Prediction of primary production in Laguna de bay using artificial neural network

Jonah John G. San Pedro
Institute of Environmental Science and Meteorology
University of the Philippines - Diliman

ABSTRACT

Understanding the link between water quality and primary production plays a crucial role in managing lake ecosystems effectively. While frequent and rapid measurement of primary production is desirable, the method, however, is costly in terms of labor, time, and financial resources. This study proposes a practical scheme to predict primary production using non-linear autoregressive with exogenous input (NARX) neural network coupled with principal component analysis (PCA) based on water quality data inputs. The three main bays in Laguna de Bay (west, central, and east bay) are considered as the study site. NARX networks are constructed based on water quality data inputs selected from PCA where it reveals weak linear correlation between water quality variables and primary production using the data collected from 1999-2016. Nonlinear interaction may exist among these factors and therefore nonlinear methods such as NARX are needed to capture this complex relationship. The dataset was also categorized according to seasonality (dry and wet) and whole-year round data to determine the influence of seasonality on predictive modeling. During the input selection process, PCA identified the significant variables that suggestively represent water quality data in the lake. Using these extracted inputs from the three different dataset, NARX networks are developed for west, central, and east bay through repeated training and testing until such time that minimal performance difference is observed between the two phases in terms of mean square of errors (MSE) and root mean square of errors (RMSE). These networks are then applied for one-step-ahead and multiple-step-ahead prediction using unseen input values and errors associated with predictions are evaluated.

NARX developed from dry season data produced relatively more accurate predictions of primary production for central and east bay than networks developed using wet season and annual data, while NARX developed using annual data produced the most accurate prediction for west bay. Accurate predictions are recorded when the standard deviation (z-score) of the target value is less regardless of the prediction length. In addition, based on MSE, RMSE, and correlation of target and outputs, NARX networks outperform Multiple Linear Regression (MLR) models for three bays which validates previous findings indicating ANNs and NARX in particular are more effective in predicting non-linear dynamic systems than traditional approaches. This study demonstrates applicability of NARX to water quality forecasting and specifically finds the proposed scheme as a useful alternative to primary production in-situ methods which can help lake managers to immediately interpret and/or estimate primary production response to changes caused by natural or manmade pollution.