Dear Dr. Anthes,

We deeply appreciate your careful reading, editing and insightful comments and suggestions, which have greatly improved our manuscript. We have revised our manuscript significantly to address your comments and suggestions. Please see our point-by-point response (in blue) to your comments/suggestions as follows.

**Major Comments:**

1. The potential use of this statistical technique to estimate the likelihood and magnitude of refractivity biases in individual RO observations for data assimilation (DA) could be mentioned after line 108. However, most global models assimilate bending angles (BA) rather than refractivity. Did the authors try their statistical model to estimate BA biases? If so, they could summarize what they found. If not, this could be a topic for further study.

Thank you very much for this important comment. We now clarify the motivation of this study in this study.

- We have mentioned the potential use of the statistical technique proposed in this study in line 106-110 to address reviewer's comment.

  Our study focuses on the refractivity bias (REFB) in the lower troposphere for two reasons. First, we would like to use the estimation model to understand the characteristics of REFB and how they link to LSW and thermodynamic condition in the lower atmosphere, particularly in the planetary boundary layer. The bias estimation can be used to calibrate RO refractivity, which can be applied to improving the products of temperature and moisture profiles retrieved from the refractivity in the moist lower troposphere or estimating precipitable water vapor (Yeh et al. 2024). Second, the estimated REFB can be used for data assimilation purposes. With the DA systems that assimilate the RO-REF profiles, it is expected that the RO data in the lower atmosphere can be better exploited by using the bias estimation as a QC flag or assimilating the calibrated REF profiles. Although most global models assimilate the bending angles, the data has large uncertainties in the moist lower troposphere or is unavailable below the ducting layer. It is unclear whether the assimilating bending angle can positively impact low-level moisture accuracy. If the RO refractivity can be calibrated, they can likely be used in these conditions!

Here, we show an example of the RO profile over the ocean with moist PBL. The ECMWF 12-h forecast (echPrf) suggests the potential existence of ducting. The RO refractivity seems negatively biased below the ducting layer (black dashed lines), while the RO bending angle has significant variations in the low levels and quickly decreases to a very small value near the surface.

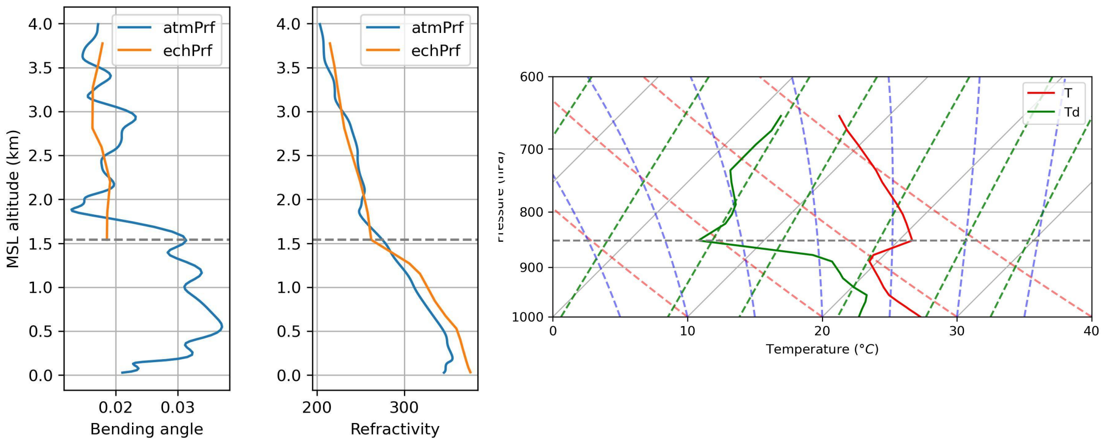


**Figure A An example of a RO profile with the existence of ducting over South China Sea. The atmPrf and echPrf of (a) bending angle and (b) refractivity. (c) Vertical distribution of the temperature and due-point temperature at this location from the ECMWF 12-h forecast.**

2. A key part of this paper is the regression model for REFB vs. LSW or T and q. It is not clear how the two regression equations for LSW and for T and q were obtained. Why is Eq. (5) a quadratic in LSW and not some other relationship (e.g. linear, cubic, or higher order)? Why is Eq. (7) quadratic in Q (normalized specific humidity) plus the product of normalized Q and T? What other polynomials were tested? Presumably some of the process in selecting the optimum polynomial is described in lines 181-188, but additional detail would be useful.

We apologize for the unclear description of our methodology in the previous version of the manuscript. Different combinations of the order of the predictors are evaluated to define the optimal equation (polynomial) to best represent the real REFB. This section (section 2.2) has been revised significantly to clarify and address your comments. In the following, we explain in more details about how to find out the best formula of each participated variable and their associated coefficients.

The feature engineering and model selection tools in Scikit-learn (Python 3.1) are used to evaluate the fitting performance with the polynomial regression model. The evaluation is conducted for the 'degree' parameter ranging from 1 to 6 with the metrics of R-squared ($R^2$) and mean squared error (MSE).

The following figure shows the $R^2$ of the LSW-based polynomial with different degree (order), and the corresponding computational time. Although a higher $R^2$ is obtained with the higher degree of polynomial, the computational cost is increased. The figure shows that $R^2$ significantly increases from a linear to quadratic form of the polynomial and becomes saturated. In comparison, the computational cost increases linearly as the degree increases. This suggests that the evident gain of the performance is obtained from the linear to quadratic form and saturates for higher order. To avoid overfitting and consider the computational time, the quadratic form was chosen for the LSW-based REFB model.

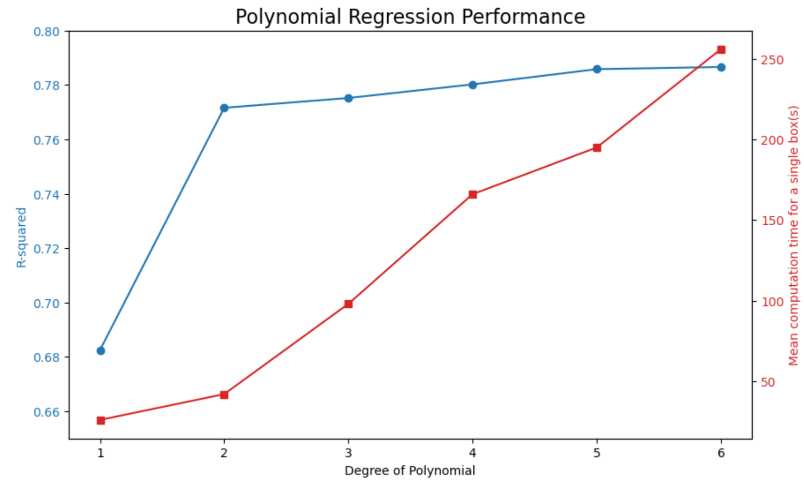


**Figure B The performance and computation time with respect to different degree of polynomial regression using LSW as the predictor**

For the TQ-based model, there are two variables ($T$ and $Q$) used in the regression model and the product term ($TQ$) is included to consider the joined effect (interaction) between these two variables. The following table lists $R^2$ and MSE of polynomials with different variable combinations. Results show that the formulas including $Q^2$, $Q$ and $TQ$ terms gives the best fitting performance. In particular, including the quadratic term of moisture is essential. The R-squared value increases from 0.535 with the $Q$ and $TQ$ terms to 0.732 with the $Q^2$ term. However, $TQ^2$ or $T^2Q$ only mildly increases $R^2$ but degrades MSE (Table A). This suggests that the high order interactive terms do not add significant benefit to the fitting. Therefore, we choose the formula with $Q^2$, $Q$ and $TQ$ terms.

We have briefly included the discussion in line 184-187, 191-192.

**Table A: $R^2$ and MSE of polynomials with different variable combinations**

| Equation forms | Order of variables: 1 | | | | | Order of variables: 2 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T and Q | TQ and const | Q, TQ | $Q^2$, Q | $Q^2$, Q, TQ | $Q^2$, Q, $TQ^2$ + TQ | $Q^2$, $T^2Q$, $TQ^2$, TQ, Q | $Q^2$, $T^2$, $T^2Q$, $TQ^2$, TQ, Q, T | $Q^2$, $T^2$, $T^2Q$, TQ, T | $T^2$, $T^2Q$, TQ, T | $T^2$, T, $T^2Q$ | $T^2$, T | $T^2$, $Q^2$ |
| $R^2$ | 0.399 | 0.421 | 0.535 | 0.481 | 0.732 | 0.734 | 0.734 | 0.689 | 0.695 | 0.245 | 0.115 | 0.111 | 0.211 |
| MSE | 77.267 | 72.115 | 37.044 | 42.886 | 26.610 | 26.611 | 26.611 | 36.614 | 34.205 | 94.336 | 101.591 | 109.460 | 98.629 |

3. In Section 2.3 the authors say that they derived the statistical models using the data for five different subsets of the data and chose the ones with the best fits as their model. I am not an expert in statistics, but why not use the entire sample of data for their statistical model? The "best" model for one subset may not be the "best" model overall? And if more subsets (e.g. 10) were chosen, a different "best" model would be obtained. This issue should be discussed.

Splitting the data into training and testing subsets is commonly adopted for constructing a statistical model. The training data is used to build the regression model. The testing data, independent of the training data, is used to evaluate the performance of the regression model. The fitting performance may be very good with the training data but poor with the testing data (i.e. an overfitting case). Therefore, the testing data is required to avoid overfitting and ensure the robustness of the derived model.

Theoretically, we need sufficient (randomly generated) training data to capture the general behaviors of the data. Given that the sample size is limited, we try to find the most representative regression model by repeating the training procedure with different data portions (80%) and evaluating the fitting performance with the rest of the data (20%). The regression model that can best fit the testing data is selected. With 20% data as the testing data, the replacement of the testing data is repeated five times so that the regression model is eventually applied to the entire data set.

The applications of the regression models are conducted in two parts.

(1) All profile data (80% for training and 20% for testing) is used to determine the order of the LSW-based regression model and the optimal combination of the multi-variable (T and Q) regression model (Given two variables $y$ and $z$, there are different combinations of order and interaction terms as $\sum_{m=0}^{m=M}\sum_{l=0}^{l=L} b_{m,l}y^m z^l$, where $m$ and $l$ are the order of variable $y$ and $z$, respectively, and $b_{m,l}$ is the regression coefficient).
(2) The forms of the polynomial regression are then applied regionally. From 45°S to 45°N, we define 72 x30 boxes and each box is 5° longitude x 3° latitude degree. The boxes are defined by considering the number of available RO profiles below 1.5km should be sufficient for regression testing. With the 3 months of data used in our study, choosing testing data lower than 20% of the total sample results in a very coarse resolution of the boxes. Choosing any number larger than 20% would sacrifice the amount of data that can train a reliable regression model. In each box, the regression fitting is also repeated five times to derive the optimal regression coefficients.

The discussion above has been included in the section of methodology (section 2.2).

**Specific comments:**

1. Line 123: Fig. 1 is labeled profile density, but it is actually profile counts. The label should be changed.

   We have corrected this, thank you!

2. Line 128: The issue with possible biases in the reference (ERA5) should be mentioned here.

   Thank you for pointing this issue. We have addressed this comment accordingly in line 136.

   "Nevertheless, it is possible that ERA5 may carry its own biases, which will not be discussed in this study"

3. Sometimes N-REFB is used and other times REFB is used. Please be consistent. Since you are only discussing refractivity biases, I suggest using just REFB. When referring specifically the negative biases, I suggest saying "negative REFB."

   We apologize for the inconsistency. We now only use REFB and replace N-REFB to negative REFB as the reviewer suggested.

4. 5 caption: I suggest emphasizing that the REFB, LSW, q and T in Fig. 5 are all averages over the lowest 1.5 km MSL. Add to the caption: "The values of REFB, LSW, specific humidity and temperature are averages over the lowest 1.5 km MSL of the atmosphere." And in line 251 write "the averaged value of REFB below 1.5 km"

   Thank you for your great suggestion. This is emphasized in the caption of figure 5 and wherever it is necessary.

5. 6 caption: "Vertical cross section of refractivity bias over the ocean as a function of height and average values over the lowest 1.5 km of (a) LSW/2, (c) specific humidity and (e) temperature over land……….

   Thank you for your suggestion. We have corrected the caption of Figure 6.

6. Lines 131-173 (Section 2.2)

   This section contains some incorrect or misleading statements and is unnecessary for this paper. For example, in Lines 144-146: "Normal" is not well defined; nonspherically symmetric conditions are common. Spherically symmetric means no horizontal variation of refractivity on a constant level surface, either small-scale turbulent variations in T and q or larger-scale horizontal gradients of T and q. In line 146, a large vertical gradient of refractivity does not necessarily imply nonspherical symmetry, which depends on horizontal variations of N not vertical gradients. I suggest a much shorter simplified summary that refers to more complete discussions of the causes of negative refractivity biases. Here is an example:

"Negative refractivity biases can arise in the atmosphere from multiple causes, as summarized by Feng et al. (2020) and Wang et al. (2020). A common cause (but not the only one) of negative biases in the lower troposphere is ducting or superrefraction (Sokolovskiy 2003; Ao et al. 2003). When the vertical gradient of refractivity $\partial N/\partial z$ exceeds a critical value of -157 N units per km, ducting occurs and rays are trapped inside the ducting layer. This leads to a negative bias in N and there are an infinite set of bending angle profiles that correspond to the observed refractivity profile."

We deeply appreciate your corrections and clarification for the nonspherical symmetry and vertical gradient of refractivity. Since this section is greatly reduced, we merged the main sentences to the introduction to review the causes of negative refractivity bias (line 59-75).

7. Line 175: Why not use LSW as a predictor instead of LSW/2? The correlations should be the same.

The LSW is the standard deviation of bending angle, which defines the data uncertainty. In Liu et al. (2018), LSW/2 is used as a predictor and the rescaling factor is to represent the standard deviation of a Gaussian distribution. This is now clarified in line 147-148.

8. Line 194—what 1D-Var algorithm was used? The original CDAAC wetPrf or the new CDAAC wetPf2 (Wee et al. 2023)? Or some other one?

The wet product we used in the study is the 1D-var wetPrf2. The wetPf2 refractivity is derived from the bending angle profile of atmPrf and through a variational regularization of Abel transform (Wee, 2018), instead of the traditional Abel inversion. We have clarified that the T and Q obtained from the 1D-Var analysis of the RO wet products in line180.

Wee, T.-K.: A variational regularization of Abel transform for GPS radio occultation, Atmos. Meas. Tech., 11, 1947–1969, https://doi.org/10.5194/amt-11-1947-2018, 2018.

9. Lines 255-257: there is no direct relationship between large vertical gradients of N and nonspherical symmetry, which is caused by horizontal variations in N. Large horizontal variations in N and corresponding large LSW may occur with small vertical gradients of N. I suggest deleting the two sentences in lines 255-257; the previous sentence is sufficient.

We apologize for the misleading statement. Thank you very much for your correction. These sentences are deleted in the revised manuscript.

10. Which is land and which is ocean in Fig. 8? The caption should provide more details, i.e. explain the dots, explain the surface (is it a fit to the dots?)

Thank you for your comments. For better illustration, we have re-plotted Figure 8 to be function of normalized *Q* and *TQ* over ocean (Fig. 8a) and land (Fig. 8b) in Southern Hemisphere. We also included Figs. 8c and 8d to show the same plot for Northern Hemisphere. Details about this figure are provided in the caption.

11. Lines 311-312 and Fig. 9: Why aren't the "Real REFB" the same for the training and testing data? Is this a sampling issue?

The training and testing data are 80% and 20% of the total data, respectively. The real REFB in Figs. 9a and 9b are calculated with the training and testing data, respectively. Therefore, it is expected that the pattern should be very similar but have some differences.