HOKKAIDO UNIVERSITY

Modelling Arctic Coastal Plain Lake Depths Using Machine Learning and Google Earth Engine

Ram Avtar, Ph.D.

Associate Professor

Graduate School of Environmental Science

Hokkaido University, Japan

ram@ees.hokudai.ac.jp

Motivation

Introduction

- Numerous shallow thermokarst lakes in the Arctic coastal plains recently shows a change in lake abundance and area owing to global warming.
- Lakes and ponds can occupy more than 20%–40% of the landscape in Arctic lowland regions (Grosse et al., 2013; Muster, et al., 2017)
- IPCC and other recent studies reported that the high latitudes warming is much higher than the global annual average warming
- Grosse et al. 2013 estimated that more than half of the lakes found in permafrost regions are likely of thermokarst origin

https://news.uaf.edu/arctic_lakes_july2014/

Introduction

- These lakes undergo various phases of formation, growth, drainage, and reformation (Jone et al., 2015)

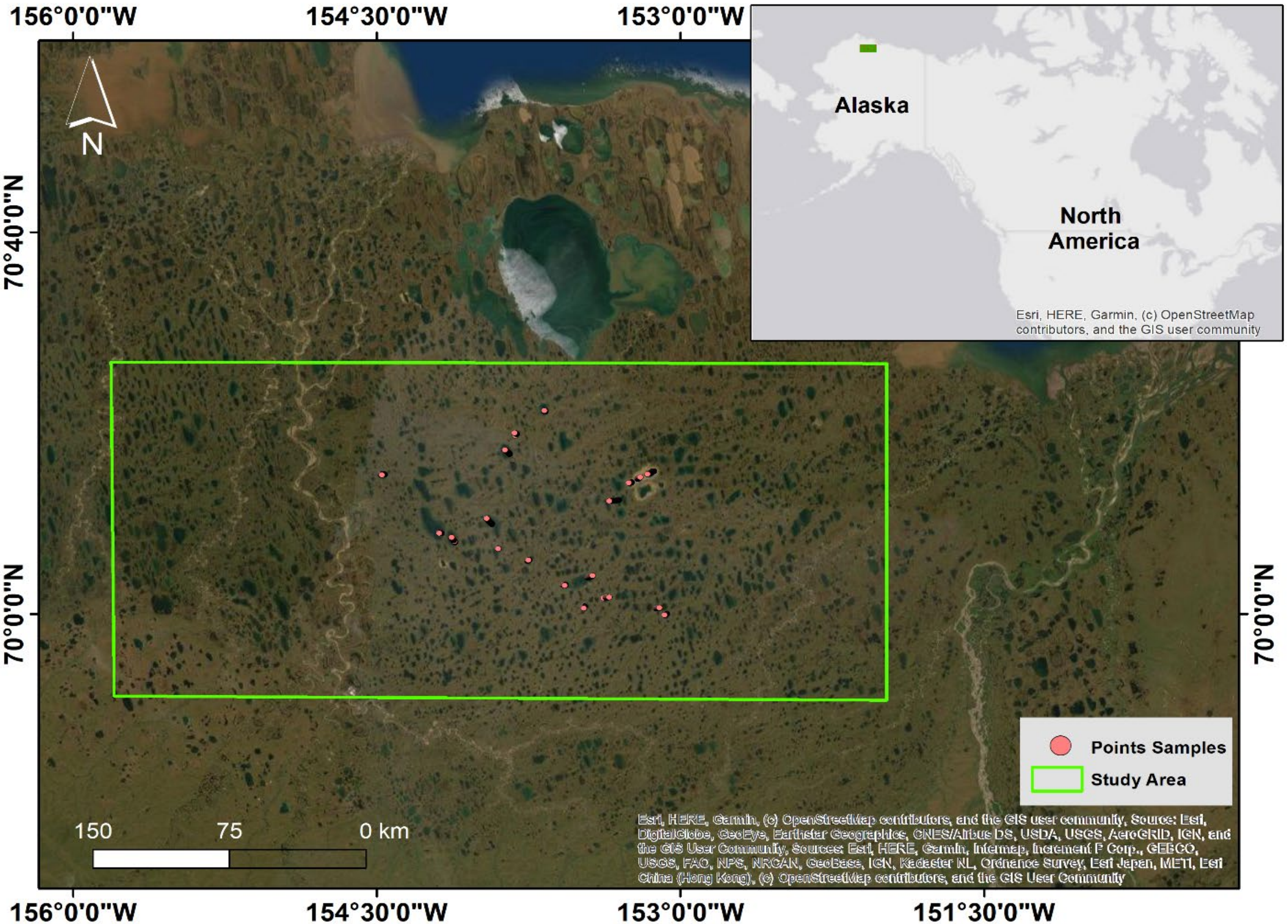
- With a rapidly warming in Arctic there is dramatic changes in lake hydrology, lake ice characteristics, and permafrost degradation (Liljedahl et al., 2016)

- Therefore, monitoring of lake-rich Arctic regions at high temporal and spatial resolution is crucial for understanding their response to climate change (Paltan et al., 2015)

- To better understand the degradation of lake bodies due to the climate change in the Arctic environment, a comprehensive lake water volume analysis over multiple time period and a larger spatial scale is necessary.

- However, no quantitative bathymetric studies in a larger scale has been attempted in the region primarily owing to the logistic issues, and harsh climatic environment.

Study Location



Data & Methods

- Simpson and Arp, (2017) measured the discrete depth points of 19 lakes in the study area using a HumminBird 798ci HD SONAR instrument mounted on a floating plane during the summer of 2017.

- The dataset is published as an open inventory and can be retrieved from Arctic Data Center portal (https://arcticdata.io/).

- To adequately represent the depth samples, they selected the lakes which are larger than 1 km^2 in surface area. The minimum measured depth value is 0 m, maximum is 21.3 m, mean is 5.26 m and standard deviation is 4.77 m.

Regional-scale inventories of lake bathymetric mapping in inaccessible terrains is possible with the help of satellite remote sensing images (based on Reflectance)

- We used Landsat-2017 (corresponding to the field dates) to model the bathymetry for the study area.
- As some of the Landsat-8 OLI images of July to September are cloudy and partial cloudy, we used the 'median' image of Landsat surface reflectance collection (ee.ImageCollection('LANDSAT/LC08/C01/T1_SR')) acquired between July and September of 2017.
- Landsat SR product is the atmospherically corrected surface reflectance from the Landsat 8 OLI/TIRS sensors having contain 5 visible and near-infrared (VNIR) bands, 2 short-wave infrared (SWIR) bands, and two thermal infrared (TIR) bands. The clouds in the image scenes have been masked based on the 'pixel_qa' band of Landsat 8 SR data.

BATHYMETRY RF_with_factor

```
Imports (4 entries)
 var table: Table users/yunuscool/LAKE_BOUNDARY
 var table2: Table users/yunuscool/TRAIN_Bathy
 var table3: Table users/yunuscool/ALASKA_2014_2019_Bound
 var geometry: Polygon, 4 vertices

function maskL8sr(image) {
  // Bits 3 and 5 are cloud shadow and cloud, respectively.
  var cloudShadowBitMask = 1 << 3;
  var cloudsBitMask = 1 << 5;

  // Get the pixel QA band.
  var qa = image.select('pixel_qa');

  // Both flags should be set to zero, indicating clear conditions.
  var mask = qa.bitwiseAnd(cloudShadowBitMask).eq(0)
      .and(qa.bitwiseAnd(cloudsBitMask).eq(0));

  // Return the masked image, scaled to reflectance, without the QA bands.
  return image.updateMask(mask).divide(10000)
      //.select("B[0-9]*")
      .copyProperties(image, ["system:time_start"]);
}

// Make a cloud-free Landsat 8 TOA composite (from raw imagery).
var l8 = ee.ImageCollection('LANDSAT/LC08/C01/T1_SR')
          .filterDate('2017-07-01', '2017-9-1')
 .map(maskL8sr)
.filterBounds(table)

var image = l8.mean();

var ratio = image.select('B2').divide(image.select('B3')).rename(['ratio']);
var withratio = image.addBands(ratio);
print(withratio);

// Use these bands for prediction.
var bands = ['B1', 'B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'B10', 'B11', 'ratio'];

// Load training points. The numeric property 'class' stores known labels.
var points = ee.FeatureCollection(table2);

// This property stores the land cover labels as consecutive
```

Machine Learning Models

- CART introduced by Breiman refer to decision tree (DT) algorithms that are useful for classification or regression predictive modeling problems.

- RF is a supervised machine learning based on the decision trees but formed by many number of individual DT's that function as an ensemble

- The advantage of RF over CART is that the former decrease the variance of the resulting output, and are less prone to overfitting, thus gives an robust results.

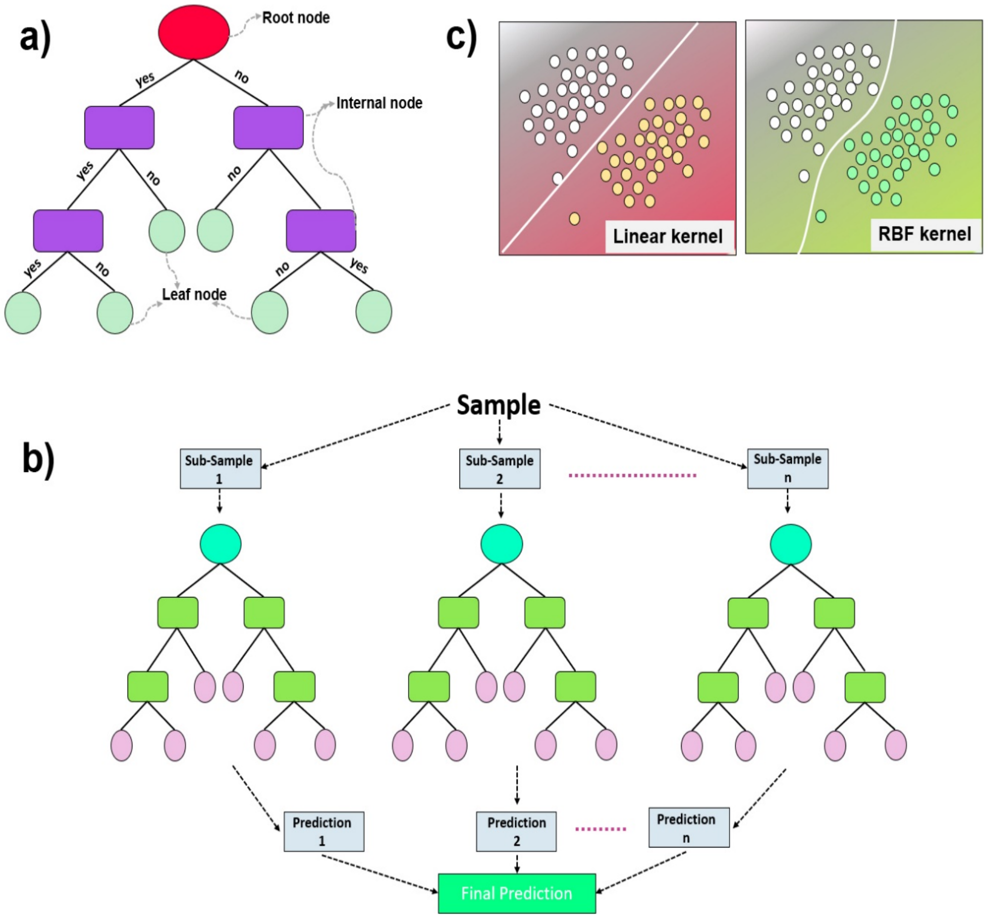


Fig 2. Schematics of ML models employed in this study: (a) classification and regression tree, (b) random forest, and (c) support vector machines with different kernels.

Results

Decision Trees

Random Forest

Support Vector Machines

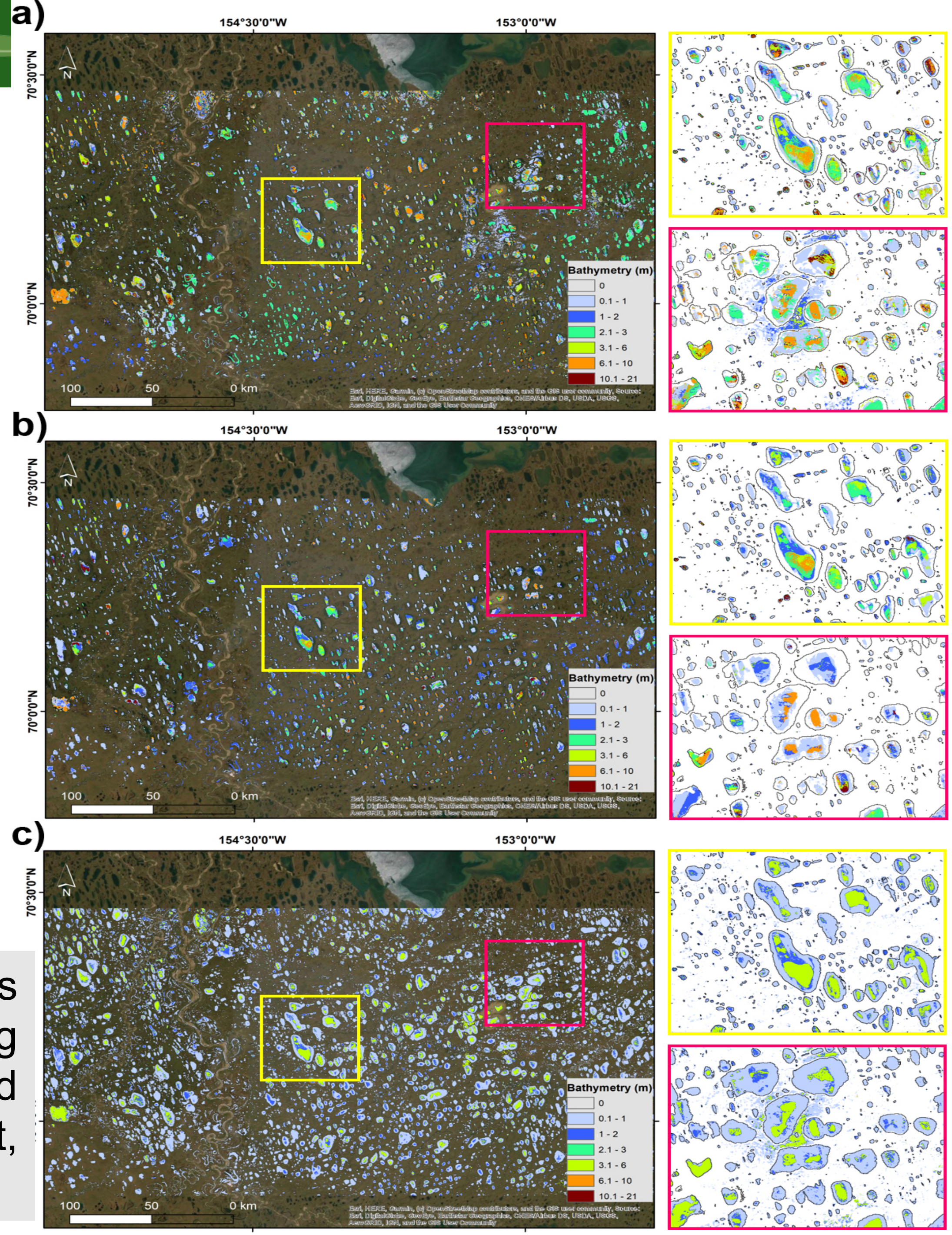


Fig 3. Lake bathymetric maps derived from three machine learning models (a) classification and regression trees, (b) random forest, and (c) support vector machines.

Accuracy Assessment

Evaluation Matrix	Decision Tree	Random Forest	Support Vector Machines
ACC	0.88	0.88	0.32
MAE	0.55	0.53	3.12
RMSE	0.9	0.86	5.22
R^2	0.977	0.979	0.21

The hold out data (5734) of the depth points (validation data) was used to prepare the ACC, MAE, RMSE and R^2.

- RF has the lowest error (RMSE = 0.86), while the SVM method again showed the lowest performance (RMSE = 5.22). In terms of Accuracy, both CART and RF shows a similar values of ACC (88%), whereas ACC for SVM is rated the poorest (32%).

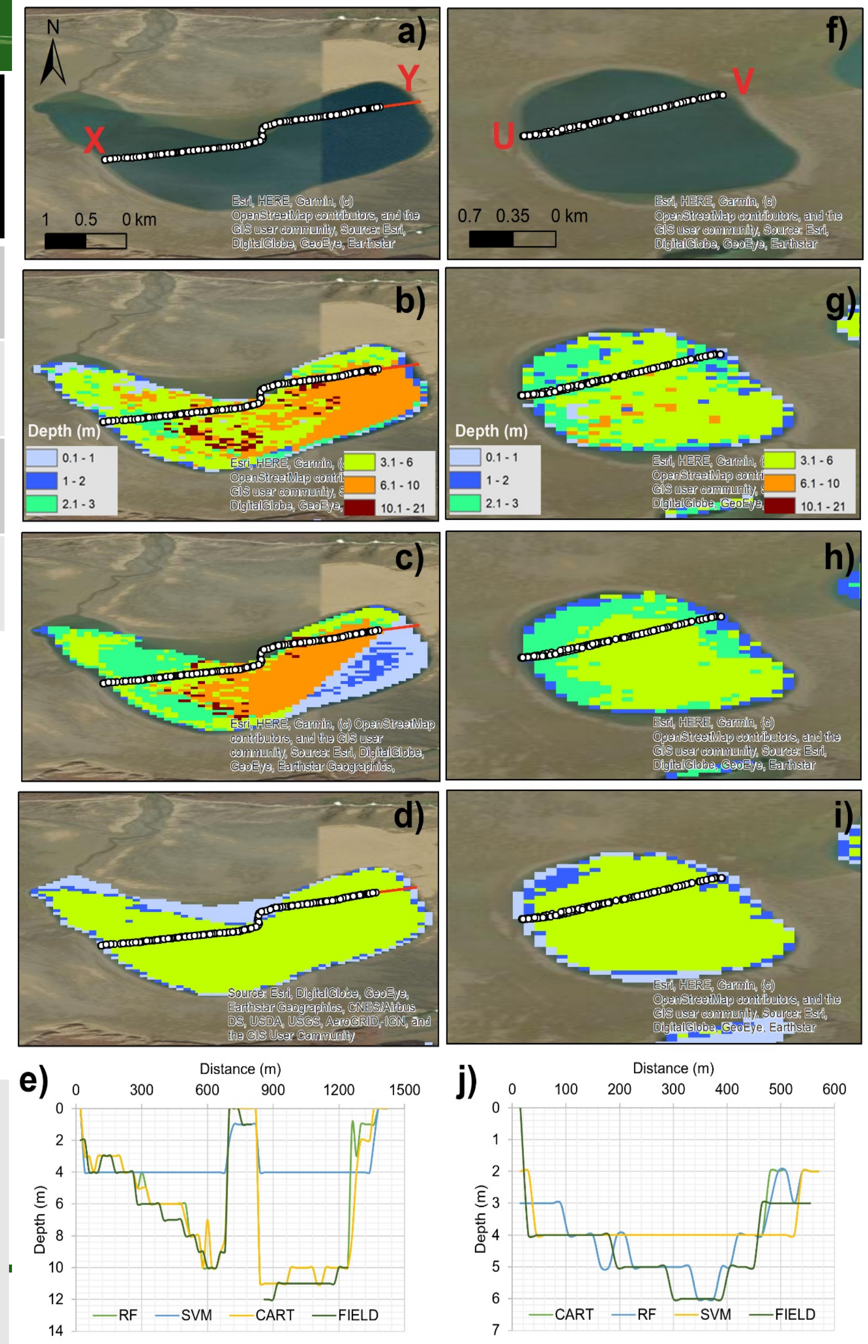


Figure 5. Bathymetric profiles of some studied Arctic lakes: (a) and (f) are the locations of field samples for X-Y and U-V profiles; (b) and (g) are CART derived bathymetric maps, (c) and (h) are RF derived bathymetric maps, (d) and (i) are SVM derived bathymetric maps for the two lakes shown in (a) and (f); (e) and (j) are corresponding profile plots for X-Y and U-V profiles respectively.

Future Work

- One of the potential application of satellite derived bathymetry is that we can model the bathymetry for any time period as long as cloud free images are available.
- Figure presents the comparative analysis of bathymetric maps for 2017 and 2016 derived from the random forest model. It can be seen that there is a considerable amount of under estimation in the depth estimation when applying the models trained from 2017 to 2016 despite having used the same time period images and model.
- We are exploring the possibility to overcome this limitation by applying more effective atmospheric correction algorithms.
- We will apply this algorithm to other areas of Arctic.

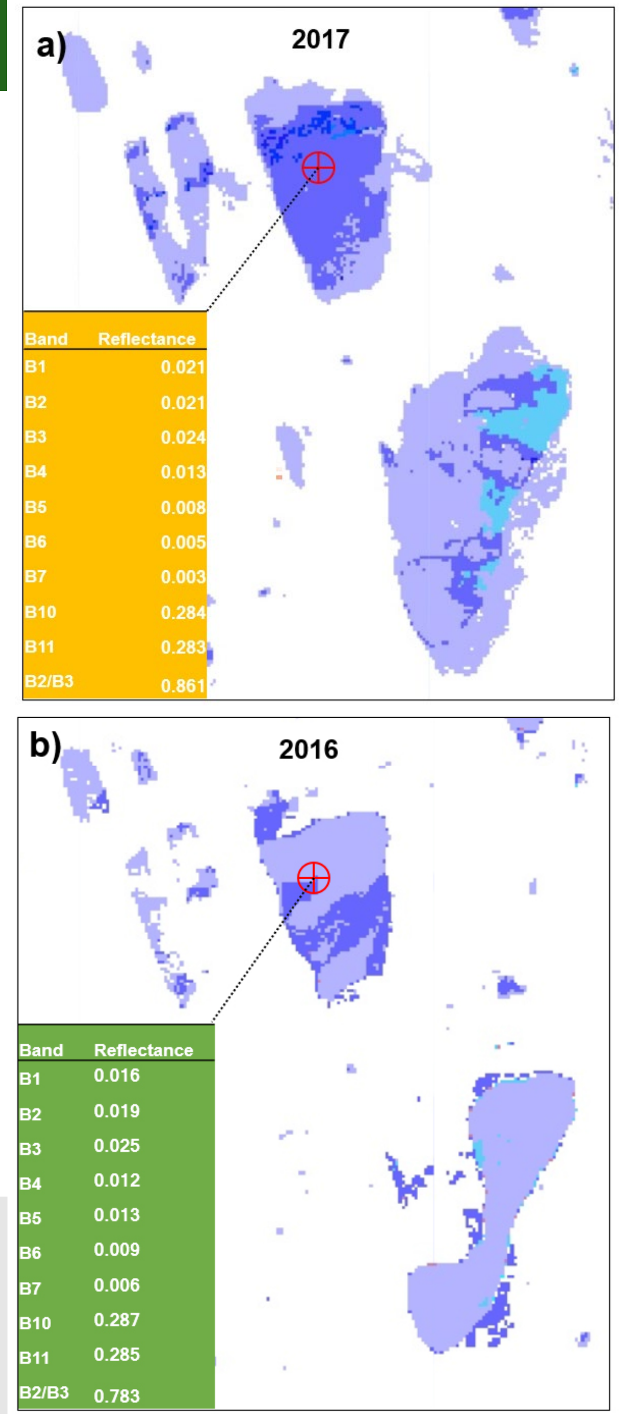


Figure 5. Applying the trained data from 2017 bathymetric model to 2016 images: (a) bathymetric map of 2017, (b) bathymetric map of 2016. Note that the reflectance values of same location (red cross) is different for both years.

Conclusions

- High-quality bathymetric data of water bodies in climate sensitive areas are important for understanding glacier mass balance, water storage and the pace of warming induced degradation.

- In this study, we employed in situ data from Simpson and Arp [26] to derive the bathymetric information of hundreds of thermokarst lakes in Arctic coastal plains in the Northern Alaska. Landsat 8-OLI surface reflectance product of 2017

- The best method for bathymetric retrieval from OLI images was RF method, which out classed both CART and SVM

- Therefore, RF method and their implementation in GEE is recommended for future lake depth retrievals in shallow coastal lakes, especially for large scale analysis.

THANK YOU

for your kind attention