Reply on Comment 1

Dear Reviewer 1,
Thank you for your valuable comments and suggestions regarding our manuscript titled "Bias Correction and Application of Labeled Smartphone Pressure Data for Evaluating the Best Track of Landfalling Tropical Cyclones". In response to your concerns, we have taken significant steps to ensure that the manuscript now meets the necessary requirements for publication of AMT. Specific attention was given to:

- Adding 3 sets of unlabeled bias correction on tropical cyclone MSLP to fully demonstrate the advantages of the labeled bias correction methods.
- Improving explanatory descriptions, especially those related to user privacy protection, and labeled and unlabeled bias correction methods.
- Improving figures, tables and words.

This paper presents interesting results related to atmospheric pressure measurements using crowd sourcing from smartphones in China. The authors have published similar work before, and hence this is not that new or innovative. The new part of the research is a methodology for bias correction of the smartphone pressure data using labeled (identified) smartphone users, unlike their previous work using un-labeled users. While the results do show an improvement, I have a few issues that need to be addressed before publications.

REPLY: Thanks for your positive consideration and valuable comment to our study. We have tried our best to reply each of your concern and revised this manuscript.

First, due to privacy issues (particularly in Europe where EGUsphere is published) it will be very difficult for researchers in the future to use labeled data for research. Hence, this issue may not be available in the future. For this reason, I think the results presented in the paper for the 3 tropical cyclones should also present (maybe in a Table) the MSLP for the labeled and unlabeled bias corrections, compared with the best track pressure measurements. How different are the two methods compared with the standard station data today?

REPLY: We have found similar considerations in the work of other researchers, and we agree that it is undoubtedly right to protect user privacy. We have added more description about user privacy in Section 2.1(2) of the revised manuscript: Line 76 (Line 133 in the track-changes file): *The data is provided by users who have signed a data sharing agreement, and each pressure record carries an encrypted user ID that helps to distinguish the source of the data*; Line 81 (Line 141 in the track-changes file): *All research data in this study have been legally verified to comply with all provisions of the Personal Information Protection Law of the People's Republic of China issued on August 20, 2021 (https://www.gov.cn/xinwen/2021-*

08/20/content_5632486.htm), which was confirmed by the legal department of Moji Weather company.

In our view, regardless of the country, we can always let the public choose whether or not to help with research. Besides, since we choose to build a machine learning model for every single labeled user, the bias correction process can also be carried out locally on the users' smartphones with increasing computing capability, as long as we distribute weather station data to the user, which is not difficult. In this way, the users can submit unlabeled pressure with high quality, and the needs of both user privacy protection and unbiased data are met. We keep optimistic about the future availability of smartphone pressure data.

Furthermore, we have adopted your suggestion to add experiments with unlabeled data as part of Section 4, and revised Fig. 9 (Fig. 10 in the revised manuscript) can show the main results. Supplementary explanations about unlabeled MSLP are provided in Line 228 (Line 431 in the track-changes file) of the revised Section 4: *The unlabeled MSLP clearly exhibits a significantly positive bias compared with both labeled and station MSLP, which is consistent with the previous conclusions.*

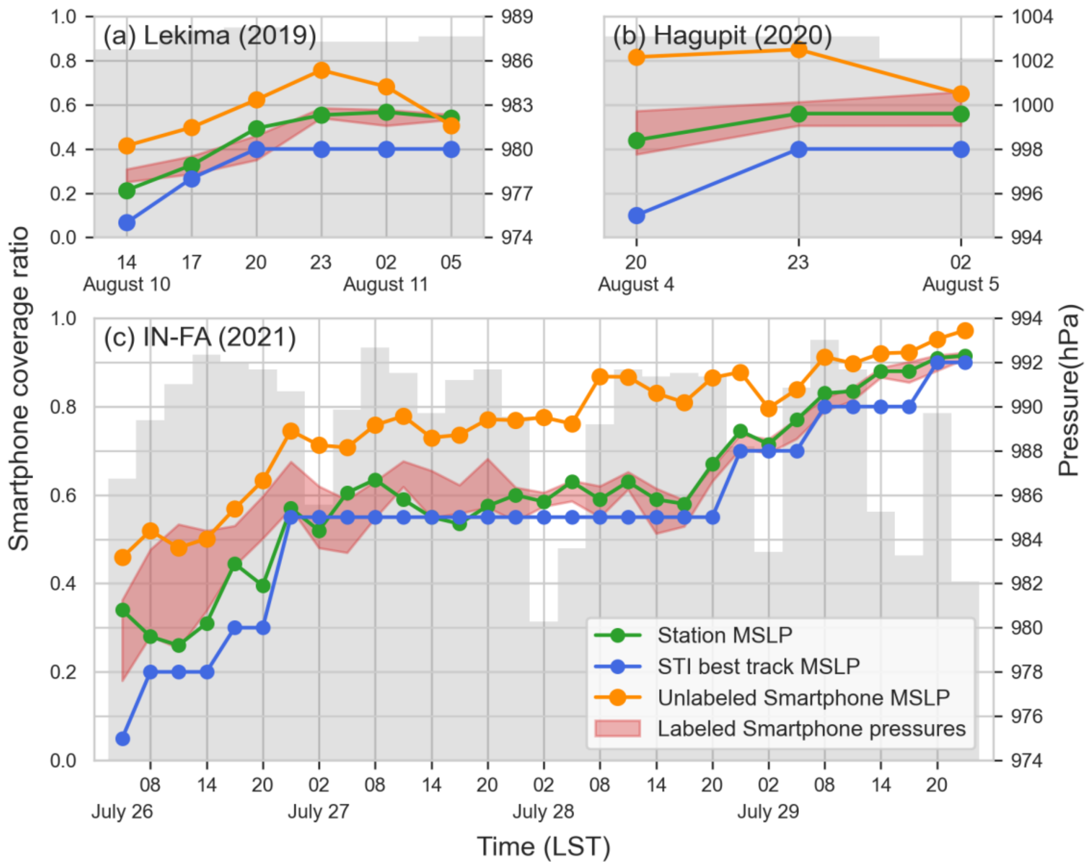

Figure 10 Variation of the MSLPand smartphone coverage ratio during (a) TC Lekima from 14:00 LST on August 10 to 05:00 LST on August 11, 2019, (b) TC Hagupit from 20:00 LST on August 4 to 02:00 on August 5, 2020, and (c) TC IN-FA from 05:00 LST on July 27 to 23:00

LST on July 29, 2021. Green, blue and orange dots represent the MSLP from weather stations, STI best track and unlabeled smartphones, with a temporal resolution of 3, 3 and 6 hours respectively. Red shaded areas represent the lowest 10% labeled smartphone pressure. Gray bars represent smartphone coverage ratio. All the statistics were done in the area of 1.2° x 1.2° surrounding the TC center.

Second, the machine learning model is not completely clear to this Reviewer (Table 1). Why does only the unlabeled data have land-use type, while not for the labeled data?

REPLY: The reference pressure is always certain, while the unlabeled smartphone pressure is not, as we calculate the mean and standard deviation of the smartphone pressure for each smartphone site, which means that data of similar horizontal positions are calculated together, regardless of the altitude. Supplementary explanations about the relationship between land-use type and unlabeled data are provided in Line 119 (Line 204 in the track-changes file) of the revised Section 2.1(4): *Since the data obtained from the same smartphone site in urban high-rise buildings can exhibit a significant degree of uncertainty, whereas the opposite holds true for rural areas, it's helpful to introduce land-use types into machine learning models for describing the acceptability of uncertainty for unlabeled data.* In contrast, the labeled smartphone pressure data is fed into the machine learning models without statistical transformation, so it is always certain and doesn't require an uncertain description.

Why does the labeled data need "week" if we know the date? I think "date" should be "day, and "moment" should be "time". Correct?

REPLY: Perhaps we should change the word " week" to "day of the week" to avoid confusion. Most people's lives are structured around a week-long cycle, such as working 5 days and then having 2 days off, so "day of the week" is helpful for judging their user behavior, even if we already have "date". Moreover, "date" marks the timeline (not the day of the month, so we think "date" is better than "day"), and "time" means the exact time of the day (We agree that "time" is better). We have revised Table 1 in the manuscript.

Table 1 Descriptive features of the two machine learning models

Unlabeled data	Labeled data
Longitude	Longitude
Latitude	Latitude
Month	Month
Date	Date
Moment	Time
Land-use type	Day of the Week
Gridded pressure	Smartphone pressure
Observations number	
Pressure standard deviation	

For those not familiar with ML, I would explain the parameters in Table 2.

REPLY: The parameters in Table 2 are all from the function "RandomForestRegressor" of the Scikit-learn machine learning library in Python (Pedregosa et al. 2011). We have added parameter descriptions to Table 2 in the revised manuscript.

Table 2 Hyperparameter settings of the two machine learning models

	Unlabeled data	Labeled data
max_depth	9999	9999
max_samples	0.7	0.7
min_samples_leaf	1	1
max_features	log(M+1)	M
n_estimators	100	30

All parameters are from the function "RandomForestRegressor" of the Scikit-learn machine learning library in Python (Pedregosa et al. 2011).
max_depth: The maximum depth of the tree (also known as "the base estimator").
max_samples: The proportion of samples to draw from the training set to train each tree when bootstraping.
min_samples_leaf: The minimum number of samples required to be at a leaf node.
max_features: The number of features to consider when looking for the best split. M represents the number of features used by the model.
n_estimators: The number of trees in the forest.

Minor points:
1. Line 1: ....have demonstrated significant potential to complement traditional surface pressure.....
REPLY: Text Revised.

2. Line 4: We propose a....
REPLY: Text Revised.

3. Line 11: observations
REPLY: Text Revised.

4. Line 15: data are crucial
REPLY: Text Revised.

5. Line 30: statistically
REPLY: Text Revised.

6. Line 54: Maybe you need to say more about user privacy issues of Moji Weather App.
REPLY: Relevant changes have been described in the previous paragraphs.

7. Line 99: intervals centered
REPLY: Text Revised.

8. Line 145: See comments above about Table 1 and 2
REPLY: Table Revised.

9. Line 185: retention
REPLY: Text Revised.

10. Line 211, Figure 9 caption: why is the resolution 1.2x1.2 and 0.6x0.6 different to the previous analysis resolution? I suggest being consistent, or explaining why you use these different resolutions.

REPLY: Thanks for your comment. Fig. 9 (now Fig. 10 in the revised manuscript) was revised. Now we use same resolution, which is 1.2x1.2.

11. Figure 10: Why is the number of stations (triangles) so different in region southwest of Lake Taihu in the two examples? Did the number of stations change so dramatically over 2 years?

REPLY: The city to the southwest of Taihu Lake where the number of stations changes dramatically is Huzhou in Zhejiang Province (119°14'E-120°29'E, 30°22'N-31°11'N), and we have counted the number of weather stations in this domain providing hourly pressure observations in revised Fig. 10 (now Fig. 11 in the revised manuscript). In fact, the number of weather stations increased steadily. Supplementary explanations about the changing station number are provided in Line 239 (Line 456 in the track-changes file) of the revised Section 4: *This instance* (2019-08-10 23:00LST) *happened to fall in a period when some stations with lower maintenance levels failed to measure and upload data steadily(Fig. 11c), which shows that smartphone pressure observations are valuable for filling some of the gaps created by unstable weather stations.*

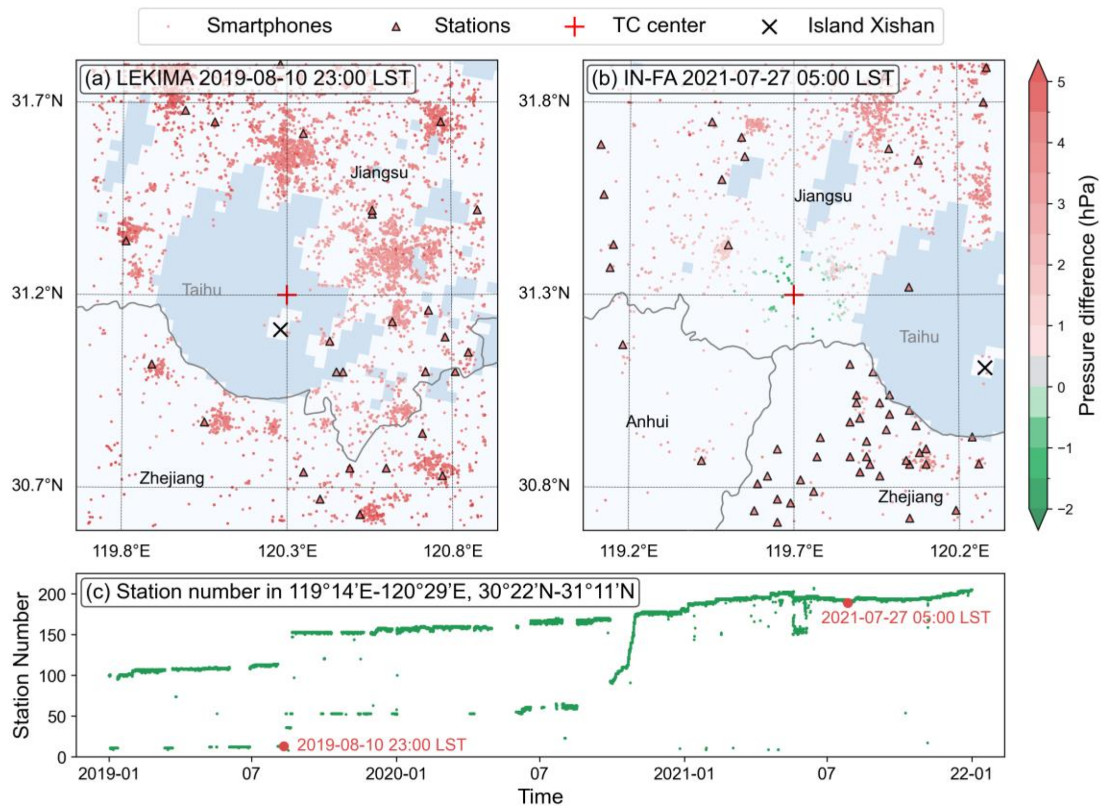

Figure 11 Distributions of weather station and smartphone observations from two examples during (a) TC IN-FA and (b) TC Lekima, in the area of 1.2º×1.2º surrounding the TC center. The coloring represents the difference between the pressure observations and the STI best track MSLP. (c) Changes in the number of weather stations providing pressure observations from 2019 to 2021, in 119°14'E-120°29'E, 30°22'N-31°11'N (the geographical scope of Huzhou, Zhejiang Province).

Reference

Pedregosa, F., et al. (2011). "Scikit-learn: Machine Learning in Python. " Journal of Machine Learning Research 12: 2825–2830.

Personal Information Protection Law of the People's Republic of China. (2021) the Xinhua News Agency, Standing Committee of the National People's Congress, Retrieved November 12, 2024, from https://www.gov.cn/xinwen/2021-08/20/content_5632486.htm

Reply on Comment 2

Dear Reviewer 2,
Thank you for your valuable comments and suggestions regarding our manuscript titled "Bias Correction and Application of Labeled Smartphone Pressure Data for Evaluating the Best Track of Landfalling Tropical Cyclones".

In response to your concerns, we have taken significant steps to ensure that the manuscript now meets the necessary requirements for publication of AMT.

- Improving explanatory descriptions, especially those related to labeled and unlabeled data and their bias correction methods.
- Adding a flowchart of the data correction steps to make the process structure clearer.

Additionally, we are especially grateful for your suggestions for the future research.

The ideas presented in this article possess a certain level of innovation and have the potential to make an impact in the industry. Smartphone barometer observations, as a vital supplementary data source for meteorological observations, have not been fully utilized due to the lack of abundant and reliable data. This article assesses the data quality of smartphone barometers and their potential in TC observation, which holds significant academic value and could even influence smartphone manufacturers.
There are some details in the article that are not clearly elaborated. I suggest the authors review and make adjustments and supplements where necessary: Throughout the article, the terms "labeled" and "unlabeled" smartphone pressure measurements are frequently mentioned (starting from line 69), but it does not specify what the label is. I believe the authors might be referring to measurements that have been corrected with exact pressure values. It is recommended to adjust the expression to avoid misunderstandings with other label content. Furthermore, around line 134, the article mentions that methods can be divided into two categories for labeled and unlabeled data, which could be misinterpreted. The authors might be trying to convey supervised and unsupervised methods.

REPLY：Thanks for your positive comment to our study. The labeled and unlabeled approach is after the study of McNicholas and Mass (2018) and Li et al. (2021). In view of the unclear description of labeled and unlabeled data, we have made adjustments in Line 76 (Line 133 in the track-changes file) of Section 2.1(2): *The data is provided by users who have signed a data sharing agreement, and each pressure record carries an encrypted user ID that helps to distinguish the source of the data. However, we clearly understand that user IDs are sometimes not available, so we also made a dataset with user IDs removed for comparative experiments. In the rest* *of this paper, we refer to data without user ID as “unlabeled data”, and correspondingly data with user ID as “labeled data”.*

Furthermore, both labeled and unlabeled data correction methods compared in this paper are supervised machine learning, as we have mentioned that the specific technique is random forest, and we treat the differences between the smartphone pressures and reference sea level pressures as the true values of the training set.

The description of the methodology section is relatively brief. I suggest illustrating the data processing procedure with an algorithm flowchart or diagram for clarity.

REPLY: We agree that the calculation steps involved in this article are complicated, so we have added a flowchart as Fig. 7. Descriptions for new Fig. 7 are provided in Line 198 (Line 361 in the track-changes file) of the revised Section 3.2: *The workflow diagram shown in Fig. 7 summarizes the process of quality control and BC from the raw smartphone pressure data to the final data we used in the study.*

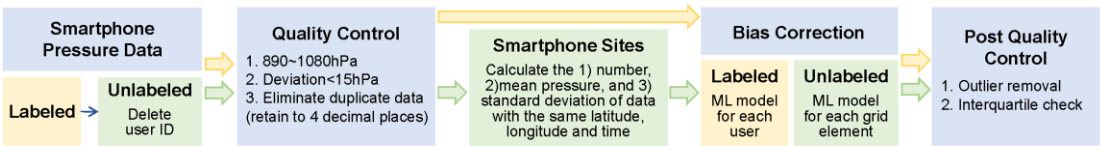

Figure 7 The work flow for smartphone pressure data quality control and bias correction.

Lastly, regarding the relationship between smartphone barometer data bias and the height of the building floors, further analysis could be conducted by screening the distribution of pressure data from different users at the same location and time. Generally, there is a linear relationship between floor height and air pressure. It is challenging to avoid situations where users are indoors on higher floors in urban settings, and this aspect could be further explored in future work to enhance the precision and usability of the data.

REPLY: Good suggestion! Though the current labeled correction method has some effect on high floor users, we have not evaluated the effect in a targeted way. Nowadays some smartphone GPS services are able to provide altitude data along with latitude and longitude, which could be very helpful to our work, but unfortunately the dataset used in our study did not include altitude information. Your concern about this issue coincides with ours, and we will continue to explore in future works.

Reference

Li, R. M., et al. (2021). "Smartphone pressure data: quality control and impact on atmospheric analysis." Atmospheric Measurement Techniques 14(2): 785-801.

McNicholas, C. and C. F. Mass (2018). "Smartphone Pressure Collection and Bias

Correction Using Machine Learning." Journal of Atmospheric and Oceanic Technology 35(3): 523-540.