16th International Conference on Environmental Science and Technology

Rhodes, Greece, 4 to 7 September 2019

Prediction of Algal Bloom Occurrence in Laguna Lake, Philippines using Artificial Neural Networks (ANN)

Esguerra G. [1,*] **, Ballesteros F.** [2]

[1,2] Environmental Engineering Graduate Program, University of the Philippines Diliman, Quezon City, Metro Manila, Philippines

*corresponding author: e-mail: givetteesguerra@yahoo.com

Abstract

Algal blooms pertain to an undesirable formation of unicellular freely-floating algal scum caused by the rapid growth of phytoplankton, which can become a hazard for the water body ecosystem. Laguna Lake serves as both a source of livelihood and water supply for the residents in the region and the risk of algal blooms should be detected for safe and efficient management. The research presents a method for predicting the amount of phytoplankton to alert the monitoring agencies of incidences of high phytoplankton as a scalable and inexpensive early-warning tool. The study focuses on the development of a prediction model based on water quality parameters measured by the Laguna Lake Development Authority (LLDA) from 2008 to 2018: nitrate, orthophosphate, water temperature, turbidity, chlorophyll-a, and phytoplankton counts. The system predicts the phytoplankton counts of the next month using three months of previous values of the water quality parameters, modeled through the multilayer perceptron neural network method. The research uses a walk-forward validation method to obtain the root-mean-square-error (RMSE) of the model. The model was used on three stations and these predicted values that had statistically less RMSE than the ordinary least square regression.

Keywords: algal blooms, phytoplankton counts, artificial neural networks, *Laguna de Bay*, water quality prediction

1. Introduction

Inland freshwater bodies are necessary sources of water supply, venues for aquaculture, and serve as essential elements for industrial processes, among others. Due to the usage and valuation of these water bodies, proper monitoring and evaluation by relevant agencies is necessary to safeguard stakeholders of possible health threats. The Philippines is home to many freshwater bodies, and one such body is the largest lake known as Laguna Lake or Laguna de Bay. The lake largely contributes to the aquaculture of the area, and recently serves as a source for water supply for nearby residents. Unfortunately, the lake has been noted to experience various periods of algal blooms of the *Microcystis aeruginosa* specie (Baldia, et al., 2003), prompting the need for a prediction model to prepare the relevant agencies of threats. Algal blooms are regarded as rapid growth of photosynthetic unicellular microorganisms, or phytoplanktons (Egerton, et al., 2014). The blooms can harm the water body either by directly containing toxins that are considered as health hazards, or by indirectly reducing the dissolved oxygen of the lake, which prompts the need to monitor and forecast their development (Stumpf, 2001). Based on the study by Baldia, et al. (2003) on the algal bloom species in Laguna Lake, bloom trends were noticed to vary due to water transparency, turbidity, water temperature, and the presence of enough nutrients. Additionally, one of the commonly-used indicators for algal biomass is chlorophyll-a concentration, which measures the amount of the primary pigment for photosynthesis (Stumpf, 2001). Knowing factors that can affect the increase in phytoplankton counts, a predictive model can be created for algal blooms. However, given the funding and capacity restrictions of agencies in developing countries, a cheap, flexible, and scalable model is necessary. Relatedly, there is a notable rise in attention to machine learning algorithms and a lot of forecasting has been done using artificial neural networks (ANN) with different algal bloom indicators (). Based on the need for a prediction for algal bloom, but using a different bloom indicator than most ANNs employ, the research aims to create an ANN model that predicts the phytoplankton counts using the following water quality parameters: nitrate, orthophosphate, water temperature, turbidity, and chlorophyll-a.

2. Methodology

Water quality historical data was requested from the agency assigned to monitor Laguna Lake, which is the Laguna Lake Development Authority (LLDA). The request comprised of 5 input water quality parameters and the phytoplankton counts for the covered years of 2008-2018, for three different stations as shown in **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**.

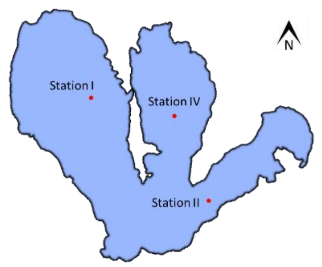

Figure 1. Sampling stations of Laguna Lake.

A baseline model using the ordinary least square (OLS) regression was employed as used by Lee & Lee (2018) and the performance of the prediction model was compared to the OLS by computing the root-mean-square error of each:

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{j=1}^{n}\left(\boldsymbol{value}_j - \boldsymbol{mean\ of\ \widehat{value}_j}\right)^2}{\boldsymbol{n}}}$$

The ANN model was setup using the Python programming language with the Keras and TensorFlow libraries. The input values were lagged in the dataset and transformed with the Quantile Transformer based on Bogner, et al. (2012). The dataset was divided into the training, internal validation, and out of sample testing validation partitions based on the walk-forward analysis (Cao & Tay, 2003) shown in **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**. The model will be trained with 50% of all the data available, while the next 15% will serve as the internal validation for the hyperparameter tuning of the training, then the next 15% will serve as the performance evaluator. The next iteration will move all the data by one timestep and the next 50% of the data will be trained, with the next 15% as validation, and the next 15% as test. The process will be repeated until the end of the data has been reached by the test validation dataset.

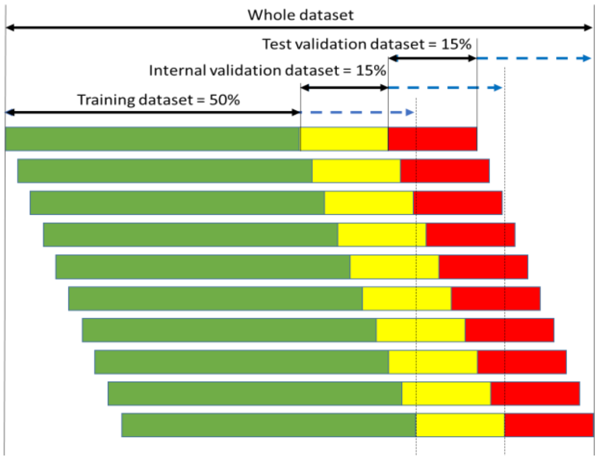

Figure 2. Walk-forward analysis setup.

3. Results and Conclusion

The results of the walk-forward validation is shown as a boxplot of RMSE values in Figure 3. The boxplot presents the OLS RMSE values in red-green, which lies above all the means of the ANN model.

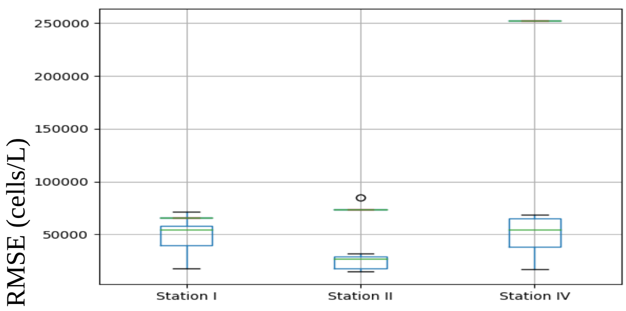

Figure 3. Boxplot of RMSE values from ANN.

The high RMSE for OLS indicates a poor performance of OLS compared to ANN. Upon further confirmation, a Student's t test comparison was made as shown in Table 1. In the table, the ANN model has significantly less RMSE than the OLS model, indicating that the ANN model performed better than the OLS model.

Table 1. Results of a Student's t test between ANN and OLS.

RMSE/Station	**I**	**II**	**IV**
OLS Mean	65,740.38	73,492.07	251,979.66
ANN Mean	49,373.24	29,101.66	49,828.06
ANN Standard deviation	14,892.58	19,518.23	17,630.41
t Statistic	-3.65	-7.54	-38.0287
P(T<=t) two tailed	0.0045	1.96E-05	3.77E-12
t Critical two tailed	2.23		

Based on the successful creation of the model and its subsequent evaluation, the research was able to establish an ANN model that can predict the next-month value of phytoplankton counts with better RMSE for the three stations as compared to the OLS regression model for Laguna de Bay, Philippines.

References

Baldia, S. F., Conaco, M. C. G., Nishijima, T., Imanishi, S., & Harada, K. I. (2003). Microcystin production during algal bloom occurrence in Laguna de Bay, the Philippines. *Fisheries Science*. doi:10.1046/j.1444-2906.2003.00594.x

Bogner, K., Pappenberger, F., & Cloke, H. L. (2012). Technical Note: The normal quantile transformation and its application in a flood forecasting system. *Hydrology and Earth System Sciences*. doi:10.5194/hess-16-1085-2012

Cao, L. J., & Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*. doi:10.1109/TNN.2003.820556

Egerton, T., Morse, R., Marshall, H., & Mulholland, M. (2014). Emergence of Algal Blooms: The Effects of Short-Term Variability in Water Quality on Phytoplankton Abundance, Diversity, and Community Composition in a Tidal Estuary. *Microorganisms*. doi:10.3390/microorganisms2010033

Lee, S., & Lee, D. (2018). Improved prediction of harmful algal blooms in four major South Korea's rivers using deep learning models. *International Journal of Environmental Research and Public Health*. doi:10.3390/ijerph15071322

Stumpf, R. P. (2001). Applications of Satellite Ocean Color Sensors for Monitoring and Predicting Harmful Algal Blooms. *Human and Ecological Risk Assessment: An International Journal*. doi:10.1080/20018091095050