The authors present a methodology to create a benchmark dataset of supraglacial hydrology across the Antarctic. Their approach takes advantage of the large amount of multi-spectral satellite imagery provided by Landsat-8 and Sentinel-2. They propose a pipeline consisting of a cascade of spectral thresholding for the creation of a binary mask and subsequent K-means clustering to further distinguish lakes and channels.

The manuscript is clear and the exposition is easy to follow and linear.

I have only one specific comment related to the validation approach that was used, particularly thinking about future use by researchers in the area of machine learning.

As stated by the authors already in the abstract, the dataset provides a partition of the supraglacial features into non-water, lakes and streams. However, from my understanding in Table 1 the authors are showing aggregated results (binary water/non-water) for the accuracy and sensitivity metrics. I think that for the future users to benefit the most from the proposed inventory, it would be beneficial to show a validation considering all the classes.