OUTLINE (for an original research paper) – Tony Cannistra

7 January 2020

**1.** General info

**A**. title: Assessing High-Resolution CubeSat Imagery and Machine Learning for Detailed, High Resolution Snow-Covered Area

**B**. coauthors: Nicoleta Cristea, [Annie Burgess?]

**C**. journal (1st choice): Remote Sensing of Environment

**D**. journal (backup):Environmental Modelling & Software

**2.** The overarching question of this paper is: “Can an emerging satellite-based remote sensing dataset be used to observe snow-covered-area at ecologically-relevant spatial and temporal scales?”

**3.** Which is important / interesting / unresolved because (*1-4 reasons*)

**A**. A particular characteristic of this dataset renders traditional “snow detection” techniques unusable.

**B**. As such, we must borrow well-developed analytical techniques from another domain to overcome this challenge, which introduces a novel method into the snow hydrology research area.

**C**. There is no remote sensing dataset of snow that has 1) 3m spatial scale 2) Daily temporal scale 3) global geographic coverage –– this approach does

**D**. A daily, high resolution snow dataset has wide-ranging applications, from enhancing the accuracy of phenological studies to filling in temporal “observation gaps” in other, lower-resolution satellite products.

**4.** To answer this question / explore this topic, I addressed the following objectives: (*NB you can have more or less than 3 objectives, but I recommend 2-4*)

**A**. Create a method, borrowing from machine learning / computer vision, to discriminate snow in this emerging remote sensing dataset.

**B**. Compare the snow-discrimination skill of this method to other “gold standard” snow-covered-area datasets and evaluate relative performance.

**C**. Assess whether methods “familiar” with one geographic area are able to “transfer” their skill to another geographic area (to assess the generalizability of the method).

**5.** I addressed these objectives: (*use lists/bullet points below*)

**A**. in the Tuolumne Basin, California and Grand Mesa, Colorado.

**B**. with the following focal/model species/model system: n/a (*yet, though planned for another publication*)

**C**. and the following approaches:

* **Method:** an “image segmentation neural network” – a machine learning-based technique to classify individual pixels in an image.
* **Data:** this method initially requires “labeled data” to create a representation of what “snow” looks like in these satellite images. “Labeled data” are remote-sensing data which have snow-pixels labeled from a third party source to serve as examples for the model to “learn” from. Once “trained,” these third-party data are no longer required. We use the Airborne Snow Observatory (ASO) snow depth data in select watersheds in California and Colorado as this third-party data source for labels.
* **Evaluation:**
  + We compare the results from our method to standard, lower-resolution snow covered area datasets that are publicly available and “state of the art.” These comparisons assess: 1) the geographic match (or mismatch) of identified snow-covered area, 2) the “classification performance” (e.g. pixel-level accuracy) of snow classification, and potentially other metrics. The other datasets compared against include: ASO, Landsat 8 fSCA, Sentinel 2 NDSI, and MODIS fSCA. This evaluation scheme is slightly more complicated than will fit in an outline, but this is the gist.
  + We also evaluate the model’s ability to “transfer its knowledge.” We do this by “training” the model (e.g., presenting it with images where we know where the snow is) in one “snowpack type” (e.g. California), then evaluating the model in another snowpack type (e.g. Colorado).

**6.** I found: (*one sentence description of main results, usually no more than 4-5 points, less ok. Main results are those you find most important/compelling, that you will discuss extensively in the discussion)*

A. The method I developed in this study identifies snow-covered area in this emerging remote sensing dataset with high accuracy.

B. Compared to other snow-covered area datasets, my method performs as well at identifying snow in individual images (if not better!), and has the added advantage of having a higher spatial and temporal resolution than any other dataset compared against.

C. The method demonstrates promise in its ability to transfer across snowpack types, performing well in regions outside of its trained domain.

D. I was able to develop this method as an open source, cloud-based software package which is freely available.

**7.** These results are interesting (or important or unexpected) because (or suggest the following critical point): (*briefly describe why each finding(s) is interesting/important/surprising, no more than one sentence per finding*)

A. This finding represents the introduction of a novel, observation-first approach to snow covered area mapping at spatial and temporal resolutions unmatched by other satellite remote sensing datasets.

B. The good performance of this method relative to the “state of the art” methods suggests that my approach is a suitable replacement for other snow discrimination techniques.

C. The transferability of the models we’ve trained is important in their general applicability, and presents a path forward for research in better understanding the ways in which snowpacks differ from a remote-sensing perspective.

D. The open-source nature of this work and (accompanying educational resources) allow other researchers to freely use and modify these techniques for their own work.