Supplementary Information for
A Comparison of Imagery Sources for Automated Detection of Retrogressive Thaw Slumps

Heidi Rodenhizer[1], Yili Yang[1], Greg Fiske[1], Stefano Potter[1], Tiffany Windholz[1], Andrew Mullen[1], Jennifer Watts[1], Brendan M. Rogers[1]
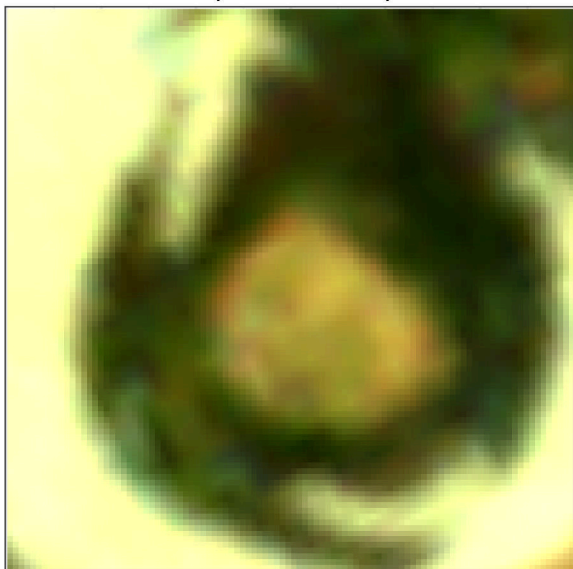
[1]Woodwell Climate Research Center, Falmouth, MA

Corresponding author: hrodenhizer@woodwellclimate.org
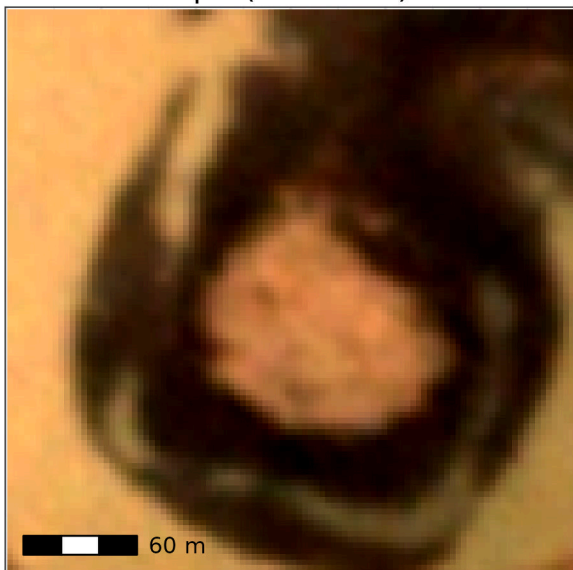
Contents:
Figures S1 to S9

a. Sentinel-2 (Reference)

b. Raw PlanetScope

c. PlanetScope (AROSICS)

60 m

d. PlanetScope (AROSICS + IRMAD)

Figure S1. An example of AROSICS geometric and IRMAD radiometric calibration of a PlanetScope image. The Sentinel-2 reference image is shown in a., the raw PlanetScope image in b., the PlanetScope image after geometric calibration in c., and the PlanetScope image after both geometric and radiometric calibration in d. In this example, this portion of the PlanetScope image was shifted slightly to the East relative to the reference, and the entire image was considerably more red than the reference.
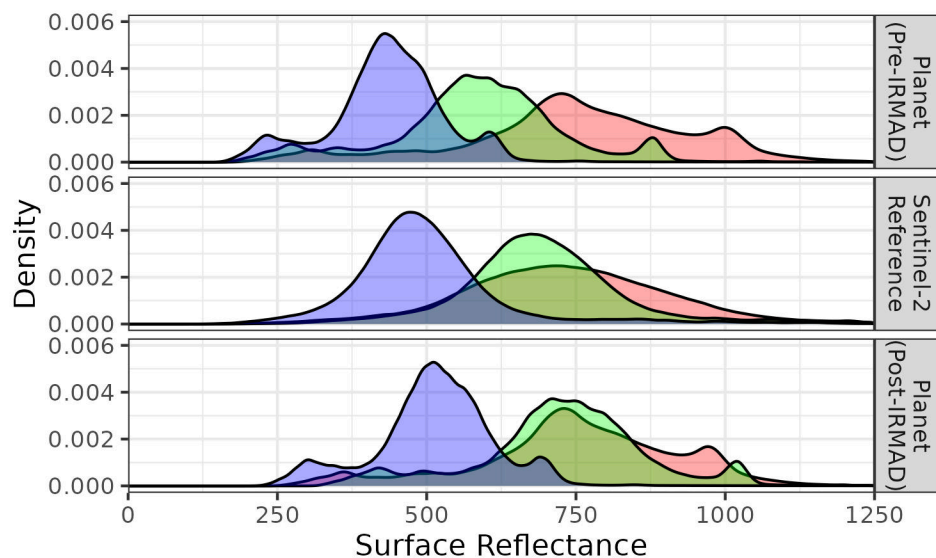
Figure S2. Surface reflectance histograms of an example PlanetScope image before and after IRMAD radiometric calibration compared to the Sentinel-2 reference image.
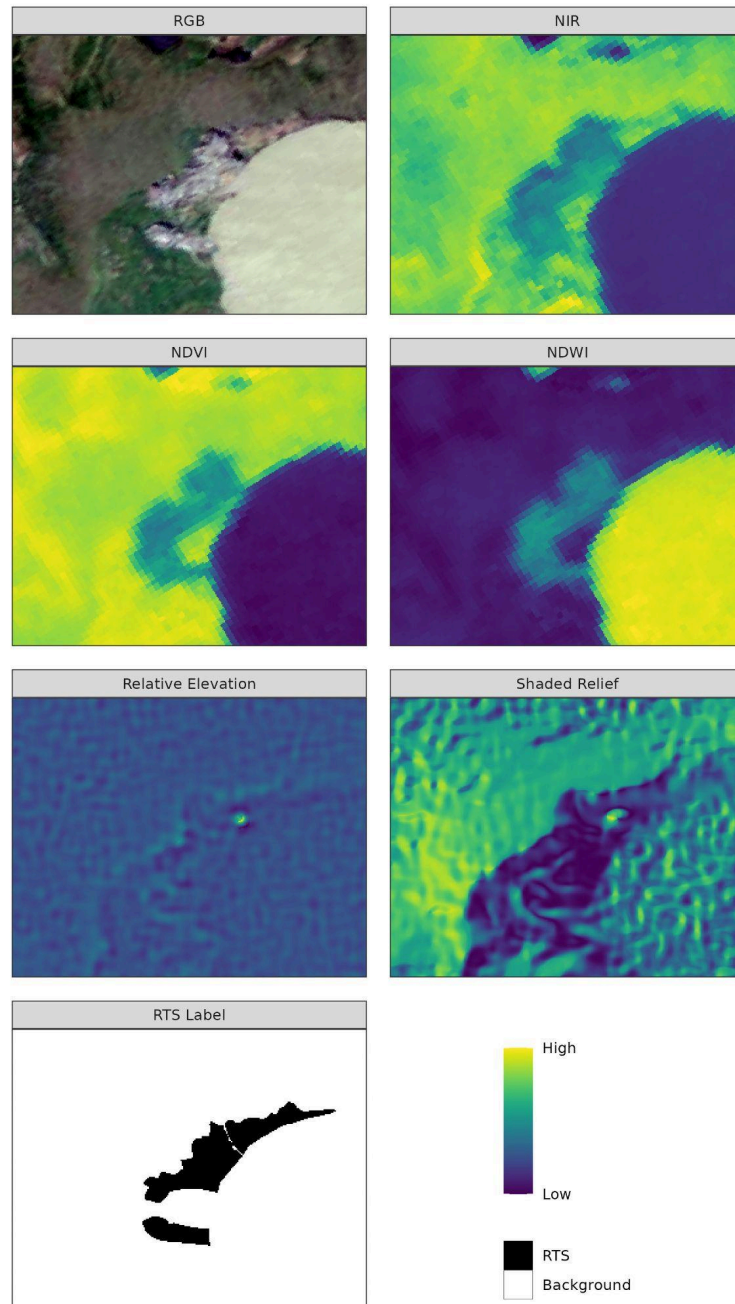
Figure S3. Input imagery bands to the Unet3+ model training. The three visible bands (RGB) were sourced from either WorldView, PlanetScope, or Sentinel-2. Near Infrared (NIR) was sourced from either PlanetScope or Sentinel-2 imagery. Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI) were sourced from Sentinel-2. Relative Elevation and Shaded Relief were sourced from the ArcticDEM. Finally, a binary layer with the ground truth RTS label was included for model training.
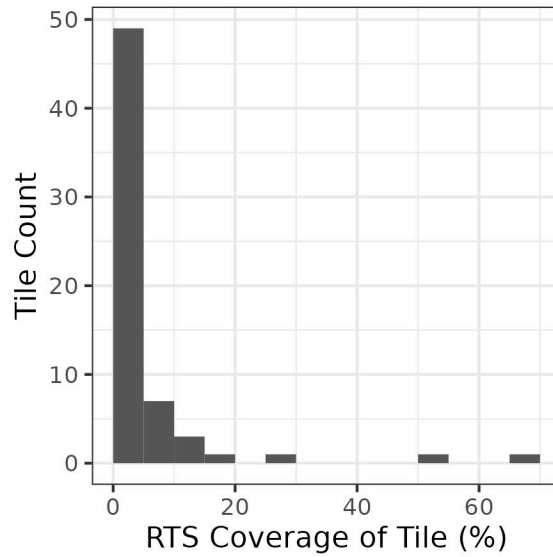
Figure S4. Percent cover of RTS pixels across the 63 testing tiles used in this study. RTS pixels comprised less than 10% of total pixels in over 85% of the tiles.
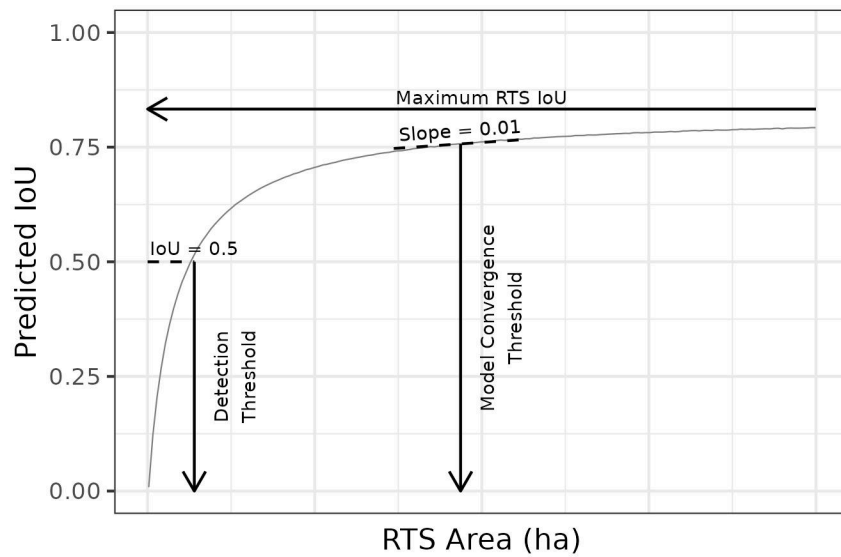

Figure S5. Conceptual diagram of the metrics used to describe the performance of the deep learning RTS models. The maximum RTS IoU is equal to the parameter A in the fit asymptotic curves. The detection threshold is the RTS area at which IoU is predicted to be 0.5. The model convergence threshold is the RTS area at which the slope is 0.01, indicating that beyond this area IoU will increase by less than 0.01 per hectare.
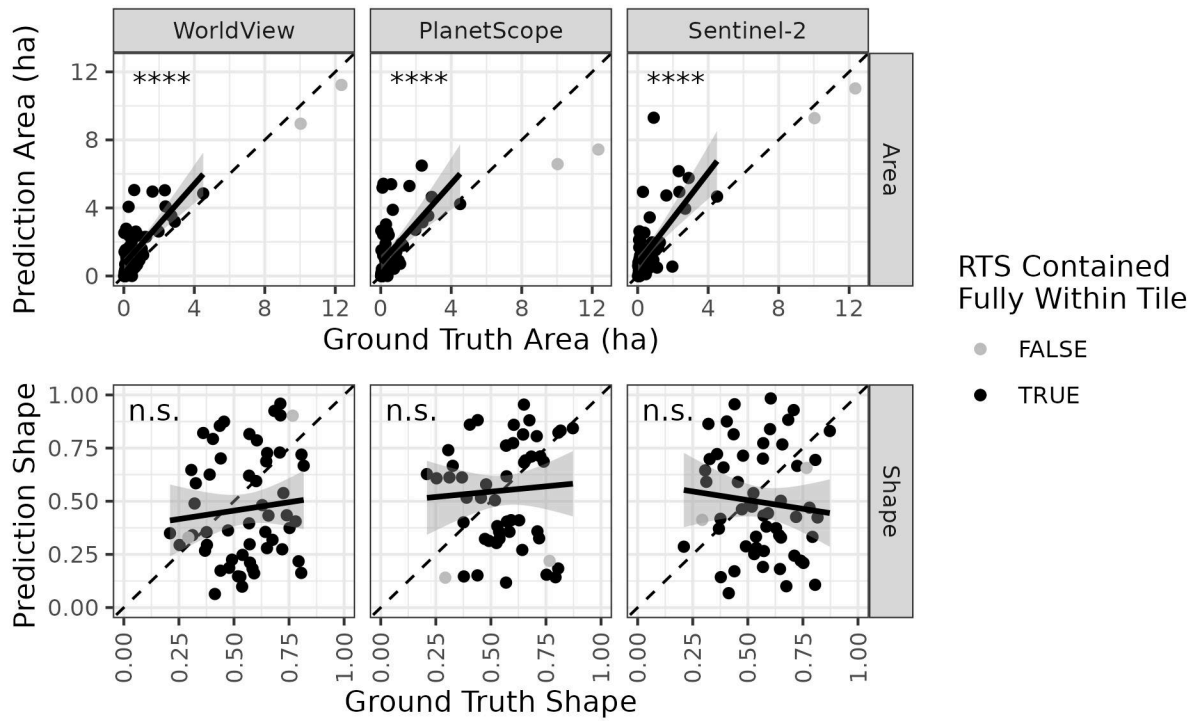
Figure S6. Linear regressions between metrics of the ground truth polygons and prediction polygons. RTS area is shown in the top row and RTS shape (Polsby-Popper test) is shown in the bottom row. Two RTS features that extended beyond the bounding box of the tiles they were located (gray) in have been excluded from the regressions. Statistical significance indicated with asterisks (not significant: 'n.s.', p-value < 0.0001: '****').
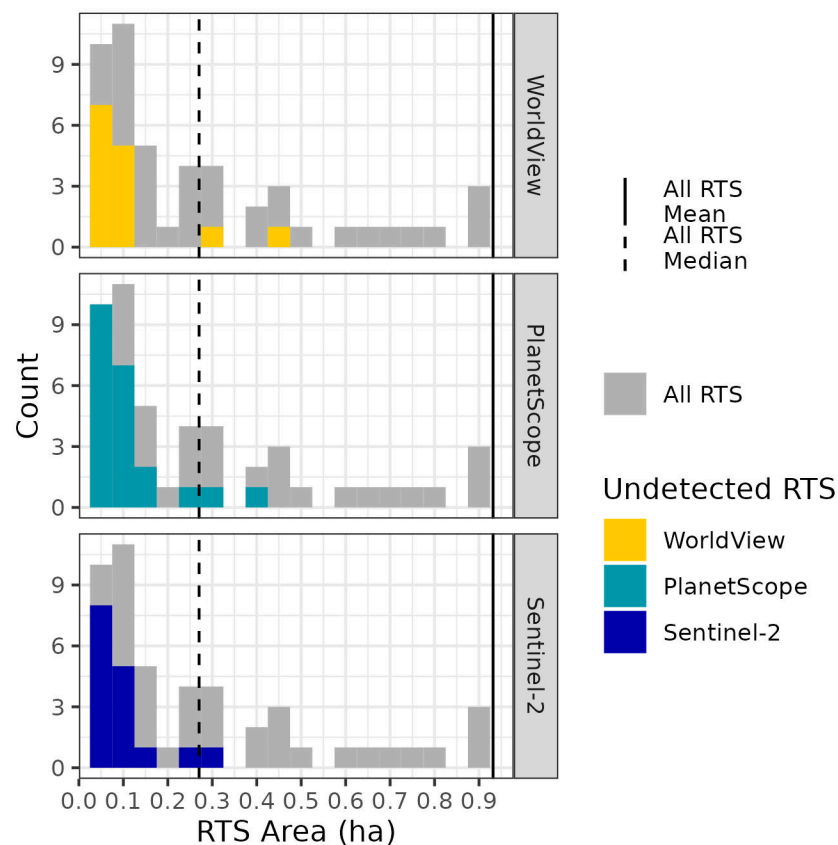
Figure S7. Count of undetected RTS by RTS area compared to the total number of RTS. The mean and median area of all RTS are shown as solid and dashed lines.
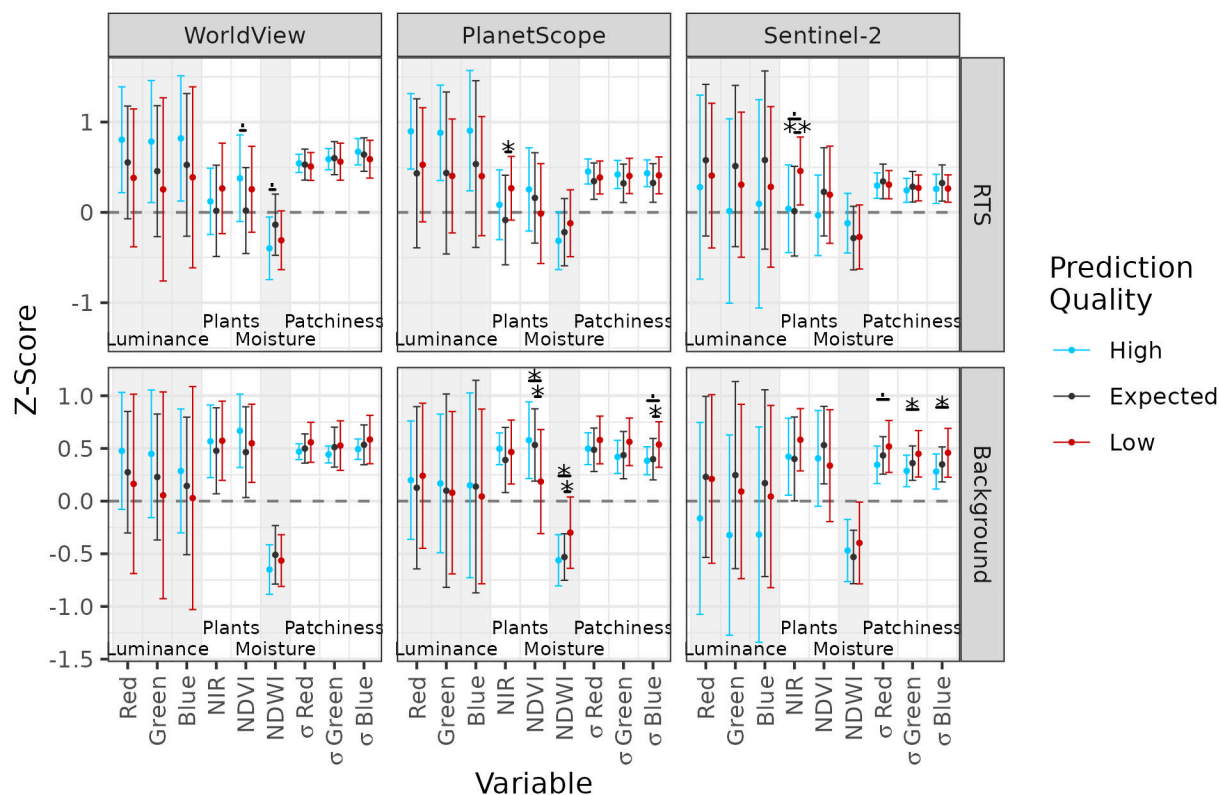
Figure S8. Average input data values in RTS and non-water background (BG) pixels across prediction quality classes. The points were calculated by first taking the mean pixel value (z-score) within each RTS class on a tile-by-tile basis and then averaging this value across all 63 testing tiles. The error bars show the standard deviation across tiles. Z-scores were calculated using all pixel values, including water pixels. Relative elevation and shaded relief are not included, as there were no discernable patterns across classes. Statistically different groups are indicated with lines between the two classes and a label for the level of significance (p < 0.1: '.', p < 0.05: '*').
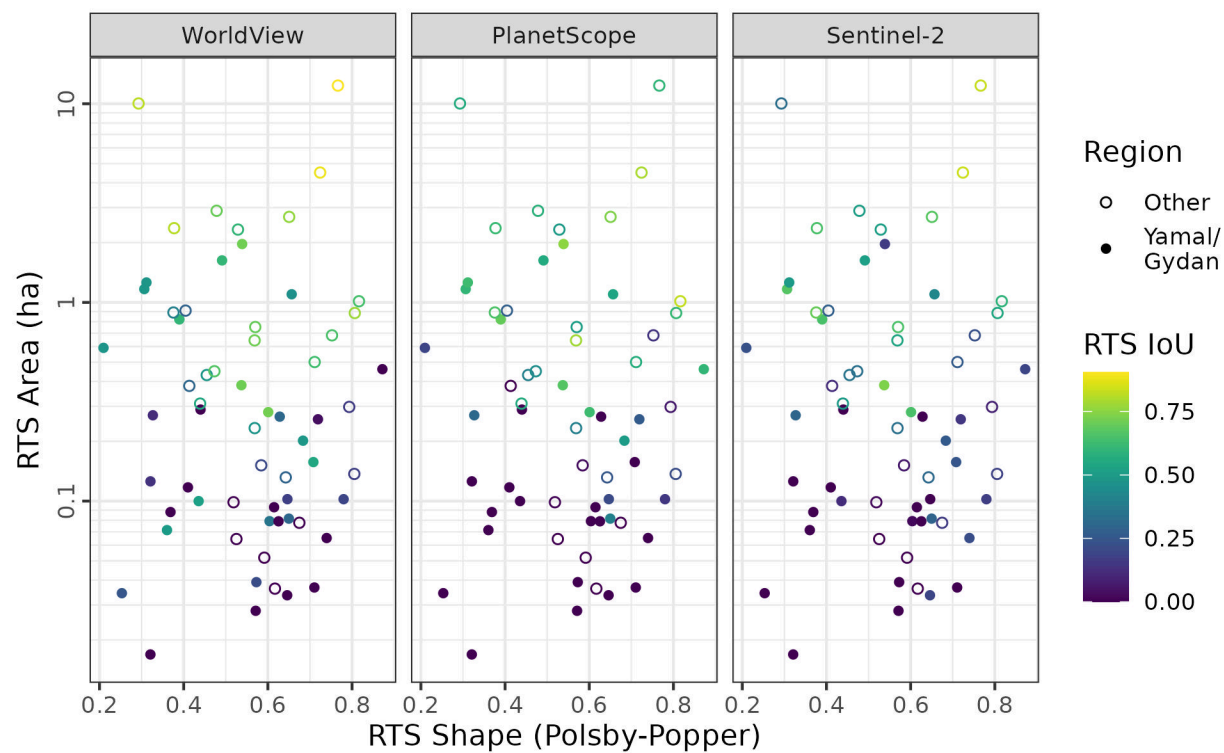
Figure S9. A scatterplot of RTS shape and RTS area. Each point represents one RTS feature. The prediction IoU of the RTS class is indicated with the point color. Solid points indicate features in the Yamal/Gydan region and open points indicate all other regions.