General Response to the Editor and Reviewers

We sincerely thank the Editor and both reviewers for their careful evaluation of our manuscript and for the detailed and constructive comments. We fully recognize that the original version of the manuscript required substantial revision. The reviewers' comments were very helpful in identifying weaknesses in the presentation, methodological description, and interpretation of the results, and they have significantly helped us improve the overall quality and clarity of the manuscript.

Following the reviewers' suggestions, we have carefully revised the manuscript throughout. All comments have been addressed point-by-point in this response document. Corresponding modifications have been clearly indicated in the revised manuscript with page and line numbers for easy reference.

The main revisions made in the manuscript include:

- Updated and consolidated author affiliations on the title page to accurately reflect current official institutional names and organizational structures (e.g., the integration of the State Key Laboratory of Physical Oceanography and the Institute of Oceanographic Instrumentation as per current official records).Clarification of the methodological framework, including the role of historical radiosonde data, prior database construction, systematic bias characterization, and the online retrieval process;
- Revision and reorganization of the forward modelling and retrieval method descriptions to improve physical accuracy and logical clarity;
- Addition of instrument performance and uncertainty information for both the microwave radiometer and the reference radiosonde observations;
- Improvement of the statistical analysis and data quality-control description, including clarification of the $\mu \pm 2\sigma$ criterion and the introduction of skewness and excess kurtosis indicators;
- Revision of figures and related descriptions to improve consistency between main panels and zoomed panels and to clarify the profile comparison methodology;
- Adjustment of wording throughout the manuscript to avoid overly strong or potentially misleading statements and to ensure a more cautious interpretation of retrieval performance;
- Correction of terminology and typographical errors identified by the reviewers.

We hope that the revisions have adequately addressed the reviewers' concerns. We would be grateful for any further comments or suggestions from the Editor and reviewers and would be happy to make additional revisions if necessary.

Below, we provide detailed point-by-point responses to all reviewer comments.

Response to RC1

Line 11 - something is missing to relate lack of measurements over the sea to microwave radiometers' ability to do this

Response:

We sincerely thank the reviewer for this valuable and insightful comment. We agree that the original abstract did not sufficiently emphasize the logical connection between the scarcity of routine in situ observations over oceanic regions and the unique capability of microwave radiometers to provide continuous atmospheric profiling.

Following the reviewer's suggestion, we have revised the abstract to more clearly establish this relationship and to better highlight the motivation of the present study.

In the revised manuscript, we have added the following sentence in the abstract (Page 1, Lines 12–14):

"Ground-based microwave radiometers offer a unique capability for continuous, all-weather remote sensing of atmospheric thermal emission, enabling routine retrieval of temperature and humidity profiles over oceanic regions."

This addition strengthens the scientific motivation of the study by explicitly highlighting the unique advantage of microwave radiometers in compensating for the lack of routine atmospheric observations over the ocean.

Line 16 - You state several times that the algorithm does not rely on historical data but later say that it relies on a limited amount of historical data. Please state clearly in the abstract, and later on, what exactly are the requirements in terms of radiosounding data.

Response:

We appreciate the reviewer pointing this out. We agree that the original wording regarding the dependence on historical data was not sufficiently precise and may have caused ambiguity.

The proposed method does not rely on large-scale historical datasets for training a retrieval model. Instead, only a limited amount of local radiosonde data is required to construct a small-sample prior database, which serves solely as an initial physical constraint for the retrieval process. During subsequent online retrieval, no historical data input is required.

Following the reviewer's suggestion, we have revised both the abstract and the methodological description to clearly and explicitly define the role and data requirements of the radiosonde observations in the proposed framework.

1. Revision in the abstract (Page 1, Line 16-19):

The abstract has been revised as follows:

"This method does not rely on large-scale historical datasets for model training and integrates platform attitude information. Our approach uses a multi-objective genetic algorithm to construct a small-scale joint prior database based on a limited amount of local radiosonde data, which serves only as an initial physical constraint for the retrieval process."

2. Additional clarification in Section 3.3.2 (Page 10, Lines 246–248):

We have added the following explanation at the beginning of Section 3.3.2:

"It should be emphasized that only a limited amount of local radiosonde data is required to construct the small-sample prior experience database, which serves solely as an initial physical constraint for the retrieval process. During subsequent online retrieval, no historical data input is required."

Through these revisions, the role of the radiosonde data and the actual data requirements of the proposed method are now clearly and explicitly defined, thereby eliminating the ambiguity regarding historical data dependence in the original manuscript.

Line 18 - 'pressure-altitude model' - you present an unreferenced empirical formula later, is this what you mean or is there something more complicated?

Response:

We sincerely thank the reviewer for this careful and important comment. We confirm that the "pressure–altitude model" mentioned in the abstract indeed refers to the empirical pressure–altitude equation presented in Section 3.3.1 (Eq. 3). We agree that the original wording in the abstract was not sufficiently precise and may have caused ambiguity.

Following the reviewer's suggestion, we have revised the terminology in the abstract to explicitly reflect the empirical nature of the equation. In addition, we have revised Eq. (3) and expanded the physical and empirical background description in Section 3.3.1 to clearly document its origin, formulation, and applicability.

1. Revision in the abstract (Page 1, Line 20):

We have revised the wording from

"a pressure–altitude model" to *"an empirical pressure–altitude equation"*

to more accurately reflect the formulation's empirical nature.

2. Revision and expansion of Eq. (3) in Section 3.3.1 (Page 10, Lines 231–244):

We have revised Eq. (3) and added a complete physical and empirical background description. The pressure profile is now explicitly defined as a weighted combination of two components:

- A polynomial regression derived from historical pressure–altitude statistics of major cities in China (regional empirical climatology);
- The standard atmospheric pressure model is defined in the national electric power industry code DL/T 5240–2010.

The revised text reads as follows:

" *Furthermore, pressure profiles were obtained using actual ground-based pressure sensor data combined with an empirical pressure-altitude equation. To obtain pressure profiles for the specific experimental region, we adopted a hybrid empirical approach that combines local statistical climatology with standard engineering modeling. The pressure P (kPa) at a given altitude H (m) is calculated as the arithmetic mean of two components eq.(3):*

$$P(H) = 0.5 \times P_{local}(H) + 0.5 \times P_{std}(H)$$

Where:

$P_{local}(H)$ *represents the regional statistical fit, derived from a polynomial regression of historical pressure-altitude data from major cities in China:*

$$P_{local} = 101.3 \times (5.3788 \times 10^{-9} H^2 - 1.1975 \times 10^{-4} H + 1)$$

$P_{std}(H)$ *represents the standard atmospheric model defined in the national electric power industry code DL/T 5240-2010(National Energy Administration):*

$$P_{std} = 101.3 \times \left[1 - 0.0255 \times \frac{H}{1000}\left(\frac{6357}{6357 + \frac{H}{1000}}\right)\right]^{5.256}$$

This weighted combination effectively mitigates the systematic bias of the general standard model while retaining the physical consistency required for vertical profiling."

In addition, we have explicitly clarified that this empirical pressure–altitude equation is a hybrid empirical model designed to correct the systematic bias of the standard atmospheric model under the local terrain conditions of the experimental region, thereby improving regional applicability.

The standard model component has now been properly referenced as:

National Energy Administration of China: Code for Design of Combustion System of Fossil Fired Power Plant (DL/T 5240-2010), China Electric Power Press, Beijing, 2010.

Line 28 - 'unprecedented challenges' - this is too strong. Explain concisely and clearly, after introducing that microwave radiometers are able to profile

atmospheric temperature and humidity, that their use in the marine environment is further complicated.

Response:

We sincerely thank the reviewer for this constructive comment. We agree that the phrase "unprecedented challenges" was overly subjective and exaggerated, and does not conform to the objective and precise tone required in scientific writing.

Following the reviewer's suggestion, we have revised the sentence by removing the subjective wording and replacing it with a more factual and engineering-oriented description that clearly conveys the additional complexity of applying microwave radiometers in the marine environment.

In the revised manuscript, the sentence on Page 2, Lines 31–32 has been modified from:

"… unprecedented challenges are presented for the precise detection of atmospheric parameters due to the unique geographical characteristics and dynamic processes of the marine environment."

to:

"… the precise detection of atmospheric parameters in the marine environment is further complicated by its unique geographical characteristics and dynamic processes."

This revision ensures a more objective, rigorous, and scientifically appropriate description of the observational challenges associated with the marine environment.

Line 31 - citation needed here

Response:

We sincerely thank the reviewer for pointing out the lack of appropriate literature support at this location. We agree that the statement regarding the complexity of marine atmospheric structures under extreme weather conditions requires authoritative observational evidence.

Following the reviewer's suggestion, we have added relevant and well-established references to support this discussion.

In the revised manuscript, we have added two authoritative references on Page 2, Line 34 to support the statement regarding the dynamical and thermodynamical complexity of the marine atmospheric boundary layer under extreme weather conditions:

- Guimond, S. R., Zhang, J. A., Sapp, J. W., and Frasier, S. J.: *Coherent turbulence in the boundary layer of Hurricane Rita (2005) during an eyewall replacement cycle*, Journal of the Atmospheric Sciences, 75, 3071–3093, 2018.
- Ahern, K., Bourassa, M. A., Hart, R. E., Zhang, J. A., and Rogers, R. F.:

Observed kinematic and thermodynamic structure in the hurricane boundary layer during intensity change, Monthly Weather Review, 147, 2019.

These studies provide comprehensive observational evidence of the complex dynamical and thermodynamical structures of the marine boundary layer under extreme weather conditions, thereby strengthening the scientific support for the original statement.

Line 36 - sparse distribution of what?

Response:

We sincerely thank the reviewer for pointing out the ambiguity in the original wording. We agree that the phrase "sparse distribution" was not sufficiently specific and could lead to misunderstanding.

Following the reviewer's suggestion, we have revised the sentence to explicitly indicate that the sparse distribution refers to the limited spatial coverage of oceanic radiosonde stations.

In the revised manuscript, the sentence on Page 2, Line 39 has been modified from: *"This sparse distribution makes data acquisition in the marine environment extremely difficult."* to: *"This sparse distribution of oceanic sounding stations makes data acquisition in the marine environment extremely difficult."*

This revision clarifies the subject of the sparse distribution and eliminates the ambiguity in the original expression.

Line 38 - This isn't necessarily true. OEM/1D-Var methods do not necessarily rely on large amounts of training data. What do you mean by 'potential interference and correction issues'? The results that you present here also involve a systematic height-dependent correction, which is based on radiosondings, so I am not sure that NOGA-II framework avoids this problem.

Response:

We sincerely thank the reviewer for this careful and constructive comment. We agree that the original wording regarding data dependence among different retrieval methods was not sufficiently rigorous, that the description of "potential interference and correction issues" was ambiguous, and that the role of the NSGA-II framework in handling systematic errors required further clarification.

Accordingly, we have revised the manuscript to address these issues in three aspects:

1. On the data dependence of different retrieval methods:

   We now clearly distinguish between physics-based retrieval approaches (e.g., OEM/1D-Var), which are physical constraint–driven frameworks and do not require large training datasets, and data-driven approaches (e.g., statistical and neural-network-based models), which rely on large historical datasets for model training.

2. On the meaning of "interference and correction issues":
   We have replaced the vague term "interference" with a clear physical description, namely, brightness temperature deviations caused by buoy-platform attitude variations (e.g., zenith-angle changes). If uncorrected, these effects introduce systematic measurement errors.
3. On the role of attitude-related systematic errors and the NSGA-II framework:
   We now explicitly state that platform-attitude effects are treated as a measurement preprocessing problem and are corrected before the retrieval stage via attitude calibration and bias correction. The NSGA-II framework is used solely as a multi-objective optimization tool during the retrieval stage and is not intended to model or eliminate attitude-related systematic errors.

In the revised manuscript, we have systematically revised the relevant text on Page 2, Lines 40–45. The original sentence:

"Specifically, these limitations include an excessive reliance on large amounts of historical data and an inability to effectively account for potential interference and correction issues. These issues are caused by changes in the attitude of the detection platform on observational signals."

has been replaced by:

"Specifically, data-driven retrieval methods such as statistical and neural-network-based approaches exhibit strong dependence on large historical datasets, which are difficult to obtain in marine environments. In addition, buoy-based platforms are subject to wave-induced motion, which leads to variations in the observation zenith angle and consequently introduces brightness temperature deviations. These attitude-related effects must be addressed through appropriate measurement preprocessing and correction procedures before the retrieval stage."

This revision clarifies the methodological distinctions, provides a physically explicit explanation of the measurement interference, and properly defines the role of the NSGA-II framework within the overall retrieval workflow.

Line 53 - You talk about research from the past 20 years and then mention research from almost 50 years ago as if this follows on from the other two.

Response:

We sincerely thank the reviewer for pointing out the chronological inconsistency in the literature discussion. We agree that mixing studies from the past two decades with work from nearly five decades ago in a single sequence created confusion and weakened the logical flow of the argument.

Following the reviewer's suggestion, we have reorganized this paragraph to clearly distinguish early pioneering studies from more recent developments, and to present the literature in a coherent temporal progression.

In the revised manuscript, the paragraph on Pages 2–3, Lines 58–64 has been reorganized according to chronological order. Early foundational studies are now introduced first (Decker et al., 1978; Guiraud et al., 1979), followed by more recent advances (Turner et al., 2007; Candlish et al., 2012), thereby establishing a clear timeline of methodological development.

The revised text reads:

"Early pioneering studies established the fundamental capabilities of ground-based microwave radiometry. For instance, Decker et al. (1978) made significant progress in ground-based detection, and Guiraud et al. (1979) revealed the detection capabilities of absorption spectra at different frequencies. However, applying these traditional methods to complex marine environments presents significant challenges. More recently, Turner et al. (2007) noted that physical retrieval methods, while accurate, are computationally inefficient for real-time applications. Furthermore, in their Arctic region study, Candlish et al. (2012) explicitly pointed out ..."

This revision ensures a clear chronological structure and improves the coherence and readability of the literature review.

Line 61 - One alternative (if feasible) is mechanical stabilisation such as Schnitt et al. (2024). If not used, please justify why and quantify the residual pointing variability.

Response:

We sincerely thank the reviewer for this valuable and technically important suggestion. We fully agree that mechanical stabilization systems (e.g., Schnitt et al., 2024) represent a feasible and effective approach to mitigating platform-motion effects. We also acknowledge the importance of justifying the chosen technical solution and of quantitatively characterizing the residual pointing variability.

Accordingly, we have added a detailed engineering justification for not adopting mechanical stabilization on a long-term autonomous buoy platform, and we have quantified the residual pointing variability together with a description of the corresponding software-based compensation strategy.

1. Feasibility of mechanical stabilization for long-term autonomous buoy operation:

Although mechanical stabilization systems can effectively reduce platform-motion-induced observation errors, their application on long-term autonomous buoy platforms faces significant challenges in terms of power consumption, mechanical complexity, maintenance requirements, and long-term reliability under harsh marine environmental conditions. The buoy platform used in this study is designed for extended unattended operation and is therefore subject to strict constraints on power budget and system robustness. Introducing an active mechanical stabilization device would substantially increase system complexity and power demand, thereby reducing operational stability and endurance. For these engineering reasons, mechanical stabilization was not adopted in the present system design.

2. Quantification of residual pointing variability and software-based compensation:

Instead, this study adopts a software-based real-time correction strategy using high-frequency attitude measurements. The buoy is equipped with a MEMS-based attitude sensor integrating a three-axis accelerometer, gyroscope, and digital motion processing unit (DMP), with a typical static accuracy of 0.1°–0.3° and dynamic accuracy of 0.3°–0.5°.

Wave-induced buoy motion leads to continuous variations in pitch and roll, resulting in dynamic changes in the observation geometry. During data processing, the effective zenith angle is calculated for each radiometer integration period using the real-time attitude measurements and is explicitly incorporated into the forward radiative transfer model. Brightness temperatures are then corrected on a sample-by-sample basis for geometric effects. In this way, residual pointing variability is treated as a measurable, computable, and correctable error source and is explicitly compensated in the retrieval process.

In the revised manuscript, we have added the following clarification on Page 3, Lines 67–74:

"These studies collectively highlight a core issue: how to effectively overcome errors caused by platform attitude and achieve high-precision, high-efficiency retrieval of atmospheric parameter profiles. Although mechanical stabilization (Schnitt et al., 2024) can reduce platform motion, it introduces significant challenges for long-term

autonomous buoy operation in terms of power consumption and maintenance complexity. Therefore, this study adopts a software-based approach using real-time zenith angle correction from attitude sensors. During field experiments, the buoy exhibited residual pitch and roll variations of approximately ±2.5° and ±3.2°, respectively. Our algorithm explicitly compensates for these residual pointing variations, enabling accurate retrieval of atmospheric parameters under dynamic marine conditions."

These additions improve the transparency of the system design rationale and provide a quantitative characterization of the residual pointing variability.

Line 63 - You mention MWRs as if they are introduced for the first time here. I don't think this is the case. Please rearrange the paragraphs so that this is in the correct place or adapt this paragraph to what has come before it

Response:

We sincerely thank the reviewer for pointing out this issue regarding the narrative logic. We agree that the original wording could give the misleading impression that microwave radiometers (MWRs) were being introduced at this point in the manuscript, which disrupts the continuity of the discussion.

Following the reviewer's suggestion, we have revised the paragraph structure and adapted the wording to ensure that it is clearly connected to the preceding discussion and reflects the fact that MWRs have already been introduced earlier in the manuscript. In the revised manuscript, we have reorganized the paragraph and modified the wording on Page 3, Line 75. The revised text now begins with:

"As discussed previously, ground-based microwave radiometers (MWR) provide a powerful remote sensing technology ..."

This revision improves the logical flow of the manuscript and ensures consistency in the introduction and subsequent discussion of MWRs.

Line 70 - What is meant by flexible buoy technology?

Response:

We sincerely thank the reviewer for pointing out the ambiguity in the expression "flexible buoy monitoring platforms." We agree that this wording was insufficiently precise and could lead to a misunderstanding of the technical meaning.

Following the reviewer's suggestion, we have revised the terminology to more clearly describe the technical characteristics of the buoy platform used in this study.

In the revised manuscript, the phrase on Page 3, Line 82-83 has been modified from:

“flexible buoy monitoring platforms” to: *“autonomous, multi-parameter buoy monitoring platforms”*

This revision clarifies that the buoy platform is designed for autonomous operation and capable of integrating multiple environmental sensing modules, thereby supporting long-term, continuous, and all-weather data acquisition in marine environments. The technical meaning of the original statement is preserved, while the description is now more explicit and accurate.

Line 85 - Do you have a citation for this radiometer?

Response:

We sincerely thank the reviewer for pointing out the need to provide appropriate references for the QFW-6000 microwave radiometer. We agree that a clear and authoritative citation is necessary to support the description of the instrument used in this study.

Following the reviewer’s suggestion, we have added both an official technical document and a recent peer-reviewed publication to substantiate the instrument specifications, design principles, and operational performance.

In the revised manuscript, we have added the relevant references on Page 4, Line 99 to support the description of the QFW-6000 microwave radiometer. Specifically, we have included:

1. The official technical documentation released by the 22nd Research Institute of China Electronics Technology Group Corporation (CETC-22): *QFW-6000 Ground-Based Microwave Radiometer User Manual*, which provides detailed information on the instrument design, technical specifications, and operational procedures.
2. A recent peer-reviewed study published in *Remote Sensing* demonstrates the successful application of the QFW-6000 radiometer in atmospheric profile retrieval.

These additions provide authoritative, verifiable technical support for the radiometer employed in this study and strengthen the credibility of the observational framework.

Line 90 - The measurements are not technically used to detect oxygen; atmospheric oxygen concentrations are usually one of the assumed quantities for detecting temperature profiles in this frequency band.

Response:

We sincerely thank the reviewer for this precise and important technical correction. We agree that the original wording was not sufficiently rigorous. The V-band channels are not used to directly detect atmospheric oxygen concentration; instead, they exploit the well-known oxygen absorption complex to retrieve atmospheric temperature profiles, assuming a constant oxygen mixing ratio.

Following the reviewer's suggestion, we have revised the corresponding description to more accurately reflect the physical radiative-transfer mechanism.

In the revised manuscript, the sentence on Page 4, Line 102-103 has been modified by replacing the original wording implying "oxygen detection" with a description emphasizing temperature retrieval based on oxygen absorption characteristics. The revised text now states that the V-band channels are used:

"Eight channels in the V-band (51.26–58.0 GHz) exploit the absorption of molecular oxygen to retrieve atmospheric temperature profiles."

This revision improves the scientific accuracy and rigor of the radiative-transfer description.

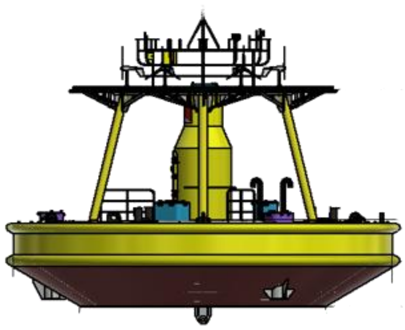
Line 101 - a zoomed out photograph of the buoy, or some kind of graphic showing the design would help the reader to visualise how it looks here

Response:

We sincerely thank the reviewer for this helpful suggestion. We agree that a visual representation of the buoy platform would significantly improve the reader's understanding of the overall system configuration and facilitate interpretation of the platform description.

Following the reviewer's recommendation, we have added a schematic overview of the buoy platform to provide a clear and intuitive visualization of its structural design.

n the revised manuscript, we have added a new schematic illustration of the overall buoy platform design (Page 5, Line 131, new Fig. 2) in Section 2.2. This figure presents a scaled structural overview of the buoy system and is accompanied by a detailed caption. In addition, we have inserted a transitional sentence in the main text to explicitly link the platform description with the newly added schematic figure.



These additions enable readers to more easily visualize the buoy configuration and gain a better understanding of the system layout.

Line 112 - 'microwave radiation equation' - do you mean radiative transfer equation?

Response:

We sincerely thank the reviewer for this professional and important correction. We agree that the original term "microwave radiation equation" was not sufficiently precise and could lead to conceptual ambiguity. The physical model referred to in this context is indeed the atmospheric radiative transfer equation.

Following the reviewer's suggestion, we have corrected the terminology to accurately reflect the adopted physical modeling framework.

In the revised manuscript, the terminology on Page 6, Line 141 has been corrected by replacing:

"microwave radiation equation" with: *"atmospheric radiative transfer equation"*

This revision improves the scientific accuracy and standardization of the radiative-transfer theory description.

Line 113 - 'improves retrieval accuracy' - it would be interesting to see how much of a difference this makes when compared to retrievals with no correction at all. I would think that this correction is vital for the accuracy, but perhaps if zenith angle perturbation is not too large, the correction is not necessary.

Line 113 - Please also discuss the error of the attitude sensor and how this affects the zenith angle error. Was a pointing angle calibration of the radiometer made when it was deployed to ensure that when the buoy is horizontal, that the antenna is pointing with a zenith angle of 0°?

Response:

We sincerely thank the reviewer for this technically important and insightful comment.

We fully agree that zenith-angle correction is a critical processing step for buoy-based microwave radiometer observations and plays a fundamental role in ensuring retrieval accuracy.

Even under relatively calm sea-state conditions where platform attitude perturbations may appear small, variations in the observation geometry still introduce systematic brightness temperature biases, particularly in the oxygen and water-vapor absorption bands, where atmospheric optical depth is highly sensitive to viewing angle. Therefore, zenith-angle correction has become a standard engineering practice for buoy-based and shipborne microwave radiometer measurements.

In this study, real-time pitch and roll measurements from the attitude sensor are used to compute the effective zenith angle for each radiometer integration period. This viewing geometry is then explicitly incorporated into the forward radiative transfer model, and brightness temperatures are corrected on a sample-by-sample basis prior to retrieval, thereby removing systematic biases induced by platform motion.

Regarding the attitude sensor accuracy, a MEMS-based attitude sensor integrating a tri-axis accelerometer, gyroscope, and digital motion processor (DMP) is employed. The typical static accuracy is 0.1°–0.3°, and the dynamic accuracy is approximately 0.3°–0.5° under typical marine operating conditions. These error levels are significantly smaller than the wave-induced pitch and roll variations of the buoy platform, and therefore do not dominate the overall zenith-angle uncertainty. The resulting residual angular uncertainty remains well within the tolerance range of the radiative transfer modeling framework.

In addition, a dedicated pointing-alignment calibration was conducted prior to deployment. When the buoy platform was placed on a horizontal reference plane, the radiometer antenna was mechanically aligned to ensure that its viewing direction corresponds to a zenith angle of 0°. This procedure establishes a physical reference baseline between the attitude sensor measurements and the actual observation geometry.

Although zenith-angle perturbations may be relatively small under calm sea conditions, long-term autonomous marine observations inevitably encounter continuously varying sea states. For this reason, zenith-angle correction is applied to all measurements in this study to ensure the stability and consistency of the retrieval results under different environmental conditions.

In the revised manuscript, we have added a detailed description of the zenith-angle correction procedure, the attitude sensor accuracy, and the antenna pointing-alignment calibration on Page 6, Lines 142–148. The added text reads:

"The effective viewing geometry is computed from real-time pitch and roll measurements and applied before inversion, thereby removing systematic brightness temperature deviations induced by platform motion. A MEMS-based attitude sensor integrating a tri-axis accelerometer, gyroscope, and digital motion processor (DMP) is used, with a typical static accuracy of 0.1°–0.3° and a dynamic accuracy of 0.3°–0.5° under typical marine operating conditions. A dedicated pointing alignment calibration was performed before deployment to ensure that the radiometer antenna corresponds to a zenith angle of 0° when the buoy is horizontally leveled."

These additions clarify the engineering implementation of the observation-geometry correction and provide a transparent description of the associated error characteristics and calibration procedures.

Line 131 - 'liquid nitrogen calibrated' - was this done on board the buoy or previously on land?

Response:

We sincerely thank the reviewer for raising this important practical and engineering-related question. We clarify that the liquid nitrogen calibration of the QFW-6000 microwave radiometer was conducted on land before buoy deployment.

Performing liquid nitrogen calibration on board a buoy platform is neither safe nor technically feasible due to continuous platform motion induced by ocean waves and the high operational risks associated with handling cryogenic liquids in a dynamic marine environment. Therefore, absolute radiometric calibration was completed under controlled laboratory conditions before installation on the buoy platform to ensure measurement accuracy and traceability.

This pre-deployment calibration provides a reliable absolute radiometric reference for subsequent long-term offshore observations and is a critical step for guaranteeing the quality and consistency of brightness temperature measurements.

In the revised manuscript, we have explicitly clarified the calibration procedure on Page 7, Line 166-168. The revised text now reads:

"In the Jiaozhou Bay area, the QFW-6000 microwave radiometer, which was modified and liquid-nitrogen calibrated on land before deployment, together with an attitude sensor, was installed on a 10-m buoy platform according to the layout shown in Fig. 4."

This revision improves the transparency of the instrument calibration procedure and enhances the engineering reproducibility of the observational system.

Line 137 - please state the instrument details for the university of Wyoming data (including resolution and measurement accuracy)

Response:

We sincerely thank the reviewer for this important and technically meaningful suggestion. We fully agree that, as the radiosonde observations are used as the primary reference for validating the retrieval results, the technical specifications of the sounding instrument should be explicitly stated to ensure transparency and reproducibility.

According to Tu et al. (2021), the Qingdao upper-air sounding system (Station ID: 54857) is based on the GTS1 radiosonde, and the sounding data used in this study were obtained from the University of Wyoming upper-air sounding archive. Based on the manufacturer's technical specifications, the GTS1 radiosonde provides continuous vertical profiles of temperature, humidity, and pressure with a typical vertical resolution of approximately 5–10 m. The corresponding measurement accuracies are approximately ±0.2 K for temperature, ±5% for relative humidity, and ±1 hPa for pressure.

In the revised manuscript, we have added the instrument model, vertical resolution, and measurement accuracies of the radiosonde system on Page 8, Lines 184–187. The added text reads:

"According to Tu et al. (2021), the Qingdao upper-air sounding system (Station ID: 54857) is based on the GTS1 radiosonde. According to the manufacturer's technical specifications, the GTS1 radiosonde provides continuous vertical profiles with a typical vertical resolution of approximately 5–10 m during ascent. The measurement accuracies are approximately ±0.2 K for temperature, ±5% for relative humidity, and ±1 hPa for pressure."

These additions improve the transparency of the reference observations and strengthen the credibility of the validation framework.

Line 147 - please reference the ITU-R model. Why was this used as opposed to the previously mentioned ones?

Response:

We sincerely thank the reviewer for this important and technically meaningful comment. We fully agree that the ITU-R model should be explicitly referenced and that the rationale for selecting this model over the previously mentioned alternatives should be clearly explained.

In the revised manuscript, we have added the formal reference to ITU-R

Recommendation P.676-13 (ITU-R, 2022) and provided a detailed justification for choosing the ITU-R model instead of the MPM-series and MonoRTM models.

Currently, the most widely used atmospheric gaseous absorption models include the MPM series, the MonoRTM model, and the ITU-R recommended model. The MPM and MonoRTM models are based on detailed molecular spectroscopy and have been extensively used in atmospheric radiative transfer research. However, in engineering practice, these models are typically implemented through encapsulated numerical libraries, where the internal spectroscopic parameters and continuum absorption formulations are not fully transparent to users, and the computational cost is relatively high.

In contrast, the ITU-R model is provided in the form of fully documented parametric formulations with publicly available coefficients and calculation procedures. It is recognised as the international standard in radio propagation engineering and offers excellent transparency, reproducibility, and standard compliance. Its layered parameterisation framework provides a favourable balance between modelling accuracy and computational efficiency, making it particularly suitable for the real-time forward modelling and retrieval requirements of buoy-based microwave radiometer systems.

Based on these engineering and computational considerations, the ITU-R P.676 model was adopted in this study for the calculation of atmospheric gaseous absorption coefficients.

In the revised manuscript, we have added the reference to the ITU-R model and expanded the explanation of the model selection on Page 9, Lines 204–207. The revised text reads:

“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 utilises a line-by-line summation method that offers an optimal balance between computational efficiency and accuracy, making it particularly suitable for the real-time retrieval requirements of this buoy-based system.”

In addition, the following reference has been added to the reference list:

ITU-R (2022): Recommendation ITU-R P.676-13: *Attenuation by atmospheric gases and related effects*, International Telecommunication Union, Geneva.

These additions improve the transparency and standardisation of the forward modelling framework and clarify the engineering motivation for the model selection.

Line 155 - 'The resulting ...' everything from here is not necessary

Response:

We thank the reviewer for this helpful suggestion. We agree that the description of the comparison between simulated and observed brightness temperatures belongs to the retrieval algorithm workflow discussed later (Section 3.4), rather than to the forward-model description.

To avoid redundancy and to improve the clarity and focus of the forward-model section, we have removed the unnecessary sentences starting from "The resulting …" in Section 3.2.

In the revised manuscript, we have removed the paragraph starting from *"The resulting ..."* in Section 3.2 (Page 8, Line 155), as this content is more appropriately discussed in the retrieval algorithm section.

This revision streamlines the description of the forward model and improves the overall structure and readability of the manuscript.

Line 157 - This section has the same name as the preceding section. The article goes from explaining which model is used for calculating absorption coefficients to trying to explain the mechanism through which microwave radiation is present at the surface. Many sentences in this section are oddly phrased, misleading, or incorrect.

Take: *"In the microwave frequency band (300 MHz–3000 GHz), electromagnetic waves exhibit both wave-like and particle-like properties."* In most literature, microwave frequencies are said to be between 300 MHz and 300 GHz (not 3 THz), particle-wave duality does not apply to microwaves alone, and it is not particularly relevant to the research presented.

Please combine this sub-section with the previous sub-section (3.2 and 3.3) and ensure that the section is clear, informative and factually correct.

Response:

We sincerely thank the reviewer for this detailed and highly constructive comment. We fully agree that the original manuscript contained duplicated subsection titles, fragmented logical structure, and several physical descriptions that were either imprecise, misleading, or not directly relevant to the scope of this study.

Following the reviewer's suggestion, we have merged the original Sections 3.2 and 3.3 into a single unified subsection and completely reorganized and rewritten the content. The new subsection is now entitled:

“Atmospheric Radiative Transfer Principle and Model”

The revised section provides a coherent and focused description of the forward modelling framework for the buoy-based microwave radiometer system, covering two core components:

- The calculation of atmospheric gaseous absorption coefficients based on the ITU-R Recommendation P.676-13 model;
- The forward simulation of brightness temperatures using the microwave radiative transfer equation.

During the revision, we have removed all unnecessary or potentially misleading physical background discussions (e.g., wave–particle duality), corrected the definition of the microwave frequency range to the standard 300 MHz–300 GHz, and rewritten the entire section to ensure that the physical descriptions, engineering logic, and remote-sensing modelling framework are consistent with established practice in the microwave remote sensing community.

In the revised manuscript, the original Sections 3.2 and 3.3 have been merged into a single subsection and rewritten (Page 9, Lines 193–211). The revised subsection now begins with the following representative statement:

“This section describes the forward radiative transfer modelling framework used in this study, including the calculation of atmospheric absorption coefficients using the ITU-R P.676-13 model and the corresponding microwave radiative transfer formulation for brightness temperature simulation.”

These changes eliminate redundancy, correct the physical descriptions, and provide a clear, informative, and technically rigorous presentation of the forward modelling methodology.

Line 188 - ‘obtain the normal distribution patterns for temperature and humidity’- do you mean to say normally distributed errors?

Response:

We sincerely thank the reviewer for this important and precise clarification request. We agree that the original expression “normal distribution patterns” was not sufficiently rigorous and could easily lead to misunderstanding.

In this study, a quasi-normal distribution assumption is only introduced as a statistical criterion for data quality control and outlier detection. We do not assume that atmospheric temperature and humidity themselves strictly follow a normal distribution, nor do we imply that the retrieval errors are normally distributed. Therefore, the original

wording was scientifically imprecise.
To avoid ambiguity and to improve the statistical accuracy of the description, we have revised the sentence to emphasize that the purpose is to obtain the statistical distribution characteristics of temperature and humidity, rather than to imply a normal distribution model.
In the revised manuscript, the sentence on Page 10, Lines 225–226 has been modified to:
"... to accurately obtain the statistical distribution characteristics of temperature and humidity."
This revision removes the implication of a normal-distribution assumption and ensures a more objective and scientifically appropriate statistical description.
Line 189 - 'deviating from the normal rage were excluded' - what do you consider the normal range? Was this the two sigma mentioned earlier?
Response:
We sincerely thank the reviewer for pointing out that the definition of the "normal range" was not sufficiently clear in the original manuscript. We agree that this ambiguity could lead to a misunderstanding of the data quality-control procedure.
In this study, the "normal range" is defined in a statistical sense as a confidence interval centered on the sample mean and bounded by ±2 standard deviations ($\mu \pm 2\sigma$), which is a commonly used criterion for radiosonde data quality control and outlier detection in atmospheric studies. Data points falling outside this interval are considered outliers and are excluded from subsequent analysis.
In the revised manuscript, we have explicitly clarified this definition on Page 10, Line 226-227. The revised text now reads:
"Based on the statistical analysis, data points falling outside the confidence interval (defined as the mean ± 2 standard deviations) were identified as outliers and excluded."
This revision improves the transparency and scientific rigor of the data screening procedure.
Line 190 - 'inconsistent measurement data at different altitudes' - Not sure what is meant by this - that the profile is not a straight line?
Response:
We sincerely thank the reviewer for pointing out the ambiguity in the original wording. We agree that the phrase "inconsistent measurement data at different altitudes" could easily be misunderstood and does not accurately describe the physical meaning

intended in the manuscript.

The original statement does not imply that the radiosonde profiles are physically discontinuous or unrealistic. Instead, it refers to the fact that, due to variations in balloon ascent rate and horizontal drift induced by wind, the radiosonde measurements are sampled at irregular vertical intervals. This non-uniform vertical sampling is a common characteristic of radiosonde observations and requires interpolation onto a unified height grid prior to subsequent analysis and retrieval.

To avoid ambiguity and to improve the physical accuracy of the description, we have replaced the original wording with a more precise expression.

In the revised manuscript, the phrase on Page 10, Lines 228–230 has been modified from:

"inconsistent measurement data at different altitudes" to: *"irregular vertical sampling intervals"*

This revision provides a clearer and more scientifically accurate description of the actual sampling characteristics of radiosonde profiles and is consistent with standard radiosonde data-processing practice.

196 - Plots to show what the temperature and humidity distributions look like would be useful here. Also, describe the skewness and excess kurtosis, and your considered thresholds of these values to consider a distribution normally distributed. You state that atmospheric temperature followed a 'perfect' normal distribution, please state what you consider to be 'perfectly' normal.

Response:

We sincerely thank the reviewer for this constructive and valuable suggestion. We fully agree that visual evidence and quantitative statistical indicators are essential to support the data quality-control strategy.

In response, we have comprehensively revised Section 3.3.2 to address all of the reviewers' concerns:

1. Clarification of "perfect normal distribution":

   We apologize for the imprecise wording in the original manuscript. The term "perfect normal distribution" has been removed and replaced with "approximate normal distribution" to reflect a statistically appropriate description.

2. Quantitative indicators and thresholds:

   We now introduce Skewness (S) and Excess Kurtosis (K) as quantitative measures to assess normality. Following Kim (2013), we adopt the criteria $|S| < 1$ and $|K| < 1$

to judge whether a distribution can be considered approximately normal.

3. Visualization:

   We have added a new figure (new Fig. 6) showing the probability density distributions of temperature and relative humidity at a representative height level (50 m) in January, including kernel density estimation (KDE) curves and fitted normal distributions for comparison.

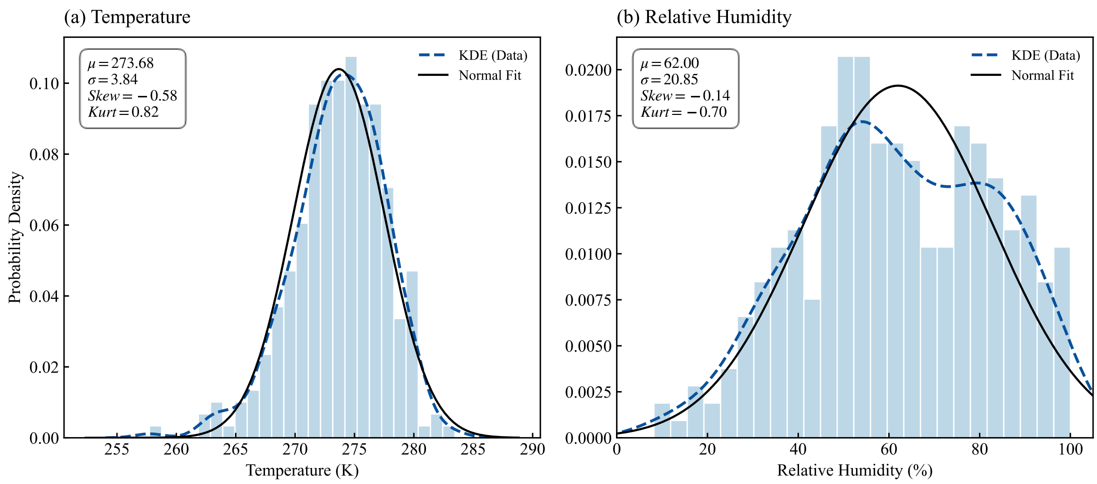
4. Quality-control rationale:

   The statistical analysis (e.g., relative humidity skewness S = −0.14) confirms approximate normality for the study region. Based on these characteristics, we apply the 2σ criterion to strictly remove outliers and ensure high data quality for the empirical database.

We have rewritten the relevant paragraph in Section 3.3.2 (Page 11, Lines 251–264) and added Fig. 6. The revised text reads:

"To quantitatively assess the normality of the data distribution, we performed a monthly statistical analysis of temperature (T) and relative humidity (H) at each height layer. We calculated the Skewness (S) and Excess Kurtosis (K) for both variables. A distribution is considered to be approximately normal when $|S| < 1$ and $|K| < 1$ (Kim, 2013).

Taking the 50 m height layer in January as an example (Figure 5), the atmospheric temperature exhibits a high degree of symmetry with $S = -0.58$ and $K = 0.82$, indicating a good fit with the normal distribution. Similarly, the relative humidity at this level also demonstrates approximate normality ($S = -0.14$, $K = -0.70$). Based on these statistical characteristics, we constructed the empirical database by filtering out outliers falling outside the 2σ interval to ensure strict data quality."



New reference added:

- Kim, H. Y. (2013). *Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis*. Restorative Dentistry & Endodontics, 38(1), 52–54.

204 - Is the NOGA-II algorithm used elsewhere? Could you please reference this if so.

Response:

We sincerely thank the reviewer for raising this point, and we apologize for the typographical error in the original manuscript. The algorithm used in this study is NSGA-II (Non-dominated Sorting Genetic Algorithm II) rather than "NOGA-II". The occurrence of "NOGA-II" was a spelling mistake.

NSGA-II is a well-established and widely used multi-objective evolutionary optimization algorithm and has been extensively applied in engineering and scientific optimization problems.

Following the reviewer's suggestion, we have corrected the spelling throughout the manuscript and added the original and authoritative reference for the NSGA-II algorithm.

In the revised manuscript, all occurrences of "NOGA-II" have been corrected to NSGA-II. In addition, we have added the following classic reference on Page 12, Line 271:

- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: *A fast and elitist multiobjective genetic algorithm: NSGA-II*, IEEE Transactions on Evolutionary Computation, 6(2), 182–197, 2002.

This revision corrects the terminology and provides an authoritative citation for the optimization algorithm used in this study.

207 - 'it implements essential enhancements in...' - this phrasing is very imprecise and is reminiscent of AI language. Please rephrase.

Response:

We sincerely thank the reviewer for this constructive comment. We agree that the original wording, "essential enhancements," was overly vague, promotional in tone, and not sufficiently technical, which is inappropriate for an engineering and methodological description.

Following the reviewer's suggestion, we have rewritten this part of Section 3.3.3 to replace the generic and imprecise phrasing with a concrete and technically explicit description of the algorithmic framework and its engineering implementation.

Specifically, the revised text now focuses on clearly describing the structural components of the proposed retrieval framework, including the construction of a small-sample prior knowledge database, the incorporation of physical constraints, the fusion of attitude sensor data, and the use of a parallel-computing architecture for efficient optimization.

In the revised manuscript, Section 3.3.3 has been rewritten on pages 12, Lines 273–293. The original vague and promotional expressions (e.g., “essential enhancements”) have been removed and replaced by a technically explicit description of the algorithmic structure and implementation.

This revision improves the technical rigor of the method description and removes language that could be interpreted as promotional or AI-generated.

209 - ‘For the first time it achieves high-precision temperature and humidity profile inversion...’ - I’m not sure you can consider measurement error of 2 K (for temperature) and 20% (for RH) high precision. Radiometers are not designed to produced highly precise measurements of temperature and humidity- their advantages are continuous remote soundings as opposed to precision measurements of atmospheric quantities.

Response:

We sincerely thank the reviewer for this professional and important clarification. We fully agree that microwave radiometers are not designed to provide high-precision in situ measurements of atmospheric temperature and humidity. Their primary advantages lie in continuous remote sensing capability, all-weather operation, and the ability to provide high-temporal-resolution atmospheric profiling, rather than in achieving the measurement precision of radiosonde observations.

We acknowledge that the original wording (“for the first time” and “high-precision”) was inappropriate and overstated the performance of radiometer-based retrievals.

Following the reviewer’s suggestion, we have removed these exaggerated expressions and revised the text to focus instead on the robustness, real-time capability, and engineering applicability of the proposed retrieval framework under sparse-data and dynamically varying marine conditions.

In the revised manuscript, we have modified by deleting the phrases *“for the first time”* and *“high-precision”*. The discussion now emphasizes the practical advantages of the proposed framework in terms of continuous observation capability, operational robustness, and suitability for long-term autonomous marine deployment.

This revision ensures a more scientifically accurate and objectively framed description of radiometer-based retrieval performance.

210 - for each of the three items mentioned here, please describe in detail how each of the steps is done. Please in particular, as highlighted above, describe what 'physical laws' are used as constraints. Please also describe the processing of the level 1 data and the attitude sensor data.

Response:

We sincerely thank the reviewer for this important and technically detailed request. We agree that the original description of the three core methodological components was too brief and did not sufficiently explain the complete data-processing workflow and engineering implementation.

Following the reviewer's suggestion, we have substantially expanded Section 3.3.3 to provide a clear, step-by-step description of the entire retrieval framework, including Level-1 brightness temperature processing, incorporation of physical constraints, and fusion of attitude sensor data. The revised framework now explicitly includes the following components:

1. Level-1 brightness temperature processing:

   The radiometer brightness temperature data are first calibrated and subjected to quality control. Real-time pitch and roll measurements from the attitude sensor are then used to compute the effective zenith angle for each radiometer integration period. This viewing geometry is explicitly introduced into the forward radiative transfer model, and brightness temperatures are geometrically corrected before the retrieval.

2. Construction of the small-sample prior knowledge database:

   Based on historical radiosonde profiles, a small-sample prior database is constructed using a statistical boundary-generation approach. This database provides physically plausible initial constraints for the optimization process and reduces dependence on large-scale training datasets.

3. Physical-law constraints:

   During optimization, atmospheric thermodynamic lapse-rate constraints are introduced as hard constraints to suppress solutions that violate fundamental atmospheric physics.

4. Attitude sensor data processing:

   Real-time attitude measurements are continuously assimilated into the forward modelling stage to dynamically compensate for platform-motion-induced zenith-angle deviations.

In the revised manuscript, these processing steps have been explicitly added to Section 3.3.3 on Pages 12, Lines 273–284. Page 14, Lines 358-363. The expanded description now provides a complete, transparent, and reproducible engineering workflow for the proposed retrieval framework.

216 - 'suitable for real-time maritime applications' - can you reference these requirements?

Response:

We sincerely thank the reviewer for this important and application-oriented suggestion. We fully agree that the term "real-time maritime applications" should be supported by authoritative operational standards and clearly defined performance requirements, rather than being used as a purely qualitative description.

In operational marine meteorological monitoring and nowcasting systems, "real-time" typically refers to data processing and product generation within minute-level latency to support rapidly updating monitoring and decision-making workflows. According to the *WMO Guide to Meteorological Instruments and Methods of Observation* (WMO, 2018), routine operational observing systems generally adopt reporting cycles of 1–10 minutes (with some integrated products reported on hourly cycles). In addition, Sun et al. (2014) indicate that convective nowcasting systems usually require update cycles on the order of 5–10 minutes to capture rapidly evolving atmospheric conditions.

Within this operational context, the proposed NSGA-II-based retrieval framework, combined with a parallel-computing architecture, achieves a stable runtime of less than one minute per retrieval under the configuration of 175 individuals, 10 generations, a crossover probability of 0.9, and a mutation probability of 0.2. This performance is significantly faster than the typical 5–10 minute update requirements of operational nowcasting systems, and therefore meets the engineering requirements for real-time marine meteorological applications.

In the revised manuscript, we have added the operational definition of "real-time maritime applications" and introduced authoritative references to support this statement on Page 12, Lines 285–289. The revised text now explicitly links the computational performance of the proposed framework to established operational requirements defined by WMO (2018) and Sun et al. (2014).

These additions provide a standardized, application-driven definition of "real-time" and strengthen the engineering justification of the proposed system for operational maritime use.

New references added:

- WMO (2018): *Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8)*, World Meteorological Organization, Geneva.
- Sun, J., et al. (2014): *Use of NWP for nowcasting convective precipitation: Recent progress and challenges*, Bulletin of the American Meteorological Society.

258 - Which data is included in the systematic bias correction? all data or only that which you include in the stats?

Response:

We sincerely thank the reviewer for this important methodological question. We fully agree that the data usage strategy for constructing and validating the systematic bias correction model must be clearly documented to avoid any ambiguity regarding model development and performance assessment.

In this study, the systematic bias correction model is not constructed using all available data. As described in the revised manuscript (Section 3.3.3, Page 14, Lines 337–341), the 38 collocated microwave radiometer–radiosonde observation pairs are randomly divided into a training set (80%) and an independent test set (20%).

The training set is used exclusively to establish the systematic bias correction model, while the test set serves as a fully independent dataset to evaluate the actual performance of the correction scheme. The statistical results presented in Table 1 and the corrected retrieval profiles shown in Fig. 7 are all derived from the independent test set after applying the bias correction model trained on the training data.

This data usage strategy ensures an objective assessment of the bias correction performance and avoids any artificial inflation of accuracy due to overlap between training and validation datasets.

In the revised manuscript, we have explicitly clarified the data usage strategy for systematic bias correction in Section 3.3.3 (Page 12, Lines 302–305). These additions eliminate potential ambiguity and ensure full transparency of the model training and evaluation procedure.

**286 - Figure 4a - I would expect the retrieved profile to be smooth and the radiosonde to have more curves. How do you explain that the opposite is true? The zoom on both graphs appears to not be very zoomed in and show different curves than on the main graph. The feature shown on the 2a zoom and 2d, where at 5km the RH has a difference of around 1% does not appear to be present on the large

graph. Why? Please also explain the criteria for valid comparisons.

Response:

We sincerely thank the reviewer for the careful inspection of Fig. 7 (formerly Fig. 4) and for these highly professional and constructive comments. We fully agree that the issues raised are critical for improving the clarity, consistency, and scientific rigor of the figure presentation and interpretation. Accordingly, we have systematically revised both the figure and the accompanying text.

1. Explanation of the apparent smoothness difference between radiosonde and retrieved profiles

In Fig. 7, the radiosonde profile represents the ensemble-averaged profile derived from 38 spatiotemporally collocated samples. This ensemble averaging effectively suppresses small-scale vertical fluctuations and therefore results in a visually smoother structure.

In contrast, the microwave radiometer retrieval is obtained through layer-by-layer optimization on a discrete vertical grid with finite vertical resolution. In altitude regions where the weighting-function sensitivity is relatively low, weak stratification structures may still be retained. This behavior is a typical characteristic of passive microwave retrieval systems and reflects the integral nature of radiative transfer and the discrete-layer inversion framework, rather than measurement noise or instability.

We have added an explicit physical explanation of this phenomenon in the revised manuscript.

2. Consistency between the main panels and zoomed-in subpanels

The reviewer correctly pointed out that some features visible in the zoomed-in panels were not clearly discernible in the main panels. Upon careful inspection, we found that the original version used different plotting workflows and display scaling for the main and zoomed panels, which introduced visual inconsistencies.

In the revised manuscript, we have standardized the plotting procedure as follows:

(1) The main panels and zoomed-in subpanels are generated from the same processed profile dataset.

(2) The zoomed-in panels differ only in axis range (coordinate scaling);

(3) No additional smoothing, resampling, or data manipulation is applied to the zoomed panels.

As a result, the zoomed-in panels now represent a true magnification of the corresponding height intervals in the main panels, ensuring that no artificial features or visual artefacts are introduced.

3. Criteria for valid profile comparison

To ensure physical comparability between the retrieved and radiosonde profiles, we adopt the following strict comparison criteria:

(1) Temporal collocation between radiometer retrievals and radiosonde launches;

(2) Interpolation of both profiles onto a common vertical height grid;

(3) Evaluation within the effective vertical resolution range of the microwave radiometer.

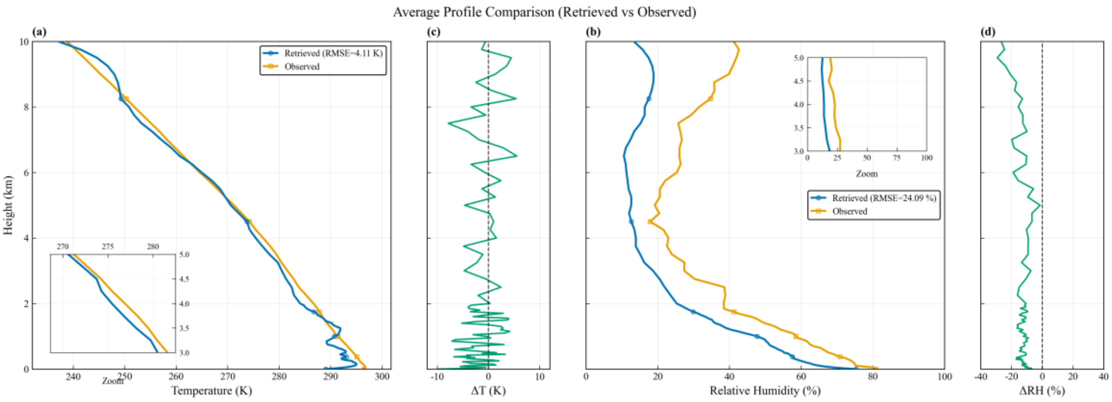
These steps ensure that the comparison reflects the intrinsic retrieval performance rather than artefacts caused by sampling differences.

In the revised manuscript (Page 15-16, Lines 387–415), we have implemented the following changes:

1. Added a physical explanation for the apparent smoothness difference between the ensemble-averaged radiosonde profiles and the discretely retrieved radiometer profiles;

2. Unified the plotting workflow for the main panels and zoomed-in subpanels in Fig. 7 to guarantee visual consistency;

3. Explicitly stated the criteria and data-processing steps used for valid profile comparison in both the figure caption and the main text.



These revisions significantly improve the scientific clarity, interpretability, and graphical consistency of Fig. 7 and strengthen the transparency and credibility of the profile comparison.

305 - Did you also retrieve the liquid water path? Please mention this if so

Response:

We sincerely thank the reviewer for this important and clarifying question. We confirm that the liquid water path (LWP) was not included as a state variable in the NSGA-II retrieval framework and was not part of the multi-objective optimization process. In this study, the retrieval variables are limited to atmospheric temperature and water-vapor profiles.

However, as an operational instrument, the QFW microwave radiometer provides liquid water path (LWP) as part of its standard Level-2 product suite. In the present work, the LWP product is used only as an auxiliary diagnostic variable for data quality control and for analysing potential error mechanisms, rather than as a retrieval constraint or optimization target.

In the revised manuscript, we have explicitly clarified the role of LWP in the retrieval framework on Page 16, Lines 408–411. The added text specifies that LWP is not retrieved within the NSGA-II framework but is used solely as an auxiliary diagnostic parameter.

This clarification eliminates any ambiguity regarding whether LWP is treated as a retrieval variable in this study.

309 - I am not sure why you mention the mesosphere here

Response:

We sincerely thank the reviewer for pointing out this important terminology issue. We fully agree that the use of the term *"mesosphere"* in the original manuscript was inappropriate and scientifically incorrect in the context of this study.

The present work focuses exclusively on the retrieval of temperature and humidity profiles within the troposphere (below approximately 10 km). The mesosphere, which typically refers to the atmospheric layer between about 50 and 85 km, is entirely outside the scope of this investigation. The original intention was to refer to the middle troposphere, where the retrieval results exhibit a more pronounced height-dependent systematic bias.

Accordingly, we have corrected this terminology in the revised manuscript.

In the revised manuscript, the term *"mesosphere"* has been replaced with *"middle troposphere"* in the discussion of the height-dependent bias characteristics (Page 16, Line 413).

This correction ensures that the terminology accurately reflects the physical altitude range investigated in this study and avoids any conceptual ambiguity.

315 - If the retrievals in these figures have been systematically bias corrected, why do the lines not entirely agree? I do not see what the point is in showing bias corrected results if the reference period for the bias correction is the same as the results being presented (it is not clear from reading the article whether this is the

case or not). Also, the zoom on figure 5b appears again to show a different curve than the main plot.

Response:

We sincerely thank the reviewer for this rigorous and constructive comment. We fully agree that the purpose, data usage strategy, and visual consistency of the bias-corrected results must be clearly documented to ensure scientific transparency and avoid any potential misinterpretation.

1. Why do the bias-corrected curves not coincide perfectly

The systematic bias correction model is designed to remove the height-dependent *mean systematic bias* rather than to force point-by-point agreement between the retrieved and radiosonde profiles. Even after bias correction, small residual differences remain due to intrinsic atmospheric variability and measurement uncertainty. Therefore, while the systematic offset between the two mean profiles is significantly reduced after correction, the curves are not expected to coincide exactly at every height level.

2. Data usage strategy for bias correction

The bias correction model is not fitted using the same samples shown in Fig. 5(new Fig. 8). Instead, it is trained on an independent calibration dataset and subsequently applied to a separate validation dataset. This avoids any circular reasoning and ensures an objective assessment of the correction performance. The manuscript has been revised to explicitly describe this training–validation separation.

3. Consistency between the main and zoomed panels in Fig. 5b（new Fig. 8b）

We agree that the earlier version contained inconsistencies between the main and zoomed panels due to differences in plotting workflows and scaling. In the revised manuscript, the plotting procedure has been unified so that the zoomed panel is now a true magnification of the corresponding section in the main panel, without any additional smoothing or resampling.

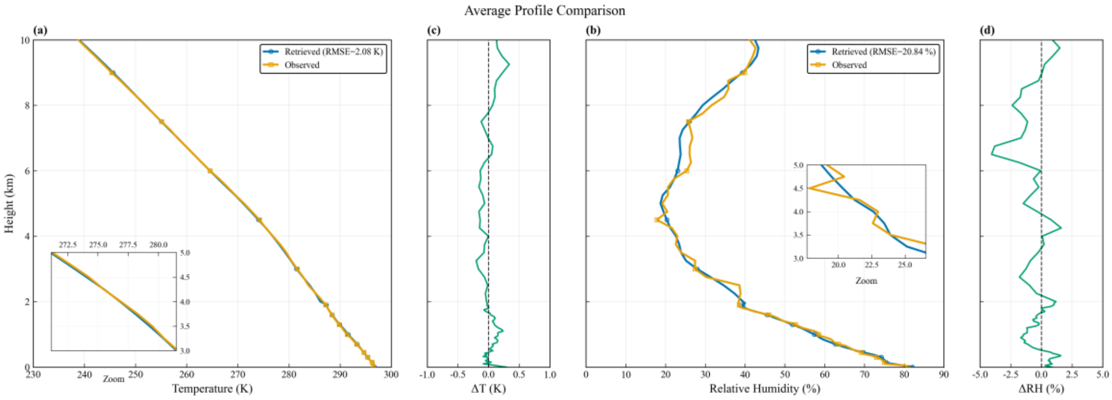
In the revised manuscript (Page 17, Lines 422–423、432-433), we have implemented the following changes:

(1) Added an explicit explanation that the bias correction removes height-dependent mean systematic bias rather than enforcing exact point-by-point agreement;

(2) Clarified that the bias correction model is trained on an independent calibration dataset and applied to separate validation cases;

(3) Unified the plotting workflow for Fig. 5b (new Fig. 8b)to ensure visual

consistency between the main and zoomed panels.



These revisions improve the methodological transparency of the bias-correction procedure and ensure that the figures accurately reflect the underlying data.

343- ‘In the upper troposphere (2-10 km)’ - the whole of this range should not be taken as the upper troposphere. I would suggest that the upper troposphere is typically around 6-12km at these latitudes.

Response:

We sincerely thank the reviewer for this precise and professional correction. We fully agree that, at the latitudes relevant to this study, the altitude range of 2–10 km spans both the middle and upper troposphere, and therefore should not be entirely referred to as the “upper troposphere.”

To avoid any physical ambiguity and to ensure an accurate description of the atmospheric vertical structure, we have revised the terminology accordingly.

In the revised manuscript, the phrase on Page 18, Line 446 has been corrected from: *“upper troposphere (2–10 km)”* to: *“middle and upper troposphere (2–10 km)”*

This revision ensures a scientifically accurate representation of the atmospheric layer structure within the studied altitude range.

345 - what values of RMSE are required for what purpose for marine monitoring?

Response:

We sincerely thank the reviewer for this important and practically oriented question. We agree that the original manuscript did not clearly specify what level of retrieval accuracy is required for different marine monitoring purposes.

In operational marine meteorological applications, microwave radiometer retrievals are primarily intended for routine environmental monitoring and synoptic-scale atmospheric structure diagnosis, especially in data-sparse coastal and offshore regions. These applications emphasize continuous, all-weather remote sensing capability and high temporal resolution rather than high-precision in situ accuracy.

To provide an objective engineering benchmark, we have added references to several representative studies on ground-based microwave radiometer profile retrievals. These studies consistently report typical retrieval accuracies on the order of:

- Temperature RMSE: approximately 1–2 K
- Relative humidity RMSE: approximately 10–30%

and regard this level of accuracy as suitable for routine operational monitoring and atmospheric structure analysis (e.g., Massaro et al., 2015; Cimini et al., 2011; Yan et al., 2020).

Within this operational context, the retrieval accuracy achieved in this study (temperature RMSE of 2.08 K and relative humidity RMSE of 20.95%) meets the practical requirements for real-time marine environmental monitoring and offshore meteorological services. At the same time, we explicitly note that higher-precision applications such as research-grade retrievals and data assimilation would require further improvement.

In the revised manuscript, we have clarified the application scope and corresponding accuracy requirements on Page 18, Lines 448–451. The revised text explicitly states that the achieved retrieval accuracy is suitable for routine marine environmental monitoring and synoptic-scale diagnosis, while also acknowledging the limitations for higher-precision research applications.

Relevant references have been added, including:

- Massaro, G., et al. (2015)
- Cimini, D., et al. (2011)
- Yan, J., et al. (2020)

These additions provide an objective engineering and operational context for interpreting the reported RMSE values.

Technical Corrections

Line 27 - change ‘voyage’ to ‘travel’

Response:

We thank the reviewer for pointing out this wording issue. We agree that *“travel”* is the more appropriate term in this context.

In the revised manuscript, the word *“voyage”* has been replaced with *“travel”* on Page 2, Line 30.

Response to RC2

L 49: what is BP?

Response:

We thank the reviewer for pointing out this issue. In the original manuscript, the abbreviation "*BP*" was not explicitly defined, which may indeed cause ambiguity. Here, "BP" refers to the backpropagation algorithm commonly used in neural networks. In the revised manuscript (page 2, line 55), we have revised its first occurrence to "*backpropagation (BP) neural network*" to avoid confusion and improve readability.

Section 2.1: I am missing an accurracy estimation about how good the microwave radiometric profiler is or at least a reference on a publication that is adressing the quality of this instrument. Please add infomartion about the typical uncertainties of the instrument or give references.

Response:

We thank the reviewer for this important comment. We agree that information on the performance and uncertainty of the microwave radiometric profiler is necessary for evaluating the reliability of the retrieval results. In the revised manuscript (Section 2.1), we have added a description of the typical measurement uncertainties for this class of ground-based microwave radiometers, including the accuracy of brightness temperature measurements and the uncertainties in retrieved temperature and humidity profiles, along with relevant literature references.

Specifically, the following text has been added in the revised manuscript (page 5, lines 116–122):

"The QFW-6000 microwave radiometric profiler belongs to the class of ground-based multi-channel microwave radiometers that have been widely used for atmospheric profiling. When properly calibrated, ground-based microwave radiometric profilers can provide brightness temperature measurements with absolute uncertainties on the order of a few tenths of a kelvin (e.g., Hewison, 2007; Cimini et al., 2006). The corresponding uncertainties in retrieved temperature profiles are generally on the order of 1–2 K in the lower troposphere, while relative humidity uncertainties are typically within 10–20 %, depending on atmospheric conditions and retrieval configuration. These ranges represent typical performance reported for this instrument class under proper calibration and favorable conditions, and may vary with calibration strategy and atmospheric conditions."

L 90: In this section, the authors provide information on the vertical resolution. It

would be helpful to further explain how this vertical resolution is defined and derived. My understanding is that it is determined by the temperature and humidity retrieval setup and may therefore reflect a choice made during the training process rather than an inherent capability of the measurement system. In this context, it would be important to clarify that the effective vertical resolution, in terms of degrees of freedom, is comparatively low. Adding this explanation would improve the interpretation of the results.

Response:

We thank the reviewer for this important suggestion. We agree that the definition and origin of the reported vertical resolution should be clarified more explicitly. In the revised manuscript (Section 2.1, pages 4, lines 104–109), we have added an explanation to state that the vertical resolution refers to the vertical grid adopted in the temperature and humidity retrieval, rather than to an intrinsic physical resolution of the microwave radiometer itself.

We further clarified that the effective vertical resolution is mainly constrained by the information content of the observed microwave channels and the shape of the corresponding weighting functions. Therefore, in terms of the degrees of freedom for the signal, the number of independent vertical pieces of information is comparatively low.

The following text has been added to the revised manuscript:

"It should be noted that the vertical resolution reported here refers to the retrieval grid used for constructing the temperature and humidity profiles, rather than the intrinsic physical resolution of the microwave radiometer itself. The effective vertical resolution is mainly constrained by the information content of the observed microwave channels and the corresponding weighting functions. As a result, the number of independent pieces of vertical information (degrees of freedom for signal) is comparatively low, and the retrieved profiles represent smoothed atmospheric structures rather than fine-scale vertical details."

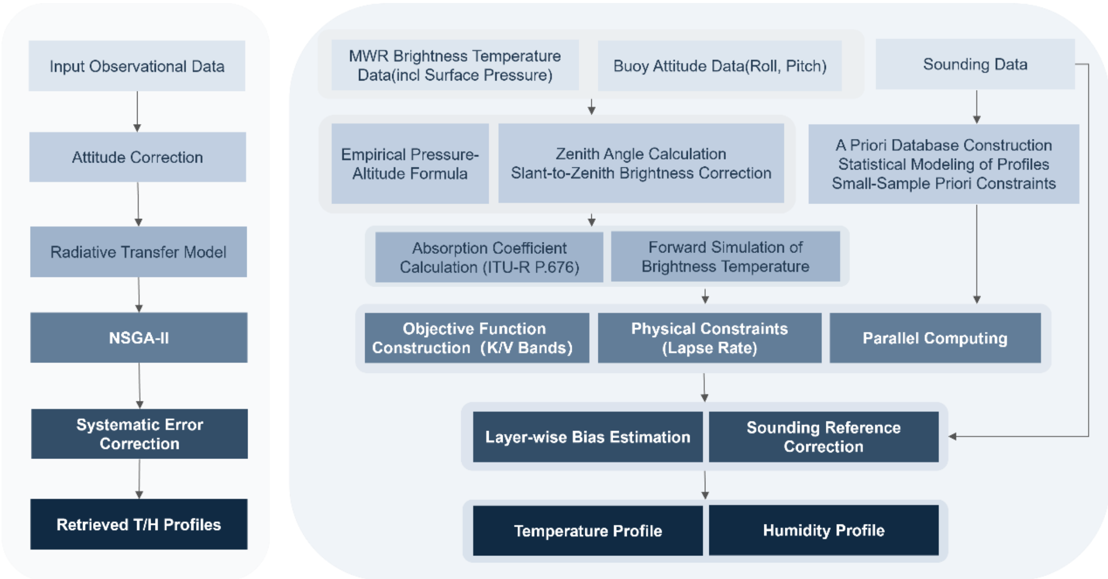
Section 3: It would be beneficial to begin this section with a concise and straightforward overview of the algorithm. A flowchart or schematic illustration could help guide the reader through the individual processing steps, clearly indicating which correction models or algorithms are applied at each stage and which data sources are used. Including such an overview would substantially improve the readability and comprehensibility of the manuscript.

Response:

We thank the reviewer for this constructive suggestion. We agree that providing an overall overview of the algorithm at the beginning of Section 3 helps improve the readability and comprehensibility of the manuscript.

In the revised manuscript, we have added a schematic flowchart of the retrieval procedure at the beginning of Section 3(pages 7-8, lines 172–181) (new Fig. 5; see Supplement), which provides a concise and intuitive illustration of the overall processing chain of the proposed method. The flowchart explicitly indicates the main processing steps, including the input data sources, the application of correction models and physical constraints, the forward radiative transfer simulation, the NSGA-II optimization procedure, and the generation of the final temperature and humidity profiles.

With this flowchart, readers can more clearly understand how the different models and algorithms are organized and interact within the entire retrieval framework.



Section 3.4.3

L 212: How does the lapse rate constraint performs or reacts under temperature inversion conditions. Please comment or show in a case study.

Response:

We thank the reviewer for raising this important question. We clarify that the so-called "lapse rate constraint" used in this study is not a strict lapse-rate constraint, but rather an inter-layer continuity (smoothness) constraint that limits the *magnitude* of the layer-to-layer differences between adjacent vertical levels (i.e., $|T_{i+1} - T_i| \leq \delta_1$ and $|H_{i+1} - H_i| \leq \delta_2$, applied to all adjacent layers in practice). Because the constraint is formulated using the absolute value, it does not discriminate between positive and negative temperature gradients and therefore does not suppress temperature inversions. As long as the layer-to-layer changes remain within the prescribed bounds, inversion layers are permitted. The primary role of this constraint is to suppress nonphysical oscillations between neighboring layers and to enhance retrieval stability, rather than to enforce a monotonic decrease of temperature with height.

In the revised manuscript, we added explicit text in Section 3.3.3 to clarify the behavior of this constraint under inversion conditions (page 13, lines 302–312). While we agree that case studies (with/without inversions) can be helpful, to keep the manuscript concise and avoid redundancy with the existing statistical evaluation, we did not add additional case-study figures at this stage. We will consider including one or two illustrative inversion/non-inversion cases in a future extension, or if the reviewer/editor considers it necessary.

L 216: here the authors mention some quality metrics, but the reader has no information about its meaning. There is no introduction to these metrics nor an interpretation about the values? Is it good or bad? Please clarify!

Response:

We thank the reviewer for this important comment. We agree that the description of the computational performance metrics in line 216 required further clarification. In the revised manuscript, we have explicitly stated that the reported "single-run inversion time" refers to the computational time required for one retrieval under the specified parameter configuration (175 individuals and 10 generations).

In addition, we have added an interpretation of this value by comparing it with the typical temporal resolution of shipborne microwave radiometer observations (on the order of minutes). This comparison indicates that the proposed algorithm can complete one retrieval within a comparable time scale and therefore has the potential for quasi-real-time operational applications. With these additions, the physical meaning and practical relevance of this performance metric are now clearer to the reader.

Accordingly, the original sentence

"Furthermore, employing a parallel computing architecture and systematic error correction mechanism, the single-run inversion time was reduced to the second level under parameter configurations of 175 individuals, 10 generations, a crossover probability of 0.9, and a mutation probability of 0.2, meeting the requirements for real-time maritime applications."

has been revised (page 12, lines 281–284) to:

"Furthermore, by employing a parallel computing architecture and a systematic error correction mechanism, the single-run inversion time is on the order of several minutes under a parameter configuration of 175 individuals and 10 generations (with a crossover probability of 0.9 and a mutation probability of 0.2). This computational cost is acceptable for operational shipborne applications and enables quasi-real-time processing of the observed data."

L 219: Up to this point, the reader has absolutely no idea what the retrieval actually looks like.

Response:

We thank the reviewer for pointing out this issue. We agree that in the original manuscript, the description of the retrieval method focused mainly on the algorithmic concept. In contrast, the overall retrieval workflow was not presented in a sufficiently intuitive manner, which may have made it difficult for readers to understand how the retrieval is actually performed.

In the revised manuscript, we have added a flowchart at the beginning of Section 3 (new Fig. 5), which systematically illustrates the complete retrieval procedure, including brightness temperature observations, attitude correction, forward radiative transfer simulation, construction of the objective functions, NSGA-II optimization, and the generation of the final temperature and humidity profiles. In addition, a corresponding textual explanation has been added at the beginning of Section 3.3.3 (pages 11-12, lines 266–269), explicitly outlining the main steps of the retrieval.

The following text has been added to the revised manuscript:

"To clarify how the retrieval is actually performed, the complete inversion workflow is illustrated in Fig. 5. 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 error correction to obtain the final temperature and humidity profiles."

With these additions, the retrieval procedure is now presented more transparently and intuitively, enabling readers to better understand how the inversion is implemented in practice rather than only at the conceptual level.

L 221: Is the retrieval an optimal estimation method? Please describe the method in more detail and give references if the method or parts of were already published.

Response:

We thank the reviewer for raising this important question. We clarify that the retrieval approach adopted in this study is not based on the Optimal Estimation Method (OEM). Instead, the temperature and humidity profile retrieval is formulated as a physically constrained multi-objective optimization problem and is solved using the Non-dominated Sorting Genetic Algorithm II (NSGA-II).

The term "NOGA-II" appearing in the original manuscript was a typographical error and has been corrected throughout the revised manuscript to "NSGA-II". In addition, following the reviewer's suggestion, we have expanded the description of the retrieval strategy in the revised manuscript (page 12, lines 269–273) and added the classical reference for NSGA-II (Deb et al., 2002) to clarify the theoretical basis of the method and to avoid confusion with OEM.

Unlike OEM, which is based on a Bayesian framework and relies on background and observation error covariance matrices, the proposed method simultaneously minimizes multiple objective functions and directly searches for a Pareto-optimal solution set using NSGA-II. This strategy allows stable retrieval results to be obtained without explicitly constructing prior error covariance matrices.

The following text has been added to the revised manuscript:

"Addressing the challenges of sparse data, significant 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."

L 223 – 226: Here the authors use a very repetitive description. Please shorten the paragraph.

Response:

We thank the reviewer for this helpful suggestion. We agree that the original description of the objective function in lines 223–226 was repetitive. In the revised manuscript, we have shortened and streamlined this paragraph by merging the repeated explanations for different frequency bands into a unified formulation, thereby reducing redundancy and improving clarity and readability.

Accordingly, the text in the revised manuscript (pages 12-13, lines 299–301) has been modified to:

"Specifically, the objective function is constructed from the residuals between simulated and observed brightness temperatures in the K-band (22 GHz) and V-band (58 GHz). In this formulation, $\tilde{T}_{ret}$ *and* T_{obs} *denote the simulated and measured brightness temperatures, respectively, and the superscripts indicate the corresponding frequency channels."*

Paragraph 2: Please add references for the density sorting and Pareto techniques or describe them in more detail.

Response:

We thank the reviewer for this important suggestion. Following the reviewer's advice, we have provided a more detailed description of the density sorting and Pareto techniques in the revised manuscript (page 13, lines 317–322).

Specifically, we clarified that Pareto dominance is used to rank candidate solutions into different non-dominated fronts according to multiple objective functions, and that a density-based sorting strategy based on the crowding distance is applied to preserve the diversity of solutions along the Pareto front. In addition, we have added the classical reference for the NSGA-II algorithm (Deb et al., 2002) to support this methodological description.

The following text has been added to the revised manuscript:

"To ensure both convergence and diversity of the solution set, Pareto dominance is used to rank individuals into different non-dominated fronts according to the multiple objective functions, while a density-based sorting strategy (crowding distance) is applied to maintain solution diversity along the Pareto front. This multi-objective selection mechanism allows the algorithm to search for stable atmospheric temperature and humidity profiles while avoiding premature convergence (Deb et al., 2002)."

L241: I kindly suggest replacing the terms "master-slave" with more modern and neutral terminology, such as "leader-follower" or "controller-peripheral". This would improve the inclusiveness of the language without affecting the technical meaning.

Response:

We thank the reviewer for the valuable suggestion regarding terminology. In the revised manuscript (page 13, lines 324–328), we have replaced the term "*master–slave*" with the more neutral and modern expression "*controller–worker*" thereby improving the appropriateness and inclusiveness of the language without changing the technical meaning of the parallel computing scheme.

In addition, we have retained the statement regarding the improvement in computational efficiency and slightly revised it to more clearly emphasize its contribution to enhancing the efficiency of the retrieval procedure.

In general, the description of the method would benefit from improved clarity. I suggest adding two figures to support the methodological explanation. The first could be a schematic overview of the processing chain, as suggested above. The second could illustrate example temperature and humidity profiles and their evolution as the retrieval is applied at each processing step. Such visualizations would make it easier for readers to follow and better understand the method.

In addition, I note that the manuscript includes a detailed presentation of the RMSE equation, which is generally well known and may not require explicit introduction, while other key aspects of the method are described only briefly. I therefore encourage the authors to reconsider the structure of the methods section and to revise it comprehensively in order to clearly and transparently present all essential components of the proposed approach.

Response:

We thank the reviewer for these helpful and constructive general comments. We agree that in the original manuscript, the presentation of the RMSE equation was relatively detailed, whereas some key components of the retrieval procedure were described too briefly.

Following the reviewer's suggestions, we have revised and reorganized the Methods section to improve its clarity and overall structure. Specifically, we have removed the explicit formula and detailed definition of the RMSE, streamlined repetitive descriptions, and added a schematic overview of the retrieval procedure at the beginning of Section 3. This new flowchart provides an integrated view of the main processing steps, including the construction of the objective functions, the introduction of physical constraints, the NSGA-II optimization process, and the generation of the final temperature and humidity profiles.

Regarding the suggestion to add a second figure illustrating the evolution of temperature and humidity profiles at different processing stages, we note that the Results section already presents representative retrieved profiles together with reference profiles to demonstrate the retrieval performance and the effect of the systematic error correction. To avoid redundancy and excessive length of the manuscript, we did not introduce an additional step-by-step evolution figure. Instead, we further strengthened the textual explanation of the retrieval procedure in Section 3.3.3 to more clearly describe how the profiles are obtained through the different stages of the inversion.

We believe that these revisions substantially improve the clarity and transparency of the methodological description and make the proposed retrieval approach easier for readers to follow and understand.

L 253ff: I would appreciate some clarification regarding the treatment of systematic errors. According to the manuscript, the retrieval output is corrected using radiosonde data. However, in an operational application over the open ocean, radiosonde observations are typically not available. It is therefore unclear whether the authors intend to apply systematic error corrections derived from the 38-sample dataset. If so, the relatively limited size of this dataset may warrant further discussion. More generally, this approach appears to rely on historical data, whereas one of the stated motivations of the method is to reduce such dependence. Clarification on this point would help to better understand the intended applicability of the proposed retrieval. If the correction is based on the 38 samples and is later applied to the bias, which is also based on the 38 samples, then the algorithm is applied to its training data. This distorts the result significantly in a positive direction. Unseen data would produce worse results. Please comment on that.

Response:

We thank the reviewer for this critical and insightful comment. We agree that estimating the systematic error and evaluating the retrieval performance on the same dataset may lead to overly optimistic results and blur the distinction between calibration and validation. In response to the reviewer's concerns regarding "unseen data", operational applicability over the open ocean, and dependence on historical data, we have clarified and extended the manuscript as follows.

(1) Independent validation on unseen data (to avoid overfitting):

To rigorously assess the effectiveness of the systematic error correction on unseen data, the 38 available samples were randomly divided into a training subset (80 %, N = 31) and an independent test subset (20 %, N = 7). The training subset was used exclusively to derive the height-dependent systematic error profile $E(h)$, while the test subset was reserved solely for validation.

On the independent test subset, the resulting performance is RMSE = 1.99 K for temperature and RMSE = 21.85 % for relative humidity, which is comparable to the statistics obtained using the full dataset (2.08 K and 20.95 %, respectively). This indicates that the systematic error correction is robust and does not exhibit clear overfitting behavior.

(2) Applicability to operational open-ocean conditions:

It should be noted that radiosonde observations are used only under land-based conditions to estimate the height-dependent systematic error profile as part of an instrument calibration procedure. In practical operational applications over the open ocean, this derived systematic error profile can be applied as a fixed correction term to the retrieval results and does not require the availability of concurrent radiosonde observations.

(3) Clarification regarding dependence on historical data:

The statement that the proposed method "reduces dependence on historical data" should be understood in contrast to statistical learning approaches (e.g., neural networks) that typically require thousands of training samples. In the present study, only a small calibration dataset (38 profiles) is used to estimate a sensor-specific systematic bias, which corresponds to a standard engineering practice for zero-bias correction. The core retrieval algorithm itself is based on a physically constrained multi-objective optimization framework (NSGA-II) and does not rely on long-term statistical priors derived from large historical datasets.

The following text has been added to the revised manuscript (page 14, lines 352–357):

"It should be noted that the radiosonde observations are used only to estimate the height-dependent systematic error profile under land-based conditions as part of an instrument calibration procedure. In operational applications over the open ocean, the derived systematic error profile can be applied as a fixed correction term and does not require the availability of concurrent radiosonde data. To avoid overly optimistic performance estimates, the systematic error profile is derived from the training subset only, while the independent test subset is used exclusively for validation."

Through these clarifications, we aim to better distinguish between calibration and validation, to explain the intended operational use of the systematic error correction, and to avoid an unduly optimistic assessment based on training data.

L 278-282: I think it is enough to mention that the RMSE gives the statistical uncertainty. It does not need to be defined here in detail. It is general knowledge. Please remove the formula.

Response:

We thank the reviewer for this suggestion. Following the reviewer's advice, we have removed the explicit RMSE formula and its variable definitions from the revised manuscript (page 15, lines 377–379) and retained only a brief textual description stating that RMSE is used as a statistical measure of retrieval uncertainty.

Since RMSE is a commonly used and widely understood evaluation metric, it is no longer introduced or defined in detail in the manuscript.

Section 4: results

I suggest beginning the results section with two illustrative case studies in which all relevant aspects required for the subsequent statistical analysis are explained in detail. For example, cases with and without temperature or humidity inversions could be selected. This would help readers develop a better intuitive understanding of the associated uncertainties and would clearly demonstrate the impact of the systematic error correction on individual cases.

Response:

We thank the reviewer for the constructive suggestion regarding the structure of the Results section. We agree that introducing representative case studies can help readers develop a more intuitive understanding of retrieval uncertainties and the effect of systematic error correction.

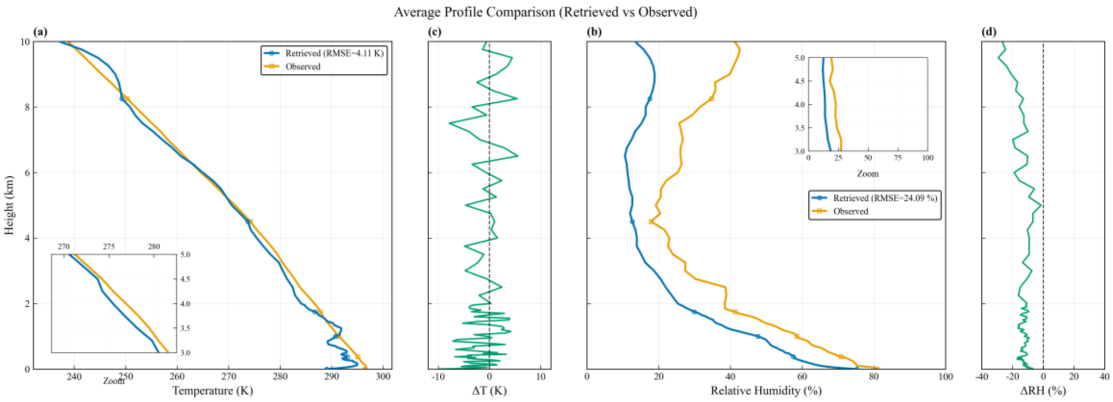
In the revised manuscript, we have strengthened the presentation of the results and now systematically show the comparison between retrieved and radiosonde temperature and humidity profiles, as well as their corresponding error distributions as a function of height, before and after systematic error correction (new Figs. 7 and new Figs. 8). These mean profile statistics already integrate a wide range of atmospheric conditions, including situations with and without inversion layers, and therefore reflect the overall statistical behavior of the retrieval under different atmospheric structures.

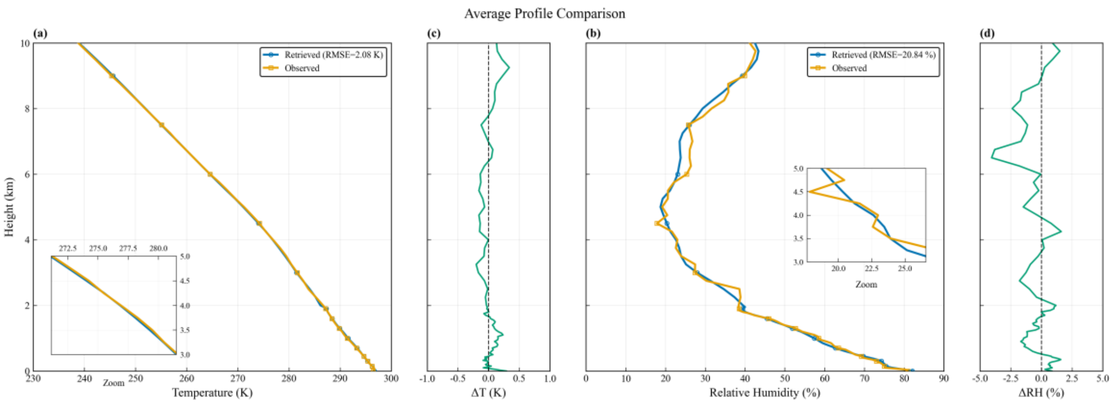
Given that the primary objective of this study is to evaluate the stability and robustness of the proposed method in a statistical sense rather than to provide a qualitative analysis of individual cases, we consider that emphasizing statistical results better supports conclusions regarding the general applicability of the method. For this reason, we did not introduce additional individual case-study figures, but instead demonstrate the impact of the systematic error correction through profile comparisons and error distributions in the existing results. Nevertheless, we appreciate the reviewer's suggestion, and we will consider adding one or two illustrative case-study figures in a future extension or if the reviewer/editor considers it necessary.

Figure 4 and 5: I suggest to move the panels c and d on the right of each prolile that all plots can share the y axis.

Response:

We thank the reviewer for this helpful suggestion. Following the reviewer's advice, we have reorganized the layout of Figs. 4(new Figs. 7) and 5(new Figs. 8). In the revised manuscript (page 15, line 380 and page 17, line 416) （see Supplement）, the difference panels (c) and (d) are now placed to the right of the corresponding temperature and humidity profile panels ((a) and (b)), and they share the same vertical (height) axis. This revised arrangement improves the visual consistency of the figures and allows a more intuitive comparison between the retrieved profiles and the radiosonde observations.





L 297: I would be cautious in describing the results as "highly consistent", as absolute deviations exceeding 5 K are evident. In this context, it would be helpful to provide information on the general uncertainty of the instrument under land-based operating conditions. Has this performance been evaluated previously, for example in an earlier publication, or has it been assessed by the authors themselves? Providing such context would allow a more informed interpretation of the reported deviations.

Response:

We thank the reviewer for this important and careful comment. We agree that describing the results as “highly consistent” may not be sufficiently rigorous when absolute deviations exceeding 5 K are present at certain altitude levels. In the revised manuscript (page 16, line 399), we have therefore weakened this wording to avoid potential overstatement. Specifically, the original phrase“…profile exhibits a trend highly consistent with the sounding data…” has been revised to“…shows an overall agreement with the sounding data…”.In addition, to maintain a conservative and scientifically rigorous description of the retrieval performance, we have further refined the wording in the Results section. In the revised manuscript (Page 17-18, Line 431-432), the phrase “excellent accuracy” has been replaced with “good retrieval accuracy”. This change avoids potential overstatement and ensures consistency with the general performance level expected for ground-based microwave radiometric retrievals.

In addition, to provide an appropriate physical context for interpreting these deviations, we have added information on the typical uncertainties of ground-based microwave radiometric profilers under land-based operating conditions in Section 2.1 (page 5, lines 116–122). This includes the typical accuracy of brightness temperature measurements as well as the expected uncertainties in retrieved temperature and humidity profiles, together with relevant literature references. These additions allow a more informed interpretation of the reported deviations in relation to the intrinsic performance of this class of instruments.

L 299: The statement “… 1.5–3 K between 3 and 6 km …” does not appear to be consistent with Figure 4c, where deviations of around –5 K can be observed. It would be helpful if the authors could clarify the origin of these values or consider rephrasing this statement to accurately reflect the data shown.

Response:

We thank the reviewer for this careful and helpful comment. We agree that the original statement indicating a deviation of “1.5–3 K between 3 and 6 km” is not fully consistent with Fig. 4(c), where both positive and negative deviations are present and local values reach approximately −5 K.

In the revised manuscript (pages 16, lines 401–404), we have therefore modified this description to more accurately reflect the data shown in Fig. 4(c)(new Figs. 7). The text now states that, in the 3–6 km altitude range, the temperature differences exhibit both positive and negative deviations, with magnitudes generally remaining within ±5 K. We

further point out that larger fluctuations around 4–5 km are associated with the reduced sensitivity of the microwave channels at these heights.
The revised text reads:
"However, with increasing altitude, the temperature differences exhibit both positive and negative deviations in the middle troposphere (approximately 3–6 km). As shown in Fig. 7(c), the deviations generally remain within ±5 K, with larger fluctuations occurring around 4–5 km, which is consistent with the reduced sensitivity of the microwave channels at these heights."
L 301: The statement "Beyond 6 km, the deviation recedes to 1–2 K but …" does not appear to be clearly supported by Figure 4c. Again, it would be helpful if the authors could clarify the source of these values or consider rephrasing this statement to more accurately reflect the data presented.
Response:
We thank the reviewer for this helpful comment. We agree that the original statement claiming that "the deviation recedes to 1–2 K beyond 6 km" is not sufficiently supported by Fig. 4(c)(new Figs 7(c)), since noticeable fluctuations are still present at these altitudes.
In the revised manuscript (page 16, lines 405–406), we have therefore removed the specific numerical range of 1–2 K and rephrased the sentence more qualitatively. The revised text now indicates that, above approximately 6 km, the deviation magnitude is reduced compared to the 3–6 km layer, while non-negligible fluctuations remain. This behavior is attributed to the decreasing sensitivity of the microwave radiometer at higher altitudes.
The revised sentence reads:
"Above approximately 6 km, the temperature differences show a reduced magnitude compared to the 3–6 km layer, although non-negligible fluctuations are still observed (Fig. 7(c))."
L 309: I don't get where the mesosphere here is coming from. I guess it is a translation error?
Response:
We sincerely thank the reviewer for pointing out this important terminology issue. We fully agree that the use of the term "mesosphere" in the original manuscript is inappropriate and inconsistent with the atmospheric physics context of this study.
The present work focuses exclusively on the retrieval of temperature and humidity

profiles within the troposphere (below approximately 10 km). The mesosphere generally refers to the atmospheric region between about 50 and 85 km altitude, which is clearly outside the scope of this study. In the original text, the authors intended to refer to the middle tropospheric height range, where the retrieval results exhibit more pronounced height-dependent systematic error characteristics.

Accordingly, in the revised manuscript, the term *"mesosphere"* has been corrected to *"middle troposphere"* to accurately reflect the atmospheric region being discussed. This correction has been made in the Results section on page 16, line 413, to ensure proper scientific terminology and to avoid conceptual ambiguity.

L 349: It appears that the authors are, in fact, applying an error correction based on historical data. Please clarify.

Response:

We thank the reviewer for raising this important point. We agree that it is necessary to clarify whether the systematic error correction relies on historical data.

In the revised manuscript (pages 18, lines 453–458), we have explicitly clarified that historical radiosonde observations are used only to characterize a height-dependent systematic bias profile, rather than to train the retrieval algorithm itself. In other words, the radiosonde data are employed solely for estimating the systematic error as part of an instrument calibration procedure.

Specifically, the systematic bias profile is derived from a training subset of the available samples, while its correction performance is evaluated using an independent testing subset. This strategy avoids applying the correction to the same data from which it is derived and thus prevents overly optimistic performance estimates.

We have added the following clarification to the manuscript:

"The NSGA-II-based retrieval framework reduces the dependence on large historical training datasets. It should be noted that historical radiosonde observations are used only to characterize a height-dependent systematic bias profile, whereas the retrieval process itself does not rely on extensive historical training samples. The systematic bias profile is estimated from a training subset, and its correction performance is evaluated using an independent testing subset. The statistical results demonstrate that this approach provides a robust solution for real-time retrieval of tropospheric atmospheric parameters in data-sparse marine environments."

This revision more clearly delineates the role of historical data and helps to clarify the intended applicability of the proposed retrieval method.

L 375: The authors indicate that the extension of the extrapolation to larger marine areas will be addressed in a future publication. This actually underscores the importance of thoroughly characterizing the instrument itself and its performance under land-based conditions, as noted in my earlier comments. Providing a clear assessment of the instrument's capabilities would strengthen the current manuscript and help contextualize its applicability to marine environments.

Response:

We thank the reviewer for this insightful comment. We fully agree that, before discussing the extension of the proposed approach to larger marine areas, it is essential to provide a clear characterization of the instrument performance under land-based conditions, in order to properly assess its applicability to marine environments.

In response to this comment, we have supplemented Section 2.1(Page 5, Line 116-122) of the revised manuscript with additional information on the typical measurement uncertainties and performance characteristics of ground-based multi-channel microwave radiometers. This includes the expected accuracy of brightness temperature measurements as well as the typical uncertainties of retrieved temperature and humidity profiles, together with relevant literature references.

These additions provide an important physical context for interpreting the retrieval deviations reported in this study and strengthen the basis for discussing the potential applicability of the proposed method to marine environments.

L 489: There is a typo in 'Wat0er'.

Response:

We thank the reviewer for pointing out this spelling error. We have corrected *'Wat0er'* to *'Water'* on page 23, line 603 of the revised manuscript.