

## **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 #1 Evaluations:**

General comments:

This study developed a state-of-the-science method to derive a global-wide PBLH dataset merging in situ observations and reanalysis dataset, which has optimized the performance of a so-called “data fusion” technology and provided critical data for climate research. There are no obvious flaws in the methodology, and the final output is informative enough to compensate for the disadvantages of current atmospheric datasets existing as the spatial-temporal discrepancy. Despite the good structure and comprehensive analysis, the authors are required to answer or address the following questions or comments. After that, I think this manuscript can be accepted for publication.

Response: We appreciated tremendously your thoughtful comments and positive review on our article. According to your nice suggestions, the concerns raised by you have been addressed in this revision as much as we can. Please see the following response and the revised manuscript for more details.

Specific comments:

Line 123: It is suggested that the authors explain a little bit more of the relationship by

a gradient of terrain or lower-tropospheric stability induced underestimation of the PBLH.

Response: Per your suggestion, the following sentence has been supplemented in the end of this paragraph:

“...Particularly, a higher terrain gradient or a more unstable troposphere generally lead to a lower PBLH in ERA5 reanalysis.”

The title of the paper is ‘...ERA5 reanalysis, and GLDAS’. However, GLDAS didn’t occur until the last paragraph. It is suggested that the authors can add some descriptions of GLDAS.

Response: We sincerely thank the reviewer for his or her careful reading. In the abstract, we give the full name of GLDAS (Global Land Data Assimilation System). Also, in Section 2.2, the description on GLDAS has been added. In the introduction section, a new phrase has been supplemented to illustrate the GLDAS product:

“...The biases between PBLH retrieved from the ERA5 and radiosonde could be represented by the land properties, near-surface meteorological conditions, among others, and further be optimized via a machine learning model. The GLDAS incorporates satellite- and ground-based observations and produces a global, high-resolution product regarding land states and fluxes (Rodell et al., 2004)...”.

Line 154: please clarify if the interpolation is based on altitude or elevation.

Response: The interpolation is based on altitude, which has been specified in Section 2.1 as follows: “all the original soundings were evenly interpolated to the profiles with a vertical resolution of 10 m by cubic spline interpolation.”

Line 158: It seems to me not correct to say spatially even coverage. The coverage in Australia is substantially not even especially in Figure 1d.

Response: Thanks for your careful checks. The statement has been modified to “...Australia have a rich geographic coverage...”

Line 173: Any reference for the definition of LST?

Response: The references have been added.

Line 207: how did the authors match the stational PBLH and gridded PBLH in the comparison?

Response: The collocation procedure between radiosonde and ERA5 grid follows Guo et al, (2021), which can be described as follows:

- (1) The grid should contain the radiosonde station.
- (2) The UTC time (hour) of grid product and radiosonde stay the same.

The above descriptions have been incorporated into the revised manuscript.

Line 259: Please specify clearly if all the data from 2011-2021 were included in the model training stage. Were they divided by the measuring time (e.g., 0000, 0006...)?

Response: All the data were used in the model training stage, and the corresponding local time of UTC was used as the input. We also tested a solution based on 24 different models that were trained by variables on different UTCs. However, the raised issue was that the boundary between different local time may has obvious gaps, due to different models used.

The above discussions have been well incorporated into the revised manuscript.

A simple question: What is the merit of ~100/200 m improvement of PBLH (compared with the raw method) considering the future application of this dataset? Any impacts on climate-scale studies?

Response: The study is the extension of our recent work (i.e., Guo et al., 2021), in which the bias between radiosonde and reanalysis has been revealed, especially in the daytime. Then the question that needs to be addressed is how to build a more robust and accurate seamless PBLH dataset. We believe a more robust and accurate PBLH dataset could be valuable for the study of some fields, such as climate model and air pollution.

Technical corrections:

Line 99 and 116: the definition of ERA-5 should be moved ahead.

Response: Amended as suggested.

Please keep it consistent by using either ERA5 or ERA-5 in the whole manuscript.

Response: Done as suggested.

Line 233: in the main text, the authors mentioned that Table 2 shows the correlation coefficients between PBLH and each variable, but the caption of Table 2 says that it is a correlation coefficient with PBLH bias between radiosonde and ERA5 reanalysis, which is easy to be misinterpreted. Please address.

Response: As suggested, the descriptions about Table 2 in the main text has been modified to be:

“...We further perform correlation analyses between the aforementioned variables and PBLH biases between radiosonde and ERA5 reanalysis, and the statistical results are shown in Table 2...”.

Line 242, please use subscripts or other notations to mark PBLH-M and PBLH-E in the equation. Otherwise, it will be easy to be recognized as a minus.

Response: Amended as suggested.