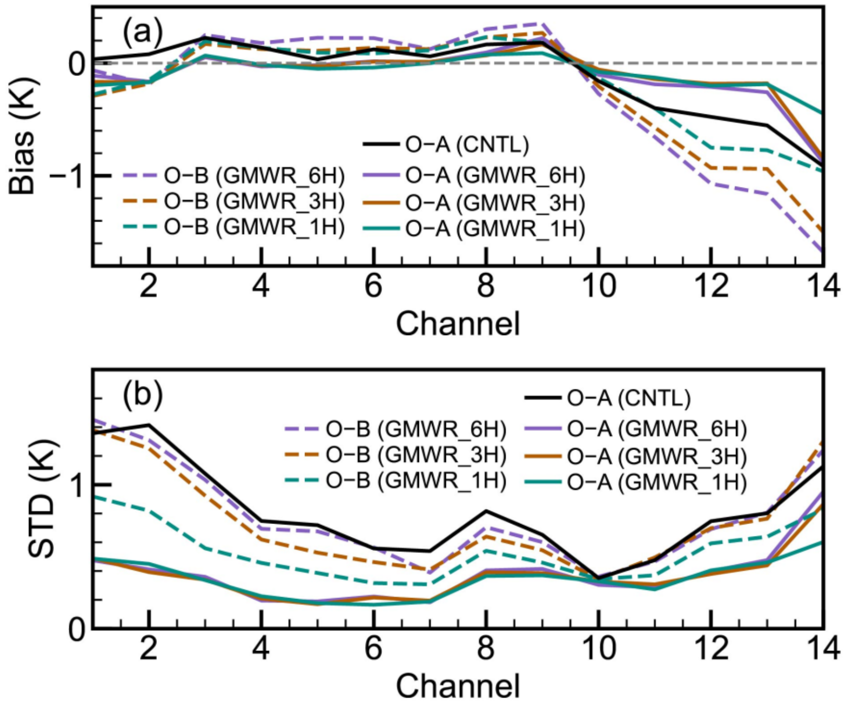

Figure 6: Verification of the initial conditions against GMWR observations, based on the ten-day assimilation experiment conducted from 13 to 22 October 2023. (a) Bias and (b) standard deviation (STD) of the observation minus background (O−B) and observation minus analysis (O−A) for the GMWR assimilation in the target region of Southwest China (blue box in Fig. 1).”

4. L480-481: This may not be true, as discussed in the major comment (2).

Response:

The corresponding modifications are as below.

“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.”

5. L498-499: A larger STD in the K-band compared to the V-band does not necessarily imply that the model’s humidity accuracy is worse than its

temperature accuracy. First, it is inherently difficult to directly compare the accuracy of humidity and temperature fields. Second, the brightness temperature STD also depends on its sensitivity to temperature and humidity. For example (using hypothetical numbers), a 1K change in K-band may correspond to 1% change in humidity, but 1K change in V-band may correspond to a much larger 10% change in temperature.

Response:

Thank you for your comment The corresponding sentence has been removed.

6. Overall, the paper includes a large amount of numerical detail (e.g., bias reductions by a few degrees or a certain percentage). If some of these values are already shown in the figures, I believe it is not necessary to restate all of them in the text. Instead, the paper could focus on highlighting the meaning and implications of these numbers. This would help make the manuscript more concise and easier to follow.

Response:

Thank you for your valuable comment. We have carefully revised the manuscript by removing redundant numerical details. These revisions enhance the manuscript's conciseness and clarity, making it easier to follow and more focused on the key findings.

Reply to Reviewer #2:

This is a high-quality and well-structured manuscript that makes a valuable contribution to the field of data assimilation. The authors present an effective and practical approach for the direct assimilation of ground-based microwave radiometer observations, supported by a carefully designed bias correction scheme. The results are clearly presented and demonstrate consistent improvements in low-level thermodynamic fields and short-term forecasts, particularly when a higher assimilation frequency is used. This study complements recent work in the field and offers useful insights for future operational applications. Nevertheless, several specific issues should be addressed before the manuscript can be considered for publication in GMD.

General comments,

1. Please discuss what is the physical significance of systematic bias? Does and offline or online BC conducted? If it is an offline bias correction, without the presence of an anchoring observation, how do you expect to differentiate the model vs observation component of the systematic biases? More physical interpretation would be helpful.

Response:

Thank you for your comments. Following Zhang et al. (2023, 2024), a nonlinear BC scheme was employed in this study. This approach is based on O–B statistics and assumes that all biases originate from the observations. Due to this limitation, it may mask model biases and fails to differentiate the model vs observation component of the systematic biases.

Nevertheless, additional experiments without BC were conducted to further investigate the impact of this BC method (In the response to Reviewer #1). Based on the 1-hour assimilation interval experiment (GMWR_1H) already presented in the manuscript, an additional experiment (GMWR_1H_noBC) without bias correction (BC) was conducted. As shown in Figure R1, the initial conditions were verified against radiosonde observations. The results indicate that the experiment with BC reduced the RMSE of temperature and water vapor in the lower atmosphere. In contrast, the experiment without BC yielded only minor improvements in temperature RMSE and even increased the RMSE for water vapor. Specifically, for temperature below 1 km, the experiment with BC reduced the RMSE by 6.32%, whereas the experiment without BC achieved only a 0.49% reduction. For water vapor below 5 km, the experiment with BC reduced the RMSE by 1.98%, while the experiment without BC resulted in an 8.47% increase. Similarly, as shown in Figure R2, the forecast fields were verified against surface station observations. The results show that the experiment with BC reduced the RMSE of 2m temperature and relative humidity at a lead time within 12 hours. In contrast, the experiment without BC degraded the forecast accuracy of 2 m temperature and relative humidity. Specifically, the time-averaged temperature RMSE decreased by 0.019 K with BC, whereas it increased by 0.001 K without BC. For relative humidity, the time-averaged RMSE decreased by 0.102 with BC, while it increased by 0.079 without BC.

The results demonstrate that, in the absence of bias correction, GMWR assimilation becomes less effective and may even negatively impact both the initial and forecast fields. As noted in major comment (2) by Reviewer #1, it remains difficult in practice to separate the sources of O–B bias. Although this limitation exists in the current BC method based on O–B statistics, it still exerts a beneficial impact on GMWR assimilation, improving both the initial conditions and subsequent forecasts. As an initial study primarily focused on the direct variational assimilation of GMWR radiance, it is admitted that this study has certain limitations. Nevertheless, we have acknowledged the limitations the BC approach employed in this study and have proposed directions for future improvement. The corresponding statement has been included in the discussion section, as shown below.

"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). Moreover, 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)."

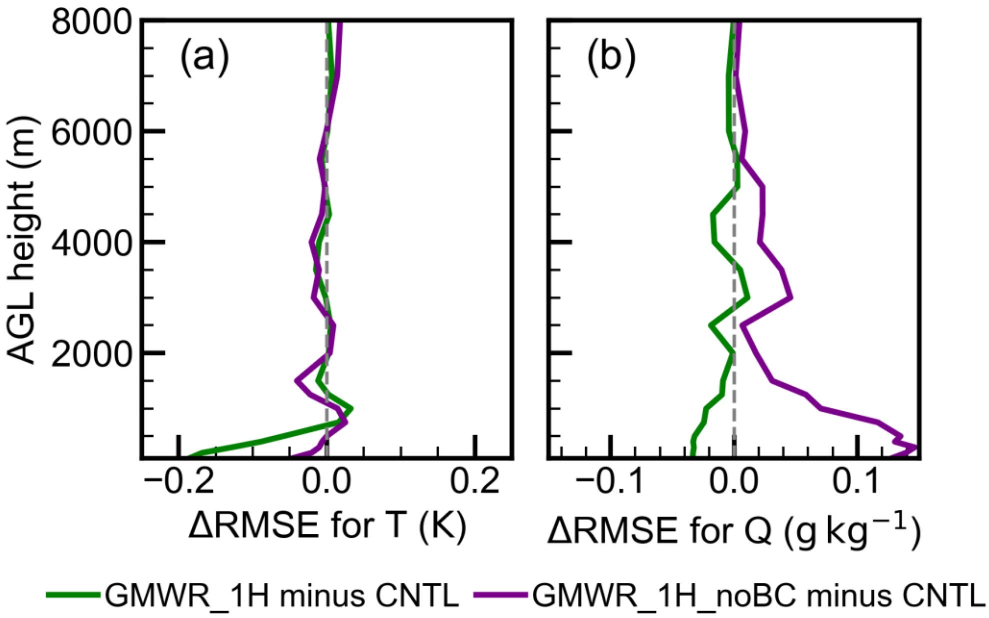

Figure R1. Verification of the initial conditions against radiosonde observations, based on the ten-day assimilation experiment conducted from 13 to 22 October 2023. Differences in root mean square error (RMSE) for (a) temperature and (b) water vapor mixing ratio between experiments.

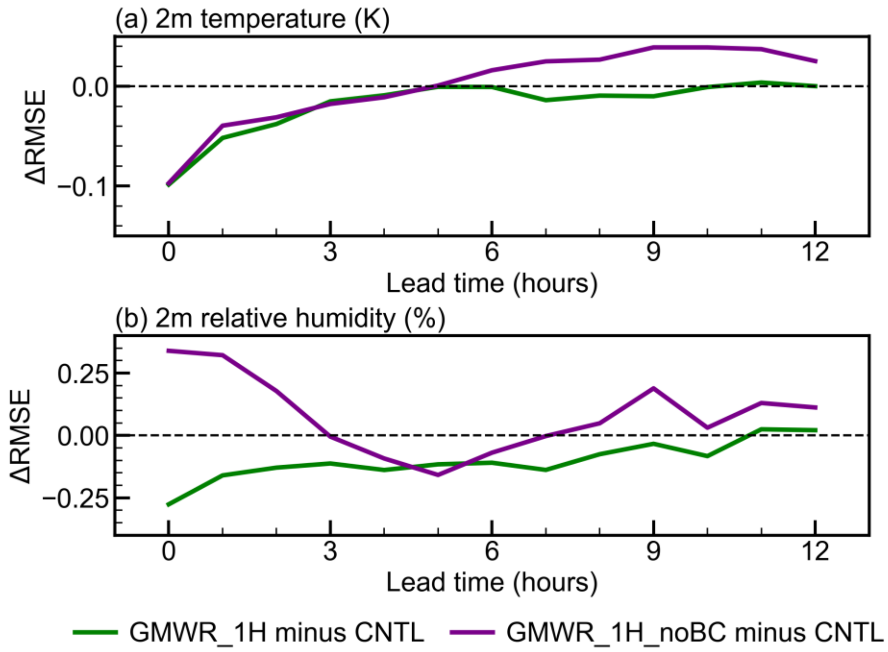

Figure R2. Verification of the forecast against surface station observations, based on the ten-day assimilation experiment conducted from 13 to 22 October 2023. Differences in root mean square error (RMSE) for (a) temperature and (b) relative humidity between experiments.

Zhang, X., D. Xu, X. Li, and F. Shen, 2023: Nonlinear Bias Correction of the FY-4A AGRI Infrared Radiance Data Based on the Random Forest. *Remote Sensing*, **15**, 1809, https://doi.org/10.3390/rs15071809.

Zhang, X., D. Xu, F. Shen, and J. Min, 2024: Impacts of Offline Nonlinear Bias Correction Schemes Using the Machine Learning Technology on the All-Sky Assimilation of Cloud-Affected Infrared Radiances. *Journal of Advances in Modeling Earth Systems*, **16**, e2024MS004281, https://doi.org/10.1029/2024MS004281.

2. Section 3.1 Bias correction scheme BASE on OMB, is not designed to remove the model biases because if model background biases generate the OmB systematic biases, you not only need to correct the biases in OmBs but also in the background. Zhu et al. (2014) stated that the error due to the background atmospheric profiles should not be removed by the bias correction. Please clarify.

Response:

Thank you for your professional comments. The bias correction scheme based on O–B statistics has some limitations: it cannot differentiate between model- and observation-related components of systematic biases. As a result, the approach may mask model biases and produce biased analyses. On the one hand, model biases should be addressed separately prior to data assimilation. On the other hand, it may be possible to consider bias correction scheme based on unbiased reference observations or adopt a constrained correction scheme, such as the constrained adaptive bias correction (Han and Bormann, 2016). However, despite these limitations, the current BC method based on O–B statistics still exerts a beneficial impact on GMWR assimilation, improving both the initial conditions and subsequent forecasts. Separating model and observation biases is a valuable research topic but beyond the scope of this initial study. In future work, we plan to explore a more effective bias correction strategy to further improve the GMWR assimilation. The relevant revision has been added to the discussion section, consistent with the response to general comment 1 above.

Han W. and Bormann N.: Constrained adaptive bias correction for satellite radiance assimilation in the ECMWF 4D-Var system, *ECMWF Technical Memoranda*, 783, https://doi.org/10.21957/rex0omex, 2016

Specific comments:

1. The description of the single-observation assimilation experiment could be more detailed, including information such as the instrument used, and which channels were assimilated.

Response:

The corresponding statements have been updated as below.

"A single-observation assimilation experiment was conducted to test the GMWR direct assimilation module. In this experiment, a pseudo radiance observation from the HATPRO sensor was assimilated in channel 6 (K band), with an assigned observation error of 1 K and an innovation of 2 K."

2. The hyperparameters of the machine learning bias correction model are important. As the figure related to hyperparameter tuning is currently omitted from the manuscript, consider including it—either in the main text or the

appendix. Alternatively, providing the tuning range and final configuration would be helpful as a reference for future studies.

Response:

The corresponding statements have been updated as below.

"During hyperparameter tuning, *n_estimators* was varied between 10 and 150, *max_depth* was adjusted from 5 to 30, *min_samples_leaf* was tested with values between 1 and 3, and *min_samples_split* was tuned in the range of 2 to 6.

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 term 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."

3. Figure 1: The legend appears to be incorrect. It should be "GMWR" instead of "MWR"

Response:

Corrected.

4. Figure A2: The instrument label is incorrect. Both Figure A1 and A2 show the sensor as "HATPRO".

Response:

The label in Figure A2 should be "MP3000A", and now is updated.

5. Line 45: Change "observation-operator" to "observation operators."

Response:

Corrected.

6. Line 48: Change "It is noted that studies began to develop fast RTMs suitable for GMWR" to "Recent studies have developed fast RTMs suitable for GMWR."

Response:

The corresponding sentences has been updated.

7. Line 98: Change “dictates” to “dictate.”

Response:

Corrected.

8. Lines 159–166: Pay attention to capitalization. In “a machine learning bias correction scheme” consider whether “a” should be capitalized to be consistent with formatting.

Response:

Corrected.

9. Line 187: Add a period before “Therefore.”

Response:

Added.

10. Line 191: Change “only calculate O−B” to “only calculating O−B.”

Response:

Corrected.

11. Lines 238–240: There should be a space between the numeric value and “hPa”. Please check the manuscript for similar issues with other units

Response:

Thanks for your comments. We have thoroughly reviewed the entire manuscript.

12. Line 287: Change “scattering patterns” to “scatter patterns.”

Response:

Thanks for your comments. The relevant sentence has already been revised in response to other comments, and this issue no longer exists

13. Line 471: An article (e.g., the or a) should be added before "GMWR assimilation module".

Response:

Added.

14. Line 474: The tense of "developed" appears to be inconsistent with "evaluate" and should be adjusted for parallel structure.

Response:

The corresponding statements have been updated as below.

"Based on this module, a three-month sample of O−B statistics was calculated to evaluate the bias and develop a BC model."

15. Line 475: Change "direct assimilation GMWR" to "direct assimilation of GMWR."

Response:

The corresponding statements have been updated as below.

"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."

16. Lines 490–491: Change "has potential to" to "has the potential to".

Response:

Corrected.