Supplement of The Cryosphere, 18, 2677–2689, 2024
https://doi.org/10.5194/tc-18-2677-2024-supplement

The Cryosphere Open Access EGU

Supplement of

Weak relationship between remotely detected crevasses and inferred ice rheological parameters on Antarctic ice shelves

Cristina Gerli et al.

Correspondence to: Cristina Gerli (cristina.gerli@northumbria.ac.uk)

S1. Supplementary

S1.1 Preliminary analysis

We perform a preliminary analysis where we examine the link between the crevasse binary map (CNN and NeRD) and the ice rate factor field A using two different ice rate factor thresholds: an A value for temperate ice and an A value 10 times large. We evaluate the confusion matrix charts (Fig. S1) and F1 scores (Table S1 and S2) to assess the performance of the model and assess whether the dataset is unbalanced. The confusion matrix chart summarises the four possible outcomes of a binary classification analysis between predicted and true classes: True positive, True Negative, False Positive and False Negative. These values are used to estimate different performance metrics such as accuracy, precision, recall and F1 score, which allow us to quantify how often the model makes correct and incorrect predictions for the presence of crevasses at a given location. The precision metric measures the proportion of true positive predictions among all positive predictions, while the recall metric is the proportion of true positive prediction among all true positive cases. The F1 score is the "harmonic mean" of these two metrics, and ranges from 0 to 1 (Rizk et al., 2019). It is useful when both false positives and false negatives are equally undesirable, and the dataset is unbalanced.

Since the crevasse datasets are highly skewed towards the non-crevassed class (99 %), with such an imbalanced distribution of classes, a "naïve" classifier will likely show a biased high accuracy, as it would predict most non-crevassed regions correctly, while performing poorly on the minority class. This is clearly reflected in the confusion matrix charts and the F1 scores reported in Fig. S1 and Tables S1 and S2, for Filchner-Ronne and Pine Island ice shelves, respectively. The confusion charts summarise the results of a binary classification analysis, between predictions and observations. In both cases, for all crevasses and heavily damaged crevasses, the non-crevassed region is correctly detected, yet the model fails to detect the crevasse class, which is a minority in the dataset. This is further reflected in the F1 score analysis. A high weighted F1 score is measured for both *A-threshold*s, which would suggest that the model identifies crevasses accurately while minimizing incorrect hits and misses. However, these scores are unreliable as the majority no-crevasses class, which dominates the sample, provides very high performances compared to the minority crevasse class (0.0074 and 0.0024, in Table S1), overall dominating the final score. This is visible on both Filchner-Ronne and Pine Island ice shelves, as most samples belongs to the non-crevassed class. As a result, the F1 scores for all crevasse maps and *A-threshold*s are considerably influenced by the majority class, making it challenging to evaluate accurately the model's performance on the minority (crevasse) class.

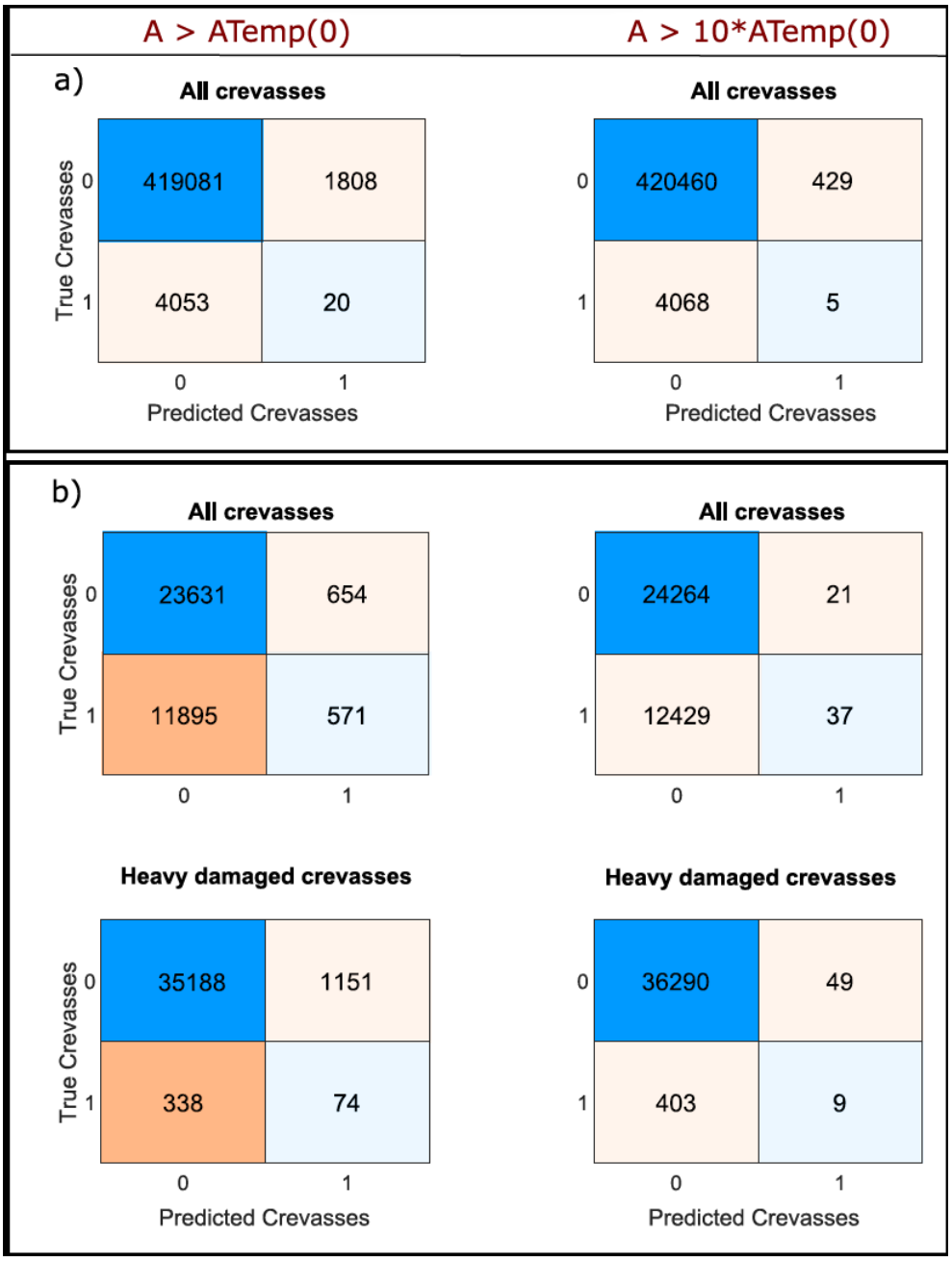

Figure S1 Confusion matrices for two classification cases, when *A* is greater than *A* for temperate ice measured by experiments and when *A* is ten times greater than *A* for temperate ice, for Filchner-Ronne Ice Shelf for all crevasses (panel a), and for Pine Island Ice Shelf for all and heavily damaged crevasses (panel b). All matrices summarize the predicted and true class labels. Each table is divided into four quadrants, depicting the four outcomes of a binary classification analysis: true negative, true positive, false positive and false negative. In both cases (panels a and b), the non-crevassed region is correctly detected, thus providing a high "accuracy", which is biased and incorrect since the non-crevassed class is dominating the dataset.

Table S1 Display of F1 scores for Filchner-Ronne Ice Shelf when relating the inverted ice rate factor field *A* and all crevasses as mapped by the CNN of Lai et al., (2020). The high weighted F1 score (0.9841 and 0.9854), for the two *"A-threshold"* cases (A(0°C) and 10*A(0 °C)) suggests that the model identifies crevasses accurately while minimizing incorrect hits and misses. These scores are unreliable because the no-crevasses class still dominates the sample (Weight_NoCrev, 99 %), resulting in high F1 scores (0.9936 and 0.9950) compared to the F1 scores of the minority crevasse class (0.0074 and 0.0024).

Filchner Ronne Ice Shelf		Weight_Crev	Weight_NoCrev	F1_Crev	F1_NoCrev	F1_Weighted
All Crevasses	A(0 °C)	0.0097	0.9903	0.0074	0.9936	0.9841
	10*A(0 °C)	0.0097	0.9903	0.0024	0.9950	0.9854

Table S2 Display of F1 scores for Pine Island Ice Shelf for all crevasses and heavily damaged crevasses; Overall, a weighted F1 score of 0.53-0.55 and 0.96 -0.98 indicates that the model's overall performance across all classes is moderate to high. However, the performance of the model seems to be imbalanced towards the No-Crevasse class (which define 66 % and 99 % of the distribution respectively). As the F1 score for crevasses does not exceed 0.11, it suggests that the performance of the current crevasse detection model is weak when classifying crevasses and that certain techniques are necessary to balance the dataset.

Pine Island Ice Shelf		Weight_Crev	Weight_NoCrev	F1_Crev	F1_NoCrev	F1_Weighted
All Crevasses	A(0 °C)	0.34	0.66	0.10	0.78	0.55
	10*A(0 °C)	0.34	0.66	0.02	0.80	0.53
Highly Damaged Crev	A(0°C)	0.01	0.99	0.08	0.97	0.96
	10*A(0 °C)	0.01	0.99	0.11	0.99	0.98

S1.2 Ice rate factor field *A* and satellite image for Filchner-Ronne and Pine Island ice shelves.

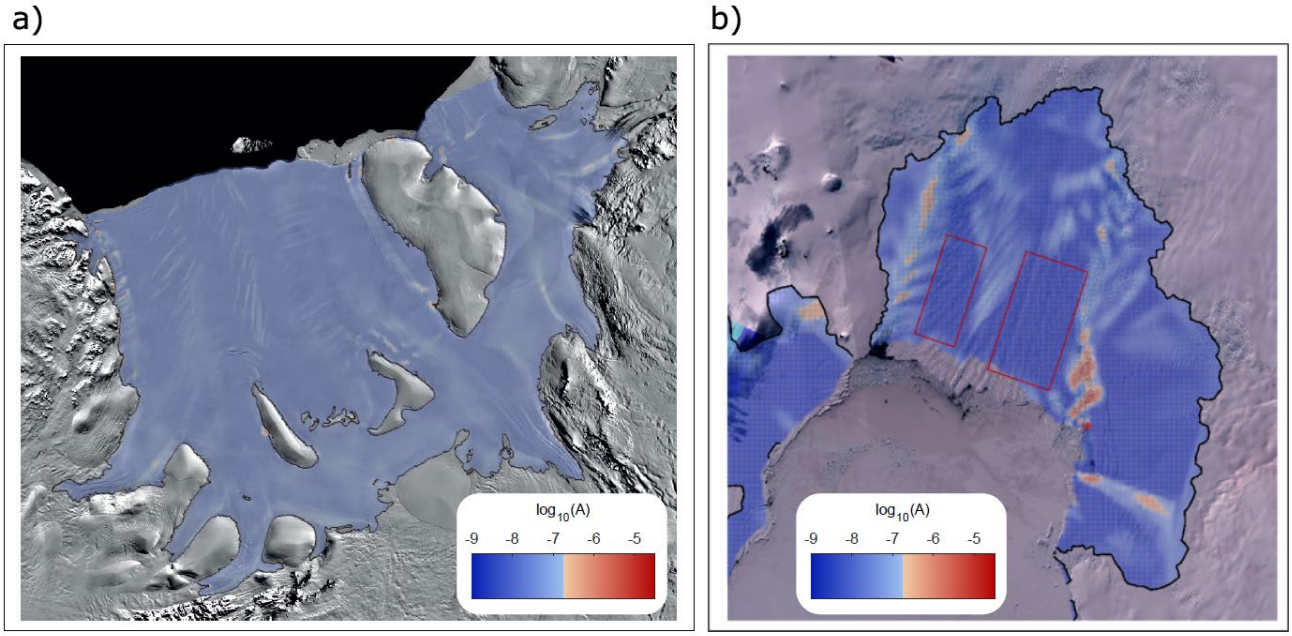

Figure S2 Detailed display of the inverted ice rate factor field *A* for Filchner-Ronne (a) and Pine Island Ice Shelf (b), using colour scale described in Fig. 1. Red boxes in b) display a discrepancy between areas displaying an ice rate factor well below the reference *A* for temperate ice and the abundance of remotely sensed crevasses depicted from the composite Sentinel (S2) image of austral summer 2019-2020. The MODIS MOA2009 mosaic underlaid in panel (a) is available at the National Snow & Ice Data Center (NSIDC): https://nsidc.org/data/nsidc-0593/. The Sentinel 2 composite in panel (b) was obtained from Izeboud et al., 2023 processed in the Google Earth Engine, for which the code is available here: https://code.earthengine.google.com/63b4cf06dffc7c6b5695dbd256e844fd.

S1.3 Calculation of the Optimal Operating Point

To further identify an ideal *"A-threshold"* which best compromises the FPR and TPR in this analysis, we calculate the OPTtimal operating PoiNT (OPT-PNT) on the ROC curve. The OPT-PNT on the ROC curve is calculated by finding the *"A-threshold"* which best comprises the pair of False Positive Rate (FPR) and True Positive Rate (TPR). This is performed by finding the slope S of a line which satisfies the relationship:

$$S = \frac{\mathrm{Cost(P/N)} - \mathrm{Cost(N/N)}}{\mathrm{Cost(N/P)} - \mathrm{Cost(P/P)}} * \frac{N}{P}$$

where the terms "Cost(N/P)" and "Cost(P/N)" refer to the costs incurred from misclassifying a positive class as a negative class and a negative class as a positive class, respectively. P is the total number of positive instances, which is the sum of true positives (TP) and false negatives (FN). N is the total number of negative instances, which is the sum of true negatives (TN) and false positives (FP). The optimal operating point is found by moving the straight line with slope S from the upper

left corner of the ROC plot (FPR = 0, TPR = 1) down and to the right, until it intersects the ROC curve (tangent to the curve).

S1.4 Sensitivity analysis to changes of the regularisation parameter (γ_s)

The importance of conducting a sensitivity analysis for γ_s pertains to values within the L-corner range [10^3 10^6]. In Figure S3 we present ROC curves for an extended sensitivity analysis applied to all values of $\gamma_{s,,}$ for demonstrating purposes. The regularisation parameter γ_s imposes a constraint on the gradient (the spatial distribution) of the final solution of A. As higher values of γ_s are applied, the solution of A is required to adhere more strictly to the imposed prior, resulting in a more uniform distribution of A — since the prior imposed was that of a uniform A (see Figure S5). Whenever performing a ROC analysis, for the cases of total uniformity, until the A threshold is not surpassed, all crevasses remain undetected. Once the threshold is surpassed, the model accurately identifies all crevasses, but at the cost of misclassifying all the non-crevassed areas as crevassed. This, therefore, provides a ROC curve that is a straight diagonal line from the bottom-left (0-0) to the top-right (1-1), indicating that the model's performance is as good as random (AUC-mean = 0.5 Figure 4b), and providing a 100% TPR and 100% FPR (Figure 5 and S6). For the cases in the ROC analysis where we see an S-shaped curve, i.e., γ_s = [$1x10^7$ $5x10^8$], the solution of A has not yet reached a total uniformity (Figure S4 b and S5), changing just slightly throughout the domain. The s-shaped ROC curve represents a biphasic behavior which does not mean a better classifier, but rather a two-phase discriminative ability as a function of the threshold. Changes in the magnitude of A are minimal but enough to affect the ROC curve distribution. In fact, whenever performing the ROC analysis, the predictions and the misses are not 100% yet (Figure 5 and S6), thus the estimated curve takes the form of an s-shaped one. Since a uniform A is not likely or expected any simulation of γ_s larger than $1x10^7$ becomes irrelevant to use for this body of work, since the classifier is not capable to differentiate between crevasses and no crevasses effectively. The main results presented in this work adopted a γ_s = 25000, and these findings hold true for γ_s within the L-curve range.

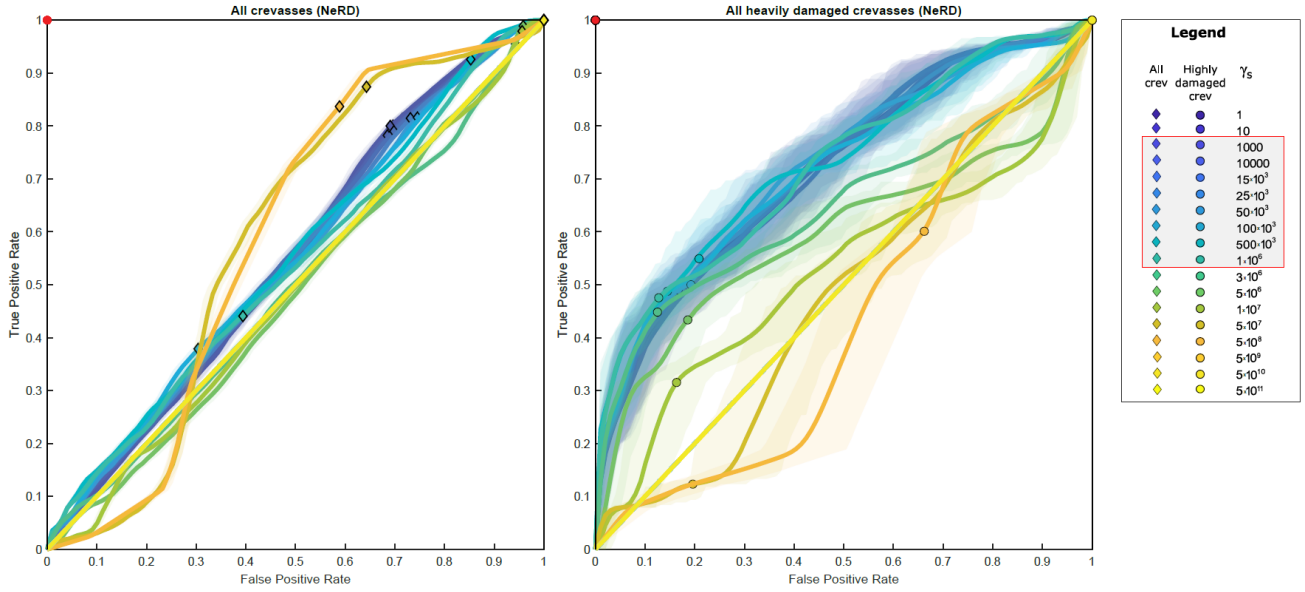

Figure S3: **ROC curve analysis for 2000 classification tests applied to Pine Island Ice Shelf, for all crevasses as mapped via NeRD (panel a) and for only heavily damaged crevasses (panel b), when testing for different regularisation values of γ_s and maintaining a constant γ_a=1. We observe a similar set of curves up until γ_s values ~ 10^6; between γ_s =10^7 and γ_s =5x10^8 the ROC curves take on an S-shaped form; for values of $\gamma_s > 10^9$, the relative ROC curves are a diagonal line, thus the performance is that of a random classifier.**

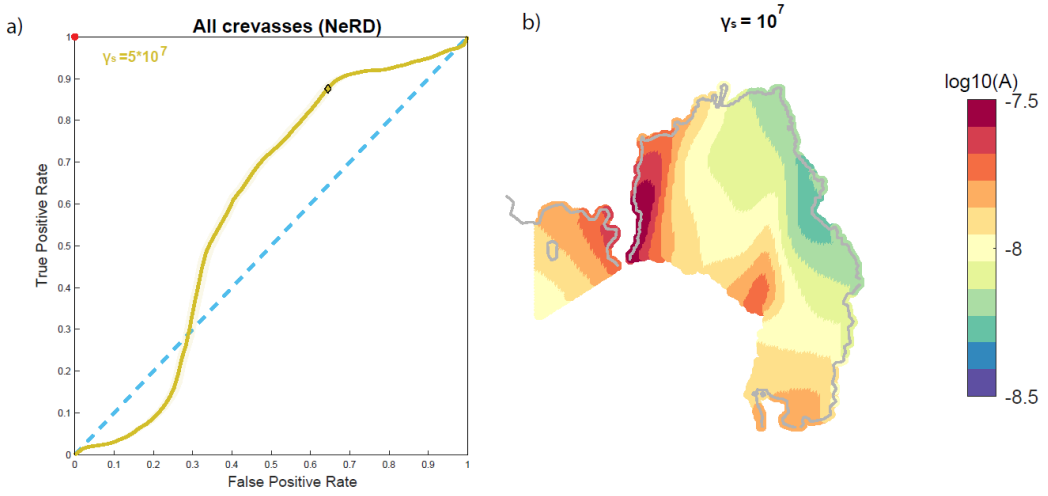

Figure S4: **a) ROC curve for Pine Island Ice Shelf (dark yellow line) when considering all crevasses and the solution of A obtained from an inversion using $\gamma_s = 10^7$. The diagonal dashed blue line represents a random classifier. b) Solution of A for $\gamma_s = 10^7$ displayed in log-scale; it is clear that A has not yet reached spatial uniformity; thus, when performing the ROC analysis, a s-shaped curve is obtained.**

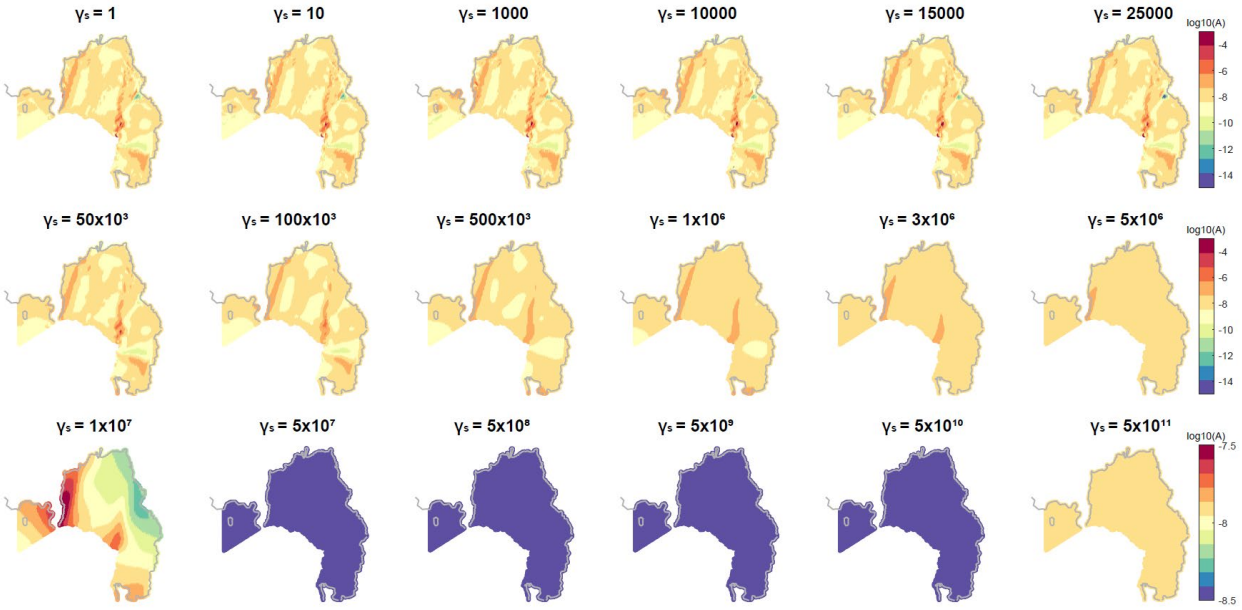

Figure S5: **Solution of A for Pine Island Ice Shelves represented in log scale for increasingly higher values of γ_s. Note that the last row of the figure adopts a different set of limits for the colour scale, to show the slight changes of A, until total uniformity is reached. By applying values of $\gamma_s > 5x10^9$, a complete spatial uniformity of A is present.**

S1.5 True Positive Rate and False positive rate for all crevasses and heavily damaged crevasses, for values of γ_s

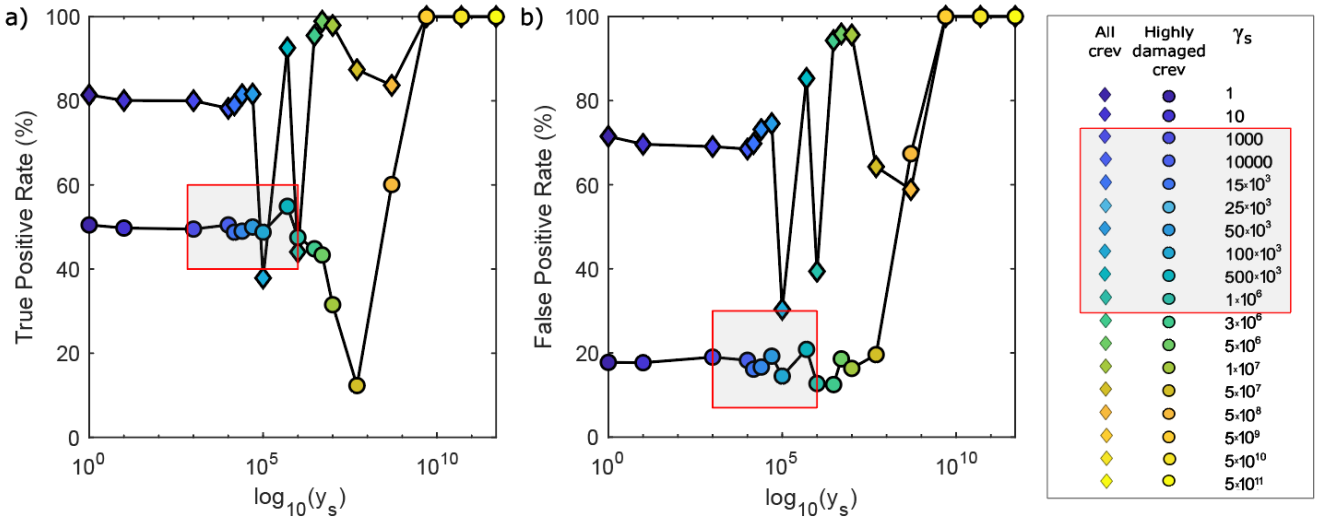

Figure S6 True Positive Rate (a) and False Positive Rate (b) for Pine Island Ice Shelf across varying regularisation values (γ_s), for all crevasses (diamonds) and heavily damaged crevasses (circles, as in Figure 5) for simulations that fitted velocities for February 2020. Within each panel, a red-outlined box filled with grey shading highlights the acceptable values for γ_s —within the L-corner range. As we increase the value of γ_s to very large values (> $5x10^9$), the solution of *A*, is spatially uniform and the TPR and FPR are both at 100%.

S1.6 Map of True Positive, False Positive and False Negative when applying the mean optimal *A-Threshold* estimated from the ROC curve.

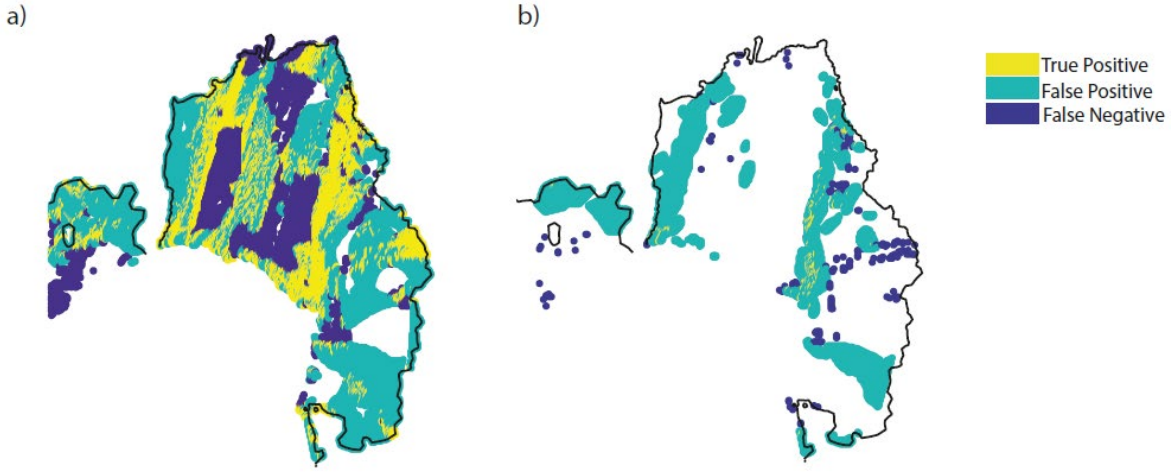

Figure S7 True Positives (yellow, crevasses that are correctly detected as crevasses), False Positive (green, crevasses that are incorrectly predicted as crevasses) and False Negatives (blue, crevasses that were missed) for Pine Island Ice Shelf, when applying the mean optimal *A-Threshold* estimated from the ROC curve analysis, for all crevasses (a), and for heavily damaged crevasses (b). Simulations inverted for *A* by fitting velocities for the month of February 2020, with a regularisation parameter of γ_a = 1 and γ_s=25000.

S1.7 Map of True Positive, False Positive and False Negative when applying the mean optimal *A-Threshold* estimated from the ROC curve for Filchner Ronne Ice Shelf

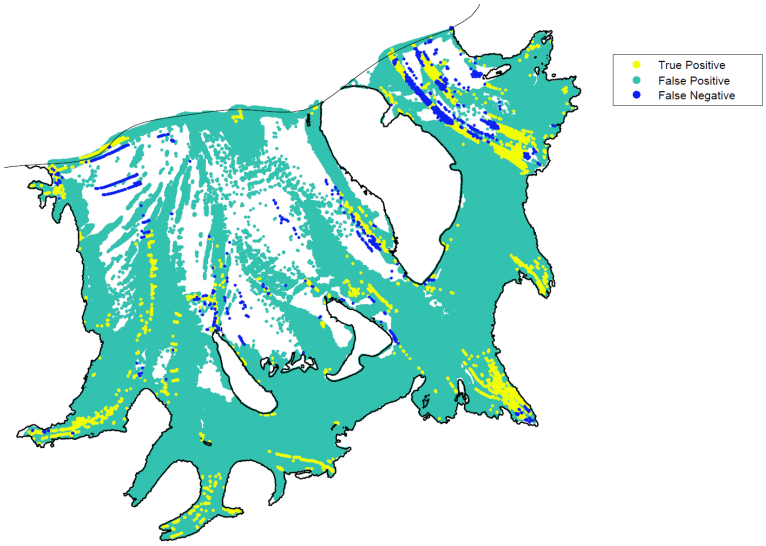

Figure S8 True Positives (yellow, crevasses that are correctly detected as crevasses), False Positive (green, crevasses that are incorrectly predicted as crevasses) and False Negatives (blue, crevasses that were missed) for Filchner Ronne Ice Shelf, when applying the mean optimal *A-Threshold* estimated from the ROC curve analysis, for all crevasses advected downstream for 5 years, to match velocities of 2014. Results are shown for a simulation with regularisation parameters of γ_a = 1 and γ_s=25000.

S1.8 Simulations fitting velocities for the month of November 2019 and February 2020 for Pine Island Ice Shelf for different values of γ_s

We conduct a comparative analysis evaluating the L-curve and ROC plot results for simulations whose ice rate factor fitted either the November 2019 (red line) or the February 2020 (black line) velocities observed on Pine Island Ice Shelf. Both sets of simulations display a similar L-curve, with ROC plots results being insensitive to the value of the regularisation parameter γ_s when within the L-corner range (grey-shaded box). The 2020 velocities provide a generally higher AUC-mean for heavily damaged crevasses in the L-corner range, compared to the 2019 velocities.

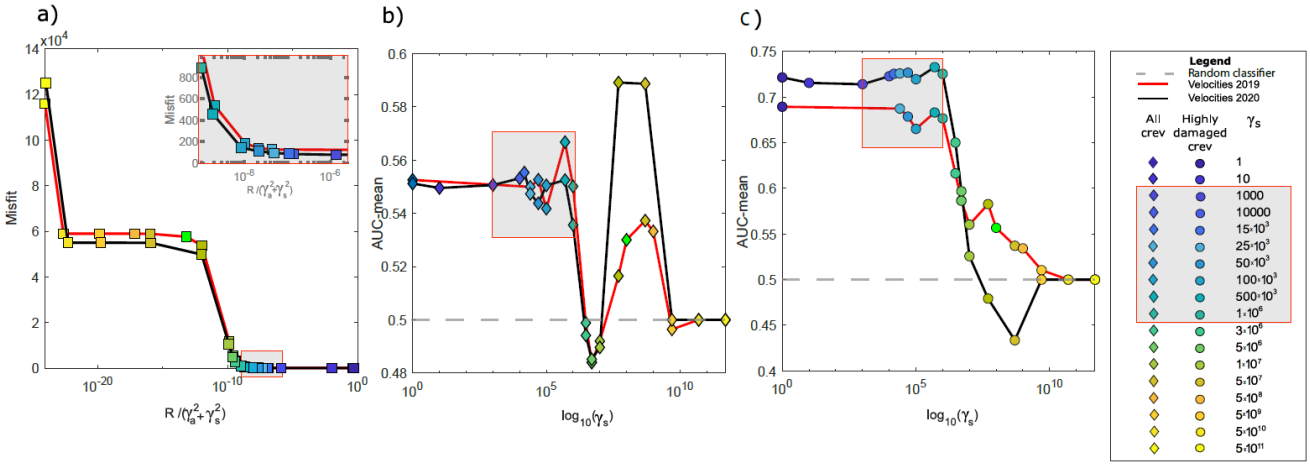

Figure S9 a) L-curve analysis comparing simulations when fitting velocities from November 2019 (red line) and February 2020 (black line). A zoomed-in plot of L-corner range is displayed as a shaded grey box, with red contouring. b) AUC-mean across varying regularisation values (γ_s) for all crevasses as depicted by the NeRD method for Pine Island Ice Shelf, for November 2019 and February 2020; Values within the L-corner range are highlighted by the grey-shaded box. c) AUC-mean across varying regularisation values (γ_s) for all heavily damaged crevasses as depicted by the NeRD method for Pine Island Ice Shelf, for November 2019 and February 2020; Values within the L-corner range are highlighted by the grey-shaded box.

S1.9 Sensitivity analysis to γ_a: comparison of ROC plots for a solution of A inferred with $\gamma_a = 1$ or $\gamma_a = 0$, and a constant $\gamma_s = 25000$, for Pine Island Ice Shelf, fitting 2019 Velocity Data.

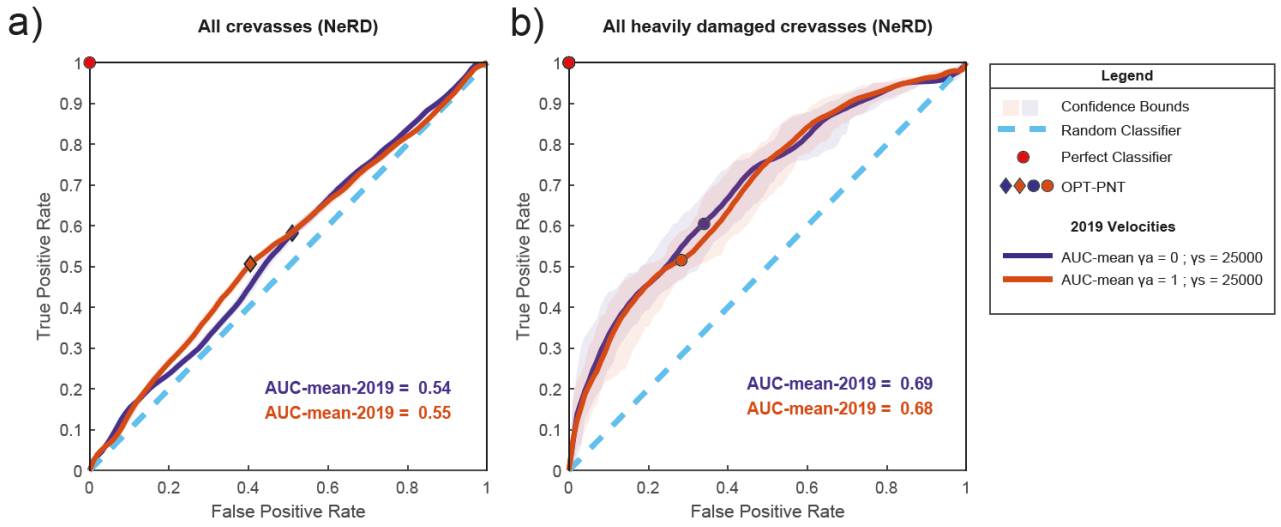

Figure S10: ROC curve analysis for 2000 classification tests applied to Pine Island Ice Shelf, for all crevasses as mapped via NeRD (a) and for only heavily damaged crevasses (b), fitting velocities of 2019. We test the sensitivity to changes in results when adopting $\gamma_a = 1$ (in orange) and $\gamma_a = 0$ (in blue) and a constant $\gamma_s = 25000$. Both simulations provide analogous AUC-means for both cases (all crevasses and heavily damaged crevasses) thus suggesting that the sensitivity to γ_a is negligible.

S2 Solution of A inferred with $\gamma_a = 1$ and $\gamma_s = 25000$ and velocity map for Pine Island Ice Shelf fitting 2019 and 2020 Velocity Data.

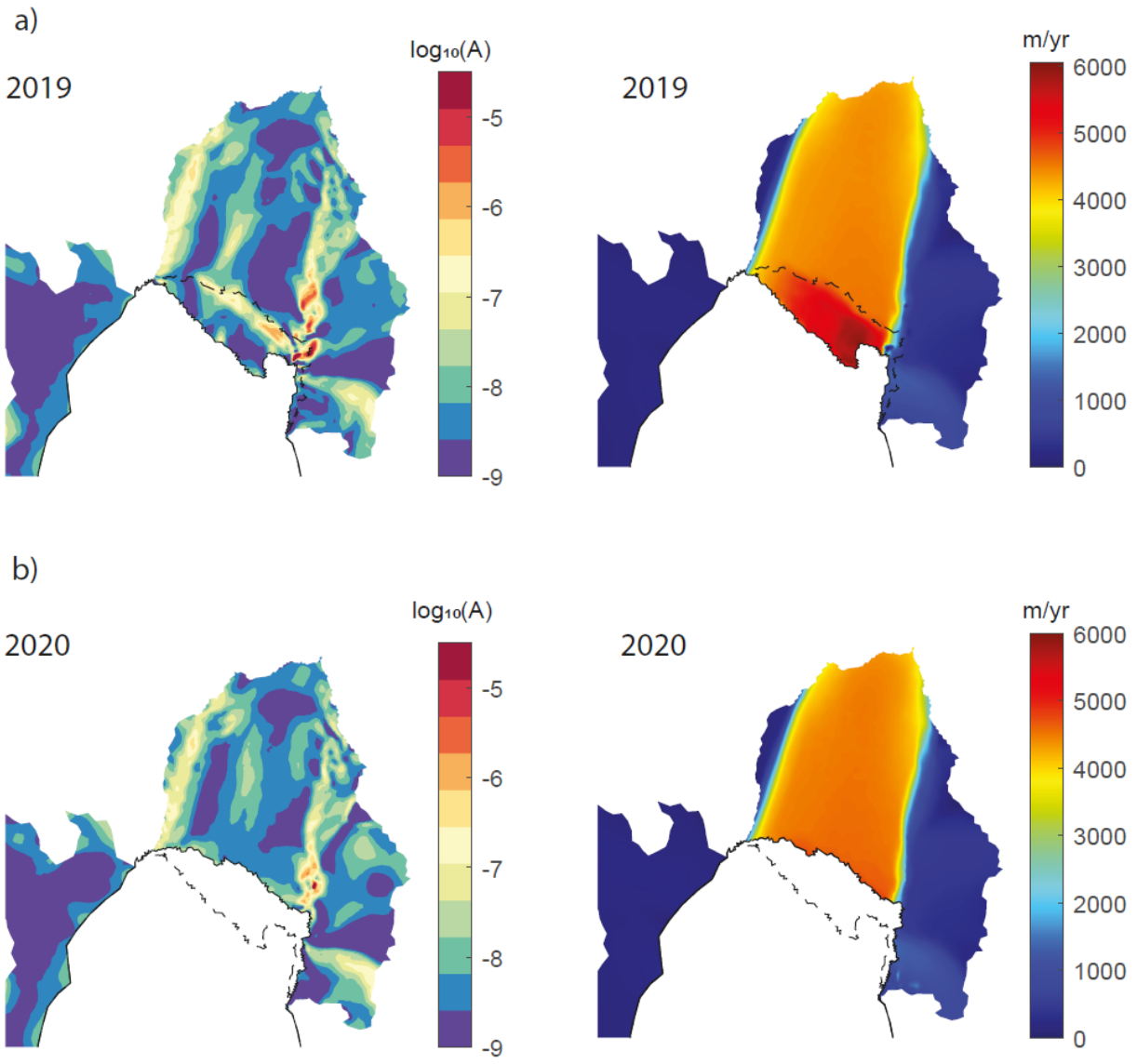

Figure S11: Inverted ice rate factor in log-scale (left-panels) and relative velocities (m/yr) for Pine Island ice shelf for November 2019 (a) and February 2020 (b). The boundary of each domain is reported as a uniform black line; dashed lines represent ice shelf front before/after calving event. Clearly visible in panel a) is the weakening region along the rift that a couple of months later would calve away a chunk of the ice shelf. Weakening is also a visible in both simulations along the ice shelf margins.

180

References

Lai, C. Y., Kingslake, J., Wearing, M. G., Chen, P. H. C., Gentine, P., Li, H., Spergel, J. J., and van Wessem, J. M.: Vulnerability of Antarctica's ice shelves to meltwater-driven fracture, Nature, 584, 574–578, https://doi.org/10.1038/s41586-020-2627-8, 2020.

185 Rizk, Y., Hajj, N., Mitri, N., and Awad, M.: Deep belief networks and cortical algorithms: A comparative study for supervised classification, Appl. Comput. Informatics, 15, 81–93, https://doi.org/10.1016/J.ACI.2018.01.004, 2019.