Implementation & Comparison of Different Water Quality Indices on Sentinel 2a Images using Python

Submitted for the mini project of Statistics and Programming for Geodata paper of M.Tech. 2019 – 21

by,

Chintan Maniyar(M.Tech PRSD)
Srishti Pandey(M.Tech WRD)
Anushree Jain(M.Tech PRSD)
Ravi Kant (M.Tech GID)
Vishnu Bajpai (M.Tech PRSD)

Indian Institute of Remote Sensing
Indian Space Research Organisation
Department of Space
Govt. of India

Table of contents

1. Introduction

Water quality study is the process of determining the chemical, physical and biological characteristics of water bodies and identifying the possible contamination sources that degrade the quality of water. [1]. In this project we have taken two in-land water bodies(lakes) having different type of water content and calculated different water quality indices namely NDWI, NDCI, Chlorophyll-a. After calculation of these water indices the results are analysed and compared with each other.

The normalized difference water index (NDWI), is proposed as remote sensing derived indexes related to liquid water from space. NDWI can be defined or calculated in one of two ways, one is by using NIR and SWIR bands and another by using Green and NIR bands. The former is used to monitor changes in water content of leaves while the latter is used to monitor changes in water content of the water bodies. With the help of NDWI temporal changes in the water bodies of the area could be calculated. Similarly to measure the algal blooms and growth of cynobactreia in water bodies Chlorophyll-a(chl-a),which is an empirical quantity derived from NDCI by coefficient adjustment . Normalized difference Chlorophyll Index is calculated using 708 nm and 665 nm bands and is further used for calculation of Chl-a.

The NDWI index used to monitor changes of the water bodies uses NIR and Green bands as the water body has strong absorption and low radiation in the range from visible to infrared wavelengths. The NDWI can enhance the water information effectively in most cases. It is sensitive to built-up land and vegetation and often results in over-estimated water bodies. Whereas the NDWI used for the water content in leaves uses NIR and SWIR bands as the SWIR reflectance reflects changes in both the vegetation water content and structure in vegetation canopies. The NIR reflectance is affected by leaf internal structure and leaf dry matter content, but not by water content. Hence, this index is used to analyse changes in water bodies containing and surrounded by plants as it does not exclude information about vegetation having a high water content. Similarly for NDCI the bands are chosen such that the change in chlorophyll can be analysed and hence the change in the chlorophyll content of the water bodies are observed through this index.Various empirical formulas are derived from NDCI such as Chl-a which further gives us the information about algal bloom, cynobacteria and various other vegetation growth inside a water body. Hence, the different indices calculated for a water body is used to analyse different aspects of a water body and the difference between these indices could be clearly visualized.

2. Objective

The overall objective of this study is to compare the nature of water bodies and the surrounding water bodies using various water quality indices, using Sentinel 2, level 2A ortho-rectified images and conclude the behaviour of each index.

Further, the sub-objectives of the study are as below:

- To explore various temporal datasets
- To download and use the dataset in python
- To explore various python libraries like gdal, shapley,folium
- To do the band math and generate the single band images of the indices
- To analyse and compare different indices for different dataset using QGIS

3. Dataset and software used

For this project two inland water bodies were chosen with different water quality and content, and at two different climatic conditions, so that the comparison of different indices can be done. The area containing Chilika lake and Okeechobee lake were chosen for the dataset, the properties of these lakes are described below. The dataset of Sentinel-2a was chosen for this project. The Sentinel-2a dataset contains 12 bands which includes SWIR,NIR,Green and various other bands of images, which are used in the calculation of indices.

Chilika lake: Chilika Lake is a brackish water lagoon, spread over the Puri, Khurda and Ganjam districts of Odisha state on the east coast of India, at the mouth of the Daya River, flowing into the Bay of Bengal, covering an area of over 1,100 km. It is the largest coastal lagoon in India and the second largest brackish water lagoon in the world[2]. Water depth of the lake varies from 0.9 ft (0.3 m) to 2.6 ft (0.8 m) in the dry season to 1.8 m (5.9 ft) to 4.2 m (13.8 ft) in the rainy season.Microalgae, marine seaweeds, and sea grasses also flourish in the brackish water of the Chilika Lagoon[3].

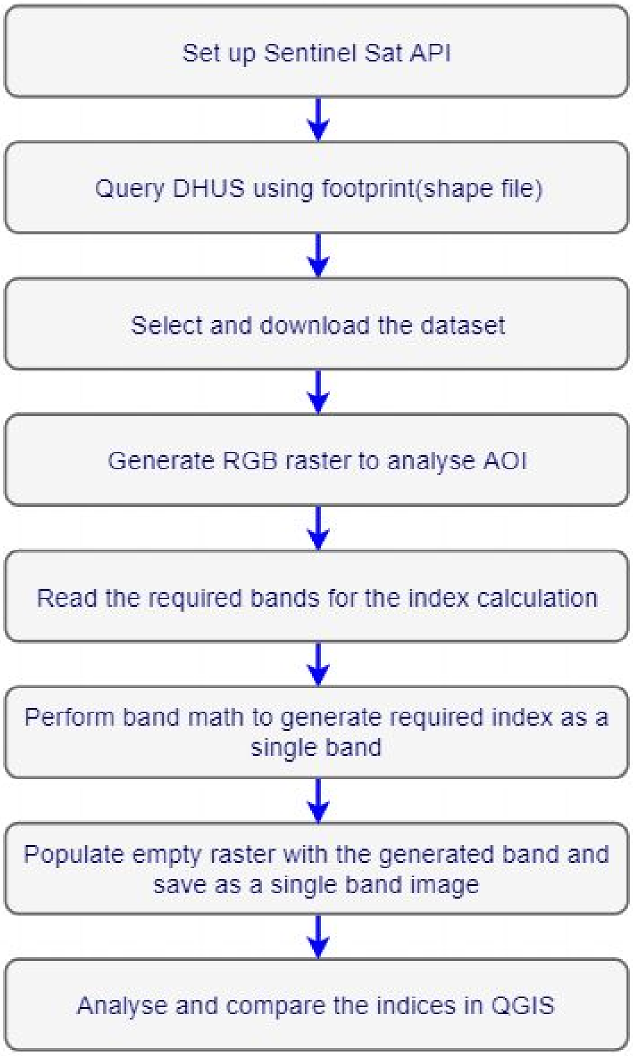
Okechobee Lake: Lake Okeechobee, also known as *Florida's Inland Sea*, is the largest freshwater lake in the state of Florida. Okeechobee covers 730 square miles (1,900 km^2), and is exceptionally shallow for a lake of its size, with an average depth of only 9 feet (2.7 metres). The Kissimmee River, located directly north of Lake Okeechobee, is the lake's primary source[4]. Every year there is an excessive growth of algae called cynobacteria in the month of july and august, recently in years 2018 and 2019 algae blooms have been reported in this lake.

The following software patches and tools were used as a part of this study:

- Python v3.6.4
- Anaconda Navigator
- Jupyter Notebook
- QGIS 3.8.

4. Methodology

The following flowchart briefly explains the methodology adopted followed by the description.



The overall system architecture is built around the open-source python libraries for remote sensing data analysis and the open source software for processing digital satellite images. Python v3.6.4 along with the following Python libraries and packages were used to perform this study:

1. sentinelsat
2. gdal
3. geopandas

4. rasterio
5. shapely
6. matplotlib

Following steps explain the methodology in detail:

1. Data Acquisition

The data was downloaded within and using the python script only. Python package 'sentinelsat', which has been distributed as under the GNU license as an open source library for the python platform. This library helps set up an api which allows us to communicate with the Copernicus DHUS platform [5], which requires us to authenticate the connection using the DHUS account credentials before we can query for products.

To be able to query the api for products, we need to provide the following necessary but not mandatory parameters:

1. Footprint shape file (which was generated using QGIS by underlaying an Open Street Map)
2. Date range/ Sensing period
3. Platform
4. Processing Level of Product
5. Cloud Cover Range

Upon querying the api with the footprints for Chilikha Lagoon and Okeechobee Lake, we get a list of products from which we choose the required products based on cloud cover to finally download the product data.

2. Data Processing and Image Generation

Once the data is successfully downloaded, the organisation and segregation of data was performed. The data comes with separate ortho-rectified files for each band (Level 2A). These band wise files are then read individually to generate the new bands for our target indices, and also the rgb raster image.

To generate the rgb image, an empty raster was defined with the dimensions and crs of the chosen band files of the same resolution, which was then populated with the data from the band files.

The following indices were generated and saved as a single band images. Following table as the legend for the band aliases for Sentinel 2A platform:

Name	Units	Min	Max	Scale	Resolution	Wavelength	Description
B1				0.0001	60 meters	443.9nm (S2A) / 442.3nm (S2B)	Aerosols
B2				0.0001	10 meters	496.6nm (S2A) / 492.1nm (S2B)	Blue
B3				0.0001	10 meters	560nm (S2A) / 559nm (S2B)	Green
B4				0.0001	10 meters	664.5nm (S2A) / 665nm (S2B)	Red
B5				0.0001	20 meters	703.9nm (S2A) / 703.8nm (S2B)	Red Edge 1
B6				0.0001	20 meters	740.2nm (S2A) / 739.1nm (S2B)	Red Edge 2
B7				0.0001	20 meters	782.5nm (S2A) / 779.7nm (S2B)	Red Edge 3
B8				0.0001	10 meters	835.1nm (S2A) / 833nm (S2B)	NIR
B8A				0.0001	20 meters	864.8nm (S2A) / 864nm (S2B)	Red Edge 4
B9				0.0001	60 meters	945nm (S2A) / 943.2nm (S2B)	Water vapor
B11				0.0001	20 meters	1613.7nm (S2A) / 1610.4nm (S2B)	SWIR 1
B12				0.0001	20 meters	2202.4nm (S2A) / 2185.7nm (S2B)	SWIR 2

Source: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR

1. Normalised difference water index (NDWI_lmc) for leaf moisture content

   NDWI_lmc = (B8A - B11)/(B8A + B11)

2. Normalised difference water index (NDWI_wc) for water content

   NDWI_wc = (B3 - B8)/(B3 + B8)

3. Normalised difference chlorophyll index (NDCI)

   NDCI = (B5 - B4)/(B5 + B4)

4. Index for Chlorophyll-a concentration (Chl-a)

   Chl-a = 14.035 + 86.115*NDCI + 194.325*NDCI*NDCI

3. Image Interpretation

Once the required images were generated as GeoTIFF files, they were analysed and interpreted using QGIS. The RGB images were constructed using 3 bands for each study site, and the images for the water quality indices - NDWI_lmc, NDWI_wc, NDCI and Chl-a were generated as grayscale images, since there was only one band for every index. Pixel level analysis and image interpretation was done by following the standard practices of analyzing a single-band image, the results of which are discussed in the next section.

5. Results and Discussion

The results obtained are in the form of TIFF files which can be accessed using softwares like QGIS, SNAP, etc. For each index calculated, an image is generated for both the lakes under study. An RGB image is also generated for each lake, making a total of five images for each lake.

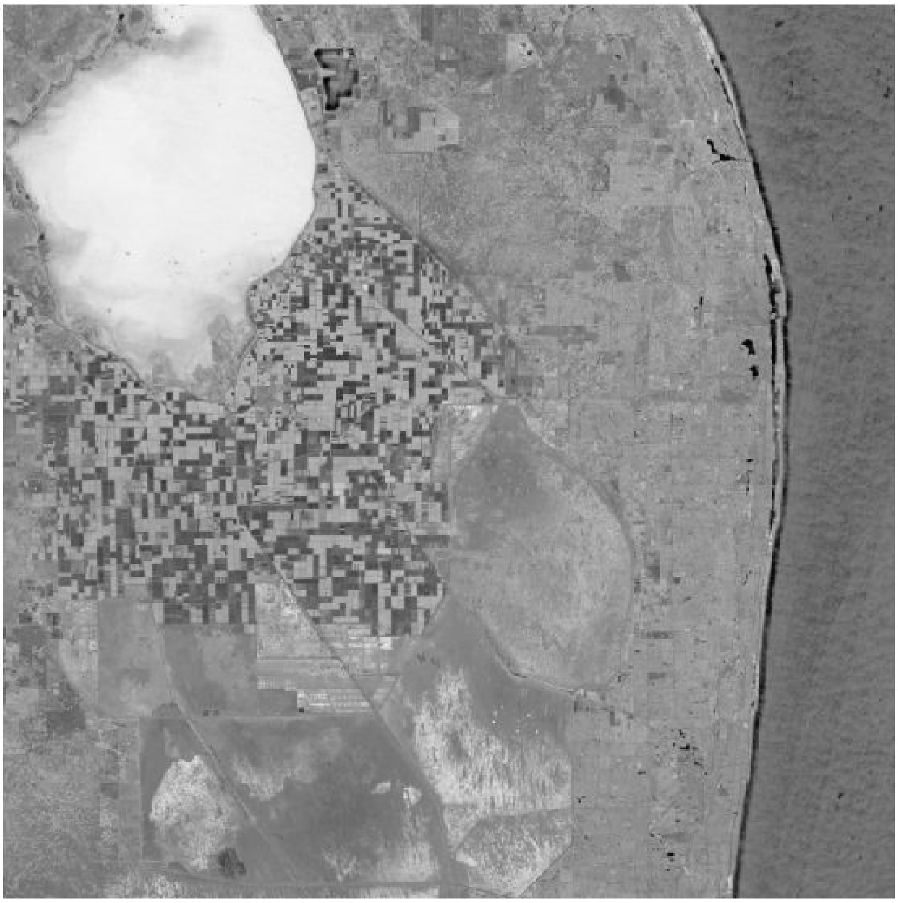
As is known about the indices, the brighter values represent the feature of the index while other features appear with darker values. Here, a comparison is made between two pairs of indices to observe the variation that occurs due to different indices.

Okeechobee Lake, Florida:

NDWI

Two types of NDWI have been calculated. One for leaf moisture content and another for water content. The lake can be seen on the upper left corner of the image. In the image for Okeechobee lake for water content NDWI, the lake looks slightly bright, less brighter than the artificial water body of Martin County. This shows that as the Martin County lake has freshwater it appears brighter while Okeechobee lake with algal bloom even in the month of December has a lower value of NDWI. The results obtained for NDWI-leaf moisture content depict areas with high leaf moisture with brighter values and low values for lower leaf moisture content. As shown in the image, the boundaries of the lake do not appear as well defined as seen in NDWI-water content because of the natural vegetation around the periphery of the lake. Brighter values can be observed in the middle part of the lake due to stagnation of water that leads to higher concentration of algae. While the artificial water body of Martin County appears mostly dark due to the absence of any vegetation as the body is kept maintained.

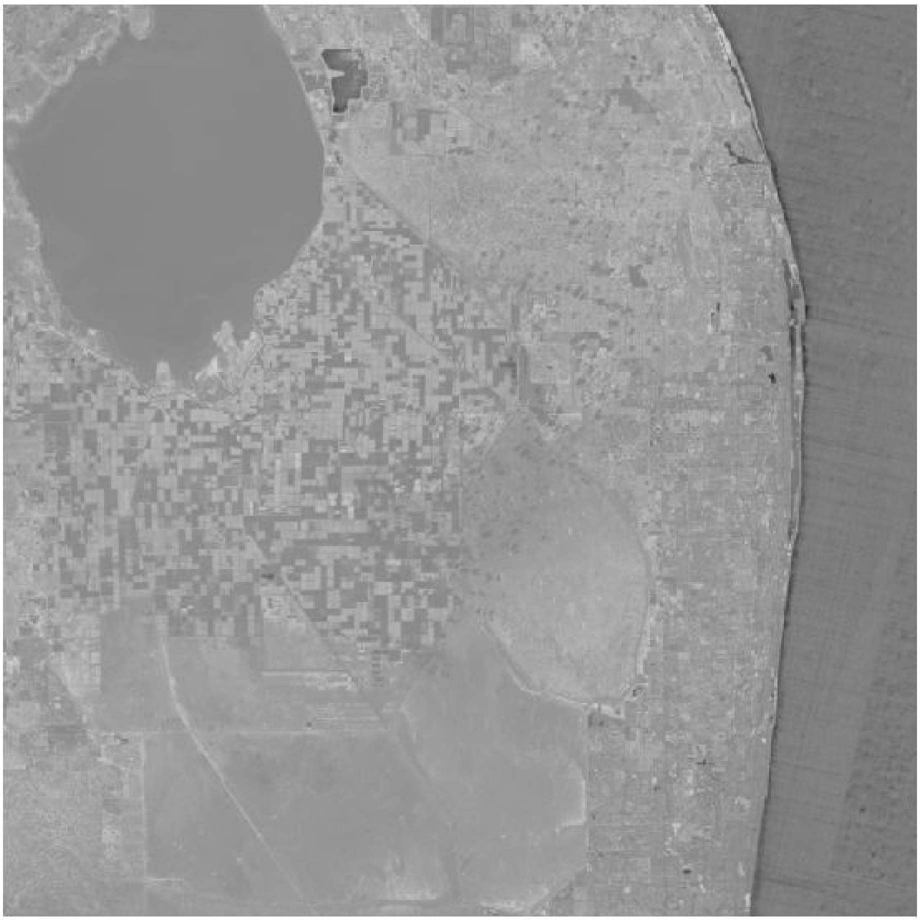
NDWI image for water content for Okeechobee lake



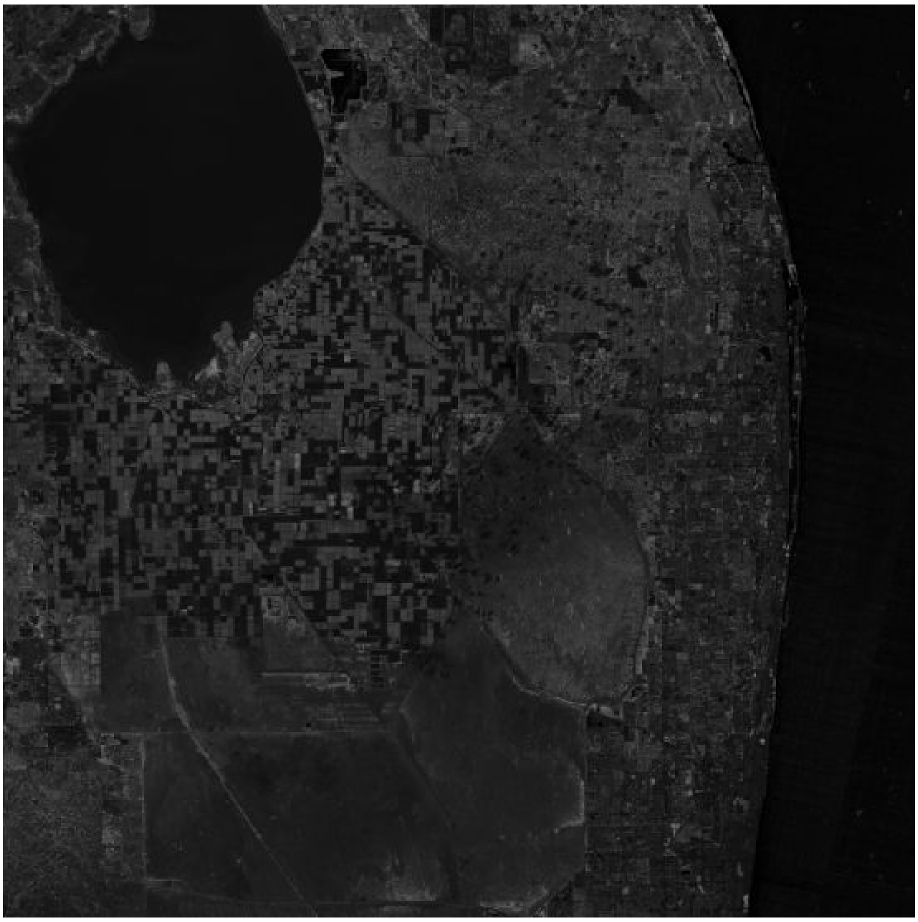
NDWI image for leaf moisture content for Okeechobee lake

NDCI and Chl-a

Normalised Difference Chlorophyll Index shows brighter values for high chlorophyll content and darker values for lower content. The NDCI values for Okeechobee lake are seen to be medium dark due to the algal bloom present while the artificial water body of Martin County appears to be darker due to the absence of much algae. The vicinity can be seen to be on the brighter side due to the vegetation present such as the roads can be seen surrounded by a bright linear feature that shows the vegetation on the sides of the roads. Chl-a index used here in the study also depicts such results with a higher contrast. The lake can be seen to be medium dark with the surrounding areas brighter due to the presence of vegetation.



NDCI image for Okeechobee lake

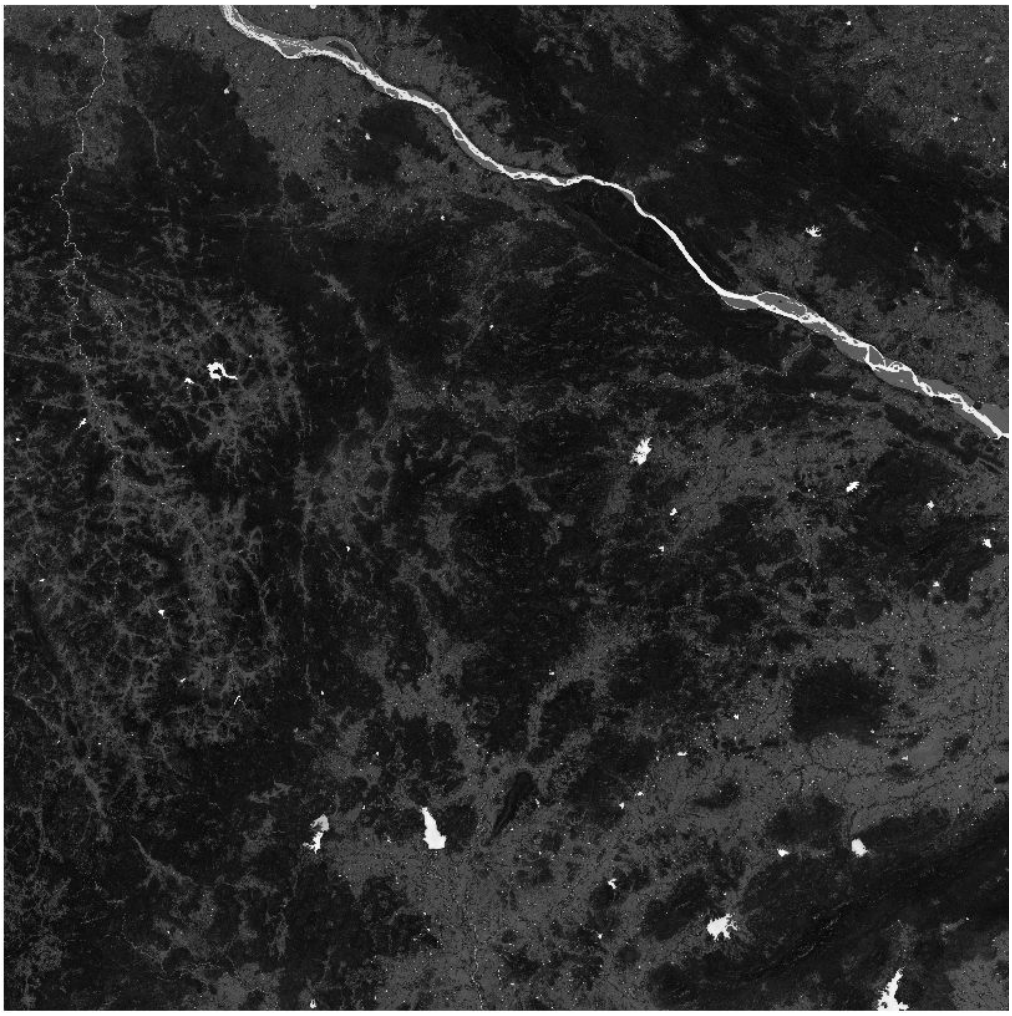


Chla-a image for Okeechobee lake

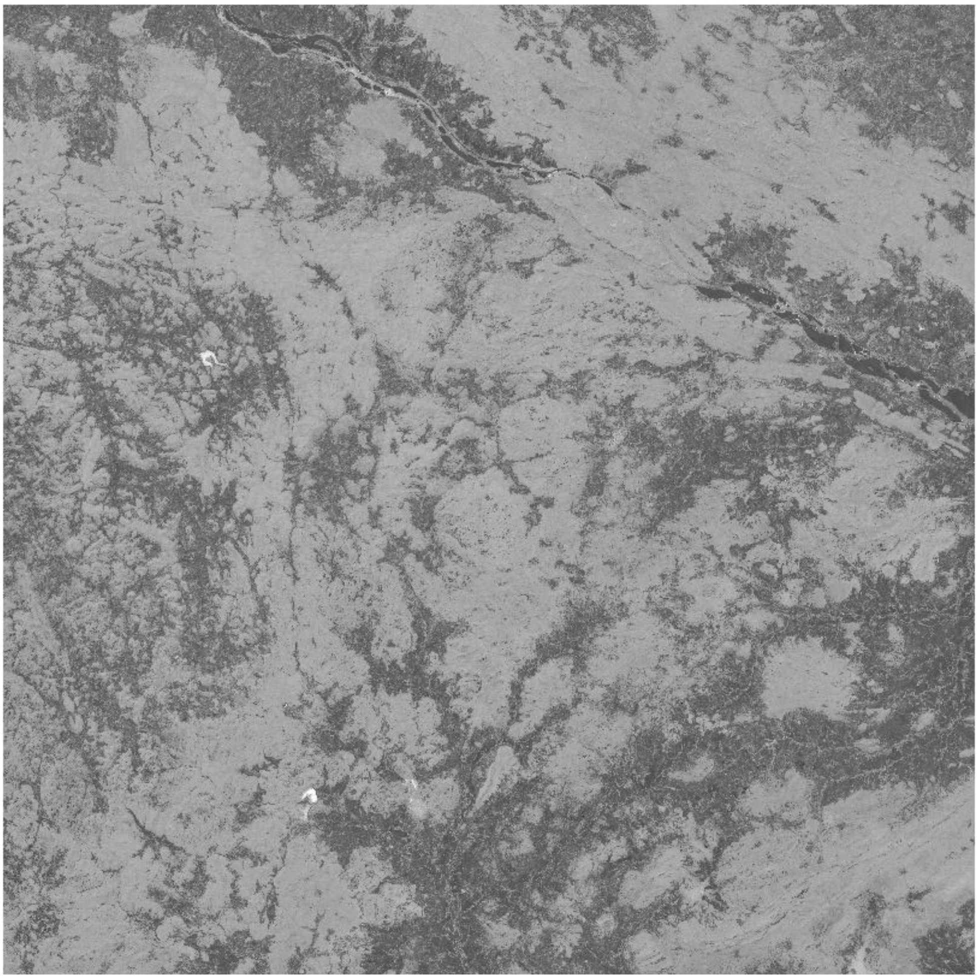
<u>Chilikha lake, Odisha</u>

NDWI

The NDWI for water content shows the lake as well the Mahanadi river as bright due to the nature of the bodies demarcating the water bodies in the image. In the NDWI for leaf moisture content, the lake appears to be medium bright due to the presence of vegetation in the lake while the river appears darker as it has running water which does not promote the growth of vegetation. Very bright areas can also be seen which are part of the Eastern Ghats forests.



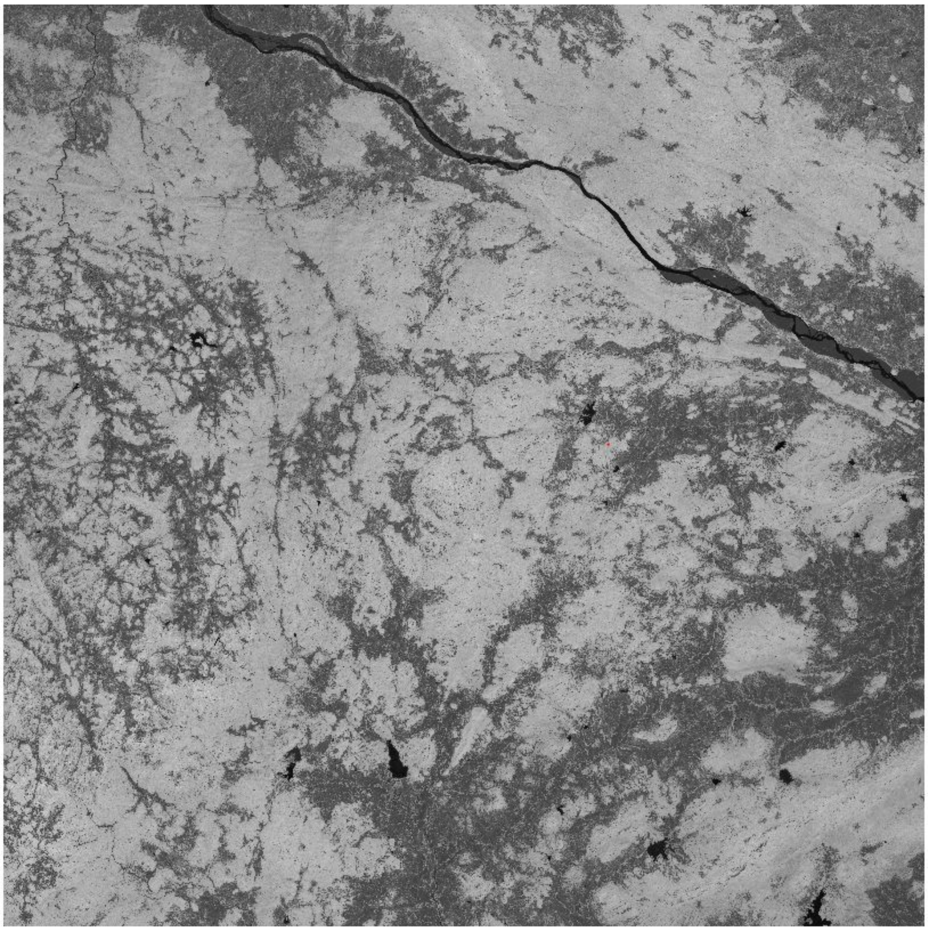
NDWI image for water content for Chilikha lake



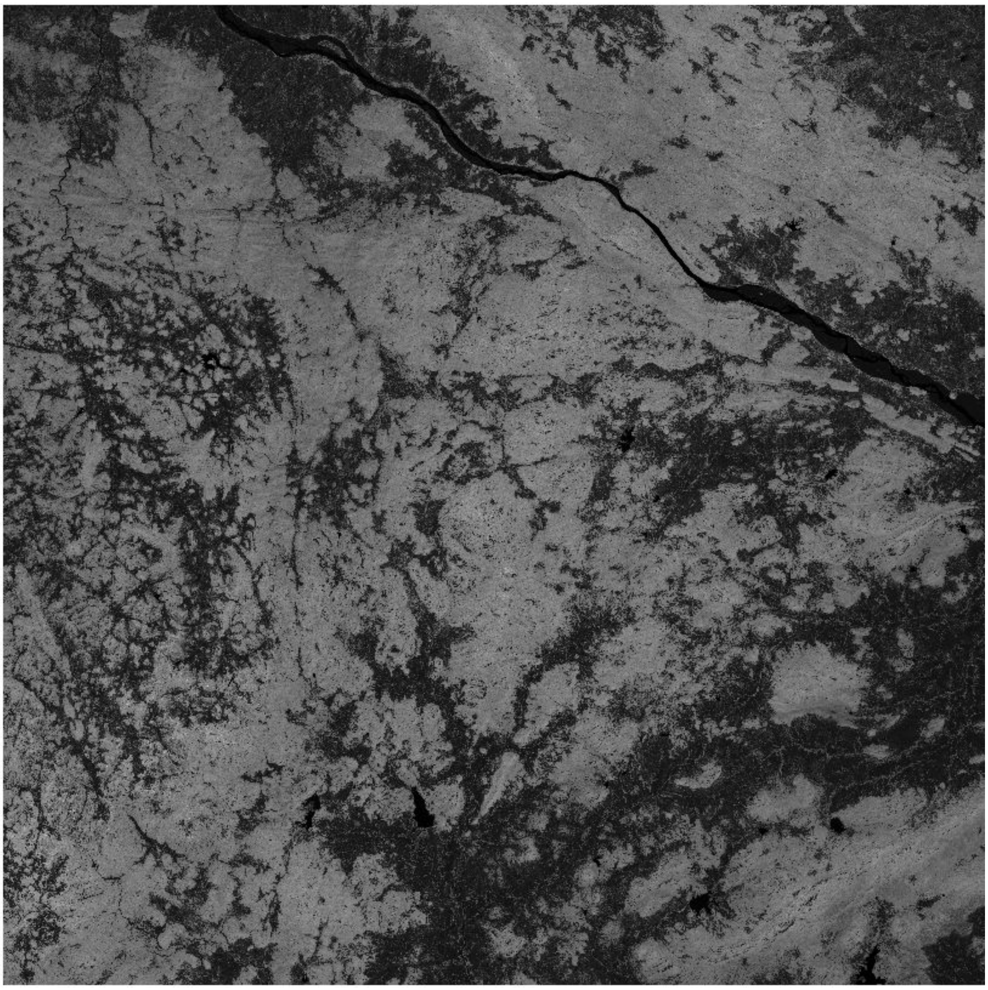
NDWI image for leaf moisture content for Chilikha lake

NDCI and Chl-a

The NDCI values for the lake appear to be dark depicting that the lake contains minimum or no vegetation. Most water bodies in the image appear to be dark including the Mahanadi river. As the study area is mostly surrounded by forests and vegetation, most of the surrounding locations appear to be on the brighter side. The Chl-a index shows similar results with higher contrast showing the lake and the river dark with surrounding areas as medium bright.



NDCI image for Chilkha lake



Chl-a image for Chilikha lake

6. Future Scope

This work can further be continued by developing kernels of size 3x3, 5x5, 7x7 and 9x9 for each of the indices described in the study. This can be done by the methods of histogram quantification, coefficient adjustment, spatial resampling and matrix interpolation.

The quantification of these indices as kernels can prove highly useful as it wil then make the generation of these indices independent of the platform and the spatial resolution of the satellite/ aerial images. The kernel can directly be convoluted over any rgb satellite/ aerial image to generate the masked/ single-band image of the concerned index.

References

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5017463/
2. https://en.wikipedia.org/wiki/Chilika_Lake
3. https://en.wikipedia.org/wiki/Chilika_Lake#Geography_and_topography
4. https://en.wikipedia.org/wiki/Lake_Okeechobee
5. https://scihub.copernicus.eu/dhus/
6. Bo-cai Gao, NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space, Remote Sensing of Environment, Volume 58, Issue 3, 1996, Pages 257-266,
7. T J Malthus, A G Dekker, First Derivative Indices for the Remote Sensing of Inland Water Quality using High Spectral Resolution Reflectance, Environment International, Volume 22, Issue 2, 1995, Pages 221-232