Anonymous Referee #1

This paper presents a circum-Antarctic iceberg database using Sentinel-1 SAR images in the Google Earth Engine platform. Their image segmentation and random forest classifier seem to work successfully in capturing the spatiotemporal distributions of icebergs, including their number and sizes, across the Southern Ocean. However, the authors need to provide more details about their iceberg detection model. While the authors mentioned that they used an ensemble random forest classifier with four different RF classifiers, based on different input features, they did not provide any details about this ensemble result (i.e., weights to each classifier, importance of statistical features, histogram features, and texture features). I encourage the authors to provide the details of their ensemble process to support the robustness of their method.

Thank you for your comments and suggestions. In the original manuscript's Method, we did not describe the model's specific parameters in detail. In the revised manuscript, we have added more detailed information on the model integration methods (L158-L172). Below are our point-by-point responses to your comments:

L146-147: How are these three subsets divided? Randomly or by any other criteria?

The sample set is randomly divided into three subsets, and we have added the relevant explanation in the revised manuscript (L155).

L210: Maybe it would be better to use 40 m, instead of 0.04 km, as already used throughout the manuscript (L69 and L216).

Thank you for your suggestion. To maintain consistency with the surrounding text, it is better to use "40 m".

L241: "Based on this analysis, we selected an average thickness of 232 m for the icebergs" -> It is not clear how this value of 232 m is derived.

This study used 19,945 iceberg freeboard measurements from the Altiberg v3.2 dataset recorded by the CryoSat SARIn mode during 2018–2021 (the latest year available in this version), yielding an average freeboard of approximately 40 m, which was adopted as the representative freeboard for all icebergs in this study. Following previous research, we set the seawater density to 1025 kg/m^3 and the iceberg density to 850 kg/m^3. According to Archimedes' principle, the relationship between iceberg freeboard height h and total thickness H is:

$$H=\frac{\rho_w}{\rho_w-\rho_i}h$$

Substituting the above parameters into the equation, the average total thickness of icebergs is calculated to be 232 m.

In the revised version, instead of assigning a uniform fixed thickness of 232 m to all Antarctic icebergs, we assign thickness values according to iceberg area (L205-213), based on the Volume/Area scaling parameterization of Iceberg Classes Model in Stern et al. (2016), thereby making the thickness attribute of individual icebergs more physically meaningful.

L256-259: Then, does it mean that 2018 data was included in training for all iterations but not tested at all, and 2023 data was never used for training? If so, I don't think this is a fair training strategy because the model could be biased to 2018 data. Would it be better to conduct 6-fold cross-validation (or so-called Leave-One-Out cross-validation), for example, 2018 data as test data and the remaining years as training data for iteration 1, 2019 data as test data and the remaining years as training data for iteration 2, and so forth? The authors mentioned that they used this strategy to “adapt to the time-series nature of the data while minimizing the risks of overfitting” (L256), but I’m not sure how the current strategy can achieve this.

Thank you for the reviewer’s thorough comments. We fully agree that leave-one-year-out cross-validation ensures a fair assessment of the model’s performance on each year’s Sentinel-1 imagery. Accordingly, in the revised manuscript we have adopted a six-fold, leave-one-year-out cross-validation scheme, using approximately 400 manually annotated superpixel samples from each year as the test set and the remaining years’ data for training, thereby ensuring that each year both tests and contributes to the training. We have replaced the original each year evaluation and rolling window validation results in the main text (L262-265) and in Table 3. The new table presents the Accuracy, Precision, Recall, and F1 score of the ensemble incremental random forest classifier with optimal parameters for each year from 2018 to 2023, as well as their averages, to more comprehensively demonstrate the model’s accuracy and cross-year generalization ability in circumpolar Antarctic iceberg detection.

Tables 3 and 4: The authors conducted performance evaluations twice: (i) evaluation for each year (Table 3) and (ii) evaluation with rolling window validation (Table 4). I’m not sure that these two different evaluations are really necessary. To evaluate the model performance, I believe cross-validation in Table 4 is enough.

Thank you for the reviewer’s suggestion. In the original manuscript, we employed both annual evaluation and rolling-window validation: the former to demonstrate classification performance on each year’s Sentinel-1 imagery, and the latter to illustrate the model’s robustness as historical data accumulate. To streamline the manuscript, we have adopted the reviewer's recommendation, revised the evaluation strategy to leave-one-year-out cross-validation, and retained only these results in the main text.

L263-264: So, what model is finally used for building the iceberg database? The database is built each year separately based on the random forest model in Table 3, or does the entire database use a single model trained from the final iteration in Table 4?

We ultimately adopted the random forest models from Table 3, using distinct parameter settings for each year, and constructed the iceberg database separately for each year.

Section 4.1: The authors should have provided a detailed performance of their "ensemble" RF model. In L150-154, the authors mentioned that they used four RF classifiers and assigned weights to these classifiers, but the manuscript lacks details about this process. It is necessary to specify the performance of these four classifiers and how the authors select the weights between these models.

Thank you for the referee's valuable suggestion. Indeed, the original manuscript lacked a detailed description of the model's specific parameters, so we have supplemented this information in the revised manuscript (L158-172). Below, we take October 2018 as an example to illustrate how we determined the parameters for our ensemble random forest classifiers.

Based on the Sentinel-1 SAR imagery, we applied the SLIC algorithm to generate superpixels and then manually selected approximately 2,000 superpixel samples per year, with roughly half representing icebergs and the remainder non-icebergs. The sample set was then randomly divided into three subsets: an initial training set, a validation set, and a test set, in a 6:2:2 ratio. The training set was used to train the RF classifier, the validation set was used to evaluate the model's performance and optimize parameters, and the test set was used for final evaluation of the model's generalization ability and reliability.

Taking October 2018 as an example, we detailed how we determined the parameters for our ensemble of random forest classifiers and performed an incremental training procedure within each 5°×5° grid cell. We constructed four independent random forest models: RF1 trained on statistical features, RF2 on histogram features, RF3 on texture features, and RF4 on all combined features. By analyzing out-of-bag error (OOB) curves under various hyperparameter settings, we identified the configurations that converged stably with minimum OOB: 200 trees/3 features for RF1, 100 trees/5 features for RF2, 250 trees/7 features for RF3, and 150 trees/3 features for RF4 (Fig. S1). Each model was then evaluated on the validation set to compute accuracy, precision, recall, and F1 score (Table S1), these four metrics were normalized to generate candidate weight schemes reflecting different perspectives on sub-model importance (Table S2).

For the ensemble, we multiplied each model's iceberg probability by its corresponding weight and summed the results to obtain a combined discriminant score for each superpixel. We scanned decision thresholds from 0 to 1 in steps of 0.01 on the validation set, plotting precision–recall and ROC curves for each weight scheme. The scheme that maximized the sum of P–R Area Under the Curve (AUC) and ROC AUC was selected as optimal, yielding weights of 0.218, 0.271, 0.246 and 0.265 for RF1–RF4, respectively. Finally, we searched for the threshold that maximized the F1 score on the validation set and set 0.783 as the final decision threshold for iceberg detection.The same procedure was applied to the remaining years to obtain the optimal parameter configurations for each respective year.

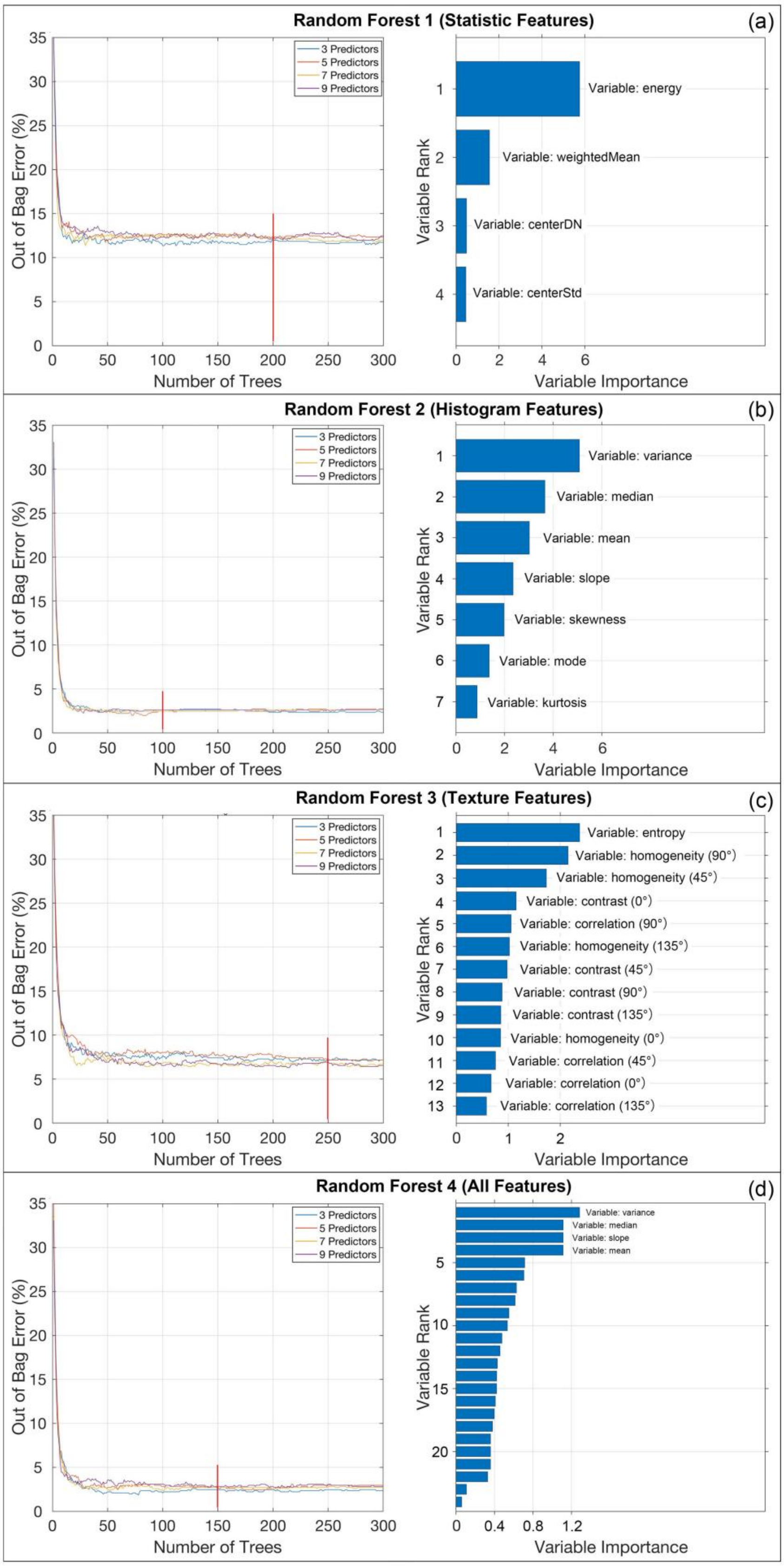

Figure S1. Out-of-bag error and parameter importance of random forest classifiers based on different feature sets.

Table S1. Performance metrics of random forest classifiers based on different feature sets

classifier	ACC	Precision	Recall	F1
RF1	0.9207	0.9505	0.8872	0.9178
RF2	0.9872	0.9847	0.9897	0.9872
RF3	0.9488	0.9534	0.9436	0.9485
RF4	0.9872	0.9948	0.9795	0.9871

Table S2 Normalized weights of random forest classifiers derived from different evaluation metrics

weight	RF1	RF2	RF3	RF4
ACC	0.2293	0.2636	0.2435	0.2636
Precision	0.2396	0.2571	0.2410	0.2623
Recall	0.2176	0.2709	0.2462	0.2653
F1	0.2282	0.2641	0.2437	0.2640

L300: "several tens of kilometers: This is too ambiguous. Please provide specific numbers.

We thank the referee for their careful correction. The specific values here are 44.08 km and 32.28 km, and we have amended this in the revised manuscript (L304).

L301-303: I would like to ask the authors to provide more details about why the BYU/NIC database cannot capture so many > 5 km icebergs. Does it intentionally skip relatively small icebergs (near 5 km size), or does its iceberg detection algorithm, by itself, have limitations in capturing near-5-km icebergs? What about much larger icebergs, for example, > 10 km?

Taking 2021 as an example, we downloaded iceberg trajectory data (Statistical Database [v7.1]) from the official Brigham Young University website (https://www.scp.byu.edu/data/iceberg/default.html), comprising 192 records representing the observed iceberg trajectories during the same period. we extracted the entries in the csv file whose “date” field corresponds to October 2021, obtaining 53 records for comparison with our study’s iceberg spatial distribution data. to ensure comparability between the two datasets, we also selected from our database all icebergs with a major axis exceeding 5 km (292 in total, of which 88 exceed 10 km in major axis). during the comparison, we followed the iceberg position and shape reports published by the u.s. national ice center and applied a one-to-one matching approach to rigorously verify each iceberg’s spatial location and shape characteristics.

The results show that all 50 icebergs recorded by BYU/NIC for October were matched in our database (black boxes in figure 2), while three icebergs (C36, B46, and UK324) were not detected (red boxes). further analysis indicates that C36 and B46 were located in the sentinel-1 SAR EW scan-mode blind zone, which remained uncovered even after image mosaicking. for UK324, no iceberg with a major axis exceeding 5 km was found in the corresponding mosaic or original single sentinel-1 images, suggesting potential positioning or identification errors in the BYU/NIC record. to improve the completeness and accuracy of our dataset, we supplemented the EW blind zone with sentinel-1 IW-mode data (Fig. S2 C63).

Fig. S3 presents the spatial distribution of icebergs with a major axis larger than 5 km detected in our study but not recorded in the BYU/NIC database, overlaid on the sentinel-1 mosaic image used in our analysis. it is evident that no duplicate counts or incorrectly merged icebergs occurred. figure 4 further illustrates the distribution characteristics of these icebergs in terms of area and major axis length: the number of icebergs decreases markedly as area and major axis increase, with most icebergs having an area between 0 and 20 km² and a major axis within the 5–9 km range. spatially, these icebergs undetected by BYU/NIC are mainly located in front of ice shelves and are typically accompanied by sea ice cover.

According to Budge and Long (2018), the BYU/NIC database has several limitations that can result in the omission of even large icebergs with major axes exceeding 5 km. First, the database primarily relies on passive microwave and scatterometer data for tracking. These sensors have relatively low spatial resolutions, typically on the order of several to tens of kilometers, so in areas with dense sea ice cover or high iceberg concentrations, the signal from an individual iceberg can easily blend with surrounding targets, leading to missed or false detections. Second, both the automatic and manual identification processes in the BYU/NIC database can be affected by cloud cover, wind waves, and other anomalous electromagnetic scattering conditions. In particular, in the complex environments in front of ice shelves or along coastlines, the signals from large icebergs may be obscured by sea ice, making them difficult to distinguish. In addition, due to temporal gaps in observational coverage, to maintain consistent measurement intervals in the consolidated database, researchers perform piecewise cubic interpolation of iceberg positions between consecutive observations, while no interpolation is conducted for observation gaps longer than two weeks. Although this approach can partially fill short-term data gaps, it may lead to inaccurate position estimates or even omissions from the records for large icebergs that drift rapidly or disintegrate within a short period of time.

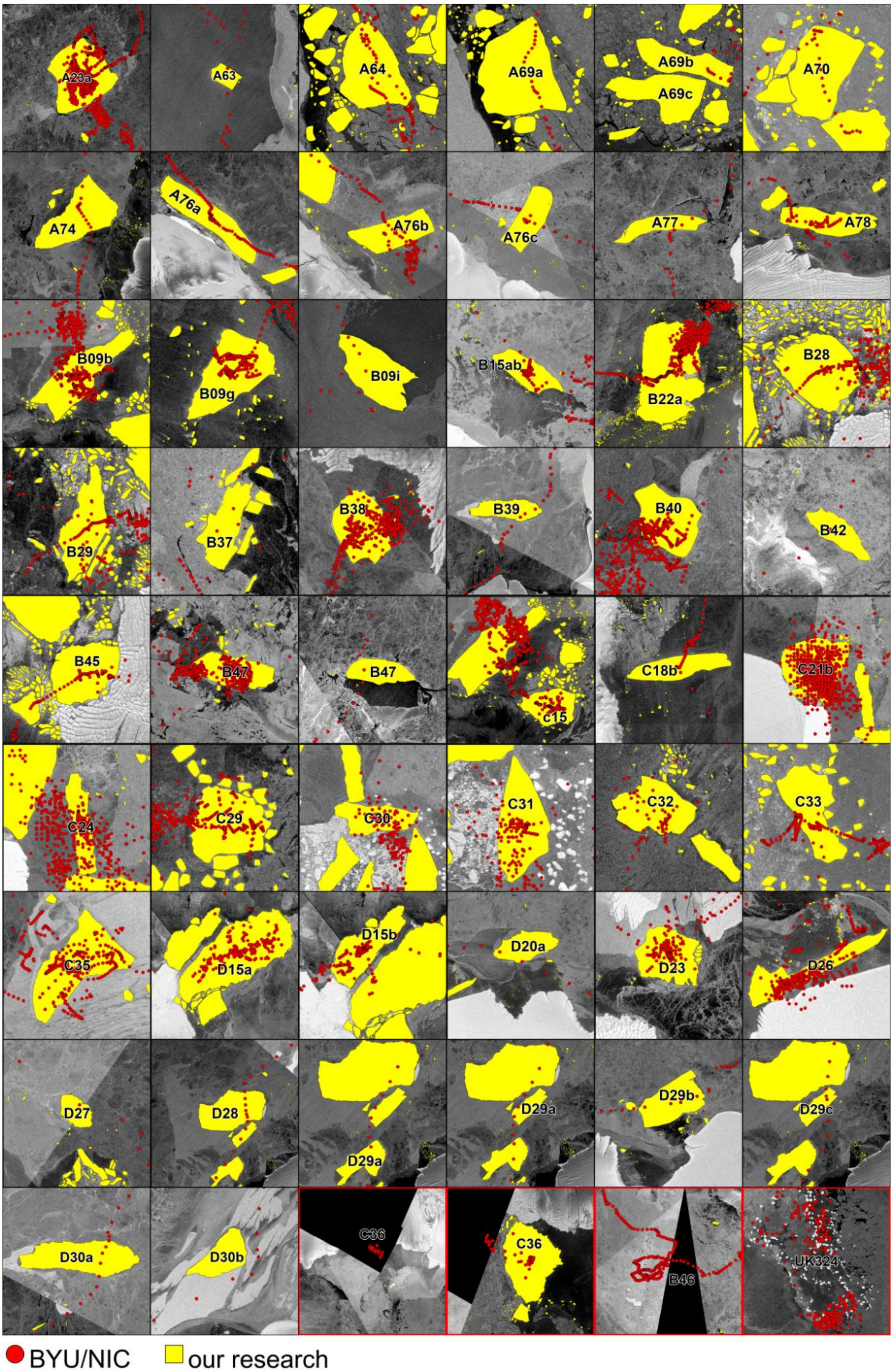

Figure S2. spatial matching results of icebergs between the BYU/NIC database (red dots) and our dataset (yellow polygons): black boxes denote successfully matched icebergs; red boxes denote unmatched icebergs.

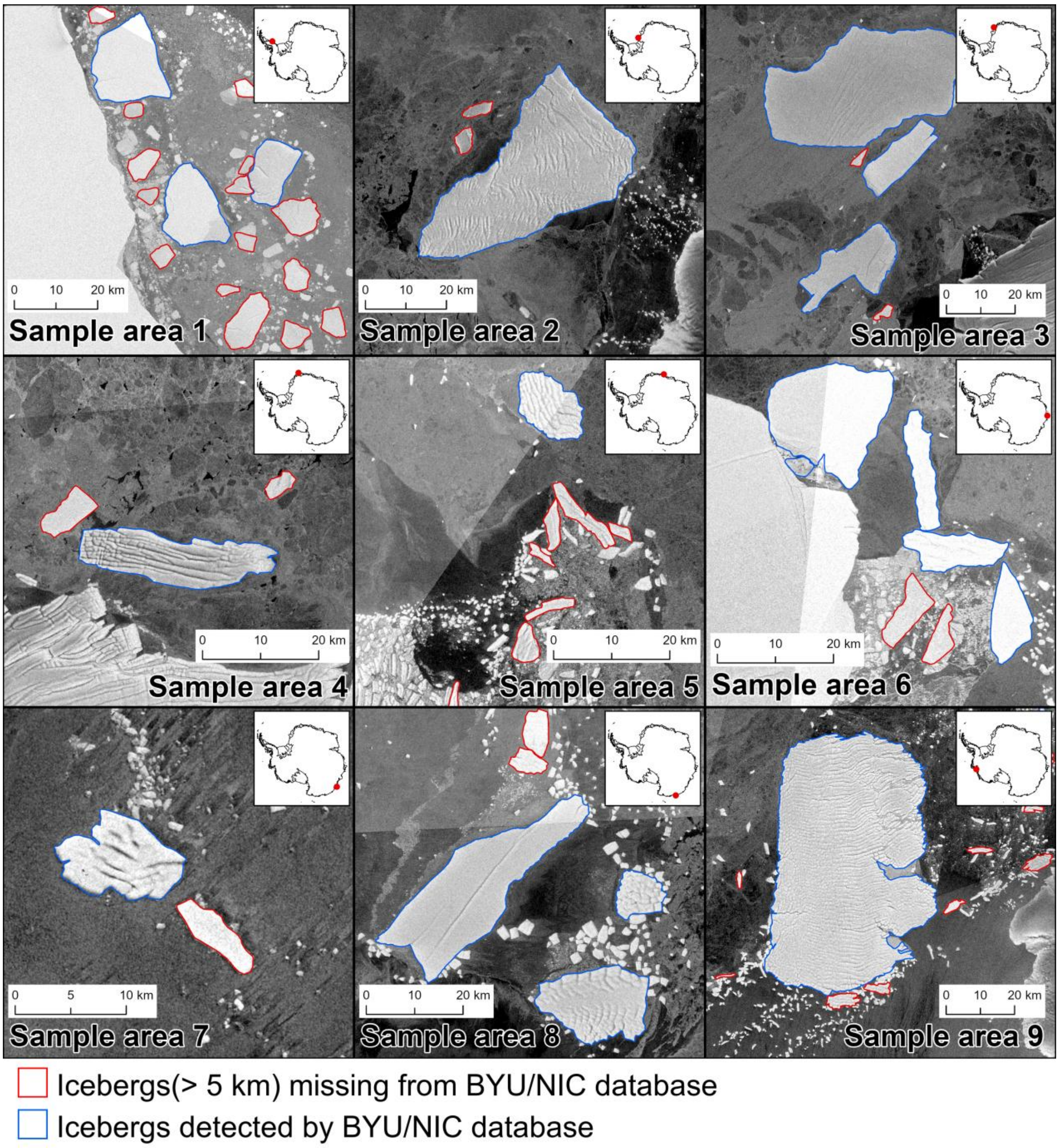

Figure S3. Examples of icebergs (>5 km) detected (blue) and missed (red) by the BYU/NIC dataset, with Sentinel-1 mosaics as background.

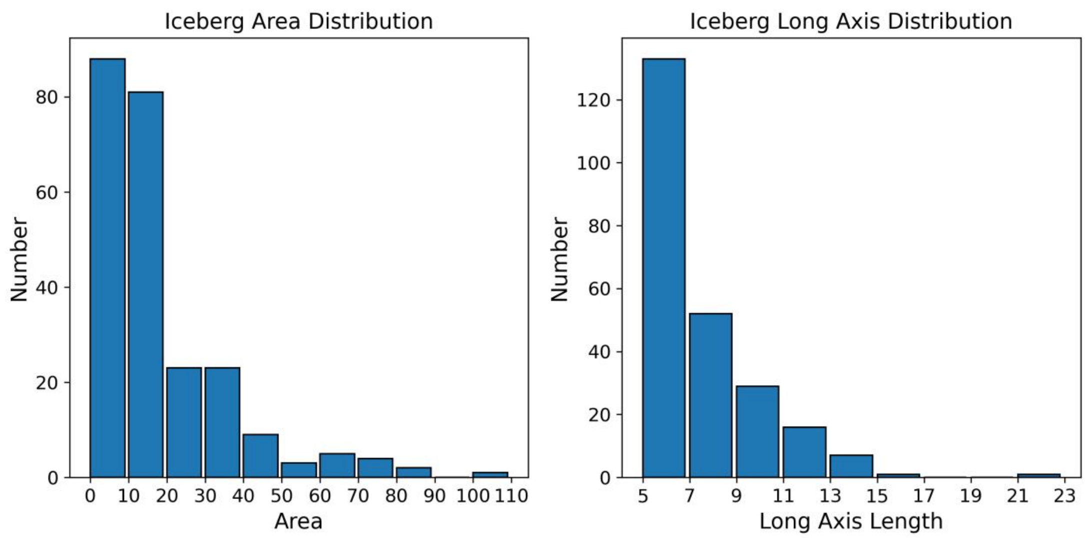

Figure S4. Histogram distribution of the area and major axis length of icebergs (>5 km) missed by the BYU/NIC dataset.

L339-349: I wonder if the total number of icebergs here and in Table 5 is the "true" number of icebergs. That is, if an iceberg is detected in two different Sentinel-1 scenes, how is this iceberg counted? This iceberg could be counted in duplicate, as the methods proposed in this study can only "detect" icebergs but cannot "track" identical icebergs. This could not be so significant because the authors used mosaiced data, but there is a possibility that the same icebergs are detected in duplicate (or some icebergs are missed) due to their drift even over a short period. It would be worthwhile to mention this issue and include any relevant discussion about it.

We greatly appreciate the referee's valuable suggestions! In this study, we believe that the total iceberg counts listed in lines 339–349 of the original manuscript and in Table 5 already reflect the real situation as accurately as possible. First, during the image-acquisition stage we sorted all Sentinel-1 HH-band images within each tile in ascending order of acquisition time and then mosaicked them sequentially using the mosaic()

function in Google Earth Engine. Later-acquired images overwrite valid pixels in earlier images, filling voids at the beginning of the month and producing a synthetic layer that is both spatially continuous and representative of the month's most recent observations. Because of this time-ordered mosaicking approach, the intervals between dates of the images composited into any single tile are generally small.

Taking 2021 as an example, we analyzed all 360 Antarctic tiles (280 after excluding no-data tiles) in terms of the number of distinct acquisition dates and the span between the earliest and latest dates (Figure S5). The results show that most tiles contain 2–4 images from different dates: 53.21 % of tiles have a maximum date span of ≤ 5 days, and 91.07 % have a maximum span of ≤ 10 days. In iceberg-dense regions (65–80°S), 56.47 % of tiles span ≤ 5 days and 92.35 % span ≤ 10 days; in less dense regions (55–65°S), 89.09 % of tiles span ≤ 10 days. We have added the relevant explanation in the revised manuscript in L99-101. Referring to Koo et al. (2023), who reported that most icebergs in the Amundsen Sea sampling area drift at < 0.2 km/day (Figure S6), the short inter-image intervals and limited drift speed yield mosaics with good boundary and texture continuity. At this rate, the cumulative 10-day displacement is <2 km—negligible relative to the tile dimensions—so bergs are unlikely to exit a tile, and repeated detection of the same iceberg is unlikely.

We also considered the rare cases of fast-moving small icebergs being detected on adjacent dates, and we explain in the revised manuscript (L198–199) that these were removed by manual correction. For the two typical repeat cases shown in Figure S7, because icebergs tend to drift together in a relatively stable spatial arrangement under the combined influence of wind and currents, we retained only the set of icebergs with the most complete outlines (e.g., those in the red box of Sample Area 1 and the yellow box of Sample Area 2). As for the very few small icebergs counted twice due to rapid drift, their impact on the total number and area estimates for all Antarctic icebergs is negligible and can be ignored. In this way, the final iceberg count should truthfully and reliably reflect the actual distribution of icebergs in the study area.

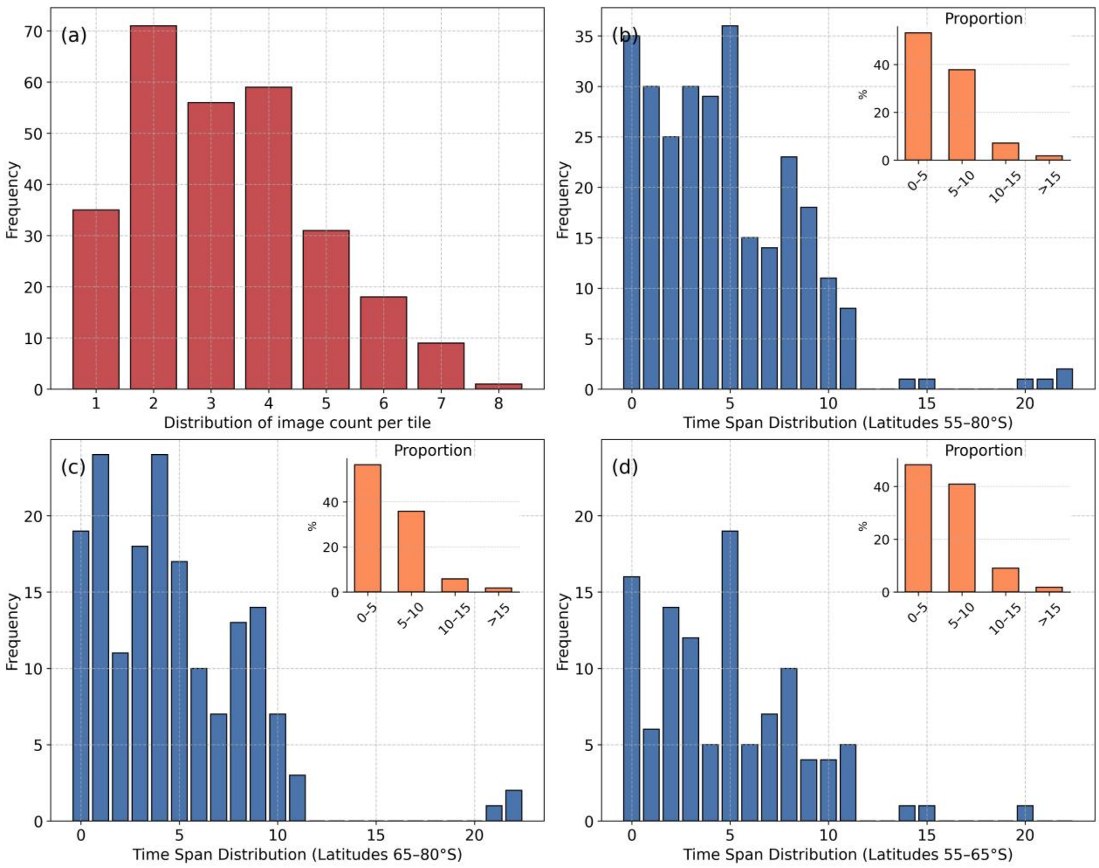

Figure S5. Panel (a) distribution of the number of Sentinel-1 images per tile. Panels (b–d) histograms of the time span between acquisition dates for tiles in different latitude bands (55°S–80°S, 55°S–65°S and 65°S–80°S).

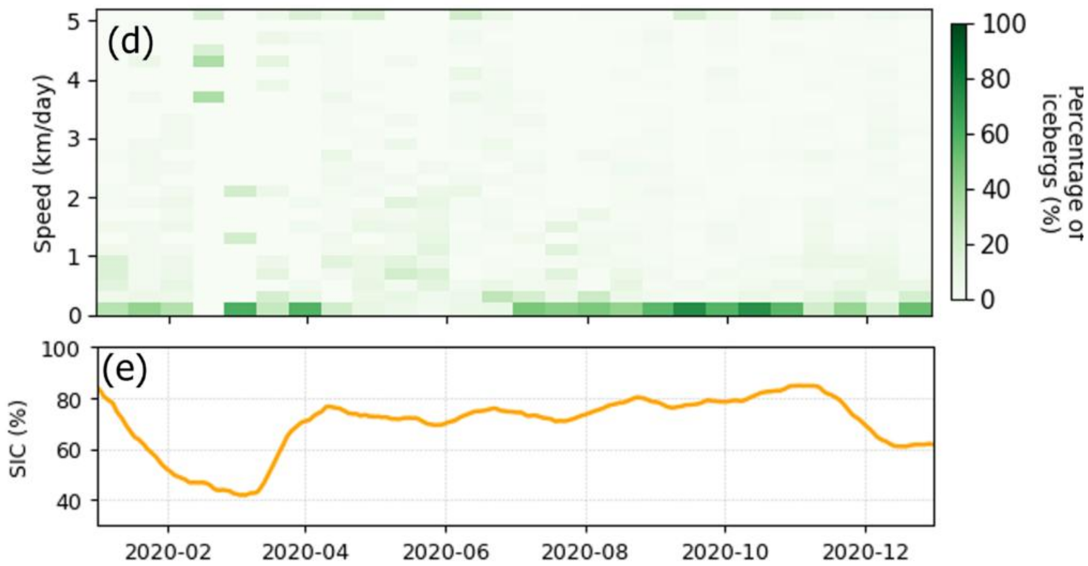

Figure S6. Drift speed distribution (biweekly average): most icebergs drift at speeds below 0.2 km/day, indicating relatively slow short-term movement (Koo et al., 2023).

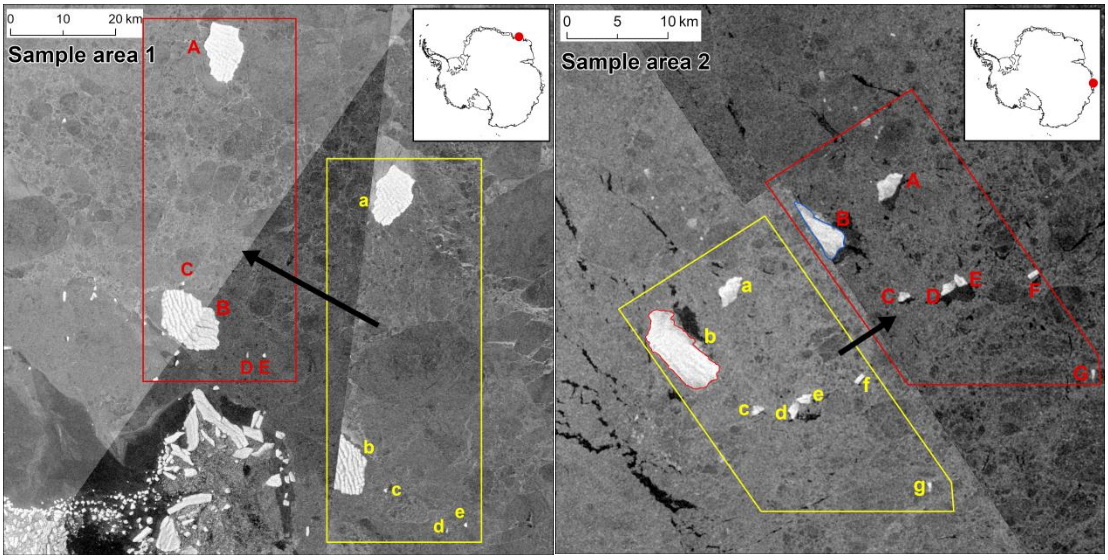

Figure S7. Examples of fast-moving icebergs appearing twice in the mosaic imagery of the same tile.

L347: We -> we

We thank the referee for their careful correction and we have amended this in the revised manuscript.

L355-356: “in the West Antarctic region and in the East Antarctic region” -> It would be better to only specify Thwaites and Doston ice shelves and Holmes and Mertz ice shelves, without mentioning too ambiguous “West and East Antarctic regions”.

We thank the referee for their valuable suggestions. In the revised manuscript, we have replaced “West Antarctic region” and “East Antarctic region” with the specific ice shelves Thwaites, Dotson, Holmes, and Mertz (L359).

L379-382: “In the Ross Sea sector, the iceberg proportion remained stable at around 16 % in 2018 and 2019, … remained relatively stable at approximately 20% over the six-year period.” In those sentences, the “iceberg proportion” may indicate “the number of icebergs in each sector / the number of total icebergs in the Southern Ocean.” However, I feel like this term “iceberg proportion” can be confused with “how much area (in percentage) is covered by icebergs (i.e., iceberg area / total ocean area of each sector).” Please consider rephrasing these sentences to clarify the meaning of the iceberg proportion. It could be good to discuss just the numbers (in Figure 11a), rather than the proportions (in Figure 11b).

We thank the referee for their correction. The term “iceberg proportion” in the text refers to the share of each region’s iceberg count relative to the total number of icebergs in the Southern Ocean. To avoid ambiguity, we have clarified this definition (L383) and revised the wording accordingly in the revised manuscript.

L387: This is similar to the previous comment; please clarify the meaning of “total area.” I believe this means the total area of icebergs.

We thank the referee for their correction. The term “total area” here refers to the cumulative iceberg area, and we have clarified this definition in the revised manuscript (L391).

L394-401: I'm not sure that this part really "validates" the small iceberg formation mechanism. The authors just present the distance from large icebergs, and it does not provide any direct clues for the small iceberg formation mechanisms. I don't think this part is necessary.

We thank the referee for their valuable suggestions. We acknowledge that our study does not directly "validate" the formation mechanisms of small icebergs; therefore, in the revised manuscript we have modified the statement to: "In analyzing the distances between small and large icebergs, we further arrived at conclusions consistent with the formation mechanisms of small icebergs proposed by Tournadre et al. (2016)." The spatial distribution pattern of distances between small and large icebergs obtained in this study closely matches the findings of Tournadre et al. (2016), and we have additionally included the distribution near the Antarctic coastline to provide more comprehensive support for the formation mechanisms of small icebergs.

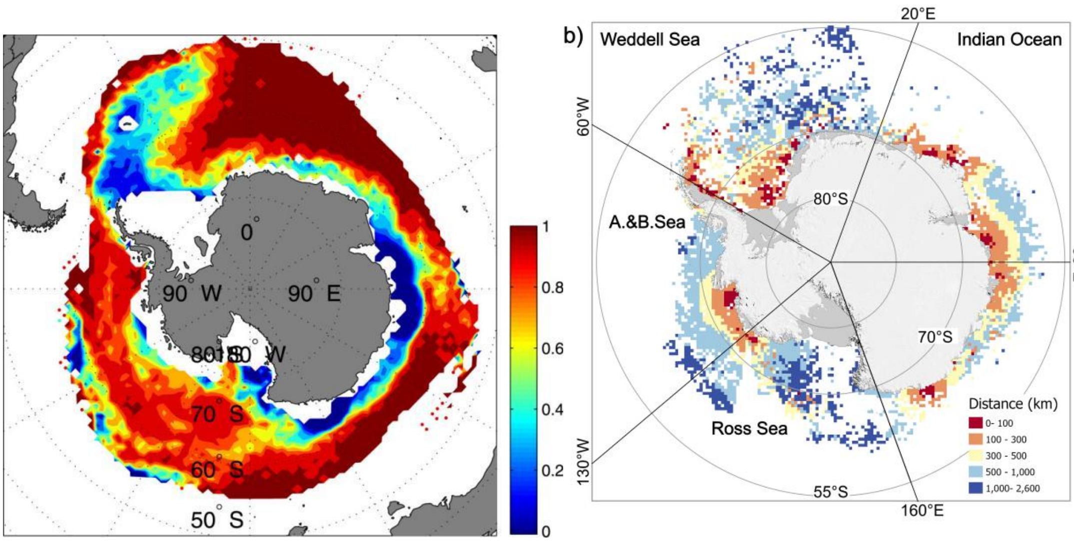

Figure S8. Average distance from icebergs in each grid to the nearest large iceberg. left panel: results from Tournadre et al. (2016); right panel:

results from our research.

References

Stern, A. A., Adcroft, A., and Sergienko, O.: The effects of Antarctic iceberg calving - size distribution in a global climate model, JGR Oceans, 121, 5773 – 5788, https://doi.org/10.1002/2016JC011835, 2016.

Koo, Y., Xie, H., Mahmoud, H., Iqrah, J. M., and Ackley, S. F.: Automated detection and tracking of medium-large icebergs from Sentinel-1 imagery using Google Earth Engine, Remote Sensing of Environment, 296, 113731, https://doi.org/10.1016/j.rse.2023.113731, 2023.

Tournadre, J., Bouhier, N., Girard - Ardhuin, F., and Rémy, F.: Antarctic icebergs distributions 1992 – 2014, JGR Oceans, 121, 327 – 349, https://doi.org/10.1002/2015JC011178, 2016.