

Supplement: Large changes in visible crevasse extent on Svalbard glaciers from 2017 to 2025, mapped with satellite embeddings

Erik Schytt Mannerfelt^{1,2} and Thomas Schellenberger²

¹Arctic Geology, The University Centre in Svalbard, Norway

5 ²Department of Geosciences, University of Oslo, Oslo, Norway

Correspondence: Erik S. Mannerfelt (erikm@unis.no)

S1 Training labels and point samples

We add polygons for training and validation by manually identifying them through a qualitative search of Sentinel-2 L1C images (Fig. S1). Points are drawn inside each labeled polygon with a capped linear area weight, bounded between 8 and 120 samples per polygon (the polygons themselves spanning 0.35–100 km²). The cap prevents the large polygons from dominating the point samples while still letting them contribute more than the smallest ones. Each of the 30 models draws points from an independent random 70% subset of polygons for training, with the remaining 30% used for validation, under a fixed random seed to ensure reproducibility.

S2 Validation

Table S2 scores the 50 m vote-fraction rasters against the label polygons in 2024 (the withheld year) over RGI7.0 glacierized area. This is the primary validation quoted in the main text.

Table S3 shows validation to the full 30% test polygons per model. This overestimates total performance since polygons from the same year are used for validation, but informs on the within-year performance and the spread between models.

We split the positive class by whether the labeled field is large and clear or small and/or ambiguous. The purpose is to observe whether large crevasse fields that are easy to spot skew the validation to the point of missing more nuanced locations. Recall is $91 \pm 2\%$ on crevasse fields that are labeled large and clear, and is $89 \pm 2\%$ on all others (cf. $91 \pm 2\%$ for all positives). The fact that recall is similar to within one standard deviation tells that both clear and less clear locations are resolved to a similar degree.

S3 Elevation dependency of visible crevasse change

We locate visible crevasse change in elevation with the 50 m S0 terrain model (Norwegian Polar Institute, 2014), resampled bilinearly onto the prediction grid. Every glacier pixel is assigned to a 100 m elevation band and to a tenth of its glacier's own elevation range, the positive area of each model is summed per band and year for the significant non-surging glaciers of each trend sign, and a trend is fitted per band and model. The band areas sum exactly to the glacier totals used in the main text. Glaciers with a significant increase gain $+175$ (155 – 215) km² a⁻¹ in the upper 60% of their elevation range and lose -12 (-16 to -3) km² a⁻¹ in the lower 40%. Part of that low-elevation loss is likely tidewater termini retreating within the fixed RGI7.0 outlines, which replaces crevassed ice with open water, and it is indeed concentrated on tidewater glaciers.

30 S4 Further limitations

The annual embedding contains data from an entire year, while training labels are drawn from late-summer scenes when snow cover is at its seasonal minimum and crevasses are most visible. A polygon that is correctly labeled positive from one August image can therefore carry an annual embedding signature dominated by firn, which we expect to be the main contributor of false positives in accumulation areas.

35 The label dataset is from a qualitative coverage effort rather than a stratified sample. Some regions in Svalbard contain more polygons than others (though part of this is a real crevasse distribution signal), and some years carry more labels in some regions while less in others. Expanding this to a global or fully comprehensive scope would require a more comprehensive labeling environment.

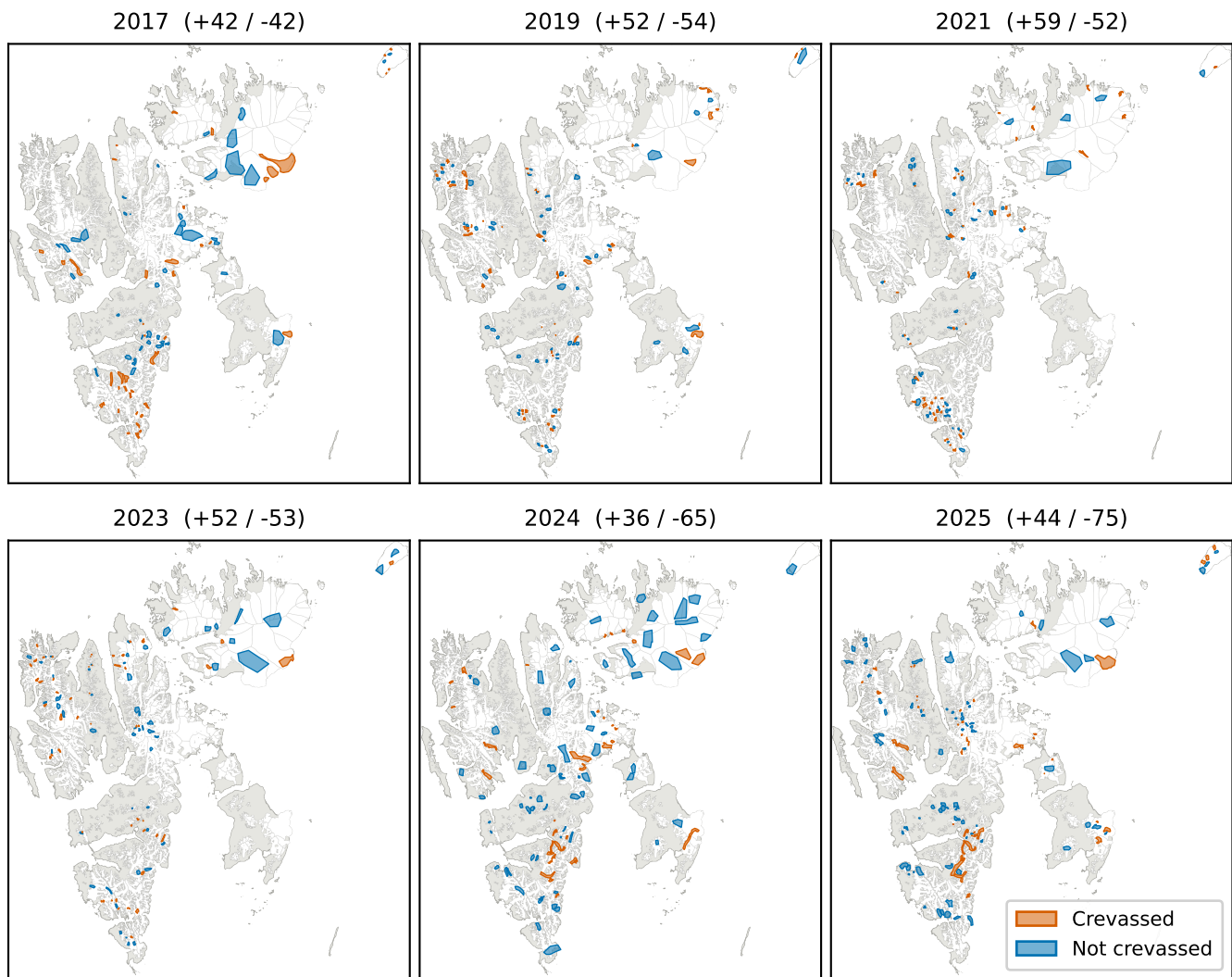
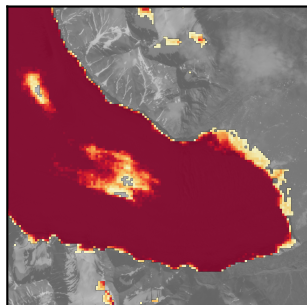
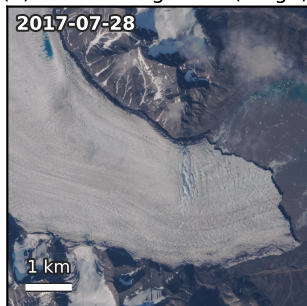
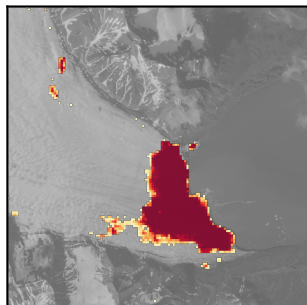
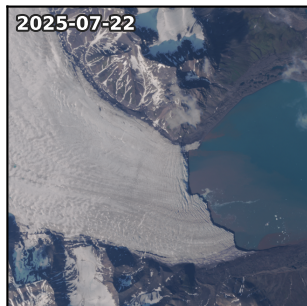


Figure S1. Spatial and temporal coverage of the hand-digitized training labels, faceted by the six years that carry labels. Label polygons are filled by class over land (gray) and RGI7 glacier outlines (white), at the same extent as Fig. 2 of the main text. Panel titles give the number of positive and negative label polygons in that year. Note that 2024 is the withheld year, shown for completeness rather than because any model was trained on it.

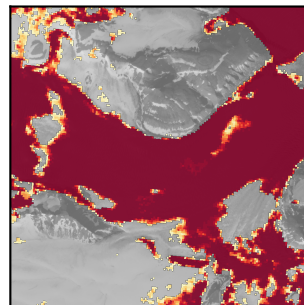
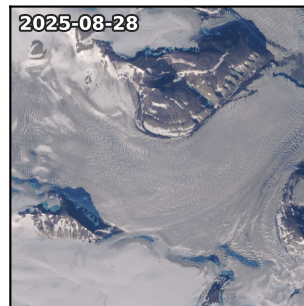
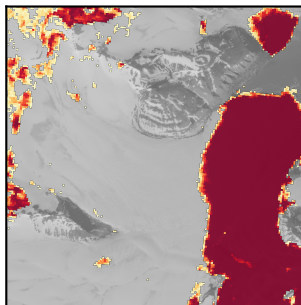
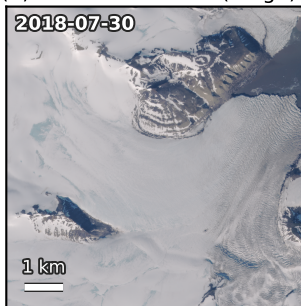
(a) Wahlenbergbreen (surge)



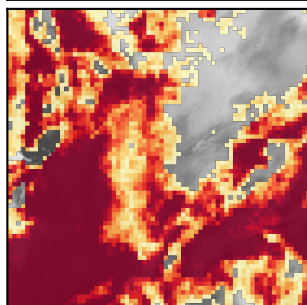
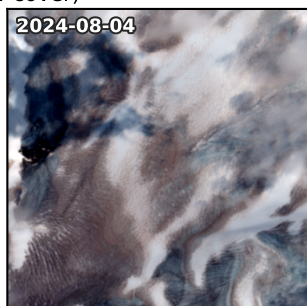
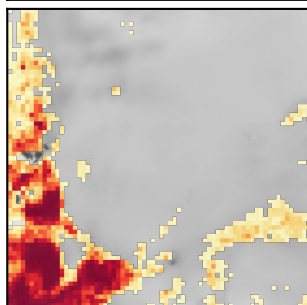
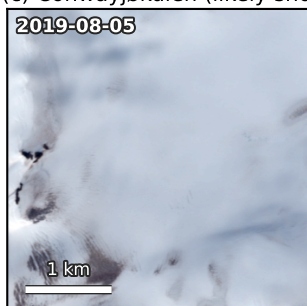
(b) Petermannbreen (surge)



(b) Petermannbreen (surge)



(c) Conwayjøkulen (likely snow cover)



(d) NE Kvitøygjøkulen (likely snow cover)

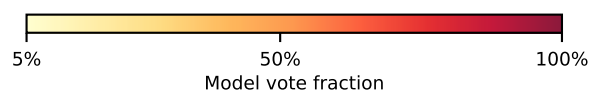
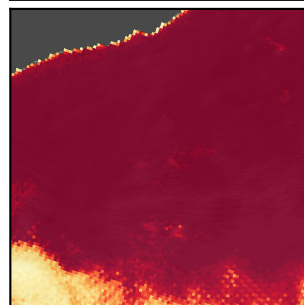
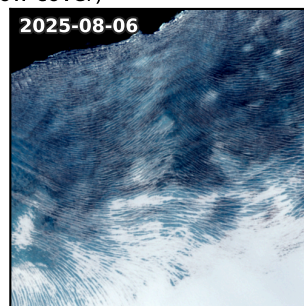
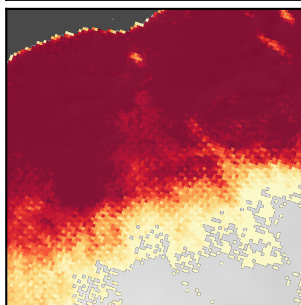


Figure S2. (Figure on previous page.) Sentinel-2 L1C images beside the predicted visible crevasse extent for four example areas. Each block shows an early and a late image (top) and the fraction of the 30 ensemble members voting positive in the same year (bottom). **a:** surge-related decrease on Wahlenbergreen (78.49 ° N, 14.20 ° E). **b:** surge-related up-glacier spread on Petermannbreen (78.49 ° N, 18.65 ° E). **c:** Conwayjøkulen, a tributary of Mittag-Lefflerbreen (78.85 ° N, 17.19 ° E). **d:** the northeastern margin of Kvitøkjøkulen (80.27 ° N, 32.52 ° E). Surge attribution in **a** and **b** follows Harcourt et al. (2026); **c** and **d** are likely affected by the lack of snow cover, but real concomitant crevasse change can also not be ruled out. Vote fractions below 5% are transparent. Images are Copernicus Browser “Highlight Optimized Natural Color” renderings.

Table S1. Model prediction scores over the withheld 2024 validation year in RGI7.0 glacierized area based on which input was provided. Ensemble mean ± standard deviation across the 30 models is shown. Unlike Table S3, these columns are directly comparable.

Predictor	Accuracy	F1	Precision	Recall	Specificity	FPR
AlphaEarth embeddings	93% ± 2%	86% ± 3%	82% ± 5%	91% ± 2%	94% ± 2%	6% ± 2%
Sentinel-1 backscatter	87% ± 2%	77% ± 2%	68% ± 4%	88% ± 2%	87% ± 2%	13% ± 2%

Table S2. Ensemble validation on the withheld 2024 validation year over glacierized area. The average row uses the standard deviation as the spread estimator.

Model	Accuracy	F1	FNR	FPR	Precision	Recall	Specificity
0	88%	78%	9%	13%	68%	91%	87%
1	93%	86%	11%	6%	83%	89%	94%
2	94%	88%	8%	5%	84%	92%	95%
3	92%	85%	7%	8%	78%	93%	92%
4	94%	88%	11%	4%	86%	89%	96%
5	94%	88%	10%	5%	85%	90%	95%
6	93%	85%	11%	6%	82%	89%	94%
7	94%	86%	11%	5%	84%	89%	95%
8	94%	88%	7%	5%	84%	93%	95%
9	93%	86%	10%	6%	82%	90%	94%
10	93%	85%	10%	6%	81%	90%	94%
11	94%	87%	10%	5%	84%	90%	95%
12	95%	89%	7%	5%	85%	93%	95%
13	93%	87%	7%	6%	81%	93%	94%
14	93%	86%	10%	6%	83%	90%	94%
15	88%	77%	10%	13%	68%	90%	87%
16	94%	88%	9%	5%	84%	91%	95%
17	94%	86%	11%	5%	84%	89%	95%
18	91%	83%	9%	9%	76%	91%	91%
19	93%	86%	10%	6%	82%	90%	94%
20	92%	83%	11%	7%	79%	89%	93%
21	94%	87%	9%	5%	84%	91%	95%
22	94%	88%	8%	5%	85%	92%	95%
23	92%	84%	9%	7%	79%	91%	93%
24	93%	85%	7%	8%	79%	93%	92%
25	92%	83%	13%	7%	79%	87%	93%
26	93%	85%	10%	6%	81%	90%	94%
27	94%	87%	12%	4%	87%	88%	96%
28	93%	87%	8%	6%	82%	92%	94%
29	95%	89%	10%	4%	87%	90%	96%
Average	93% ± 2%	86% ± 3%	9% ± 2%	6% ± 2%	82% ± 5%	91% ± 2%	94% ± 2%

Table S3. Scores for each model’s 30% split test data for the same years as training. This is more optimistic and not comparable with Table S2. The $\pm$ values in the average show the standard deviation.

Model	Accuracy	F1	FNR	FPR	Precision	Recall	Specificity
0	83%	82%	25%	9%	89%	75%	91%
1	89%	89%	16%	6%	94%	84%	94%
2	90%	90%	14%	6%	94%	86%	94%
3	88%	88%	13%	10%	89%	87%	90%
4	88%	87%	19%	6%	93%	81%	94%
5	90%	90%	13%	7%	93%	87%	93%
6	88%	88%	16%	7%	92%	84%	93%
7	89%	89%	14%	8%	92%	86%	92%
8	89%	89%	15%	6%	94%	85%	94%
9	89%	89%	16%	5%	95%	84%	95%
10	90%	89%	15%	6%	94%	85%	94%
11	88%	87%	17%	8%	92%	83%	92%
12	86%	85%	21%	7%	92%	79%	93%
13	92%	92%	9%	7%	93%	91%	93%
14	89%	88%	17%	5%	94%	83%	95%
15	91%	90%	11%	8%	91%	89%	92%
16	89%	89%	13%	9%	91%	87%	91%
17	87%	86%	21%	4%	95%	79%	96%
18	87%	85%	22%	4%	95%	78%	96%
19	91%	91%	11%	7%	93%	89%	93%
20	87%	86%	20%	6%	93%	80%	94%
21	89%	88%	16%	7%	93%	84%	93%
22	89%	88%	18%	4%	95%	82%	96%
23	89%	88%	18%	5%	94%	82%	95%
24	89%	88%	17%	4%	95%	83%	96%
25	91%	91%	13%	4%	95%	87%	96%
26	90%	89%	14%	7%	92%	86%	93%
27	88%	87%	20%	4%	95%	80%	96%
28	90%	90%	8%	12%	89%	92%	88%
29	88%	87%	20%	4%	95%	80%	96%
Average	89% $\pm$ 2%	88% $\pm$ 2%	16% $\pm$ 4%	6% $\pm$ 2%	93% $\pm$ 2%	84% $\pm$ 4%	94% $\pm$ 2%