Dear Editor and Reviewers,

We sincerely thank you for your careful evaluation of our manuscript and for your valuable comments. Following the reviewers' comments and the editorial formatting and compliance suggestions provided previously, we have conducted a systematic revision and check of the entire manuscript. The main revisions include: rechecking and unifying the calculation procedures and descriptions for Fig. 7, Fig. 9, and the related statistical results; replacing the previous single train–test split for systematic bias correction with a leave-one-out cross-validation (LOOCV) framework, with the corresponding figures, tables, text, and statistical results updated accordingly; revising the radiative-transfer equation and the expression of the optical-depth term; replacing the inaccurate term "lapse-rate constraint" with "adjacent-layer continuity constraint" and further clarifying its physical meaning; and presenting the attitude correction, the role of the constraints, the effect of systematic bias correction, and the study limitations in a more cautious and consistent manner.

In addition, following the reviewers' suggestions, we condensed and coordinated the Discussion and Conclusion sections to reduce repetition and revised overly strong statements regarding retrieval accuracy, so that the conclusions are better aligned with the evidential scope of the present dataset. We also further checked the figure and table numbering, equation formatting, textual expression, and citation style in accordance with the editorial formatting requirements. The reference list has been reorganized and standardized in alphabetical order by the first author's surname, as required by the journal.

Overall, in this revision, we have not only responded point by point to each reviewer comment, but also consistently updated the relevant parts of the manuscript concerning methodological description, result presentation, statistical interpretation, and conclusion wording. These changes were made to ensure that the revised manuscript is more coherent in logic, formatting, and expression. Our detailed point-by-point responses are provided below.

Response to RC1

Comment 1

It remains unclear to me why there are so many oscillations in the figure of the temperature retrieval bias, compared to the plots of mean temperature retrieval and mean temperature observations. Also, it is not clear exactly what is presented in the results with the humidity/temperature bias correction applied.

Response:

Thank you for pointing this out. We re-examined the plotting procedure and the accompanying description of Fig. 7 in the original manuscript and found that the statistical relationship between the mean profiles and the bias curves had not been explained clearly enough. This could have caused confusion regarding the physical correspondence between the two. Following the reviewer's comment, we re-plotted Fig. 7 and ensured that the temperature and relative-humidity bias curves are now calculated directly as the differences between the corresponding mean retrieved profiles and mean radiosonde profiles.

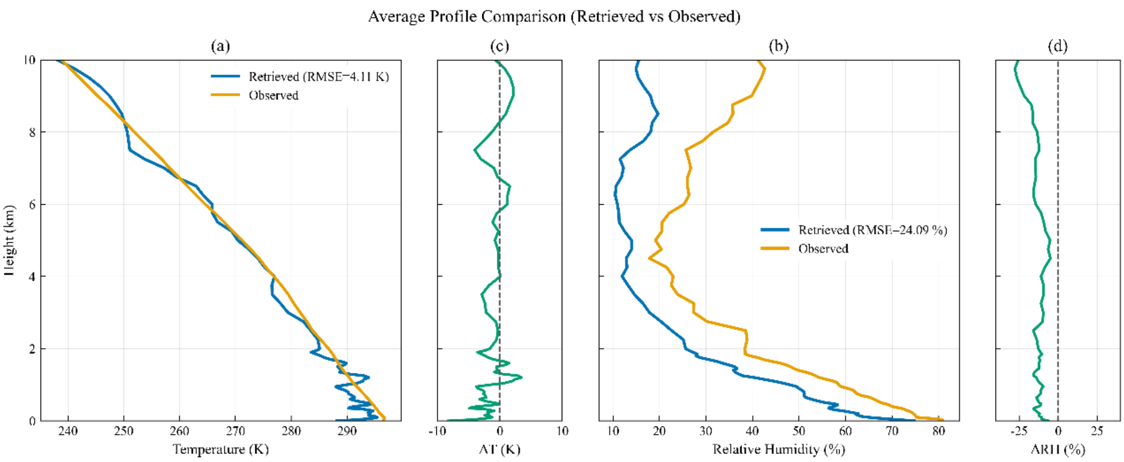
In addition, we would like to clarify that the purpose of introducing systematic bias correction in this study is not to assume that all errors originate from the brightness-temperature measurements of the instrument itself. Rather, the correction is intended to reduce the stable height-dependent bias exhibited by the retrieved profiles relative to the radiosonde reference profiles within the present observation–simulation–retrieval chain. This bias may reflect the combined net effect of instrument measurement errors, forward-model errors, and representativeness errors. Based on this consideration, we estimate the mean systematic-bias profiles from the matched samples and apply them as correction terms to the retrieval results in order to reduce the systematic-bias component, rather than to indicate a complete removal of all error sources.

At the same time, we have further clarified the sample scope and statistical interpretation of the results shown in Fig. 9 in the revised manuscript, in order to avoid the ambiguity caused by the previous unclear expression "bias correction applied results." In the revised manuscript, Fig. 7 and the corresponding text have been revised on Page 16, Lines 404–448. The specific revised text is as follows:

"Figure 7: Statistical comparison of atmospheric profiles between the uncorrected NSGA-II retrievals and the collocated radiosonde observations for 38 matched cases collected during the Jiaozhou Bay field experiment. Panels (a) and (b) present the mean temperature and relative humidity profiles, respectively, after time matching and interpolation onto the common retrieval grid. Panels (c) and (d) show the corresponding profile differences, defined directly as the mean retrieved profiles minus the mean radiosonde profiles shown in panels (a) and (b)."

…

"Given that the magnitude of the mean ΔRMSE in this subset remains small, this result should be interpreted with caution and should not be generalized beyond the present dataset. Overall, these findings suggest that while the average impact of attitude correction is modest under weak-motion conditions, it remains physically necessary and is expected to become more important under conditions with larger viewing-angle deviations."



Comment 2

The discussion and conclusion section now seem to repeat many of the same points. Could this be condensed into a single section to avoid repetition? Also the conclusion has not been changed since the previous review round, and still makes claims such as the 'high precision' retrieval. Please bring this into line with the rest of the paper.

Response:

Thank you for the reviewer's careful evaluation and constructive suggestion. We agree that the Discussion and Conclusion sections in the previous version contained some overlap, and that some statements in the Conclusion were still too strong and not fully consistent with the more cautious wording adopted elsewhere in the revised manuscript. Following the reviewer's suggestion, we have condensed and coordinated these two sections to make their functions clearer. The Discussion now focuses mainly on interpreting the results, clarifying the limitations of the present study, and outlining future work. The Conclusion has been shortened into a concise summary of the main findings, with reduced repetition relative to the Discussion.

At the same time, we revised stronger expressions in the Conclusion. For example, wording such as "high-precision retrieval" was replaced with more objective and cautious expressions, such as "good retrieval accuracy" or "operationally useful retrieval performance under the present conditions," so that the conclusion is better aligned with the evidential scope of the dataset and the overall tone of the manuscript.

In the revised manuscript, the Discussion and Conclusion sections have been condensed and rewritten on Pages 22–24, Lines 557–595. The Discussion is now mainly used to interpret the results, describe limitations, and discuss future work, whereas the Conclusion has been shortened to summarize the main findings and remove or soften overly strong expressions such as "high precision." The specific revised text is as follows:

"The sea-trial results obtained in Jiaozhou Bay support the effectiveness and feasibility of the proposed buoy-based retrieval framework under the present nearshore observational conditions. For temperature, the retrieval contains a relatively stable height-dependent bias component that can be effectively reduced through the present LOOCV-based systematic bias correction. For relative humidity, the improvement is more limited and less vertically uniform, indicating that the residual error is still strongly influenced by case-dependent variability, reduced channel sensitivity aloft, and spatial mismatch between the buoy and the radiosonde station."

…

"Overall, the present study demonstrates the feasibility of buoy-based continuous remote retrieval of marine atmospheric profiles under sparse-data and dynamically varying observational conditions. However, the current results should be interpreted within the limits of the tested nearshore dataset and should not be generalized directly

to broader offshore or open-ocean environments without further validation."

Comment 3

Line 212: Some things here are badly formatted. Equation 2 seems badly rendered and does not match the other equations. T(Z) is not well aligned.

Line 213: here you mention tau, but this does not feature in the equation. You integrate physical temperatures and absorption coefficients over a height, but the optical depth/opacity does not figure in this.

Response:

Thank you for pointing this out. After re-checking the manuscript, we agree with the reviewer that the display format of Equation (2) in the original version was problematic, and that the notation in the equation was not fully consistent with the variable definitions in the following text. In particular, the text referred to the optical depth τ, but this term was not clearly represented in the equation, which could indeed cause confusion.

Following the reviewer's suggestion, we have reformatted and revised Equation (2) in the revised manuscript. The alignment and presentation of $T(z)$have been corrected, and the optical-depth term τ has now been explicitly included in the radiative-transfer expression. We also revised the variable definitions below the equation to ensure that all symbols appearing in the formula are clearly and consistently defined.

In the revised manuscript, the display and description of the equation have been revised on Pages 9–10, Lines 209–224. The explanatory text below the equation has also been updated accordingly. The specific revised text is as follows:

"Under the Rayleigh–Jeans approximation, which is valid for microwave frequencies (300 MHz–300 GHz), the brightness temperature T_B observed by a ground-based radiometer can be expressed by the radiative transfer equation (RTE):"

…

"Under the present non-precipitating conditions, the forward model focuses on gaseous absorption, while liquid-water effects are not explicitly retrieved."

Comment 4

Line 325: "In this study, the adjacent-layer limits are set to δ1 = 8 K for temperature *and $\delta 2$ = 60% for relative humidity."*

When this is applied to all heights regardless of the grid resolution, I would suspect that it does not have much effect. Your highest resolution (smallest spacing) on the vertical grid is 25m, and as you only prevent jumps of 8K between adjacent levels, this only prevents jumps of more than +/- 320K/km (compared to typical tropospheric lapse rates of -6 to -7 K/km). I suspect that this limit is not having so much effect in the retrieval... I don't think that this can reasonably called a lapse rate constraint either.

Response:

Thank you for this professional and physically insightful comment. We agree that it was not accurate to refer to this condition as a "lapse-rate constraint" in the original manuscript, because the present constraint is not defined according to the thermodynamic lapse rate and does not enforce a monotonic decrease of temperature with height. Instead, it only limits the absolute magnitude of temperature and relative-humidity differences between adjacent retrieval levels. Its main purpose is to suppress unrealistic layer-to-layer jumps and non-physical oscillations under limited observational constraints.

Following the reviewer's suggestion, we have revised the relevant wording from "lapse-rate constraint" to "adjacent-layer continuity constraint" throughout the revised manuscript. We also further clarify in the text that this constraint is not a strict lapse-rate condition and does not exclude local inversions or gradient-transition structures. In addition, we further checked the terminology used in the schematic workflow and revised the label in Fig. 5 from "Physical Constraints (Lapse Rate)" to "Physical Constraints (Adjacent-layer continuity)" to ensure consistency between the figure and the revised text.

In the revised manuscript, the terminology in Fig. 5 and the definition and physical meaning of the adjacent-layer constraint, as well as its distinction from a strict thermodynamic lapse-rate constraint, have been revised on Page 8, Fig. 5; Page 10, Lines 229–232; Page 12, Lines 279–289; and Pages 13–14, Lines 325–337. The specific revised text is as follows:

"The retrieval aims to simultaneously minimize the residuals between simulated and

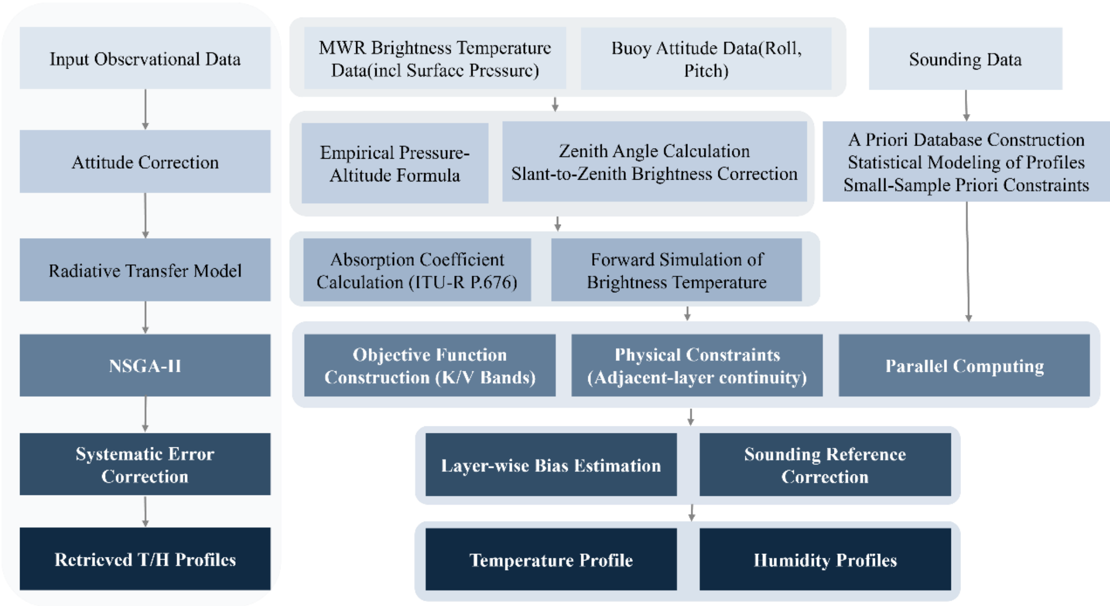
observed brightness temperatures in the temperature-sensitive V-band channels and the humidity-sensitive K-band channels under level-dependent admissible bounds and adjacent-layer continuity constraints. The forward model is based on the atmospheric microwave radiative transfer equation and the ITU-R P.676-13 absorption model."

"Based on this framework, the retrieval procedure consists of the following main steps: construction of the objective functions from brightness temperature observations, introduction of physical constraints, multi-objective optimization using the NSGA-II algorithm, and subsequent systematic bias correction to obtain the final temperature and humidity profiles. Addressing the challenges of sparse data, dynamic platform disturbances, and the high computational cost of traditional inversion methods requiring large training datasets in ocean observations, this study builds upon the NSGA-II (Deb et al., 2002) framework. It should be clarified that the proposed retrieval method is not based on the Optimal Estimation Method (OEM), but is formulated as a constrained multi-objective optimization problem solved using NSGA-II. Specifically, a small-sample prior database is constructed to reduce dependence on large training datasets through statistical boundary generation and physical boundary constraints, as described in Section 3.3.2. Physical constraints, including adjacent-layer continuity constraints and physically admissible temperature–humidity bounds, are introduced as hard constraints to suppress inversion results that violate basic physical plausibility."

"is constructed from the residuals between simulated and observed brightness temperatures in the K-band (22 GHz) and V-band (58 GHz)."

…

"Rather, it limits only the absolute magnitude of differences between neighbouring retrieval levels and does not enforce a monotonic decrease of temperature with height. Therefore, local inversions and gradient-transition structures are not excluded a priori. Under the current threshold setting, the practical role of this constraint is mainly to avoid non-physical discontinuities rather than to impose a canonical tropospheric lapse-rate structure."



Comment 5

Line 415: "The radiosonde profiles shown here represent ensemble-averaged means over 38 matched cases"

What is the meaning of an ensemble-average mean in the case of observations? What is the difference from a regular mean?

Response:

Thank you for pointing this out. We agree that the phrase "ensemble-averaged means" was not accurate in this context and could lead to ambiguity. Our intention was simply to indicate that the profiles shown in the figure are the mean profiles of the 38 matched cases after interpolation onto a common vertical grid, rather than a special statistical quantity different from a regular mean. Following the reviewer's suggestion, we have revised this wording to a more direct and standard expression.

In the revised manuscript, the relevant text has been rewritten on Pages 16–17, Lines 410–420. The specific revised text is as follows:

"Figure 7 presents the statistical comparison between the uncorrected NSGA-II retrievals and the collocated radiosonde observations for the 38 matched cases."

…

"The corresponding bias profile exhibits clear height-dependent variations, indicating the presence of structured systematic deviations prior to bias correction."

Comment 6

Figure 7: As previously mentioned, the zoom plots on here do not give any insight into the plots. For Temperature: either remove, or make the zoom in the first 2km of the plots where there are details that may not be fully appreciated by the reader. For the humidity plot: we see even less and it seems that there are fewer points in the zoom. I would advise that you remove the zoom from this plot. For plot c, the bias seems to cross the 0K lines at several points below 1km, but on the two mean plots the retrieved only exceeds the observed between ~1.3 and ~1.6km. However the bias is calculated, it should equal the difference between these two curves. This leads the reader to conclude that a) the bias was calculated from different data or b)additional processing was applied to one plot that wasn't on the other. Could the authors please explain this?

Response:

Thank you for the reviewer's careful reading and specific suggestions. We agree that this is an important point. After rechecking the plotting workflow of Fig. 7 in the previous manuscript, we found that the presentation of the mean profiles and the bias curves was not sufficiently clear, which could have led readers to question the correspondence between them. Following the reviewer's suggestion, we have re-plotted Fig. 7 and unified the calculation and plotting procedures for the main profile panels and the difference panels.

In the revised Fig. 7, the temperature and relative-humidity bias curves are calculated directly as the differences between the corresponding mean retrieved profiles and mean radiosonde profiles. In other words, the bias curves and the two mean curves in the main panels are based on exactly the same sample set, the same common vertical interpolation grid, and the same data-processing workflow. Therefore, no additional independent processing or different statistical treatment is involved. This revision makes the correspondence between Fig. 7(c) and Fig. 7(a), and between Fig. 7(d) and Fig. 7(b), clearer.

In addition, following the reviewer's suggestion, we removed the zoomed subpanels from Fig. 7 to make the figure more compact and clearer, and to avoid redundant visual information that might distract from the main comparison.

With these revisions, we hope that the revised Fig. 7 more directly and accurately presents the relationship between the mean profiles and the corresponding bias curves, and improves the interpretability of the results.

In the revised manuscript, Fig. 7 and the related text have been revised on Pages 16–17, Lines 405–448. The specific revisions have already been described in the preceding response and are therefore not repeated here.

Comment 7

Table 2: Could you comment on why the RMSE goes up for attitude angles above 2.5 ° ? Does this represent much of the dataset?

Response:

Thank you for this detailed question. We agree that the positive mean temperature ΔRMSE (+0.13 K) in the >2.5° attitude-angle group in Table 2 requires further explanation.

After rechecking the results, we consider that this result does not indicate that the attitude correction itself becomes ineffective under larger attitude-angle conditions. Rather, it more likely reflects the net statistical outcome of multiple error sources under limited-sample conditions. Physically, larger attitude angles imply more pronounced viewing-geometry deviations, and therefore attitude correction is theoretically more necessary. However, in practical marine observations, larger attitude angles are often accompanied by stronger platform motion and a more complex observational environment. In such cases, in addition to geometric deviation, short-term atmospheric variability, instrument noise, temporal–spatial mismatch, and changes in the air–sea environment may also contribute to the retrieval error. These factors may partly offset the geometric improvement introduced by attitude correction.

In the present dataset of 38 matched cases, the ≤1.5°, 1.5–2.5°, and >2.5° attitude-angle groups contain 4, 22, and 12 samples, respectively. Therefore, the >2.5° group is not dominated by a single outlier case. Nevertheless, considering that the magnitude of the mean temperature ΔRMSE in this group is only +0.13 K, this result should still be interpreted as a statistical outcome under the current limited-sample conditions and should not be generalized as a universal pattern. Following the reviewer's suggestion, we have added the sample numbers for each group and revised the related discussion in a more cautious manner.

In the revised manuscript, the sample numbers of the attitude-angle groups in Table 2 and the explanation for the slight RMSE increase in the >2.5° group have been added on Pages 17–18, Lines 427–448. The specific revised text is as follows:

"The retrieval results shown in Fig. 7 already incorporate attitude correction in the forward modelling process. To further isolate and evaluate the effect of attitude correction, an additional comparison was conducted between retrievals with and without applying the attitude correction, prior to the introduction of systematic bias correction."

…

"Given that the magnitude of the mean ΔRMSE in this subset remains small, this result should be interpreted with caution and should not be generalized beyond the present dataset. Overall, these findings suggest that while the average impact of attitude correction is modest under weak-motion conditions, it remains physically necessary and is expected to become more important under conditions with larger viewing-angle deviations."

Comment 8

Figure 9: you state here 'based on 38 valid matchups', so is this figure made with all cases? If so then I am not sure how to interpret the systematic bias correction described in lines 357-360, where you state "The 38 collocated samples were randomly divided into a training set (80%) and an independent testing set (20%). The training set was used to derive the systematic bias correction model, while the testing set was used exclusively for independent validation of the correction performance."

Shortly afterwards you state (line 366): "After the systematic error model is established using this method, it is applied to correct all retrieval profiles."、

Is the plot in figure 9 made by applying the bias correction, that was made with 30 out of 38 cases, to all 38 cases? If so I struggle to see the value of this. It would be better to compare the stats from the 8 testing cases before vs after bias correction. Even though it is not a big sample size, it is more meaningful.

Response:

Thank you very much for this important and constructive comment. We agree that the description of the systematic bias correction procedure and the presentation of Fig. 9 in the previous manuscript were not sufficiently clear. In particular, the original description based on a single 80 % / 20 % train–test split could easily lead readers to understand that the bias-correction term estimated from only part of the dataset was then directly applied to all 38 samples, thereby creating ambiguity in the statistical interpretation of Fig. 9.

In response to this concern, and also in accordance with the other reviewer's suggestion to adopt a more robust validation strategy, we have replaced the previous single train–test split with a leave-one-out cross-validation (LOOCV) framework. Specifically, for each of the 38 collocated samples, one sample is held out as the validation case, while the remaining 37 samples are used to estimate the height-dependent systematic-bias profiles for temperature and relative humidity. The estimated bias profiles are then applied only to the held-out sample in that validation round. This procedure is repeated until all 38 samples have been independently held out once, and the final statistics are obtained by aggregating the corrected held-out predictions from all validation rounds.

Therefore, the revised Fig. 9 is indeed based on all 38 valid matched samples, but each corrected profile shown in the figure is a held-out prediction generated within the cross-validation framework. In other words, the bias-correction term for each sample is estimated without using that sample itself. Thus, the revised Fig. 9 no longer represents the case in which a bias correction derived from 30 samples is directly applied to all 38 samples; instead, it summarizes the LOOCV-based corrected results for all 38 samples while preserving sample independence in the correction procedure.

We believe that this revision removes the ambiguity in the statistical interpretation of Fig. 9 and makes the evaluation of the systematic bias correction more robust and meaningful. Accordingly, we have revised Section 3.3.3, the Fig. 9 caption, and the related Results text.

(1) In the revised manuscript, the description of the systematic bias correction method in Section 3.3.3 has been revised on Page 15, Lines 358–387. The previous single 80 % / 20 % train–test split has been replaced by the LOOCV strategy. The specific revised

text is as follows:

"To obtain a more robust estimate of the correction performance under limited-sample conditions, the systematic bias correction was evaluated using a leave-one-out cross-validation (LOOCV) strategy over the 38 collocated samples. In each validation round, one sample was held out as the validation case, while the remaining 37 samples were used to estimate the height-dependent systematic bias profiles for temperature and relative humidity."

…

"The corrected held-out profiles obtained in all LOOCV rounds were then aggregated to form the final LOOCV-based corrected dataset. In the present study, radiosonde observations are used only within the validation framework to estimate the height-dependent systematic bias profiles and to assess the correction performance. For future operational applications over the open ocean, the transferability and stability of the derived bias characteristics will require further validation under broader environmental conditions."

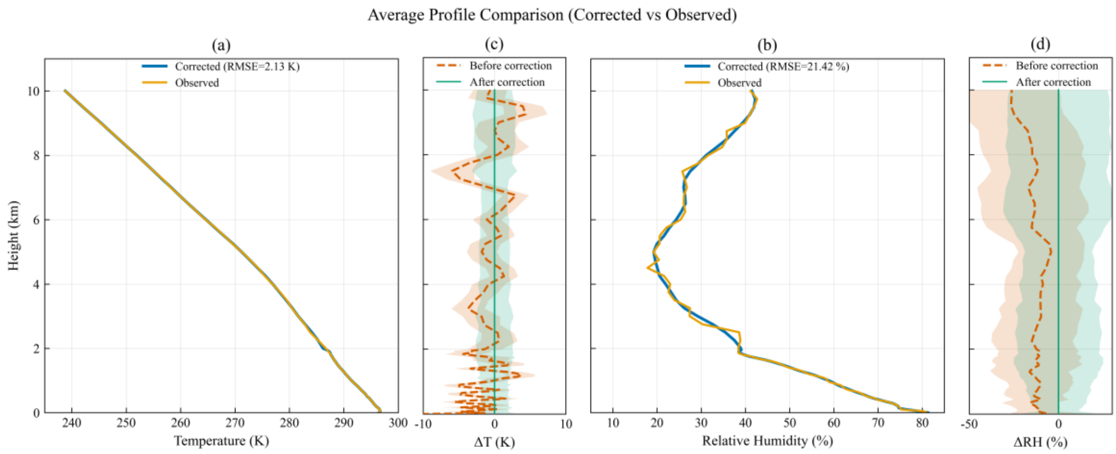
(2) In the revised manuscript, the Fig. 9 caption and the corresponding Results text have been revised on Pages 20–21, Lines 479–522. The numerical values and descriptions in Table 4 have also been updated accordingly, and the corrected results are now explicitly described as held-out predictions under the cross-validation framework. The specific revised text is as follows:

"Figure 9: Average-profile comparison between the LOOCV-corrected retrievals and the collocated radiosonde observations over 38 valid matched cases on the common vertical grid. Panels (a) and (b) show the mean corrected and mean observed temperature and relative humidity profiles, respectively. Panels (c) and (d) show the corresponding mean bias profiles before and after correction. In panels (c) and (d), the shaded regions denote ±1 standard deviation of the profile errors across the 38 matched cases."

…

"In summary, the sea-trial results support the effectiveness and feasibility of the proposed method under the present nearshore observational conditions. By

constructing a small-scale prior experience database and integrating platform attitude information, the NSGA-II-based retrieval framework reduces the dependence on large historical training datasets. Under the LOOCV-based correction framework, the retrieval results after systematic bias correction should be interpreted primarily as evidence of retrieval feasibility and of the usefulness of height-dependent bias mitigation under the tested buoy-based nearshore conditions, rather than as a generalized validation across all marine environments."



Comment 9

Figure 10: It would be more informative to see the effects of constraints if the plot showed a retrieval with and without constraints (as well as observations). Now that it has been explained that the 'lapse rate constraint' simply did not allow 8K jumps, the plot as it is not so informative.

Response:

Thank you very much for this valuable and constructive suggestion. We agree that the previous version of Fig. 10 was not sufficiently informative, because it only compared the constrained retrieval profile with the observed profile and did not directly show the effect of the constraint itself on the retrieval result.

Following the reviewer's suggestion, we have revised Fig. 10. In the revised manuscript, the temperature retrieval results obtained with and without the adjacent-layer continuity constraint are both shown, together with the corresponding radiosonde profile. The revised figure therefore provides a more direct illustration of the influence of the adjacent-layer continuity constraint on the retrieved profile structure.

At the same time, we further clarified in the text that the purpose of introducing this constraint is not to guarantee a lower RMSE for every individual case. Rather, it is mainly intended to suppress unrealistic layer-to-layer oscillations and jumps and to maintain basic profile continuity under limited observational constraints. Therefore, in some individual cases, the unconstrained solution may happen to be numerically closer to the radiosonde profile and thus yield a lower RMSE, whereas the constrained solution exhibits stronger structural regularization. In this sense, Fig. 10 should be interpreted primarily as an illustration of the structural effect of the continuity constraint on the retrieved profile, rather than as evidence that the constraint necessarily improves the retrieval accuracy of every individual sample.

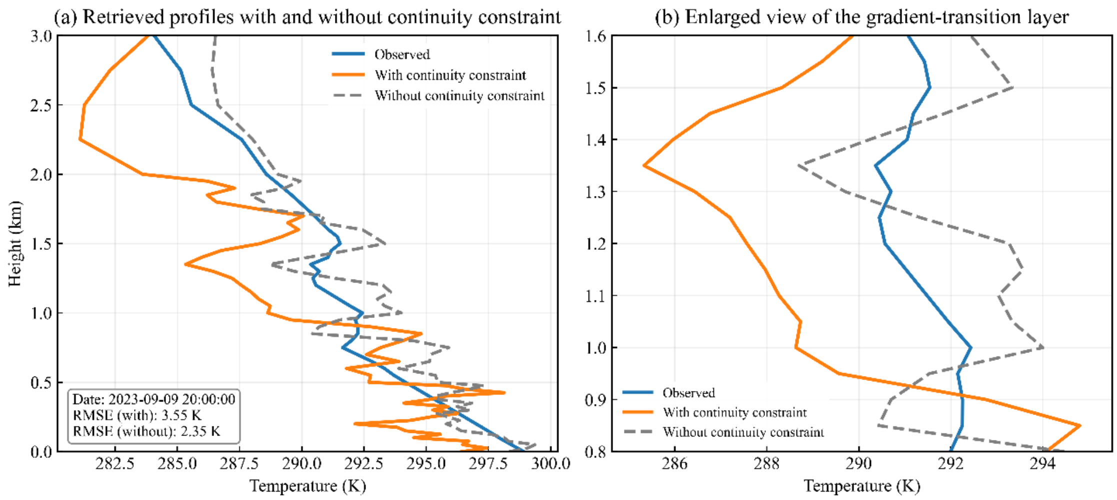
Accordingly, we have revised the corresponding text and the Fig. 10 caption to make them consistent with the revised figure.

In the revised manuscript, Fig. 10 on Page 22 has been updated to show the comparison between temperature retrievals with and without the adjacent-layer continuity constraint. The corresponding text has also been revised on Pages 21–22, Lines 529–542, to clarify that the constraint mainly affects the structural smoothness and layer-to-layer variability of the retrieved profile. The figure caption now explicitly states that the constraint does not necessarily reduce the RMSE for every individual case. The specific revised text is as follows:

"To more directly illustrate the structural effect of the adjacent-layer continuity constraint on the retrieved temperature profile, Fig. 10 presents a representative case comparing retrievals obtained with and without the continuity constraint, together with the collocated radiosonde profile. It should be emphasized that the purpose of this constraint is not to guarantee a lower RMSE for every individual case, but to suppress unrealistic layer-to-layer variability and to maintain basic profile continuity under limited observational constraint."

…

"Under limited observational constraints, a less regularized solution may fit the collocated radiosonde profile more closely in an individual case, but such a profile may also contain stronger small-scale oscillations and weaker physical smoothness."



Response RC2

Comments 1

L436: The manuscript refers to liquid water path (LWP) as a diagnostic for cloud contamination. Does this imply that the retrieval is only valid under clear-sky (cloud-free) conditions? If so, this limitation should be stated explicitly. It would also be helpful to discuss the implications for the method's applicability over the open ocean, where such conditions may not always be met.

Response:

Thank you very much for this valuable comment. We agree that the role of liquid water path (LWP) and the applicability boundary of the present retrieval framework were not stated clearly enough in the previous version, which could have led to misunderstanding. We would like to clarify that LWP is not used as an inversion input variable in this study. Instead, it is used only as an auxiliary diagnostic indicator to help identify possible cloud contamination and the associated residual retrieval uncertainty. The current retrieval framework is mainly based on radiative transfer processes associated with atmospheric gaseous absorption. It does not explicitly retrieve cloud-liquid-water profiles, nor does it independently correct for cloud-liquid-water absorption or scattering effects. This is consistent with the original intention of the manuscript, where LWP was used only as an auxiliary diagnostic quantity rather than as a retrieval variable. Therefore, the validation results presented in this study should be interpreted as demonstrating the applicability of the method mainly under non-precipitating conditions with weak or limited cloud-liquid-water influence. They should not be directly generalized as comprehensive validation under all cloud conditions. We accept the reviewer's suggestion and agree that this applicability boundary should be explicitly stated in the manuscript, particularly with respect to the implications for broader offshore and open-ocean applications under more complex cloud conditions. This clarification is necessary because, although the Results section mentioned LWP as an auxiliary indicator, the applicable scope of the method was not sufficiently defined in the previous version.

Following the reviewer's suggestion, we have explicitly added this limitation in Section 3.2 and in the Discussion section of the revised manuscript. We further note that, for open-ocean applications under stronger cloud-liquid-water influence or precipitation conditions, the performance of the current method still requires further validation using additional observational experiments. Future work will consider incorporating explicit cloud-liquid-water treatment to improve the applicability of the method under more complex marine weather conditions.

In the revised manuscript, these changes have been made on Pages 9–10, Lines 216–224, and on Page 23, Lines 567–571. The specific revised text is as follows:

"Accurate calculation of k_α *is critical for the forward model. In this study, we selected the ITU-R P.676-13 model (ITU-R, 2022) to calculate the gaseous absorption coefficients. Compared to the MPM series and MonoRTM models mentioned above, the ITU-R model is recognised as the international standard for radio propagation engineering. It uses a parameterized line-by-line formulation that offers a good balance between computational efficiency and accuracy, making it particularly suitable for the real-time retrieval requirements of the present buoy-based system. Specifically, the K-band channels (22–31 GHz) operate near the water-vapor absorption line (22.235 GHz) and are mainly used to retrieve humidity information, while the V-band channels (51–58 GHz) operate along the oxygen absorption complex and are mainly used to retrieve temperature profiles (Philip, 1998). Under the present non-precipitating conditions, the forward model focuses on gaseous absorption, while liquid-water effects are not explicitly retrieved."*

"A further limitation is that the present retrieval framework is mainly validated under

non-precipitating and relatively weak-cloud-liquid-water conditions. In the current study, liquid water path (LWP) is used only as an auxiliary diagnostic indicator for possible cloud contamination and residual retrieval uncertainty, rather than as an inversion variable. Therefore, the present results are more appropriately interpreted as demonstrating retrieval feasibility under the tested nearshore conditions, rather than as establishing the same level of accuracy for broader offshore or open-ocean applications under all weather conditions."

Comments 2

Training and testing of the System Error Correction: It would strengthen the analysis to apply a cross-validation approach (e.g., as implemented in scikit-learn: https://scikit-learn.org/stable/modules/cross_validation.html). This method is straightforward to implement and generally improves the robustness of performance estimates. Relying on a single train–test split can lead to results that are sensitive to the particular data partition (lucky pick) and may introduce bias due to a favorable or unrepresentative split.

Response:

Thank you very much for this valuable and constructive suggestion. We fully agree that, under limited-sample conditions, relying on a single train–test split may make the performance evaluation sensitive to the specific data partition. If the split happens to be favourable or unrepresentative, the resulting evaluation may also be biased to some extent.

Following the reviewer's suggestion, we have replaced the previous single 80 % / 20 % train–test split with a leave-one-out cross-validation (LOOCV) framework. Specifically, for each of the 38 collocated samples, one sample is held out as the validation case, while the remaining 37 samples are used to estimate the height-dependent systematic-bias profiles for temperature and relative humidity. The estimated bias profiles are then applied only to the held-out sample in that validation round. This procedure is repeated until all 38 samples have been independently held out once, and the final statistics are obtained by aggregating the corrected held-out predictions from all validation rounds. The results throughout the manuscript have also been updated accordingly to ensure consistency.

We believe that this revision substantially improves the robustness and interpretability of the evaluation of the systematic-bias-correction performance, because each sample is validated without being used in the estimation of its own bias-correction term. Accordingly, we have rewritten the systematic-bias-correction section under the LOOCV framework in the revised manuscript. We have also clarified in the Fig. 9 caption and the related Results text that the corrected profiles shown in the figure are held-out predictions under the cross-validation framework, rather than results obtained from a single train–test split.

(1) In the revised manuscript, the description of the systematic bias correction method in Section 3.3.3 has been revised on Page 15, Lines 358–387. The previous single 80 % / 20 % train–test split has been replaced by the LOOCV strategy. The specific revised text is as follows:

"To obtain a more robust estimate of the correction performance under limited-sample conditions, the systematic bias correction was evaluated using a leave-one-out cross-validation (LOOCV) strategy over the 38 collocated samples. In each validation round, one sample was held out as the validation case, while the remaining 37 samples were used to estimate the height-dependent systematic bias profiles for temperature and relative humidity."

…

"The corrected held-out profiles obtained in all LOOCV rounds were then aggregated to form the final LOOCV-based corrected dataset. In the present study, radiosonde observations are used only within the validation framework to estimate the height-dependent systematic bias profiles and to assess the correction performance. For future operational applications over the open ocean, the transferability and stability of the derived bias characteristics will require further validation under broader environmental conditions."

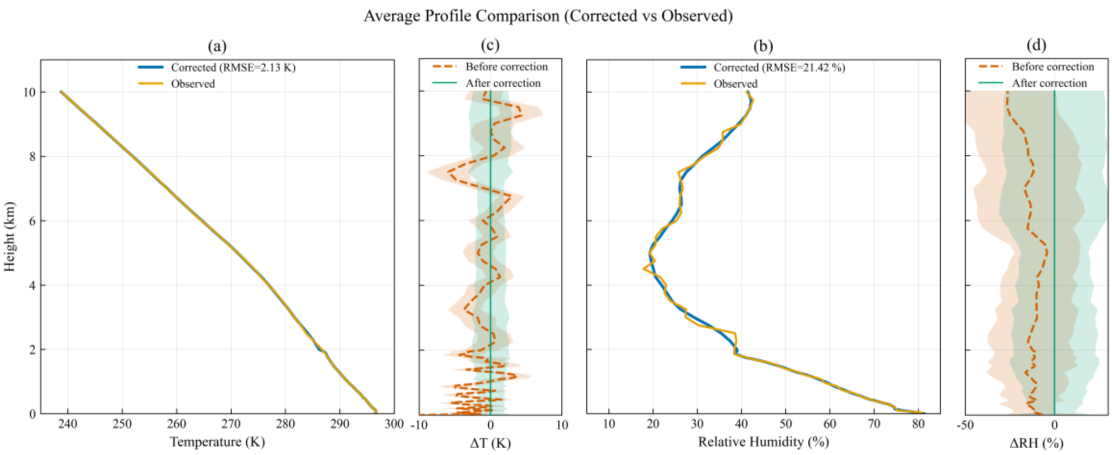
(2) In the revised manuscript, the Fig. 9 caption and the corresponding Results text have been revised on Pages 20–21, Lines 479–522. The numerical values and descriptions in Table 4 have also been updated accordingly, and the corrected results are now

explicitly described as held-out predictions under the cross-validation framework. The specific revised text is as follows:

"Figure 9: Average-profile comparison between the LOOCV-corrected retrievals and the collocated radiosonde observations over 38 valid matched cases on the common vertical grid. Panels (a) and (b) show the mean corrected and mean observed temperature and relative humidity profiles, respectively. Panels (c) and (d) show the corresponding mean bias profiles before and after correction. In panels (c) and (d), the shaded regions denote ±1 standard deviation of the profile errors across the 38 matched cases."

…

"In summary, the sea-trial results support the effectiveness and feasibility of the proposed method under the present nearshore observational conditions. By constructing a small-scale prior experience database and integrating platform attitude information, the NSGA-II-based retrieval framework reduces the dependence on large historical training datasets. Under the LOOCV-based correction framework, the retrieval results after systematic bias correction should be interpreted primarily as evidence of retrieval feasibility and of the usefulness of height-dependent bias mitigation under the tested buoy-based nearshore conditions, rather than as a generalized validation across all marine environments."



Comments 3

Figure 6: I can't find any reference to this illustration in the manuscript.

Response:

Thank you for the reviewer's careful reading and for pointing this out. We agree that the reviewer is correct. After checking the manuscript, we found that this was indeed a figure-numbering error. In the text discussing the probability-density characteristics of temperature and relative humidity at the 50 m height level in January, the corresponding illustration was incorrectly referred to as "Figure 5," whereas it should have been cited as "Figure 6."

Following the reviewer's comment, we have corrected this figure reference to Figure 6 in the revised manuscript, so that the text description is consistent with the figure numbering and readers will not be confused.

The specific revision is as follows:

In the revised manuscript, on Page 11, Line 269, the figure reference has been corrected to Figure 6.