

Supplementary information for "Multi-sensor deep-learning mapping of surface and basal fractures across Antarctic ice shelves during the last four decades"

Kaian Shahateet^{1,2,*}, Colin Prieur^{1,*}, Romain Millan^{1,*}, Lucille Gimenes¹, Andrea Sofie Haastrup-Vang³, Jaime Roas³, Raphaëlle Charrassin⁴, Emil Giraldi³, Romain Santiard¹, and Maaïke Izeboud⁵

¹Université Grenoble Alpes, CNRS, INRAE, IRD, Grenoble INP, IGE, Grenoble, France

²Department of Water Resources and Environment, Aeronautics Institute of Technology, São José dos Campos, Brazil

³Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen K, Denmark

⁴Department Earth System Science, University of California Irvine, Irvine, 92697 CA, USA.

⁵Department of Water and Climate, Vrije Universiteit Brussel, Brussels 1050, Belgium

*These authors contributed equally to this work.

S1 Loss function sensitivity analysis

To assess whether the Jaccard loss function is actually preferable to the more conventional weighted categorical cross-entropy (WCCE) loss, we retrained the same architecture on the same data set split using WCCE instead, with class weights $\alpha_c = 1 - D_c$ (normalized): $\alpha_{background} = 0.03$, $\alpha_{basal} = 0.48$, $\alpha_{surface} = 0.49$, following the inverse label-frequency scheme commonly
5 used to counteract class imbalance in cross-entropy losses. The results using the Jaccard loss were better in every metric except recall (see Table S1). Interestingly, the WCCE model shows a nearly 30% gap between recall and precision, in contrast to the nearly balanced precision and recall of the Jaccard-trained model. This behavior is consistent with how the two losses handle class imbalance: because $\alpha_{background}$ is more than an order of magnitude smaller than α_{basal} and $\alpha_{surface}$, the fixed per-pixel weighting of WCCE strongly discourages background predictions, pushing the model towards an aggressive, over-inclusive
10 detector that recovers more true damage pixels at the cost of a much higher false-positive rate.

Table S1. Metrics comparison (in %) between the soft Jaccard loss and weighted categorical cross-entropy (WCCE), both trained on the same configuration (e.g., architecture, data split, and epochs) and evaluated on the same 19-tile held out test set as Table 1.

Loss	Accuracy	Precision	Recall	F1-score	IoU	IoU _{indif}
Jaccard	95.4	64.7	64.2	64.4	47.5	49.4
WCCE	93.6	50.5	80.0	61.9	44.8	46.6

In conclusion, for the same labeled data, Jaccard results in a more accurate outcome, so we adopted it for all results reported in this study.

S2 Additional Tables

Table S2. Characteristics of the Landsat images used during training.

Platform	Band used	Pixel size (m)	Start operation	End operation	Total n° of scenes
Landsat 1	Near-infrared (Band 7)	57 × 79	1972	1978	3
Landsat 4	Near-infrared (Band 4)	30 × 30	1982	2001	7
Landsat 5 (MSS)	Near-infrared (Band 4)	30 × 30	1984	2013	2
Landsat 7	Panchromatic (Band 8)	15 × 15	1999	2025	5
Landsat 8	Panchromatic (Band 8)	15 × 15	2013	Present	11
Landsat 9	Panchromatic (Band 8)	15 × 15	2021	Present	10

Table S3. Cloud-related flags available in the Landsat QA band, by sensor.

Sensor	Cloud	Cloud shadow	Cirrus
MSS	✓		
ETM+	✓	✓	
OLI	✓	✓	✓

Table S4. Mean per-tile IoU by test region.

Region (<i>n</i> tiles)	DR0	DR1	DR2	DR3
Ronne (7)	19.2	24.1	36.4	39.4
Dotson (6)	24.0	39.3	44.3	46.5
Larsen-F and -G (6)	2.1	4.6	15.3	25.9

Table S5. Statistics of the comparison between all the human annotators using the IoU metric in percentage (%). The diagonal values are the intra-delineators comparison, which only delineator A, B, and C performed.

	A	B	C	D	E	F
A	62.0	50.9	47.4	54.6	52.0	52.0
B		70.2	45.0	57.2	54.5	56.2
C			55.9	50.1	51.1	47.6
D				-	58.4	56.7
E					-	59.2
F						-

S3 Additional figures

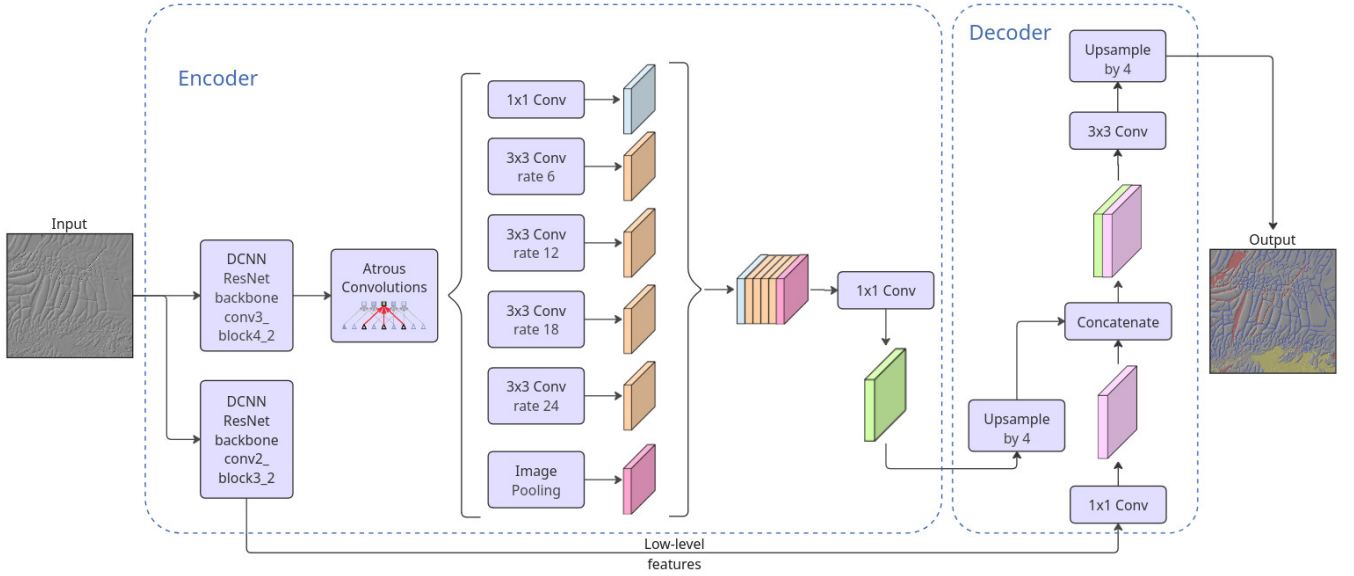


Figure S1. Schematic of the DeepLabMini architecture used in this study, adapted from DeepLabv3+ following the Keras implementation of Chen et al. (2017), with a ResNet-50 (He et al., 2015) encoder trained from scratch (no ImageNet pre-training). (1) **Encoder**: the input patch is passed through the ResNet-50 backbone, tapped at two depths. High-level features are extracted at `conv3_block4_2` (output stride 8) and fed to an Atrous Spatial Pyramid Pooling (ASPP) module, which applies six parallel branches to the same feature map, a 1×1 convolution, four 3×3 atrous convolutions at dilation rates 6, 12, 18 and 24, and a global average-pooling branch, each producing 128 feature maps. The six outputs are concatenated and fused by a final 1×1 convolution. Low-level features are extracted separately, earlier in the backbone, at `conv2_block3_2` (output stride 4), and projected to 48 channels by a 1×1 convolution to limit their weight relative to the higher-level features. (2) **Decoder**: the ASPP output is bilinearly upsampled to match the spatial resolution of the low-level features, concatenated with them, and refined by two successive 3×3 convolutions (64 filters). The result is bilinearly upsampled to the input resolution and passed through a final 1×1 convolution with softmax activation to produce per-pixel probabilities for the three output classes (background, basal crevasse, surface crevasse). Every convolution outside the ResNet-50 backbone (i.e., in the ASPP module and the decoder) is followed by batch normalization and a LeakyReLU activation, rather than the standard ReLU used inside the backbone itself.

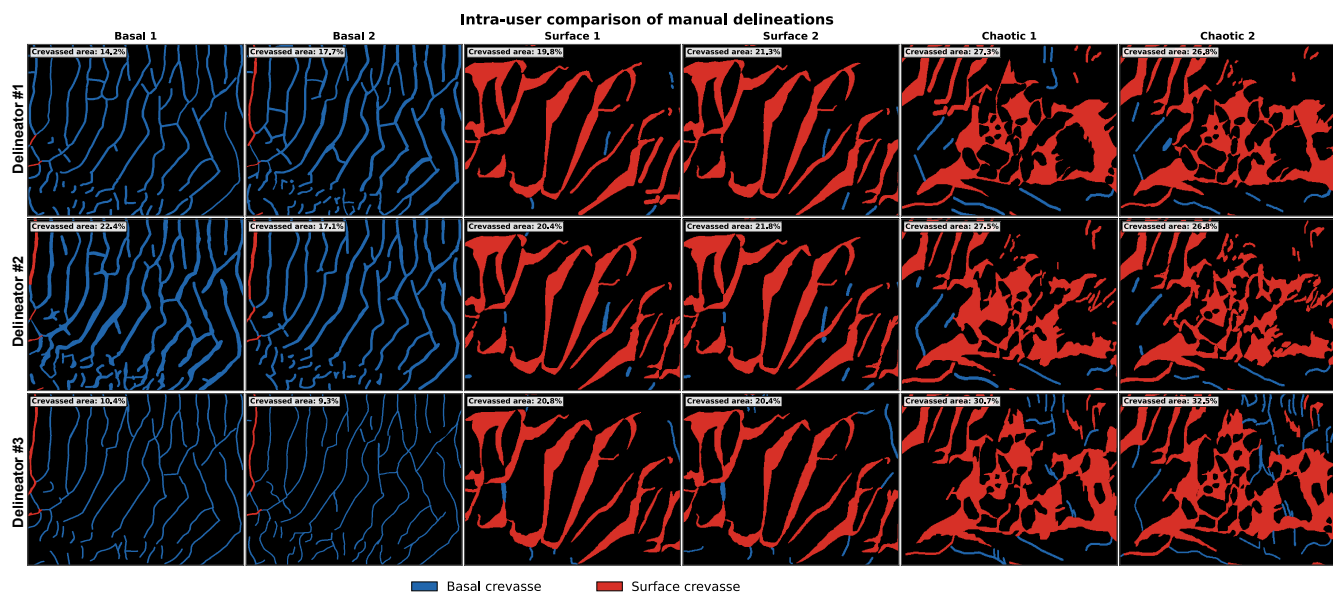


Figure S2. Comparison of manual delineation performed twice by the same delineators. Each row is a mapping made by a different user, while column represents the areas that have been mapped (see 3).

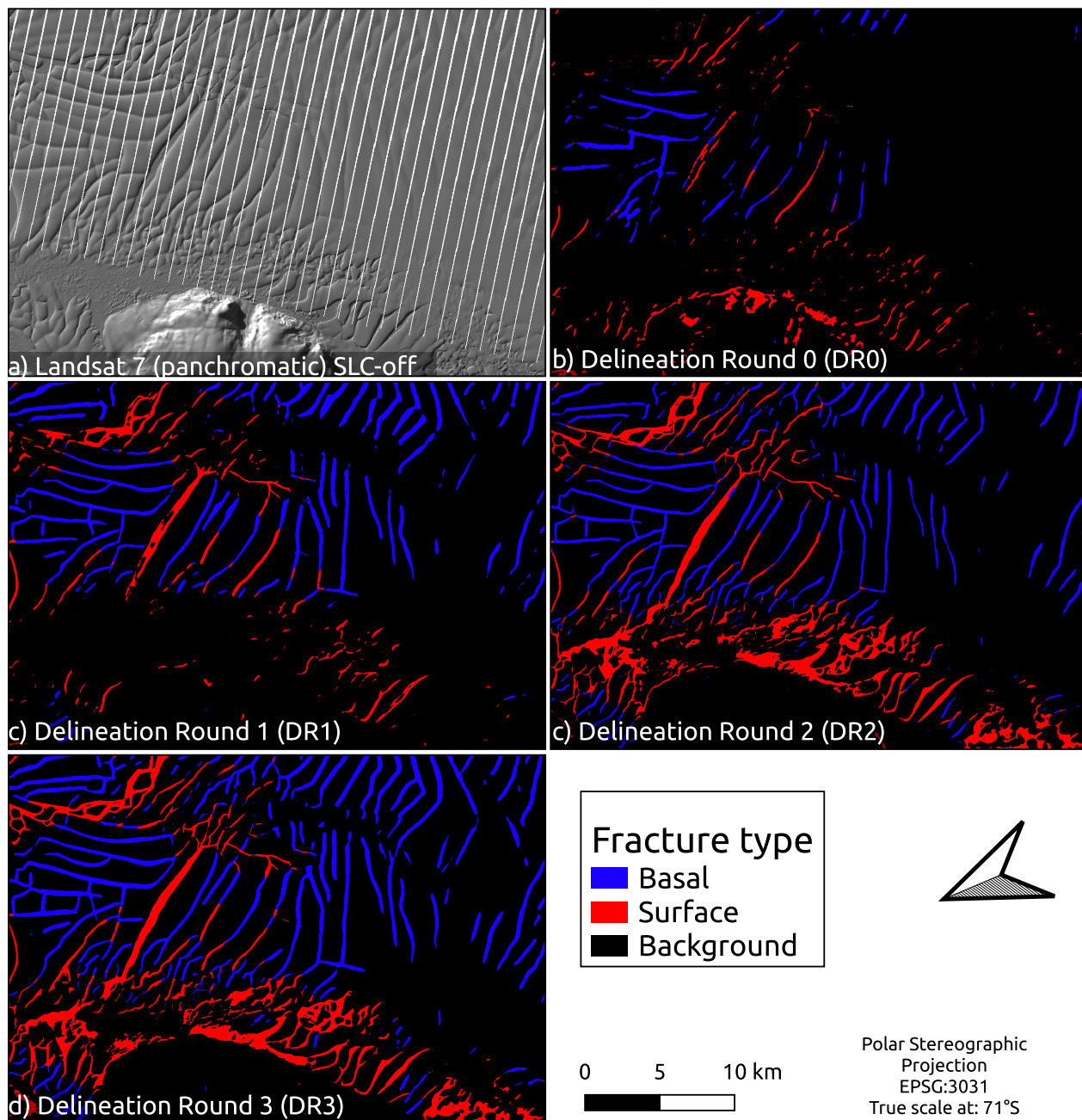


Figure S3. Comparison of the capability of the model to identify the fractures in different stages of the human-in-the-loop approach. Example is provided over the Crosson ice shelf.

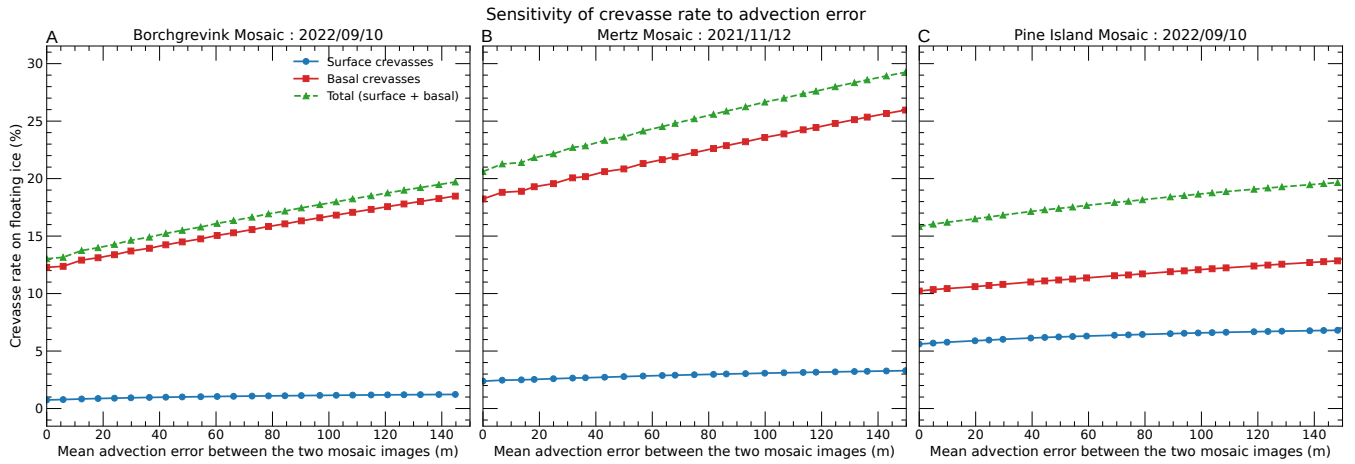


Figure S4. Impact of advection error in the mosaicking of fracture maps for (A) Borchgrevink, (B) Mertz and (C) Pine island ice shelf.

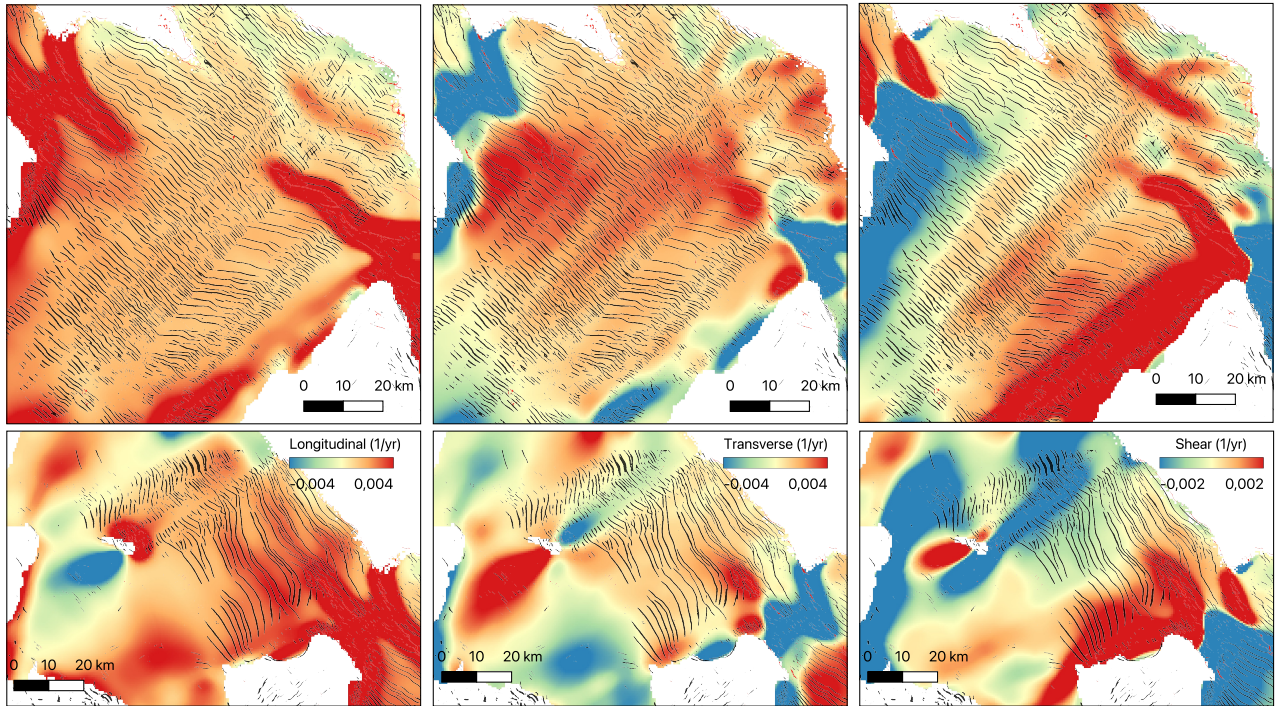


Figure S5. Spatial distribution of fractures over the Borchgrevink Ice Shelf in 2019, overlaid on longitudinal (left), transverse (center), and shear (right) strain rates. Basal and surface crevasses are shown in black and red, respectively. Strain rates are represented using a linear blue-to-red color scale, with blue indicating negative strain rates and red indicating positive strain rates.

15 **References**

- Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H.: Rethinking Atrous Convolution for Semantic Image Segmentation, <https://arxiv.org/abs/1706.05587>, 2017.
- He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, <https://arxiv.org/abs/1512.03385>, 2015.