Response to Reviewers Comments

We thank the associate editor, editor and two anonymous reviewers for their thoughtful and constructive comments and suggestions, which significantly help us to improve the quality of the manuscript. In this revised manuscript, we have tried our best as much as possible to address all concerns and have revised the manuscript accordingly. Below, we indicate the original comment of the respective reviewer in blue and our point-to-point response is denoted in black.

Before addressing the comments, we would like to express our sincere gratitude to the reviewers for their exceptionally informative, constructive, and detailed comments.

Reviewer #2 Evaluations:

The authors have compared radiosonde-based PBL height estimates with PBL heights derived from the ERA5 reanalysis for over 300 available land stations, showing a significant bias in ERA5 PBL heights. A machine learning routine is developed to predict the ERA5 PBL height bias based on numerous input parameters, and this bias is subtracted from the ERA5 PBL height to produce a corrected dataset. This produces an immediately useful and relevant dataset that can be applied in many future studies. The work is novel, well-constructed, and succinctly explained in the paper. There are a few non-structural fixes that could improve the manuscript, but no major issues with the work, so I would only call these minor revisions.

Response: Thank you for your thoughtful comments and valuable suggestions that are crucial for us to improve the quality of our manuscript. The raised concerns have been modified as much as we can.

Notes:

Line 203: can you add some detail on what you mean by the ‘second level’?

Response: As suggested, the statement has been rephrased as:

"…(1) Ri(z) in Eq. (1) exceeds 0.25, where z is the second level of radiosonde measurement…".

In equation 2, it appears that PBLH-M and PBLH-E are mis-formatted as PBLH – M and PBLH – E (where M and E are variables being subtracted) This is probably a formatting error, but is initially very confusing.

Response: The modification has been made. Thank you for providing these insights.

Line 260: I worry that randomly dividing the data can cause an issue if certain geographic regions are underrepresented in the training data. I would recommend dividing your stations into specific regions (for example: valleys, mountains, coastal, continental, tropical, polar…) and ensuring that a subset for each region is then randomly drawn for each training/validation pool. An easier solution may be just to show that the randomly selected data already chosen for training represents these differing types of regions using a map and/or histogram.

Response: We have reflected this comment by re-assessing the distribution of data over specific regions. For a random process, the index of raw dataset (increasing from 0 to n with an increment of 1, where n is the total number of dataset) is expected to be randomly perturbed. The raw index and the randomly perturbed one are assumed to be poorly correlated. As shown in Figure A, the relationship between the raw data index in training set and the randomly perturbed index over mountain (a), plain (b), tropical island (c), and artic pole (d) regions indicate that the training dataset over different region are nearly randomly distributed. This finding indicates that arrays over different regions closes to be randomly selected.

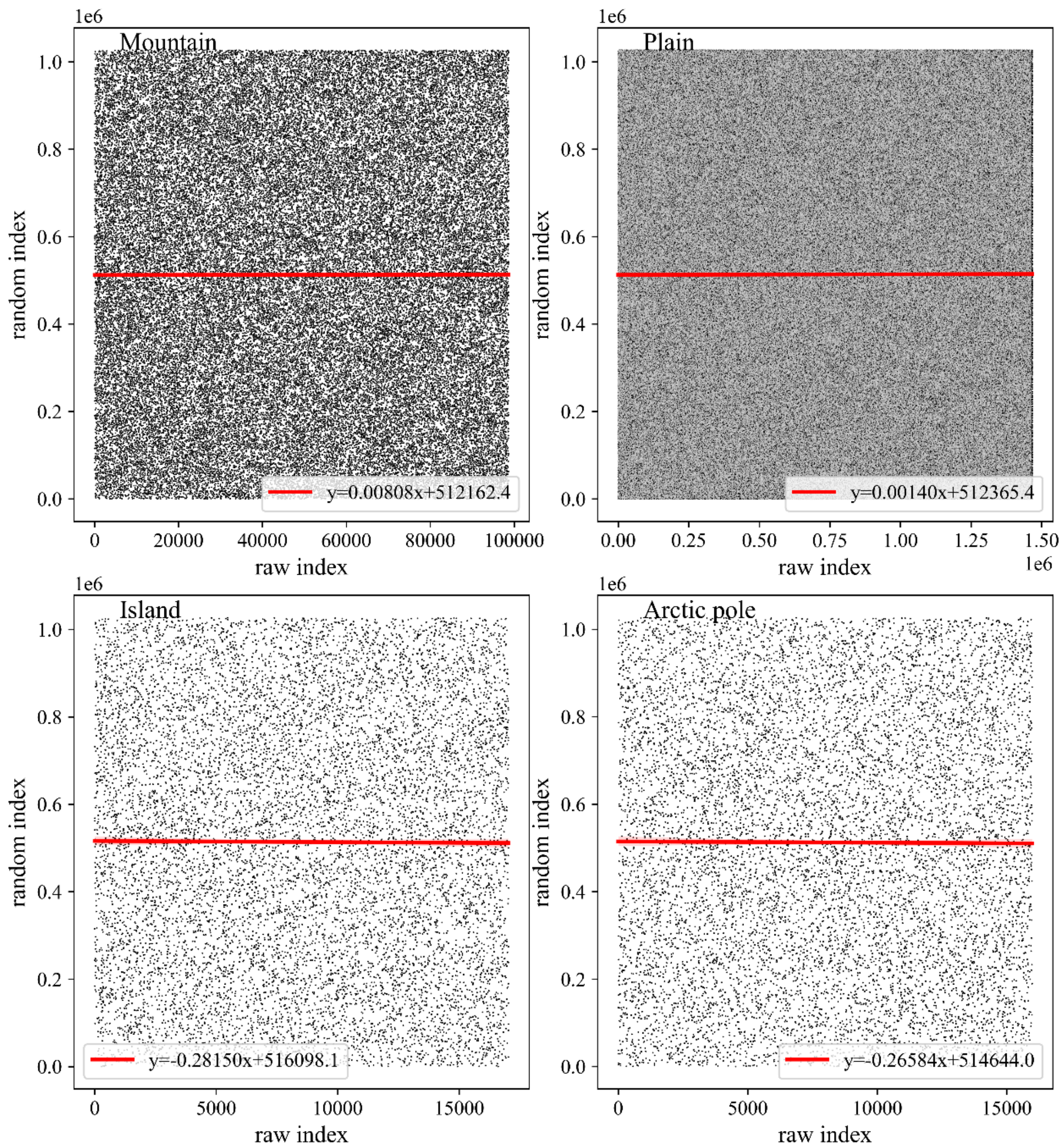

Figure A. the relationship between raw data index in training set and random index over mountain (a), plain (b), island (c), and artic pole (d) regions.

Figure 2: Dividing by calendar season for stations on both sides of the equator is not recommended, since you are lumping winter with summer, autumn with spring, etc… It would be better to combine similar seasons, so that southern hemisphere DJF is combined with northern hemisphere JJA, etc… This would better illustrate seasonal biases.

Response: We agree with you and have incorporated this suggestion throughout our paper. In the revised version, the mean bias in summer season raised from 94.15 m to

95.15 m, and from 72.46 m to 71.23 m in winter season. For the merged dataset, it is from -8.14 m to -7.88 m in summer, and it is from -0.56 m to -0.87 m in winter.

Figure 8: the panels are too small for a meaningful comparison between PBLH-R and PBLH-M (comparing the dots to the shading). I recommend making larger maps available as supporting material, or showing this comparison some other way

Response: Amended as suggested. Only 0000 UTC and 0012 UTC have been kept in the revised version. And we hope these figures in this revised manuscript could be acceptable for you.