ESTIMATION OF CHLOROPHYLL-A CONCENTRATION IN LAGUNA DE BAY USING SENTINEL-3 SATELLITE DATA

M. Conopio [1], R. K. Japor [1], A. C. Blanco [1,2], A. M. Tamondong [1,2]

[1] Department of Geodetic Engineering, University of the Philippines – Diliman, Quezon City 1101, Philippines -
(mmconopio@up.edu.ph, rknjapor@gmail.com, acblanco@up.edu.ph, amtamondong@up.edu.ph)
[2] Training Center for Applied Geodesy and Photogrammetry, University of the Philippines – Diliman, Quezon City 1101, Philippines

Commission IV

KEY WORDS: OLCI, Laguna Lake, Regression Model, Water Quality, Microalgae

ABSTRACT:

The mission of the European Space Agency's Sentinel-3 satellite is to provide data for land and ocean monitoring purposes. Sentinel-3's Ocean and Land Colour Instrument (OLCI) data are being used largely for monitoring offshore and coastal waters but can also be used for inland waters including lakes. It has a spatial resolution of 300 meters, a temporal resolution of 2-3 days and contains 21 spectral bands. Laguna de Bay, with a surface area of around 930 km², suffers from periodic algal blooms resulting from excessive nutrient inputs from surrounding watersheds. This study aims to assess the applicability of Sentinel-3 OLCI for estimating chlorophyll-a (chl-a) concentration in Laguna Lake. Several chl-a estimation algorithms (i.e., band ratios and indices) were tested for Sentinel 3 OLCI images and compared with in-situ data obtained using a chl-a sensor. A regression model comprising of individual spectral bands, band ratios, and band indices for chl-a estimation as independent variables was developed, yielding an adjusted R^2 of 0.759 and RMSE of 1.19 ug/L. The model consists of R620, R674/R708, and RED/NIR. A map was produced showing the spatial distribution of chl-a in Laguna Lake, with most of the portion of the lake having a concentration ranging from 7.5 ug/L to 15 ug/L. This shows that Sentinel 3 OLCI images can be utilized for accurately estimating chl-a in Laguna Lake.

1. INTRODUCTION

1.1 Background of the Study

Laguna de Bay is the largest freshwater lake in the Philippines and third largest in South East Asia with a surface area of 911-949 km². It is situated east of Metro Manila. Laguna de Bay has many significant contributions to the country. The lake is a major source fishery in the Philippines that accounts for 17% of the national fisheries production, 10% of the national production for aquaculture and 44% of the national production for municipal fisheries. Besides this, the lake also supports multi-purpose uses such as a transportation route between towns on its shore, a source of energy for the three power plants located in the region and also a source of irrigation for farmers and industrial use. For the past years, Laguna de Bay is under danger due to uncontrolled development, population growth and industrialization. The lake has now become the catch basin of Metro Manila's population and urbanization over-spill. The Laguna Lake Development Authority (LLDA) is an agency created for the sole purpose of rehabilitating the damaged state of the lake (Vargas, 2015).

One of the major problems concerning Laguna Lake is the excessive amounts of nutrients present (Santos-Borja, 2015). In 2016, an Environmental Health Report Card (EHRC) of Laguna de Bay was published by LLDA. The EHRC was based on a 10-year data gathered from 2004 to 2013 and was implemented by Partnerships in Environmental Management for the Seas of the East Asia Resource Facility (PEMSEA) and LLDA under the United Nations Environment Programme (UNEP)/Global Environment Facility (GEF) Project on Global Foundation for Reducing Nutrient Enrichment and Oxygen Depletion from Land-based Pollution. The indicators for water quality are Nitrates, Phosphates, Chlorophyll-a (Chl-a), Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), and total coliforms, which were compared with the Department of Environment and Natural Resources (DENR) guidelines (DAO 34). These indicators were calculated as percentages and combined as

percentage scores. The grading scale was as follows: A (91% to 100%); B (83% to 91%); C (75% to 83%); D (70% to 74%); and F (0% to 70%). Laguna de Bay has four bays namely East, West, Central, and South bay. The tests were conducted on each bay and received individual scores. West Bay got a C (76%) for water quality. All indicators for water quality were acceptable except Phosphate. Central Bay also got a C but the lowest at 71% for water quality. All indicators for water quality were acceptable except Phosphate, at 25%, and chl-a at 0%. East Bay also got a C but the highest at 81% for water quality. Almost all indicators were rated A except chl-a at 0%. Lastly, South Bay was also a C but the second highest at 77% for water quality. Again, all indicators for water quality were acceptable except Phosphate, at 63%, and chl-a, at 0% again. Laguna de Bay got an overall rating of C-, at 76%, for water quality which is suitable for fisheries and recreation as by the DENR guidelines but has very excessive chl-a, which scored 0% in 3 out of 4 bays, and Phosphate content (Tamayo-Zafaralla, 2017). High levels of phosphates in waters will trigger dense plant growth which can result in the eutrophication of the water (Perlman, n.d.). chl-a, a type of chlorophyll present in algae, is used as an indicator of eutrophication (Li, Sha & Wang, 2017). Eutrophication may lead to deteriorating water quality, dissolved oxygen depletion, and toxic phytoplankton bloom, which produces toxins that can destroy aquatic life in affected areas (Smith, Tilman, & Nekola, 1999).

As these harmful factors affecting Laguna de Bay progresses, it calls for a faster, real-time, and efficient way of monitoring its water quality, in which technology can be of great help. Remote Sensing is the science of using electromagnetic radiation as a medium in detecting features within the earth with the use of various instruments such as cameras and satellites (V, Navalgund, & Roy, 2007). The use of remote sensing is a cost-efficient approach in the monitoring and assessment of water bodies, including lakes. It gives water quality data for a handful of areas at a time. It can also help water resources researchers and decision makers to monitor water bodies more effectively and

efficiently since remote sensing techniques can be used to measure different qualitative parameters (Gholizadeh, Melesse, & Reddi, 2016).

There are numerous satellite missions launched by different countries that are being used to capture images, which will then be used in remote sensing for whatever geospatial purpose they may serve. One is the Sentinel-3 mission that is jointly operated by the European Space Agency (ESA) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). One of its objectives is to measure ocean and land surface color, in which it has a payload specifically made for that purpose called Ocean and Land Colour Instrument (OLCI). It is a pushbroom imaging spectrometer that measures the solar radiation reflected by the Earth. It has a spatial resolution of 300 meters and has 21 spectral bands (European Space Agency, 2013). Each band has a specific function where it can be used most efficiently. OLCI is advantageous in water quality monitoring because it has products restricted to detecting water bodies only. Also, a variety in its 21 bands is specifically for chlorophyll detection.

1.2 Statement of the Problem

Pollution caused by industrialization is a major problem that the Philippines is currently facing, and Laguna Lake is a manifestation of that problem. One way to improve the condition of the lake is by establishing effective management which can be achieved by an accurate assessment and monitoring of the area. Being able to visualize aspects of the lake such as the concentration of chl-a will contribute significantly on the rehabilitation of the lake by providing knowledge on factors that will lead to the creation of an effective management system, and this can be achieved by remote sensing. Remote sensing has become an essential part of environmental monitoring since it provides an efficient way of analyzation with the use of satellite images. Satellites such as Sentinel-3, with its coastal monitoring capabilities, can be utilized to provide an accurate estimation of chl-a in Laguna Lake but further research needs to be taken to assess its ability on inland waters.

1.3 Objectives

In general, the study aims to determine the capability of Sentinel-3 satellite in the estimation of chl-a concentration in Laguna Lake. Specifically, the study aims to achieve the following:
- To utilize Sentinel-3 data to estimate chl-a in Laguna de Bay.
- To produce a map that shows the chl-a concentration estimation in Laguna de Bay.

1.4 Scope and Limitations

Cloud cover is out of hand of the researchers and will be masked out. There are ways to still get data under these clouds but those include getting data from other satellite images taken from other dates and just patching those masked areas. However, this method imposes problems on the consistency of data since in model development, in situ measurements taken from a certain date should be cross-matched with the satellite image taken from the same date. The satellite in concern, Sentinel-3, passes over the Philippines in regular period or revisit cycle. Preparations for field works are made prior to the satellite overpass and the field work proper is made during. Hence, knowing whether the area concerned will be cloudy is not a certainty.

Bands of the Sentinel-3 OLCI are specifically used for monitoring land and ocean color. Which means, it is used for a very large area. Its drawback for the concerned area is its spatial resolution. Even though Laguna de Bay is around 900 km² in area, the spatial resolution of OLCI is 300 m. During field works, data are acquired by the sensor every second while the position is recorded every 10 seconds. This means that a handful of points will fall in a single pixel only and will have the same corresponding reflectance value.

2. REVIEW OF RELATED LITERATURE

2.1 Chlorophyll-a (Chl-a)

Chl-a is a type of chlorophyll found in algae which allows them to photosynthesize (Clark, 2011). According to Saeed Balali, the presence of chl-a means that nutrients such as phosphorus and silicate are present in the waters. Due to this, chl-a is most commonly used for water quality monitoring because too much concentration of it will lead to harmful algal bloom. Algal Bloom has many effects and the most commonly occurring one is the absorption of huge amount oxygen in the water, leaving other aquatic life to suffocate. Due to the importance of monitoring chl-a in waters, many researches and studies have been undertaken, including the use of remote sensing. According to a book The Light-Dependent Reactions of Photosynthesis, chl-a is a type of pigment which absorbs only a certain wavelength. Figure 1 shows the absorption spectra of chl-a which shows that it mostly absorbs blue and red light and reflects green.
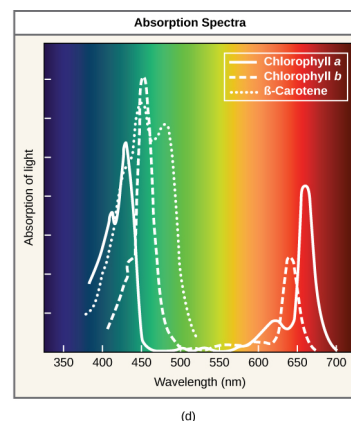

Figure 1. Absorption Spectra of Chlorophyll. Adapted from Avissar, Y., Choi, J., DeSaix , J., Jurukovski, V., Wise, R., & Rye, C. (2012) Biology.

2.2 Chl-a Concentration Estimation Model

Chl-a concentration is very important in assessing the water quality of bodies of water and in predicting or serving as proxies for characterizing harmful algal blooms (HABs). A study made by Acheampong in 2018 utilized Sentinel-2, together with Moderate Resolution Imaging Spectroradiometer (MODIS), to derive algal concentration for the algal bloom monitoring in the Arabian Gulf. MODIS Ocean Colour data was used in place with in-situ measurements because it has been successfully used in mapping chl-a concentration globally and it has good temporal resolution. In the study, Sentinel-2's high resolution is integrated with MODIS Level-2 chl-a products to derive chl-a products of high temporal and spectral resolution of MODIS and high spatial resolution of Sentinel-2. Three existing semi-empirical algorithms were used to retrieve chl-a namely the Red Tide Index

(RI) and Red Tide Index Chlorophyll Algorithm (RCA), Maximum Chlorophyll Index (MCI), and the Normalised Difference Chl-a Index (NDCI), which are based from band ratios.

In the research conducted by Lins et.al, the researchers used Sentinel-2 and Sentinel-3 products to assess chl-a remote sensing algorithms in tropical estuarine-lagoon system. The models used to estimate chl-a from reflectance bands are the Blue-Green Ratio Model, the Two-Band NIR-Red Ratio Model, the Three-Band NIR-Red Model, and the Four-Band NIR-Red Model.

Aside from the existing semi-empirical algorithms widely used for the detection and identification of chl-a concentration, there are still various band ratio-based empirical algorithms that were being tested and used. A study conducted by Toming et. al. aimed to find relationships between water quality parameters such as chl-a, TSM, and CDOM, retrieved from the Sentinel-3 satellite image and in situ measurements of these parameters acquired in Baltic Sea

2.3 Sentinel 3

One of the missions of Sentinel-3 was to monitor the coastal areas of the earth through its Ocean and Land Colour Instrument (OLCI) on board. "Carrying a suite of instruments, Sentinel-3A is arguably the most complex of all the Copernicus Sentinels (Esa, n.d.)", the expectation for Sentinel-3 to provide accurate and reliable data is high. One of the main features of Sentinel-3 is its 21 bands whose band interval is smaller compared to other satellites.

Band No.	λ center	Bandwidth	Function
Oa1	400	15	Aerosol correction, water constituent retrieval
Oa2	412.5	10	Yellow substance and detrital pigments
Oa3	442.5	10	Chl absorption max., biogeochemistry
Oa4	490	10	High Chl, other pigments
Oa5	510	10	Chl, sediment, turbidity, red tide
Oa6	560	10	Chl reference
Oa7	620	10	Sediment loading, Phycocyanin
Oa8	665	10	Chl, sediment, yellow substance, vegetation
Oa9	673.75	7.5	For improved fluorescence retrieval
Oa10	681.25	7.5	Chl fluorescence peak
Oa11	708.25	10	Chl fluorescence baseline
Oa12	753.75	7.5	O2 absorption, clouds, vegetation
Oa13	761.25	2.5	O2 absorption band/aerosol corr.
Oa14	764.375	3.75	Atmospheric correction
Oa15	767.5	2.5	O2A for cloud top pressure, fluorescence
Oa16	778.75	15	Atmos. corr. ,aerosol corr.
Oa17	865	20	Atmos. corr. ,aerosol corr., clouds
Oa18	885	10	Water vapour absorption ref. band
Oa19	900	10	Water vapour absorption, vegetation
Oa20	940	20	Atmos./ aerosol corr.
Oa21	1020	40	Atmos./aerosol corr.

Table 1 Sentinel-3 OLCI Spectral Bands. Adapted from ESA. (n.d.). Sentinel-3A Earth colour data released

3. METHODOLOGY

3.1 General Methodology

Figure 2. General Methodology

Figure 2 shows the general flowchart of the methodology which consists of 4 major steps namely Data Gathering, Data Processing, Data Analysis, and Data Output. Data Gathering refers to the collection of in situ and image data. Data Processing involves the image pre-processing and integration of image and field data. Data Analysis involves the correlation testing, regression modelling and validation. And finally, Data Output refers to the creation of the final model and map. Figure 3 shows the detailed methodology.
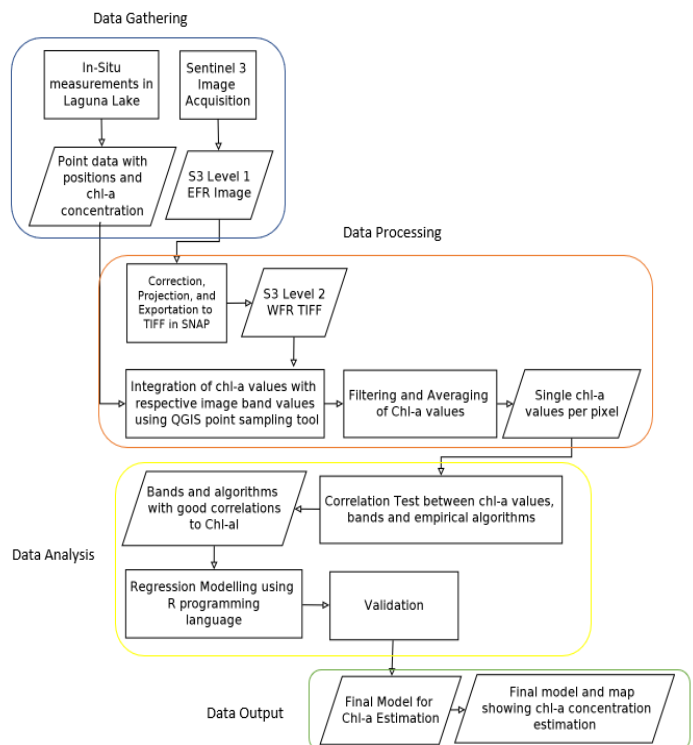

Figure 3. Detailed Methodology

3.2 Data Gathering

Field surveys were conducted on the different portions of Laguna Lake. The lake consists of 3 parts namely the east bay, center bay and the west bay (Figure 4). The dates of the field survey were chosen to coincide with the Sentinel-3 satellite overpass of the study area. These dates were December 11, 2018, and January 11, 2019. Chl-a was measured continuously throughout the area using Infinity-CLW chlorophyll and turbidity data logger attached on the boat. A handheld Global Positioning System (GPS) device was used to take note the locations of the boat which records every 30 seconds. A measurement made by the logger that intersects with GPS data was considered one data point with chl-a value and location .Full resolution (300m) level 1 OLCI images were downloaded corresponding with the dates of the field survey acquired from the ESA (European Space Agency) data hub.
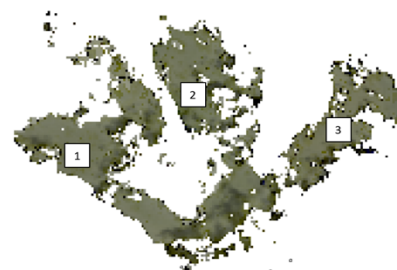

Figure 4 True Color Sentinel-3 Image of Laguna Lake on December 11, 2018:
West Bay [1], Center Bay [2] and East Bay [3]

3.3 Data Processing

3.3.1 Image Pre-Processing

All the image pre-processing was done using SNAP 6.0 (Sentinel Application Platform, a software created by the ESA for sentinel image processing. The OLCI level 1 data were corrected using a tool called iCOR, a plug-in tool in SNAP that is used for atmospheric correction, especially for images of water bodies, both inland and coastal. The correction also automatically masked out other pixels besides water. After applying the correction, reflectance values of the images were obtained. The images were then subsetted so that only parts of the Laguna Lake were showing and were then reprojected so that it will coincide with the field data. After the pre-processing, the final images were exported as Geotiff format.

3.3.2 Data Integration

The integration of the image data and the field data were done using the QGIS software. After both the data were opened in QGIS, a plugin tool called Point Sampling tool was used to collect the information from both data, chl-a from the field data and reflectance value, with respect of the location of the field data, from the image data. The result was a single file with the information of both the field and image data.

The extracted attribute table in QGIS included all the data points acquired in the field work and by using the Point Sampling Tool, band reflectance values of corresponding data points were also extracted. Since the satellite images were masked, some of the points landed on pixels that have no band reflectance. These points appeared to have a value of "nan," means "not a number," in their band reflectance were removed. Also, negative band reflectance values were also removed.

The usual values of chl-a in Laguna de Bay from March 2018 to December 2018 acquired by UP Multi-platform and Cross-sensor Water Quality Monitoring (MCWQM) project researchers were used to filter outliers from the field data. Due to the spatial resolution of Sentinel-3, which is 300 meters, several data points fell under a single pixel. Data points that fell within a single pixel were averaged in order to get a unique pixel data.

3.4 Data Analysis

3.4.1 Correlation Testing

The filtered data points, which includes the chl-a values and band reflectance values of all the bands of Sentinel-3 Level 1 (Bands 1 to 12, 16 to 18, and 21), were tested for correlation. Spearman correlation test was done between the chl-a values against each band in XLSTAT. Existing empirical algorithms and band ratios stated in Chapter 2.2 were also tested against the chl-a values. Scatter plots of band ratios and indices with high correlation to chl-a were also produced using Python.

3.4.2 Multiple Regression Analysis

Multiple regression models were produced using the Stepwise Regression Algorithm and were done in RStudio, an application programming interface (API) of R, a programming language mainly used for statistical processes. The regression model takes the linear form $y = A_0x_0 + A_1x_1 + ... + A_nx_n + e$, where y is the predicted chl-a, A is the coefficient, x is the regressor, and e is the error or intercept. The procedure for Stepwise Regression are as follows: a regressor x_i will be included in the model if it met the criteria of a p-value less than for some fixed ; another regressor will be included in the model if it met the criteria, p-values of the regressors already in the model will adjust and if one adjusted in a manner that its p-value does not meet the criteria anymore, it will be removed from the model; this process will iterate until no more regressor can be entered or removed.

The researchers considered two types of dependent variables, the raw chl-a and its natural logarithm. The reasons for considering the natural logarithm form of the data is to include the possibility that the relationship is not completely linear, and it is the most common transformation. Also, two types of data sets were used in the model making, one is using all of the data points of a certain date and the other is using 40% of the data points of a certain date to make the model and 60% for validation. Three models were made using the former type of data set (one for each data set per date). As for the latter type of data set, the 40% of data points to make the model were chosen randomly and were done in 3 trials per data set per date. So, 3 sets of randomly chosen points, with their corresponding validation points, per data set per date were generated and used. There was a total of 18 models (3 data sets, 3 trials per data set, 2 types of dependent variable per trial) made for this type of data set. All the models made were compared to arrive at the best model.

3.4.3 Validation and Determination of Best Model

The determination of the best model was based on the key measures namely adjusted R^2, Akaike Information Criteria (AIC), and Root Mean Square Error (RMSE). R^2 values closer to one (1) indicates a better linear relationship between the observed and predicted values. Lower AIC indicates a better relationship between regressors or variables in the model. RMSE shows how deviated the values are to the mean so lower RMSE indicates that the values do not spread out far from the mean.

Models produced using all of the data points of a certain date were applied to the data of other dates and combined dates. They were validated through ordinary least squares (OLS) simple linear fit and using the key measures adjusted R^2 and RMSE. On the other hand, models produced using 40% only of the data points of a certain date were validated to the remaining 60% of the said data points using the same method as above. They were also validated to the data of other dates and combined dates.

The band values of the pixel per different chl-a concentration estimation were taken for spectral validation. Spectral plots per concentration were produced from the values and compared to one another and to the usual spectral signature of chl-a.

3.5 Data Output

The models were then applied to the satellite image using the Semi-Automatic Classification Tool in QGIS. The equations of the models were inputted in the Band Calculation function of the tool where it will generate the map that shows the chl-a concentration estimation. After that, the maps were laid out and presented.

4. RESULTS AND DISCUSSION

4.1 Data Gathering

Table 2 shows the summary of the statistics obtained from the in-situ chl-a measurement of Laguna Lake measured on December

11, 2018, and January 11, 2019, respectively. The chl-a measurements were taken using the CLW logger that was synchronized with a GPS receiver that takes position every 30 seconds. So chl-a data with position was taken every 30 seconds. Figure 5 shows the map of the data collection for December 11 and January 11 respectively.

Date	Area	# of samples	Min	Max	Mean	Median	Std
12/11/2018	East Bay	13154	-4.77	101.12	11.91743	9.52	11.44539
01/11/2019	West Bay	18035	-4.77	485.15	12.55059	10.13	23.98968

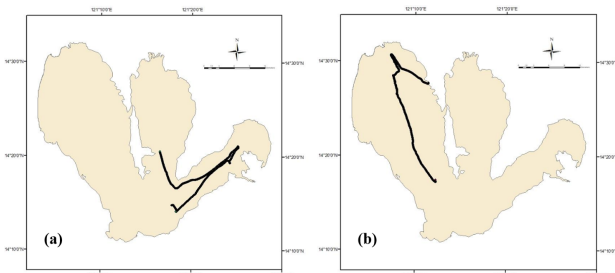Table 2. Statistics of in-situ raw chl-a measurements (ug/L)

Figure 5. Map showing the chl-a data collection: (a) December 11, 2018 and (b) January 11, 2019

4.2 Data Integration

The raw chl-a measurements were filtered to the usual values of chl-a in Laguna Lake which is between 3 ug/L to 30 ug/L determined by the field data collected by MCWQM Project from March to December 2018. Since Sentinel-3 images have a spatial resolution of 300 meters, multiple chl-a concentration data were on a single pixel. To be able to obtain one chl-a concentration for a single pixel, measurements that are within a single pixel were averaged to represent the concentration for that particular pixel. Table 3 shows a summary of the statistics of the chl-a concentration after the values were averaged.

Date	Area	# of samples	Min	Max	Mean	Median	Std
12/11/2018	East Bay	90	3.73431	21.1972	11.4718	11.3805	3.64878
01/11/2019	West Bay	103	3.718333	24.11286	14.02091	14.73	4.111342

Table 3. Statistics of chl-a measurements (ug/L) after filtering and averaging.

After obtaining one chl-a concentration per pixel, the spectra of the pixel per concentration were checked. There is a trend in the spectral signature where the higher the concentration, the lower the reflectance value compared to the ones with lower concentration. Also, peaks at 560 nm at 720 nm were observed throughout all the concentration which follows the usual trend of chl-a. Table 4 shows the spectral signature per chl-a concentration.
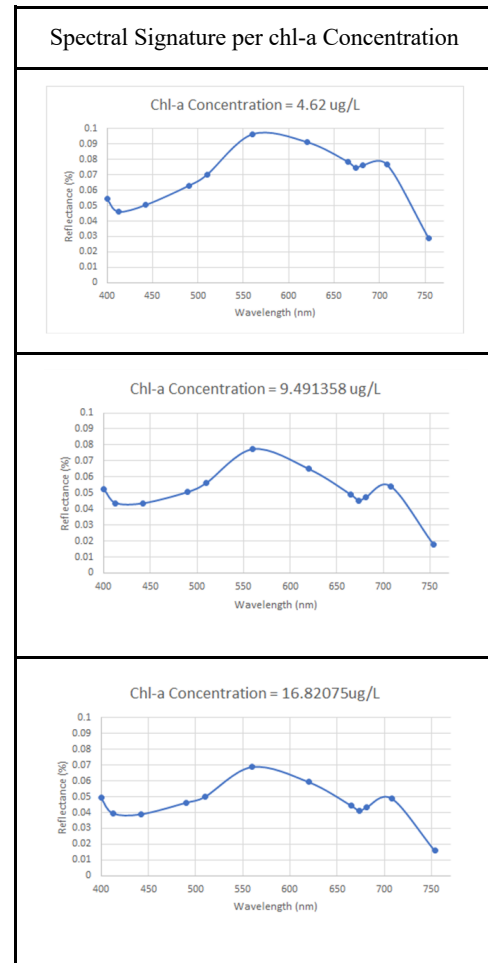
Table 4. Spectral Signature per chl-a concentration

The correlation of the image bands with the chl-a data was determined using the Spearman Correlation test. Generally, the image bands and the chl-a data have low correlations as depicted in table 5. The low correlation indicates that using a single image bands for chl-a estimation may not produce accurate results.

December 11		January 11	
Variables	Chl-a	Variables	Chl-a
R400	-0.406	R400	0.282
R413	-0.375	R413	0.285
R443	-0.355	R443	0.297
R490	-0.322	R490	0.304
R510	-0.274	R510	0.317
R560	-0.034	R560	0.335
R620	0.154	R620	0.325
R665	0.043	R665	0.310

R674	0.022	R674	0.305
R681	0.048	R681	0.309
R708	0.340	R708	0.362
R754	0.271	R754	0.267
R779	0.280	R779	0.272
R865	0.241	R865	0.209
R885	0.220	R885	0.179
R1020	-0.206	R1020	0.052

Table 5. Correlation of image bands with chl-a data per date.

The correlation of the chl-a with different band ratios and indices were also taken. This resulted in a better correlation overall compared to the bands only. Table 6 shows the correlation matrix.

December 11		January 11	
Variables	**Chl-a**	**Variables**	**Chl-a**
RI	0.15	RI	-0.315
MCI	0.337	MCI	0.363
MCI (adjusted)	0.338	MCI (adjusted)	0.362
NDCI	0.748	NDCI	-0.024
Blue/Green	-0.516	Blue/Green	-0.452
NIR/Red	0.655	NIR/Red	-0.076
R560/R665	-0.308	R560/R665	-0.214
R665/R708	-0.748	R665/R708	0.024
R665/R754	-0.374	R665/R754	0.258
R674/R708	-0.746	R674/R708	0.019
R674/R754	-0.392	R674/R754	0.263
R708/R754	-0.132	R708/R754	0.228
(1/R665 - 1/R708) x R754	0.731	(1/R665 - 1/R708) x R754	-0.096
(R490-R665)/R560	-0.377	(R490-R665)/R560	-0.401
R708 - ((R665+R754)/2)	0.650	R708 - ((R665+R754)/2)	0.466
R708 - R754	0.655	R708 - R754	0.439

Table 6. Correlation of band ratios and indices with chl-a data per date.

To check if these results yield a good linear relationship with the chl-a values, some of the band ratios and indices that showed relatively higher correlation were plotted against the chl-a values. Using R^2 value as a key measure in determining linear relationship, most of these band ratios and indices showed an R^2 value of greater than 0.5, with NDCI as the highest with 0.535 or 53.5% linearly related to the chl-a values.

Although these band ratios and indices showed high correlation and linear relationship with the chl-a values, individual bands and other band ratios and indices that have low correlation may still exhibit utility when used collectively in a single model.

4.3 Multiple Regression Modelling

The chl-a measurements, both raw and its natural logarithm, were used as the dependent variable, together with corresponding Sentinel-3 spectral bands and established band ratios and indices for chl-a estimation as regressors to create a multiple regression model using the Stepwise Regression Algorithm. The p-values of 0.1 and 0.05, or 90% and 95% confidence interval, respectively, were both used in making the models. First, using all the data points of a certain date as training points and p-value < 0.1, the best models generated are the models from December 11, 2018 data points, both raw chl-a and in its natural logarithm, based on the key measures R^2 and RMSE, as shown in Table 7. Changing the p-value from 0.1 to 0.05 resulted in almost the same results as p-value > 0.1, with the same best models.

As for the models generated using 40% of the data points of a certain date only, three sets of randomly chosen data points were used per data set per date. Since there are two types of dependent variable, raw and natural logarithm, six models were generated per data set per date. The total number of models generated were eighteen since there are three data sets, December 11, 2018, January 11, 2019, and the combined data set from both dates. The best models generated were all from December 11, 2018 data points, specifically from trials 2 and 3, with p-value > 0.1, as shown in Table 6. Similar to the models generated using all of the data points of a certain date, changing the p-value from 0.1 to 0.05 resulted in almost similar results as p-value > 0.1, with the same best models.

Model (# of points used for model - dependent variable)	R_sq	Adj R_sq	AIC	RMSE
All Data Points - Raw	0.690	0.671	396.0758	2.0922
All Data Points - Log	0.723	0.707	-38.1122	1.2062
40% of Points T2 - Log	0.724	0.698	-24.5230	1.1722
40% of Points T3 - Raw	0.690	0.661	166.7412	2.2635
40% of Points T3 - Log	0.780	0.759	-18.4514	1.1887

Table 7. Models chosen based on key measures

4.4 Statistical Validation

The best models generated per data set type were validated accordingly to determine the best model among them. As for models generated using all the data points of December 11, 2018, they do not have points left for validation so instead, they were validated against the data points of January 11, 2019 and

combined data points of December 11, 2018 and January 11, 2019 as shown in Table 8.

Model (# of points used for model - dependent variable)	Validation with			
	Jan 11 2019 data		Combined Data	
	R_sq	RMSE	R_sq	RMSE
All Data Points - Raw	0.028	5.078	0.211	3.958
All Data Points - Log	0.029	5.372	0.191	4.171

Table 8. Validation of models from all of December 11, 2018 data points with data points from other dates

As for the models generated from 40% of December 11, 2018 data points, they were validated with their corresponding validation points, the remaining 60% of data points, as shown in Table 9. All of these models yielded R^2 values of greater than 0.5, which means all of them established a good linear relationship with their validation points and can be used as models for that data set.

Trial # - Dependent Variable	Validation	
	R_sq	RMSE
T2 - Log	0.544	2.669
T3- Raw	0.599	2.354
T3 - Log	0.581	2.401

Table 9. Validation of models from 40% of December 11, 2018 data points with their corresponding validation points

To see whether these models also establishes a linear relationship with other data sets, they were validated with the data points of January 11, 2019 and combined data of December 11, 2018 and January 11, 2019. The results were summarized in Table 10 and showed that although relatively lower, linear relationship exists.

Model (# of points used for model - dependent variable)	Validation with			
	Jan 11 2019 data		Combined Data	
	R_sq	RMSE	R_sq	RMSE
40% of Points T2 - Log	0.035	5.022	0.198	4.006
40% of Points T3 - Raw	0.096	5.335	0.226	4.194
40% of Points T3 - Log	0.136	5.149	0.261	4.073

Table 10. Validation of models from 40% of December 11, 2018 data points with data points from other dates

The model that exhibited the best R^2 value and RMSE upon validation with other data points was chosen to be the best model which is the model generated from 40% of December 11, 2018 data points, specifically the third trial and used the natural logarithm of chl-a as dependent variable (40% of Points T3 - Log). This model has an adjusted R^2 value of 0.759 and AIC and RMSE of -18.4514 and 1.1887, respectively, with its training data points. Although low, it yielded the highest R^2 value among all models upon validation with the data points of other dates with 0.136 and 0.261 against January 11, 2019 data points and combined data points from December 11, 2018 and January 11, 2019, respectively. Table 11 shows the summary of statistics of the best model, together with the equation of the model.

Variable	Decision	R_sq	Adj R_sq	AIC	RMSE
R620	Enter	0.7	0.682	-9.3857	1.2196
R674/R708	Enter	0.76	0.737	-15.3758	1.1977
NIR/Red	Enter	0.78	0.759	-18.4514	1.1887
	Coefficient	Model			
Intercept	21.805				
R620	8.137				
R674/R708	-13.758	$$chl = \dfrac{e^{8.137B7+21.805}}{e^{13.758B9/B11+7.411NIR/Red}}$$			
NIR/Red	-7.411				
Overall					
R_sq	0.780				
Adj R_sq	0.759				
RMSE	1.189				

Table 11. Summary of Statistics of Best Model

4.5 Spectral Validation

Besides the statistical validation, the spectral signature was also utilized to validate the regression model produced. The spectral signatures of different concentration estimation of chl-a were taken. Table 12 shows the different spectral signature with their corresponding chl-a concentration and model.

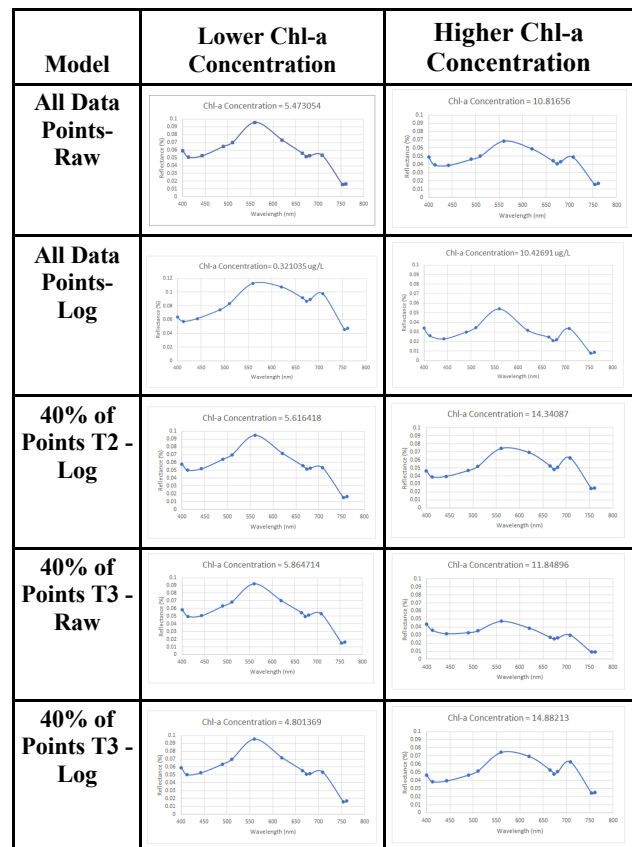| Model	Lower Chl-a Concentration	Higher Chl-a Concentration
All Data Points- Raw	Chl-a Concentration = 5.473054	Chl-a Concentration = 10.81656
All Data Points- Log	Chl-a Concentration = 0.321035 ug/L	Chl-a Concentration = 10.42691 ug/L
40% of Points T2 - Log	Chl-a Concentration = 5.616418	Chl-a Concentration = 14.34087
40% of Points T3 - Raw	Chl-a Concentration = 5.864714	Chl-a Concentration = 11.84896
40% of Points T3 - Log	Chl-a Concentration = 4.801369	Chl-a Concentration = 14.88213

Table 12. Spectral Signature of different chl-a concentration per model.

The resulting spectral signature reflects the attribute of chl-a. Generally, the ones with high chl-a concentration estimation have high absorption in the red (625-740 nm) and blue spectrum (450–485 nm) and have high reflectance value on the green spectrum (500–565 nm), denoted the peak reflectance at 560 nm. The distinct peaks at 560 nm and 708 nm are common indicators for chl-a in inland waters, these are also part of the bands whose

function is for chlorophyll detection, as stated by the Sentinel-3 team.

4.6 Visualization

After creating and choosing the model with the best results, it was then applied to produce chl-a concentration estimation images. The final model was applied to the corrected Sentinel-3 image taken on December 11, 2018. Figure 6 shows the map created from the chosen best model.
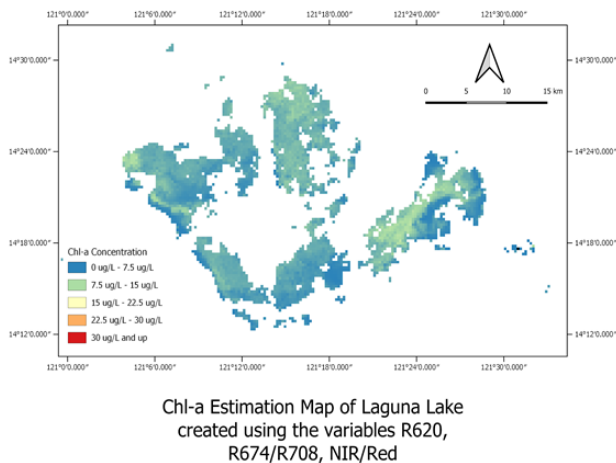

Chl-a Estimation Map of Laguna Lake created using the variables R620, R674/R708, NIR/Red

Figure 6. Final Map created from the final model (40% of Points T3- Log).

5. CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This research aimed to assess the capability of Sentinel-3 OLCI images in chl-a retrieval in inland water such as Laguna Lake using in situ data and empirical algorithms. Given the results of the study, Sentinel-3 Level 1 OLCI full resolution image can be utilized for chl-a concentration estimation of Laguna Lake since its bands showed a relationship to the in situ chl-a values. The best model was created using the in-situ measurements from December 11, 2018, composing of R620, R674/R708, and Two-Band NIR-Red Ratio Model. The model resulted in an R^2 of 0.780, an adjusted R^2 of 0.759 and an AIC of -18.4514. Also, due to this relationship, models created retained the spectral signature and trend of chl-a.

Since the model created is statistically validated and reliable, even the satellite image of Sentinel-3 with low spatial resolution, which is 300 meters, was still utilized properly to produce a map showing the estimated concentration of chl-a in Laguna Lake with known accuracy.

5.2 Recommendations

Based on the results of the study, it is recommended to explore other kinds of regression model besides the ones used in the study. Other regression models may produce a more accurate relationship between chl-a and image bands, ratios, and indices.

Due to time constraint, the research did not cover the whole Laguna Lake. In situ measurements were only taken from the east and west bay. It is recommended to be able to collect in situ measurements throughout the whole lake. Also, it is recommended to produce a model per season in Laguna Lake due to the fact that the optical properties of the water changes per season.

REFERENCES

Vargas, V. (2015, July 21). Laguna de Bay, Philippines: Environmental Literacy « IAN ... Retrieved November 15, 2018

Santos-Borja, A. C. (2015). The Laguna de Bay Ecosystem Health Report Card: An Assessment and Planning Tool for Integrated Lake Basin Management. Retrieved May 21, 2019

Tamayo-Zafaralla, M. (2017, February 2). STARScience: Saving Laguna de Bay.

Perlman, H., & Usgs. (n.d.). Phosphorus and Water

Li, X., Sha, J., & Wang, Z. (2017). Chlorophyll-A Prediction of Lakes with Different Water Quality Patterns in China Based on Hybrid Neural Networks. *Water,9*(7), 524. doi:10.3390/w9070524/

Smith, V., Tilman, G., & Nekola, J. (1999). Eutrophication: Impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environmental Pollution,100*(1-3), 179-196. doi:10.1016/s0269-7491(99)00091-3

V, J., Navalgund, R., & Roy, P. (2007). Remote sensing applications: An overview.

Gholizadeh, M., Melesse, A., & Reddi, L. (2016). A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors,16*(8), 1298. doi:10.3390/s16081298

European Space Agency. (2013, September 3). Sentinel-3 User Handbook.

Toming, K., Kutser, T., Uiboupin, R., Arikas, A., Vahter, K., & Paavel, B. (2017). Mapping Water Quality Parameters with Sentinel-3 Ocean and Land Colour Instrument imagery in the Baltic Sea. *Remote Sensing,9*(10), 1070. doi:10.3390/rs9101070

Clark, M. (2011). Handbook of textile and industrial dyeing. *Chlorophyll*. doi:10.1533/9780857093974

Avissar, Y., Choi, J., DeSaix , J., Jurukovski, V., Wise, R., & Rye, C. (2012) Biology. Retrieved from https://cnx.org/contents/GFy_h8cu@10.59:fVAf83sY@12/Preface-to-Biology

Esa. (n.d.). Sentinel-3A Earth colour data released. Retrieved from https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-3/Sentinel-3A_Earth_colour_data_released

Lins, R., Martinez, J., Marques, D. M., Cirilo, J., & Fragoso, C. (2017). Assessment of Chlorophyll-a Remote Sensing Algorithms in a Productive Tropical Estuarine-Lagoon System.

Acheampong, C. (2018, February). (pdf) Deriving Algal Concentration From Sentinel-2 Through ...