WATER QUALITY ASSESSMENT OF LAGUNA DE BAY USING MODIS SURFACE REFLECTANCE MOD09GA PRODUCT

Mark Jayson Felix (1), Gay Jane Perez (1)

[1] STAMINA4Space Program, Institute of Environmental Science and Meteorology,
University of the Philippines Diliman, Quezon City, Philippines
Email: mjbfelix@stamina4space.upd.edu.ph; gpperez1@up.edu.ph

KEY WORDS: Surface Reflectance, Water Quality, MODIS, Neural Network, Laguna de Bay

ABSTRACT: Laguna de Bay, being the largest lake in the country, is one of the primary sources of freshwater fish supply in Metro Manila and surrounding provinces. Moreover, the lake water is used for irrigation and domestic supply after being subjected to water filtration and treatment. Given these economic benefits, it is important to maintain the lake's health and overall water quality to minimize potential losses due to fish kill and pollution. Conventional water quality monitoring of the lake involves routine ground measurement of different parameters such as temperature, chlorophyll-a concentration, total suspended solids (TSS), turbidity, dissolved oxygen (DO), total phosphorus (TP), total nitrogen (TN), total chloride (TCl), and pH on pre-selected stations. However, these point measurements do not fully represent the complex dynamic water systems in the whole lake, which has an approximate area of 900 km^2. To fully understand these dynamic systems, both comprehensive spatial and temporal monitoring is needed, which can be addressed using satellite sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS). In this study, MODIS surface reflectance (MOD09GA) bands and ground measurements spanning from 1999 to 2010 were used to derive models for TSS, TN, TCl, TP, turbidity, alkalinity, and pH mapping in 4 different stations. The dataset was divided for the training and validation of the Multiple Linear Regression (MLR) and Cascade Forward-Feed Neural Network (CFF-NN) models. Results show better performance, with a lower root mean squared error (RMSE) and higher coefficient of determination (R^2), from water quality models based on the CFF-NN compared to MLR. Among the water quality parameters, TSS and turbidity derived using the CFF-NN model were observed to have the highest correlation with the surface reflectance bands. The high skill score of the CFF-NN model was found to be related in its ability to capture the variability of TSS and surface reflectance bands with season and sampling stations. With the addition of ground datasets, the water quality models' accuracy may be improved and provide vital information on the lake dynamics and insights on the occurrence of undesirable events such as fish kill.

1. INTRODUCTION

Among modern optical satellite systems, MODIS is one of the satellite sensors that provide a long historical data archive (1999 - 2020) with a track record of good radiometric accuracy suitable for terrestrial and coastal observations (Xiong, 2020; Liu, 2019; Chander, 2009). Furthermore, its revisit period of 1 to 2 days makes it advantageous over other satellite sensors designed for environmental monitoring. The MODIS sensor has a total of 16 bands in the visible-SWIR region, designed for ocean and terrestrial observations (Franz, 2006). From the 16 MODIS bands, 9 bands were particularly designed to be sensitive to the typical reflectance values of open ocean. These ocean bands, however, saturate on coastal and inland waters characterized with high suspended particulates and turbidity. For such cases, several studies had utilized the terrestrial bands for water quality assessment (Yang, 2018; Wang, 2010, Mller, 2004). Typically, water quality assessment of complex coastal and inland waters entails the use of Level-1 or Level-2 MODIS products to derive different geophysical parameters such as chlorophyll-a and suspended solid concentration. Before the period of cloud computing, these data should be manually downloaded and stored for processing. Processing a decadal time-series of these images would take a huge amount of time and storage. With the advancement of technology, cloud computing platforms are now available to process these images without the need for the actual downloading of the whole dataset. One of these platforms useful for remote sensing applications is the Google Earth Engine (GEE) (Gorelick, 2017 and Liu, 2018). The GEE supports the archiving and processing of multiple satellite products from different sensors aside from MODIS such as the OLI/Landsat-8, MSI/Sentinel-2, and OLCI/Sentinel-3.

The aim of this study is to utilize the Level-2 surface reflectance product of MODIS/Terra (MOD09GA) retrieved through the GEE platform and use it as input for the water quality modeling of Laguna de Bay, the largest lake in the Philippines. Laguna de Bay is widely used for aquaculture, supplying a large fraction of the freshwater fish demand on the island of Luzon (Cuvin-Aralar, 2001 and Santos-Borja, 2003). Being surrounded by the mega manila cities and municipalities, the lake is prone to water pollution from anthropogenic activities, making a comprehensive water quality monitoring scheme of the lake crucial in maintaining its health. MOD09GA products had been previously used to study agriculture and drought in the Philippines, however, to our knowledge, no existing study has utilized MODIS surface reflectance products for water quality assessment of different inland water bodies in the country (Hafeez, 2002 and Boschetti, 2015). In this study, different water quality parameters such as the total suspended solids (TSS), total nitrogen (TN), total phosphorus (TP), pH, and dissolved oxygen (DO) will be modeled using multiple linear regression (MLR) and cascade-forward feed neural networks (CFF-NN) implemented in MATLAB.

2. METHODOLOGY

Water quality data collected by the Laguna Lake Development Authority (LLDA) spanning from 1999 to 2010 were used in this study. Samples for DO, TN, TCl, pH, TP, and TSS were collected monthly in 4 different stations and analyzed in the laboratory. Satellite data (MOD09GA) coincident with the ground measurements were collected using the GEE platform. Spatial and spectral subsetting was conducted using the GEE platform to retrieve the image pixels coincident with the available ground data. In this manner, rapid extraction of relevant information from a decadal dataset was done.

For the water quality modeling, the total number of matching satellite and ground data were divided into training and validation dataset. Using stratified random sampling, two-thirds of the data were used for training/developing the cascade forward-feed neural network (CFF-NN) and multiple linear regression (MLR) models while the remaining one-third were used for the validation of the models. The Vis-NIR bands (band 1- 4) from MOD09GA were used as inputs in the MLR and CFF-NN models. Finally, neural network models for dissolved oxygen, pH, TN, TP, Cl, and TSS are validated and analyzed. A more detailed processing workflow is illustrated in Figure 1. Note that there are varying number of measurements per water quality parameter. Hence, the total number of training and validation points differs for the DO, TSS, TCl, TN, TP, and pH. In general, the total number of training and validation points range from 70 to 100 and 30 to 50, respectively, The Coefficient of Determination (**R^2**) and Root Mean Square Error (**RMSE**) were used as metrics for the accuracy assessment.

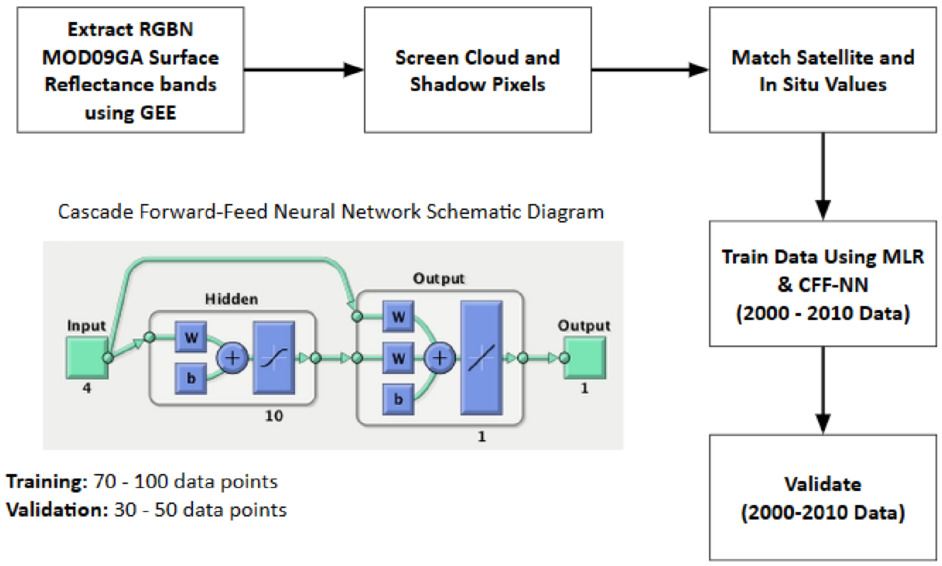

Figure 1. The processing workflow for water quality modeling and validation. The RGBN bands correspond to the Red (B1), Green (B4), Blue (B3), and NIR (B2) bands of MOD09GA.

3. RESULTS & DISCUSSION

Table 1 shows the respective coefficient of determination ($\mathbf{R^2}$) from the MLR and CFF-NN model for each water quality (WQ) parameters studied. Among the WQ parameters, $\mathbf{R^2}$ greater than 0.5 was obtained on TSS and Turbidity for both models while the least $\mathbf{R^2}$ was observed on TN. This indicates that among the WQ parameters the TSS and Turbidity is best described by both models while poor reliability on the TN models based on the four spectral bands of MOD09GA is expected. Table 2 shows the accuracy assessment of the models for each WQ parameter using the validation dataset. The CFF-NN outperforms the MLR model as observed from higher $\mathbf{R^2}$ and lower **RMSE** in all WQ parameters. While a bias may be attributed to CFF-NN due to the optimization usually conducted to train NN models, these results show the potential of neural networks to look for the best model that will describe the WQ parameters. Since the training and validation dataset comprise of ground measurements from 4 different stations for each month, season and location may be the contributing factors that produce the high skill of the CFF-NN. To verify this hypothesis, the training and validation dataset were grouped per season and location. The test was conducted on the TSS since this parameter is known to have an established relationship with the visible and NIR bands, thus, justifies the use of MLR in the grouped dataset. Figure 2 shows the correlation plot of ground TSS with the CFF-NN modeled as well as the MLR deconvolution models. The data of the deconvoluted MLR models were categorized based on the sampling stations and seasonality. For the ***MLR_Season*** model, 2 TSS MLR models were derived, corresponding to the dry (January – June) and wet (July – December) season. Meanwhile, 4 different MLR TSS models corresponding to 4 different sampling stations were derived for ***MLR_Station*** model. For the ***MLR_Season_Station*** model, a total of 8 unique TSS models were derived, corresponding to 2 different models per sampling station accounting for the dry and wet season. It can be seen in Figure 2 that categorizing the dataset according to sampling station and season significantly increase the performance of the MLR model to predict the TSS values. Comparison with the CFF-NN TSS model showed that the ***MLR_Season_Station*** has a better prediction capability, having an $\mathbf{R^2}$ of 0.88 compared to 0.80 of CFF-NN. For visualization. Figure 3 shows the TSS map of Laguna de Bay derived using the MLR and CFF-NN model.

Table 1. Statistical metrics of the training models

Training Models				
Parameters	R^2 (**MLR**)	R^2 (**CFF-NN**)	**n**	**p-value**
TSS (mg/L)	0.68	0.71	63	< 0.001
Turbidity (NTU)	0.52	0.7	62	< 0.001
Alkalinity (mg/L)	0.26	0.37	81	< 0.001
TN (mg/L)	0.16	0.24	73	< 0.1
TP (mg/L)	0.33	0.39	74	< 0.001
Cl (mg/L)	0.32	0.43	88	< 0.001
DO (mg/L)	0.24	0.42	82	< 0.001
pH	0.26	0.49	86	< 0.001

Table 2. Accuracy assessment of MLR and CFF-NN model

	Validation				
	MLR		**CFF-NN**		
Parameters	R^2	**RMSE**	R^2	**RMSE**	**n**
TSS (mg/L)	0.74	10.9	0.8	9.76	33
Turbidity (NTU)	0.73	12.51	0.81	10.86	28
Alkalinity (mg/L)	0.3	19	0.4	18.08	35
TN (mg/L)	0.02	1.11	0.08	1.08	37
TP (mg/L)	0.44	0.074	0.48	0.068	38
Cl (mg/L)	0.35	190.69	0.55	160.51	44
DO (mg/L)	0.09	1.18	0.11	1.17	43
pH	0.09	0.54	0.2	0.51	43

High correlation between TSS/Turbidity and the visible bands, particularly on the red and NIR bands was previously observed in different studies. The concentration of suspended solids is found to be directly related to the backscattering of the incident light in the red and NIR region of the electromagnetic spectrum (Novo, 1991 and Yepez, 2018). Though the TSS concentration is highly correlated with the reflectance at the red and NIR bands, results from different studies show that the linear model may vary regionally, depending on the concentration of other optically-active water constituents as well as the sediment composition and size (Martinez, 2015; Siev, 2018; Santos, 2017). On the other hand, the relationship of optical bands with other WQ parameters such as DO, pH, TN, TCl, and TP has not been fully established. In fact, there are no known spectral signature of these parameters on the optical region of the electromagnetic spectrum. WQ models of these parameters are usually generated using machine learning algorithms. Good prediction of these WQ parameters through the neural network model may be attributed to its ability to determine patterns on the ground dataset potentially originating to the spatial and seasonal variability of the water quality as evident to the test conducted for the TSS modeling. While this limits the derived WQ models to be extended for global waters, these models may provide important insights on the spatiotemporal dynamics of the local inland water body such as the Laguna de Bay. Time-series analysis of these WQ parameters may provide information useful in predicting undesirable events such as fish kill occurrences.

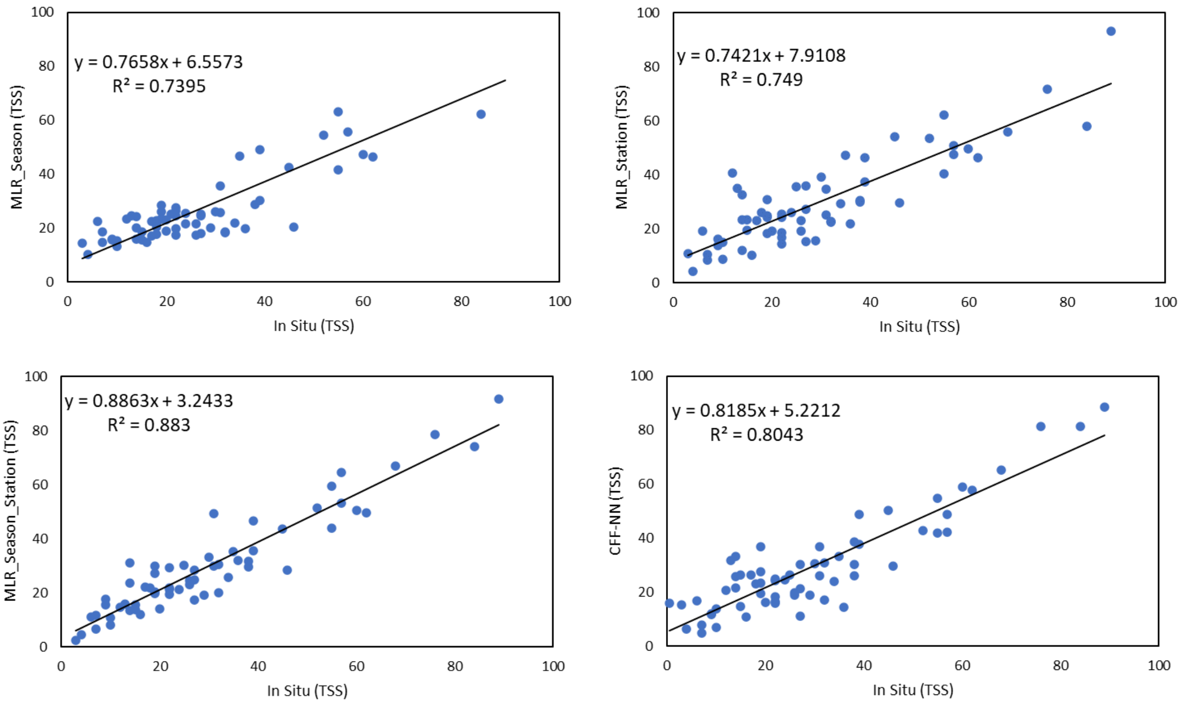

Figure 2. Accuracy assessment of the TSS training models including the deconvoluted MLR models based on season only (MLR_Season), station only (MLR_Station), and both season and station (MLR_Season_Station).

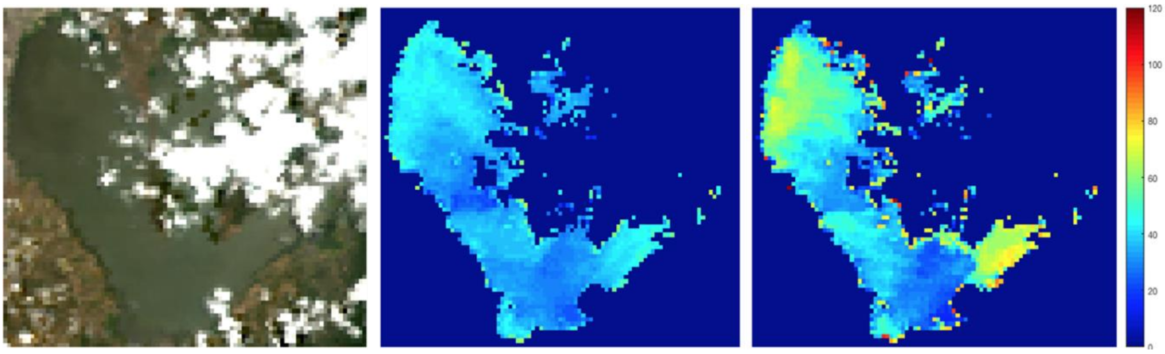

Figure 3. The (Left) RGB and TSS map of Laguna de Bay derived using the (Middle) MLR and (Right) CFF-NN model. The CFF-NN TSS map appears to be more sensitive in detecting change in the TSS values.

4. SUMMARY

Water quality parameters such as TSS, turbidity, alkalinity, TN, TP, pH, and DO were modeled using MLR and CFF-NN models. Results show optimal performance from the CFF-NN model in all WQ parameters, indicating its potential for long term monitoring of the Laguna lake. The high skill performance of the CFF-NN is attributed to its ability to identify the pattern in location and season of the dataset during the process of model optimization. Among the WQ parameters, the highest correlation with the MOD09GA spectral bands was obtained from TSS. To further improve the accuracy of the models, additional ground measurements for calibration-validation dataset is recommended.

5. ACKNOWLEDGMENT

The authors of this study would like to acknowledge the financial support of the Department of Science and Technology (DOST) through the Space Technology and Applications Mastery, Innovation, and Advancement (STAMINA4Space) Program. The authors would also like to thank the Laguna Lake Development Authority (LLDA) for the ground data of different water quality parameters used in this study.

References from Journals:

Xiong, X., 2020. MODIS and VIIRS Calibration and Characterization in Support of Producing Long-Term High-Quality Data Products. Remote Sensing 12 (19), pp. 1- 28.

Liu, B., 2019. Multi-decadal trends and influences on dissolved organic carbon distribution in the Barataria Basin, Louisiana from in-situ and Landsat/MODIS observations. Remote Sensing of Environment 228, pp. 183 – 202.

Chander, G., 2009. Monitoring on-orbit calibration stability of the Terra MODIS and Landsat 7 ETM+ sensors using pseudo-invariant test sites. Remote Sensing of Environment 114 (4), pp. 925 – 939.

Franz, B.A., 2006. MODIS Land Bands for Ocean Remote Sensing Applications, Proceedings of Ocean Optics XVIII

Yang, G., 2018. Using 250-M Surface Reflectance MODIS Aqua/Terra Product to Estimate Turbidity in a Macro-Tidal Harbour: Darwin Harbour, Australia. Remote Sensing 10 (7), pp. 1 – 26.

Wang, M., 2010. Water property monitoring and assessment for China's inland Lake Taihu from MODIS-Aqua measurements. Remote Sensing of Environment 115 (3), pp. 841 – 854.

Miller, R., 2004. Using MODIS Terra 250 m imagery to map concentrations of total suspended matter in coastal waters. Remote Sensing of Environment 93 (1-2), pp. 259 – 266.

Gorelick, N., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, pp. 18 – 27.

Liu, X., 2018. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. Remote Sensing of Environment 209, pp. 227 – 239.

Cuvin-Aralar, M., 2001. Incidence and causes of mass fish kill in a shallow tropical eutrophic lake (Laguna de Bay, Philippines). In: The 9th International Conference on the Conservation and Management of Lakes. Conference proceedings, pp. 233–236.

Santos-Borja, A., 2003. Experience and lessons learned brief for laguna de bay, Philippines. In: The Lake Basin Management Initiative ILEC/LakeNet Regional Workshop for Asia: Sharing Experience and Lessons Learned in Lake Basin Management.

Hafeez, M., 2002. Estimation of crop water deficit through remote sensing in Central Luzon,

Philippines. IEEE International Geoscience and Remote Sensing Symposium.
Boschetti, M., 2015. Rapid Assessment of Crop Status: An Application of MODIS and SAR Data to Rice Areas in Leyte, Philippines Affected by Typhoon Haiyan. Remote Sensing 7 (6), pp. 6536 - 6557.

Novo, E., 1991. Results of a laboratory experiment relating spectral reflectance to total suspended solids. Remote Sensing of Environment 36 (1), pp. 67–72.

Yepez, S., 2018. Retrieval of suspended sediment concentrations using Landsat-8 OLI satellite images in the Orinoco River (Venezuela). Comptes Rendus Geoscience 350, pp. 20–30.

Martinez, J., 2015. The optical properties of river and floodplain waters in the Amazon River Basin: Implications for satellite-based measurements of suspended particulate matter. JGR Earth Surface 120 (7), pp. 1274 – 1287.

Siev, S., 2018. Sediment dynamics in a large shallow lake characterized by seasonal flood pulse in Southeast Asia. Science of the Total Environment, 631–632, 597–607

Santos, A., 2017. Purus River suspended sediment variability and contributions to the Amazon River from satellite data (2000–2015). Comptes Rendus Geoscience 350 (1 – 2), pp. 13 – 19.