Responses to Comments

Dear Editor and Reviewers,

We sincerely thank you for your constructive comments and valuable suggestions on our manuscript entitled "*A multi-decadal dataset of surface damage on Antarctic ice shelves (1999–2024)*" (Manuscript ID: *essd-2026-414*). We appreciate the time and effort you devoted to helping us improve the quality of our paper.

As detailed in our point-by-point response below, we have taken all of the comments and suggestions into account in the revised version of the manuscript. All revisions have been tracked. We sincerely hope that this manuscript will be acceptable for publication in *Earth System Science Data*.

In the following responses, the comments are shown in "**bold**", our responses are shown in "non-bold", and the revised text in the manuscript is shown in "*italic*".

Sincerely,

Dr. Gang Qiao

College of Surveying and Geo-Informatics

Tongji University

1239 Siping Road, Shanghai, China, 200092

Editor:

Public justification (visible to the public if the article is accepted and published):

Dear Authors,

Thank you for your careful revisions. The second-round reviews are favourable overall, although a few points would benefit from further clarification.

Please address the comments from both reviewers, particularly the definition of mapped damage, the limitations associated with not distinguishing damage types, and the implications for modelling applications. Please also clarify the geographical coverage, sources of uncertainty and key metrics, and improve figure readability and the explanation of terminology.

Please submit a revised manuscript together with a point-by-point response for editorial assessment.

Response: We sincerely thank you for your constructive guidance. We have carefully addressed the comments from both reviewers and revised the manuscript accordingly. In particular, we have clarified the scope and limitations of the mapped surface damage and its potential applications in numerical modelling. We have also addressed the comments regarding geographical coverage, uncertainty estimates, statistical metrics, figure readability and terminology. Detailed point-by-point responses to reviewers are provided below.

Reviewer 3:

This manuscript presents a detailed dataset of surface damage on multiple ice shelves in Antarctica identified over the last two and a half decades through remotely sensed data and a new deep-learning algorithm trained on several data sources. The authors have responded to two individual peer-reviews in what appears to be a thorough and constructive manner. I identify one issue consistent between both reviewers and make a recommendation that the authors should consider before the manuscript is accepted.

Response:

We sincerely thank you for your constructive comments and suggestions. We have carefully considered all the comments and revised the manuscript accordingly. Our point-by-point responses are provided below.

1. Differentiation of damage types

Both reviewers identify that ice surface damage manifests from several processes, not just a single unifying source. These processes are important for ice sheet modelers to capture and differentiate, so I question whether choosing to not differentiate surface damage types in this product will ultimately be useful to the community. As a non-expert, I do not know the answer nor do I know how much extra work it would require, but both reviewers appear to highlight this as an important issue and I would want to know if it is acceptable to keep that differentiation outside the scope of this manuscript or not. Will modelers actually use this product if they do not know what source these damage features come from?

Response: We agree that distinguishing different damage types and their formation mechanisms would be valuable for understanding ice-shelf weakening and for numerical modelling. However, the available optical observations are insufficient to reliably distinguish these types and their underlying causes across the entire dataset. We therefore adopted a single damage class to prioritize the reliability and temporal consistency of the dataset, rather than introduce damage-type classifications that could not be reliably validated.

Although the dataset does not distinguish damage types, it can still provide useful information for modelling studies. Some continuum damage models represent the combined effects of fractures through a scalar damage variable, without explicitly distinguishing individual fracture types or their origins (e.g., Borstad et al., 2012). Our observations may support modelling in the following ways:

1. Comparison with inferred ice weakening. Ice-flow inversions can identify regions with anomalously high ice-softness factors, which may indicate fracture-induced weakening. Our maps could be used to examine whether these regions correspond to visible surface damage. For example, Sun and Gudmundsson (2023) compared inferred ice-softening patterns with fractures observed in Landsat imagery.

2. Evaluation of damage evolution. The annual maps from 1999 to 2024 could help evaluate whether numerical damage models reproduce observed changes in the spatial distribution and extent of surface damage. Previous studies have developed models to simulate crevasse growth and transport (Sun et al., 2017) and damage evolution over ice-flow timescales (Ranganathan et al., 2025).

Nevertheless, the distinction between observable surface damage and the damage variables used in numerical models should be considered when applying the dataset to modelling studies.

We have revised the Discussion accordingly:

The high-resolution time-series surface damage maps produced in our study document the spatial distribution and temporal evolution of observable surface damage and may help interpret patterns of ice weakening inferred from ice-flow inversions and evaluate simulated damage evolution.

…

The single-class approach was adopted to prioritize the reliability and temporal consistency of damage mapping, as further classification could not be consistently validated using the available observations.

…

For numerical modelling applications, the distinction between observable surface damage and the damage variables used in numerical models should also be noted.

References:

Borstad, C. P., Khazendar, A., Larour, E., Morlighem, M., Rignot, E., Schodlok, M. P., and Seroussi, H.: A damage mechanics assessment of the Larsen B ice shelf prior to collapse: Toward a physically-based calving law, Geophysical Research Letters, 39, L18502, https://doi.org/10.1029/2012GL053317, 2012.

Ranganathan, M., Robel, A. A., Huth, A., and Duddu, R.: Glacier damage evolution over ice flow timescales, The Cryosphere, 19, 1599–1619, https://doi.org/10.5194/tc-19-1599-2025, 2025.

Sun, S., Cornford, S. L., Moore, J. C., Gladstone, R., and Zhao, L.: Ice shelf fracture parameterization in an ice sheet model, The Cryosphere, 11, 2543–2554, https://doi.org/10.5194/tc-11-2543-2017, 2017.

Sun, S. and Gudmundsson, G. H.: The speedup of Pine Island Ice Shelf between 2017 and 2020: revaluating the importance of ice damage, Journal of Glaciology, 69, 1983–1991, https://doi.org/10.1017/jog.2023.76, 2023.

2. Data hosting

I urge the authors to consider hosting their data in a more efficient and useable online repository. Hosting static geotifs on Zenodo is fine, but when it comes time to update products, produce more results, etc., the authors will likely need to create a new data release, new Zenodo citation, etc. and community users will need to re-download the new version of these datasets. This problem is not unique to this manuscript, but I feel strongly that the Cryosphere community (and beyond) will benefit when we transition to more useful forms of data delivery. To this end, I encourage the authors to consider hosting their data as either Web Feature Services (if damage feature data can be translated to shapefile/geojson format) or Web Map Services (if the data must remain in geotif format) on an online open-source platform such as Geoserver.

This has several benefits, namely that (i) users will have a live link to the most up-to-date version of these data at all times, (ii) if the authors need to make changes and/or add data, they don't need to make a whole new data release, they simply update the WFS or WMS to reflect those changes that are then propegated to users, and (iii) these products are operable in Jupyter Notebooks and other Python IDEs through common libraries such as GeoPandas as well as Desktop GIS applications like QGIS.

Response: Thanks for pointing this out. We agree that web-based geospatial services could improve data accessibility and facilitate future updates. For the present dataset, we have chosen to retain the GeoTIFF products hosted on Zenodo, which provides open access, persistent DOIs and version management, and is consistent with the ESSD data policy. In addition, the DOI cited in our manuscript represents all versions of the dataset and always resolves to the latest version, while individual versions remain accessible through their respective DOIs. We appreciate this recommendation and will consider web-based services for future improvements to data accessibility and delivery.

Reviewer 4:

General comments:

This is a "tour de force" manuscript that presents the methodology and analysis that underpins a new data set on ice-shelf damage for several (but not all) Antarctic Ice Shelves. The manuscript is well written and well proofed.

The main focus is on methodology. While I am not an expert in any way on the methodology used to create the data, I can sense that it is authentic and at the top level of performance.

I imagine that a companion manuscript will be produced in the future that will present a more in depth interpretation of the data and relate the data to various elements of ice-shelf behavior that are critical to future evolution. Some of this was done in the manuscript, but it was preceded by a very long, and highly detailed description of how the data were produced.

I expect that this manuscript will be of great interest to the general glaciological community and that it will inspire further work along a number of lines.

Response: We sincerely thank you for your positive assessment of our work and the constructive comments and suggestions. We have carefully considered all the comments and revised the manuscript accordingly. Our point-by-point responses are provided below.

Specific comments:

1. near line 44 - It is stated: "Damage development is linked to enhanced shear, ice flow acceleration, increased strain rates and ice shelf thinning, with these interactions being climate-sensitive and becoming more pronounced in a warming climate (Izeboud et al., 2025)."

There is evidence that damage can develop from stress that is induced by ocean swell that propagates as flexural gravity waves (there are a number of Australian authors who have published this hypothesis, I don't have a specific example at hand), and also from tidal flexure. Presumably, there may be atmospheric processes that may also generate damage (I suppose a sudden melting regime that starts to move meltwater around on an ice shelf inducing flexure stresses). I have always suspected that glaciologists tend to discount the ocean and atmosphere as having a hand in driving stresses that may cause rifting, etc.

It may be worth adding some kind of attribution to the potential for damage to develop in manners that involve more than just "gravitational glaciological stress"… even if this turns out not to be true, I think at this stage it would be a mistake to completely discount them.

Response: We agree that oceanic and atmospheric processes can also contribute to ice-shelf damage development. We have revised the Introduction to acknowledge these additional influences:

In addition, oceanic and atmospheric forcing can also contribute to damage development (Massom et al., 2018; Francis et al., 2021; Marsh et al., 2025).

2. Section 2.1.1 - Just out of curiosity, why weren't the Ronne, Ross, Filchner and Larsen ice shelves included in the study? Also, it would have been interesting (and perhaps this is something to do in a future study, not in this study) to examine ice shelves that have very little damage, e.g., George VI Ice Shelf, due to, presumably, a flow-regime difference (e.g., the flow being highly compressive and constrained, as it is with the George VI Ice Shelf).

Response: Both diverse surface damage characteristics and the availability of suitable Landsat imagery were considered in the selection of ice shelves studied in this study. We aimed to produce annual damage maps covering the entire extent of each selected ice shelf. For very large ice shelves, such as Ross and Filchner-Ronne, obtaining sufficiently cloud-free imagery with relatively close acquisition dates to achieve complete coverage presents considerable challenges. Therefore, we focused on nine representative ice shelves that encompass contrasting glaciological settings and for which high-quality damage maps could be produced over the study period.

We also examined George VI Ice Shelf during the selection process. However, we found it difficult to obtain suitable Landsat imagery with sufficiently low cloud cover and relatively close acquisition dates to achieve complete coverage of the ice shelf. Therefore, George VI was not included in the present dataset. We also note that Scar Inlet, the remaining portion of Larsen B Ice Shelf, is included in our study.

We agree that extending the dataset to other ice shelves, including those with different damage characteristics and flow regimes, would be valuable for future research.

3. line 265 - Why does ice-shelf dynamics cause uncertainty?

Response: We apologize for the unclear expression. When merging damage predictions from scenes acquired on different dates, ice flow may cause surface damage features to shift in position between acquisitions, potentially introducing spatial inconsistencies in the merged map. We therefore prioritize scenes acquired close in time to minimize these inconsistencies. We have clarified the sentence as follows:

Because ice flow can displace surface damage features between acquisition dates, predictions from scenes acquired close in time are prioritized during merging to minimize spatial inconsistencies.

4. line 315 - avoid 1-sentence paragraphs….

Response: We have merged this sentence with the preceding paragraph.

5. General question: Is all "damage" area on the ice shelves, e.g., as shown in red color in figure 7, filled with ice-melange or sea ice? Are there examples of damage where the damaged area displays ice-free seawater in the rifts?

Response: The red areas in Fig. 7 represent surface damage features identified from Landsat imagery, including crevasses, rifts and densely fractured areas, rather than the materials filling individual fractures. Most of these features are fractures rather than full-thickness rifts. Wider rifts generally

appear to be filled with sea ice or ice melange in the imagery we examined. Open water can be distinguished from ice in Landsat imagery and is not classified as damage by our segmentation model.

6. Figure 9 - It is hard to see the "difference" in panel d. I suggest eliminating the black and white background and *just* showing the colored areas (reddish and bluish), or some other sort of display determined by how well it is easy to see the results (which are small features).

Response: We agree with this suggestion. We have revised the difference panels in Figs. 8–10 to display only the false-positive (red) and false-negative (blue) pixels against a uniform black background (both true-positive and true-negative), making the differences easier to identify.

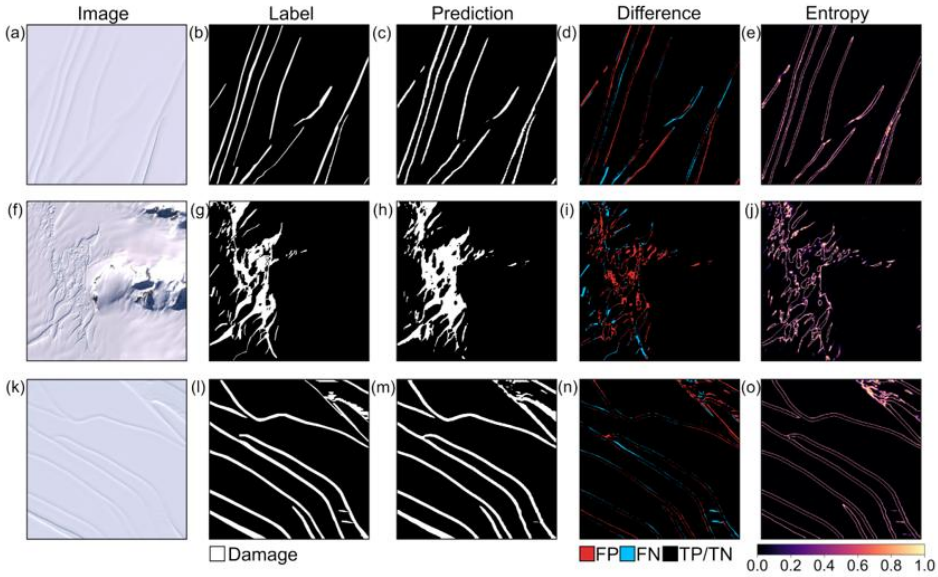


Figure 8. Examples of prediction performance on samples from the validation set. (a), (f) and (k) show image samples from Amery Ice Shelf, Crosson Ice Shelf and Larsen C Ice Shelf, respectively. (b), (g) and (l) show the corresponding manually annotated labels. (c), (h) and (m) show the corresponding model predictions. (d), (i) and (n) show the pixel-wise difference maps between the corresponding labels and predictions, where TP denotes true positive, FP denotes false positive, FN denotes false negative and TN denotes true negative. (e), (j) and (o) show the corresponding prediction entropy maps.

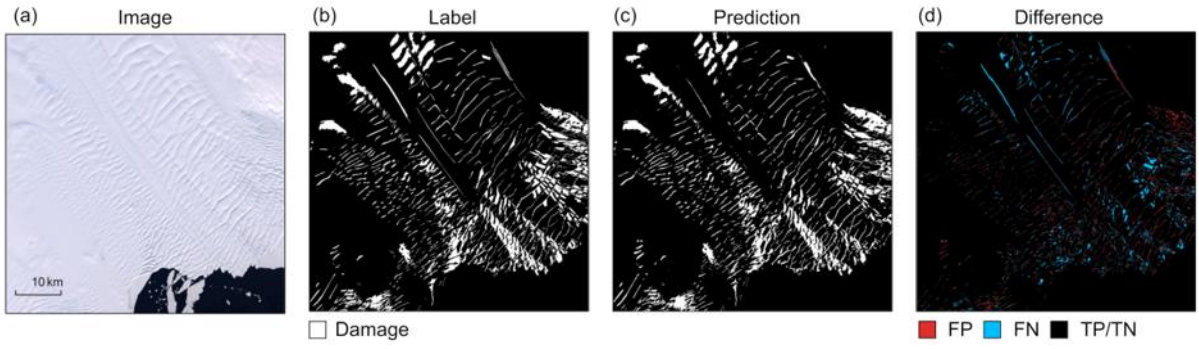


Figure 9. Prediction performance over the full contiguous Landsat 8 test scene on Totten Ice Shelf. (a) Landsat 8 image acquired on 2021-01-23. (b) Manually annotated label. (c) Model prediction. (d) Pixel-wise difference map between label and prediction, where TP denotes true positive, FP denotes false positive, FN denotes false negative and TN denotes true negative.

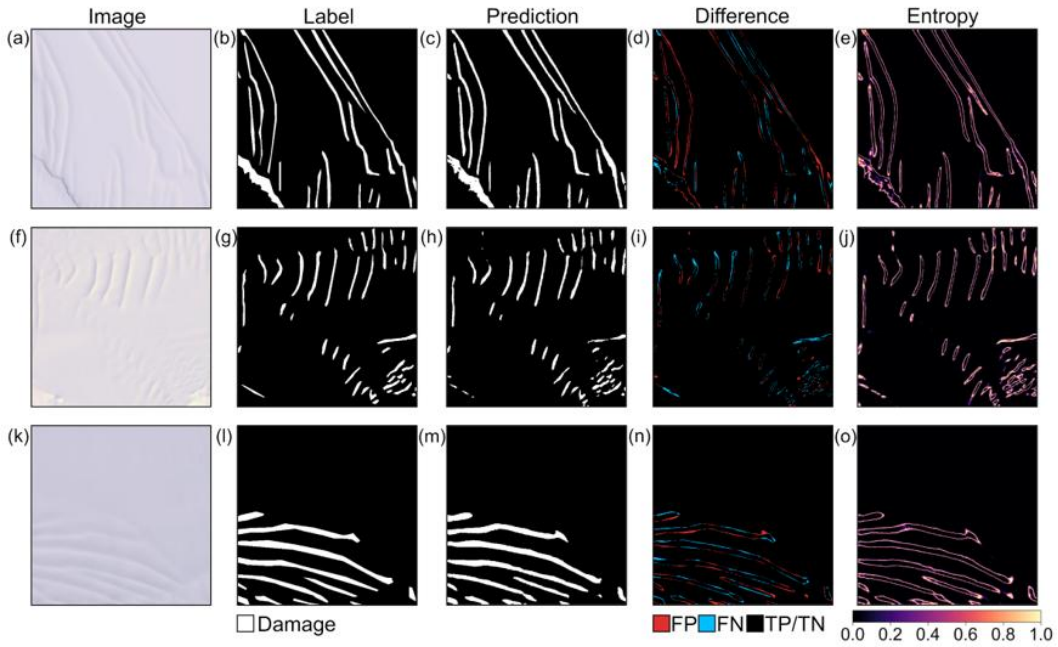


Figure 10. Examples of prediction performance on samples from the Landsat 7 SLC-on cross-sensor test set. (a), (f) and (k) show image samples from Brunt Ice Shelf, Dotson Ice Shelf and Thwaites Ice Shelf, respectively. (b), (g) and (l) show the corresponding manually annotated labels. (c), (h) and (m) show the corresponding model predictions. (d), (i) and (n) show the pixel-wise difference maps between the corresponding labels and predictions, where TP denotes true positive, FP denotes false positive, FN denotes false negative and TN denotes true negative. (e), (j) and (o) show the corresponding prediction entropy maps.

7. Figure 10 - I was surprised by the right-most panels which are labeled "entropy". I did not find where the term "entropy" was defined and discussed/motivated as an interesting diagnostic in the text (did I miss it?). I have enough trouble understanding what "thermodynamic entropy" is, so please help readers like me…

Response: Thank you for pointing this out, and we apologize that the term was not clearly defined. Prediction entropy was introduced in the manuscript to complement the map-level uncertainty estimates by providing spatial information on prediction uncertainty. We have further clarified its definition and interpretation. Specifically, entropy is calculated from the pixel-level class probabilities using the Shannon entropy formula and normalized to the range 0–1, with higher values indicating greater uncertainty in pixel classification.

The relevant description has been revised as follows:

Prediction entropy was calculated from the pixel-level class probabilities using the Shannon entropy formula and normalized to the range 0–1. Higher values indicate greater uncertainty in distinguishing damage from background, whereas lower values indicate more confident predictions.

8. I had trouble keeping up with the variables, including rs… which apparently is a Spearman's rank correlation. I wonder if it might be better to (a) reduce the number of variables defined in the paper (are they all being used and are they all needed to understand the science) or (b) include a table (or appendix) that lists the variables, gives their formula, and has a short word-based explanation of what they mean and how they are used… It may be that this suggestion isn't a good one, but I present it anyway because it occurred to me.

Response: Thank you for pointing this out. We agree that the manuscript involves a considerable number of variables and metrics, which may make it difficult for readers to follow. These variables and metrics have been defined and explained in the relevant sections, including their formulas and purposes where applicable. For example:

"Three metrics are used for multi-temporal consistency assessment, including Spearman's rank correlation coefficient (r_s), Lin's concordance correlation coefficient (CCC, ρ_c) and root mean square error (RMSE). Among them, r_s is used to assess the consistency of the rank of grid cell values in temporally adjacent damage density maps, CCC evaluates their agreement in absolute magnitude and RMSE measures their average numerical deviation. The formulae are as follows:

$$r_s = \frac{\sum_{i=1}^{n}(R_i-\bar{R})(S_i-\bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i-\bar{R})^2 \sum_{i=1}^{n}(S_i-\bar{S})^2}} \tag{10}$$

where n is the total number of valid grid cells, R_i and S_i are the ranks of the damage density values at the i-th grid cell in two temporally adjacent maps, and $\bar{R}$ and $\bar{S}$ are their respective mean ranks. Tied values are assigned average ranks.

$$\rho_c = \frac{2\sigma_{xy}}{\sigma_x^2+\sigma_y^2+(\mu_x-\mu_y)^2} \tag{11}$$

where μ_x and μ_y are the mean values of damage density for two temporally adjacent density maps, σ_x^2 and σ_y^2 are the variances of the two density maps, and σ_{xy} is the covariance between them.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{12}$$

where n is the total number of valid grid cells and x_i and y_i are the values of the i-th grid cell in two temporally adjacent density maps."

Considering the length of the manuscript and that these metrics are mainly used for dataset evaluation, we have retained their explanations in the relevant sections rather than introducing an additional summary table.

9. Figure 9, panel d - again I wonder if the black and white background should be eliminated so that the colors are seen more readily….

Response: We agree with this suggestion. As described in our response to Comment 6, we have revised the difference panels in Figs. 8–10 to improve the visibility of the classification errors.

10. Author contributions: this section contains nonsensical words that I've never heard of, i.e., "Contributions are according to CRediT." That means nothing to me, and I've never heard that word (CRediT) before. Probably this is something that is understood in some circles, but not in all, so why not communicate using regular language.

Response: Thank you for pointing this out. The original statement referred to the CRediT contributor roles taxonomy recommended by Copernicus Publications. We have removed this sentence to avoid confusion while retaining the detailed descriptions of individual author contributions.