Google earth engine application for estimating changes in water surface area of Lake Toba

F Aziz, E Kusratmoko, M D M Manessa
Department of Geography, University of Indonesia 16424 Depok, West Java, Indonesia
email: fahmiazizz50@gmail.com

Abstract. The change in the water surface area is essential to be studied as it can affect ecosystem health and function of lakes to meet human needs. The use of remote sensing techniques is to
analyze such changes in the Lake Toba is considered effective and efficient ways in terms of time and cost, moreover the application of Google Earth Engine (GEE) as an internet-based data processing platform (cloud computing) that provides various kinds of free satellite image data can simplify the process. In this study, the author aims to estimate the rate of change in the surface water area of Lake Toba in four periods, that are the period of 1990 – 1997, 1998 – 2004, 2005 – 2012, and 2013 – 2018, by using Landsat 5 TM and 8 OLI/TIRS imagery. Cloud masking methods is used to obtain cloud-free images and then using CART method in making land cover maps. After that, calculating the accuracy of CART method by using the confusion matrix to determine the Cohen’s Kappa coefficient value. The results of this study show that the highest lake surface water area is in the period of 1990 - 1997 within the area of 1133.88 km², and the lowest one is in the period of 1998 - 2004 within the area of 1119.20 km². The correlation of land cover classes with a surface water area of Lake Toba shows that the open land has high correlation value of -0.77.

1. Introduction
Google Earth Engine (GEE) is a program that provides a free facility to users to perform internet-based spatial data processing. GEE has a data bank of petabytes in which has various kinds of satellite images such as Landsat, Sentinel, MODIS, geophysical data (such as DEM) and climate data (such as CHIRPS, TRMM). This is a new breakthrough in modern science because researchers, students, governments, and public are challenged to explore further usefulness of the data available at GEE.

GEE uses programming languages such as API (Application Programing Interface) for Javascript and Python, for new users there is no need to worry to understand, because GEE has provided tutorials using Javascript and Python with the example algorithm. The presence of GEE makes data processing more effective and efficient in terms of time and cost, for example there are algorithms that have been provided by GEE to eliminate cloud interference in satellite imagery, this is very useful for research on agriculture, fisheries and especially in the current study of land cover change.

The issue of decreasing the water level in Lake Toba, North Sumatra in the past few years is allegedly due to changes in land cover around the water catchment area (DTA), increased open and built-up land area followed by a decrease area of vegetation, which makes rainwater not stored in the soil and immediately becomes runoff leading to the lake, so that when the rainy season comes the water overflows to the banks of the lake to cause flooding, on the contrary when the dry season the amount of stored ground water decreases thereby reducing water supply into the lake and causing drought in the area.
The purpose of this study is to analyze land cover changes using Landsat 5 TM and Landsat 8 OLI / TIRS satellite images in four periods, there are period of 1990 - 1997, 1998 - 2004, 2005 - 2012, and 2013 – 2018. This study uses CART method to analyze the effect of land cover changes in surface water area of Lake Toba, North Sumatra Province.

2. Study Area and Data Source

2.1. Study area

This research focuses on the Lake Toba catchment area, which is located in North Sumatra Province. The Lake Toba catchment area has a total area of 379,940 ha or 3799 km² and stretches between 2°10' N – 3° 0' N and 98° 20' E – 99° 50' E covering 7 (seven) districts including Dairi Regency, Humbang Hasundutan Regency, Karo Regency, Samosir Regency, Simalungun Regency, North Tapanuli Regency, and Toba Samosir Regency [1] dan [2].

2.2. Data sources

2.2.1. Satellite Images

This study uses Landsat 5 TM in the periods of 1990 - 1997, 1998 - 2004, 2005 - 2012 and Landsat 8 OLI / TIRS in the period 2013 - 2018, these satellite images are available in the catalog of google earth engine https://code.earthengine.google.com/

2.2.2 Water Catchment Area

The catchment area of Lake Toba is obtained from the Lake Toba Regional Ecosystem Conservation Coordinating Board (BPEKDT).
3. Methods

3.1 Cloud masking and filling
GEE has provided an algorithm to get cloud-free images, the cloud masking and filling method are used to remove clouds and fill the area with cloud-free images by using several image combinations in different periods. Algorithm used for Landsat 5 TM (https://code.earthengine.google.com/1446d79ad4edeb07dce0bb16f39a06bc) and Landsat 8 OLI/TIRS (https://code.earthengine.google.com/4030605ba0269fb82082dc7c60fe0aab) is different code, Here's an example:

This code for Landsat 4, 5 or 7 using surface reflectance QA band to mask clouds.
```javascript
var cloudMaskL457 = function(image) {
    var qa = image.select('pixel_qa');
    // If the cloud bit (5) is set and the cloud confidence (7) is high
    // or the cloud shadow bit is set (3), then it's a bad pixel.
    var cloud = qa.bitwiseAnd(1 << 5)
        .and(qa.bitwiseAnd(1 << 7))
        .or(qa.bitwiseAnd(1 << 3))
    // Remove edge pixels that don't occur in all bands
    var mask2 = image.mask().reduce(ee.Reducer.min());
    return image.updateMask(cloud.not()).updateMask(mask2);
};
// Map the function over the collection and take the median.
var collection = ee.ImageCollection('LANDSAT/LT05/C01/T1_SR')
    .filterDate('2008-01-01', '2008-12-31')
var composite = collection
    .map(cloudMaskL457)
    .median();
```

This code for Landsat 8 using QA band to mask clouds
```javascript
var maskL8 = function(image) {
    var qa = image.select('BQA');
    /// Check that the cloud bit is off.
    // See https://landsat.usgs.gov/collectionqualityband
    var mask = qa.bitwiseAnd(1 << 4).eq(0);
    return image.updateMask(mask);
};
// Map the function over one year of Landsat 8 TOA data and take the median.
var composite = ee.ImageCollection('LANDSAT/LC08/C01/T1_TOA')
    .filterDate('2016-01-01', '2016-12-31')
    .map(maskL8)
    .median();
// Display the results in a cloudy place.
Map.setCenter(98.747, 2.622, 12);
Map.addLayer(composite, {bands: ['B4', 'B3', 'B2'], max: 0.3});
```

3.2 CART (Classification and Regression Trees)
This method is most commonly used in the classification process of satellite imagery, because this method does not require parameter, so it is quite simple to do in the land cover classification process [3]. This study makes five classes, there are the water body, open land, built-up land, vegetation and shrubs.
3.2.1. Step one
Determining the study area that has been done by cloud masking and filling, after that making a training sample for each land cover class with polygon tools, then setting the import 'geometry' to 'FeatureCollection', 'property' to 'landcover' and 'value' for each class such as water body (0), open land (1), built-up land (2), vegetation (3), and shrubs (4), at last combining the classes into one feature collection.

```javascript
var classlist = twater.merge(tbare).merge(tbuild).merge(tvegetation).merge(tshrubs);
print(classlist);
```

In this study, author uses the combination of band 543 on Landsat 5 and band 654 on Landsat 8 is to make it easier to distinguish each class, such as dark green for vegetation, reddish brown for built-up land, white to dark blue for open land, and black for water.

3.2.2. Step two
Retrieving information for each class of each band using a polygon training sample.

```javascript
var bands = ['B2', 'B3', 'B4', 'B5', 'B6', 'B7'];
var trainingsample = [yourimage].select(bands).sampleRegions({
  collection: classlist,
  properties: ['landcover'],
  scale: 30
});
print(trainingsample);
```

3.2.3. Step three
Run the classification process using the CART method

```javascript
var CARTclass = ee.Classifier.cart().train(
  features: trainingsample,
  classProperty: 'landcover',
  inputProperties: bands
);
var class = image.select(bands).classify(CARTclass);
//display result
Map.addLayer(class,
  {min: 0, max: 3, palette: ['blue', 'red', 'green','yellow']},
  'classification');
```

3.2.4. Step four
Validating the sample, by undergoing the same as in the first and second steps of validating the sample using polygon tools, and then combine the sample into one feature collection.

```javascript
var vallist = vwater.merge(vbare).merge(vbuild).merge(vvegetation).merge(vshrubs);
```

3.2.5. Step five
Validating the classification results with the validation samples.

```javascript
var validation = class.sampleRegions(
  collection: vallist,
  properties: ['landcover'],
  scale: 30,
);
print(validation);
```
3.2.6 Step Six
Calculating the value of the confusion matrix, and the Cohen’s Kappa coefficient value aims to see how accurate the CART classification method is.

```javascript
var confusionmatrix = validation.errorMatrix('landcover', 'classification');
//Print the error matrix to the console
print('Validation error matrix: ', testAccuracy);
var KappaCoef = testAccuracy.kappa()
print(KappaCoef,'Kappa')
```

3.2.7. Step seven
Exporting the land cover class to ‘GeoTiff’ and save in your google drive.

```javascript
Export.image.toDrive({
  image:class,
  description: "Class90-97",
  folder: 'GEE',
  region: DTA,
  scale: 30,
  maxPixels: 1e12,
  fileFormat: 'GeoTiff'
})
```

3.3. Area calculation
The area of each class is computed using ArcMap 10.3 software, image classification which has been exported to the GeoTiff then converted to polygons, and then calculate the area by adding 'add field' in the attribute table and then run the 'calculate geometry' and select unit area (km²).

4. Result and Discussion

4.1. Land cover changes
Understanding the definition of land cover, [4] says that land is the outermost part of the Earth's surface, where interactions of the lithosphere, hydrosphere, and atmosphere elements occur, so land cover is a biophysical condition that exists on the surface of the Earth such as vegetation, soil types, buildings, water body and so on. Land cover conditions can certainly change with the impact of human activities / needs as well as the physical processes of the environment [5] [6]. Classification of satellite imagery is a way of grouping different objects on the surface of the earth, identified by the spectral value information of each pixel. Number of pixels with the same spectral value will be grouped into one class, as well as the other pixels depending on how many classes to be formed, such as water bodies, vegetation, open land and cloud [7].

The result of the CART method produces four classes, there are open land, built-up land, vegetation, and shrubs. Classification over four periods shows that the high accuracy and Kappa values ranged from 0.89 - 0.92, indicating that the land cover classification results in this study is accurate with the satellite imagery data used (Table 1).

| Period     | Overall Accuracy | Kappa Coefficient |
|------------|------------------|-------------------|
| 1990-1997  | 0.91             | 0.91              |
| 1998-2004  | 0.92             | 0.89              |
| 2005-2012  | 0.91             | 0.89              |
4.2 The effect of land cover changes on surface water area
The open land in the 1990 - 1997 period is 185.35 km², then expanded in the next period of 1998 – 2004 is 378.02 km², decrease in the period of 2005 - 2012 is 184.33 km², while the highest area occurs in the period of 2013 - 2018 is 375.95 km² (Table 2). Figure 2 shows that the open land and built-up area have the most significant changes in Toba Samosir and Samosir Regency.

Table 2. The area of each land cover class in four periods

| Period       | Bareland Km² | Built-up | Vegetation | Shrubs     | Lake Area  |
|--------------|--------------|----------|------------|------------|------------|
| 1990-1997    | 185.34       | 7.54     | 684.11     | 1,723.78   | 1133.88    |
| 1998-2004    | 378.02       | 9.65     | 596.81     | 1,639.89   | 1119.20    |
| 2005-2012    | 184.33       | 10.23    | 789.08     | 1,635.92   | 1123.87    |
| 2013-2018    | 375.95       | 68.10    | 904.00     | 1,273.41   | 1120.93    |

Figure 2. CART classification results in four periods using Google Earth Engine
Built-up land is the only land cover class increases in each period, due to the growing population every year so that the demand for living space also increases and stimulates the development of settlements in that area. Table 2 shows that in the period of 1990 – 1997, the area of built-up land is only 7.54 km², while in the later period of 2013 – 2018 the area increased by 68.10 km², from the year 1990 to 2018 the built-up land area increased by 60.56 km².

The vegetation area shows a reduction of 87.30 km² in the period of 1998 – 2004 and increased by 307.2 km² over two periods of 2005 – 2012 and 2013 – 2018. The shrubs area in all over periods shows a reduction about 450.4 km², the reduction is allegedly caused by the conversion of the built-up open land and vegetation area (Table 2).

The highest lake surface water area is in the period of 1990 - 1997 within area of 1133.88 km², and the lowest one is in the period of 1998 - 2004 within area of 1119.20 km². The correlation of land cover classes with a surface water area of Lake Toba shows that the open land has high correlation value of -0.77. A negative correlation indicates that the open land and lake surface water area have an inverse relationship, the higher open land area resulting a reduction in lake surface water area. Other land cover classes have a low correlation with lake surface water area, when built-up and vegetation area have a negative correlation of -0.40 and -0.17, when built-up and vegetation area increase, the lake surface water area decreases, while shrubs area has a positive correlation of 0.55 which indicates that shrubs area increases, the lake surface water area increases as well (Table 2).

5. Conclusion
Google Earth Engine is a new breakthrough in research especially in remote sensing as satellite imagery data processing becomes easier, more effective and efficient in time and cost. The accuracy of CART method for land cover classification is high enough, open land area has a relatively high correlation with its influence on the surface water area of Lake Toba compared to the other land cover classes. The classification maps show that Toba Samosir and Samosir Regency have the most significant changes of the open land and built-up area.

6. Acknowledgments
Thanks to Assoc. Prof. Shaun R Levic from GEARS - Geospatial Engineering and Remote Sensing lab - https://www.gears-lab.com who has provided codes and tutorial on the CART classification process in Google Earth Engine.

7. References

[1] Sihotang, H., et al., *Model konservasi sumberdaya air Danau Toba*. Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan (Journal of Natural Resources and Environmental Management), 2012. 2(2): p. 65.

[2] Nontji, A., *Danau-Danau Alami Nusantara*. Puslit Oseanografi LIPI, Jakarta, 2016.

[3] Hu, Y. and Y. Dong, *An automatic approach for land-change detection and land updates based on integrated NDVI timing analysis and the CVAPS method with GEE support*. ISPRS journal of photogrammetry and remote sensing, 2018. 146: p. 347-359.

[4] Land, A.O.o.t.U.N., W.D. Division, and A.O.o.t.U.N.I.W.G.o.L.U. Planning, *Planning for sustainable use of land resources: towards a new approach*. 1995: Food & Agriculture Organization of the UN (FAO).

[5] Jansen, L.J. and A. Di Gregorio, *Parametric land cover and land-use classifications as tools for environmental change detection*. Agriculture, Ecosystems & Environment, 2002. 91(1-3): p. 89-100.

[6] Di Gregorio, A., *Land Cover Classification System: Classification concepts. Software version 3*. 2016.
[7] Rwanga, S.S. and J. Ndambuki, *Accuracy assessment of land use/land cover classification using remote sensing and GIS*. Int. J. Geosci, 2017. 8(4): p. 611-622.