Relationship analysis of N/P ratio and phytoplankton abundance in Ranu Pakis using regression approach

E D Lusiana¹, M Musa¹, M Mahmudi¹, N R Buwono¹ and K P Utami

¹ Department of Water Resources Management, Faculty of Fisheries and Marine Science, Universitas Brawijaya, Malang, Indonesia

*Corresponding author: evellinlusiana@ub.ac.id

Abstract. Ranu Pakis is a natural lake that formed by volcanic activity. The lake waters have many benefits including as a source of irrigation, aquaculture, agriculture, and tourism. The activities around Ranu Pakis have the potential to produce organic waste into the waters. High organic waste and nutrients resulted in eutrophication or high abundance of phytoplankton which threatens biodiversity in the lake. Nutrients in the waters consist of N and P. The ratio between N and P affect the composition and abundance of phytoplankton. The purpose of this study was to analyse the relationship between N/P ratios to the abundance of phytoplankton in Ranu Pakis using quantile regression approach. The results of this study indicated that the effect of the N/P ratio was different on various quantile

1. Introduction

Lake is a basin that contains water both temporarily and permanently in which part of water system in land. Most lakes found in Indonesia are the results of volcanic activities [1]. Ranu Pakis is one of the volcanic lakes found in Ranu Pakis Village, East Java. This lake located in between Ranu Bedali and Ranu Klakah, so they are usually called as Ranu’s triangle. Ranu Pakis has been widely utilized by local communities for many purposes such as agricultural irrigation, aquaculture by using floating net cages, and tourism. It resulted direct discharge the untreated waste to the water. Therefore, the fertility of water will be affected as well as the aquatic organisms there. One of important organisms in the lake is phytoplankton [2].

Phytoplankton has crucial role as the natural feed for other organisms since it is the primary producer in aquatic food chain due to its ability to perform photosynthesis. The growth of phytoplankton supports by the availability on nutrients. Important nutrients for phytoplankton are nitrate and phosphate [3]. Therefore, the composition and abundance of phytoplankton are limited by nutrients. Nitrate and phosphate usually used as the benchmark that determines phytoplankton abundance. It is known as Redfield ratio. The common Redfield ratio to describe elemental compositions of N:P is 16:1. Lv et al., [4] stated that the phytoplankton biomass was dependent on total phosphorus (P) when N:P by weight was >17, on nitrogen (N) when N:P was < 10.

Excess nutrients will lead to excessive growth of phytoplankton population, or a bloom. A phytoplankton bloom may lead to mass mortalities of fish as happened in many different parts of the world [5]. The nature relationship between eutrophication and the expansion of algal blooms is still unclear, although in general eutrophication leads to explosions in the phytoplankton population. Increasing phytoplankton blooms coincide with nutrient enrichment. How eutrophication can stimulate the presence of toxic algal bloom species, is not yet completely understood and is still under debate.
Discharge of any organic material to the waters can lead to an overall increase in nutrient availability, and changes in nutrient composition or nutrient ratio. In general, nutrient availability is associated with increased biomass, and can lead to changes in the community. Many researchers have demonstrated quantitative approach to analyze the relationship between nutrient ratio and phytoplankton abundance by using linear regression [6,7,8]. However, this method has several disadvantages. It does not robust to outlier and merely models the data in mean basis. Therefore, it cannot to be used to estimate the nutrient limitation in order to prevent algal bloom. As the alternative, quantile regression is considered [9,10,11]. The latter method outweights linear regression in terms of dealing the outlier and modelling data in all quantile points. This study aims to apply the quantile regression to analyze relationship between N/P ratio phytoplankton abundance and to estimate nutrient limitation for high level abundance of phytoplankton in Ranu Pakis.

2. Materials and Methods
The objects of this research were nitrate and phosphate (mg/l) ratios and phytoplankton abundance (ind/l) in Ranu Pakis that collected during January-February 2019. The sampling location is displayed in Figure 1.

![Figure 1. Sampling location in Ranu Pakis.](image)

In general, the procedure of the study was summarized as the trophic status index (TSI) of Ranu Pakis. The assessment carried out included the eutrophication level, data exploration to identify outlier using Cooks’ distance criteria, comparison the result of linear regression and quantile regression, and estimation of nutrients limit related to high phytoplankton abundance. The analysis was then conducted using R version 3.4.3.

2.1. Trophic State Index (TSI)
Trophic status classification using TSI presented in Table 2 [12].

| Table 1. TSI Levels of Waters Eutrophication. |
|-----------------------------------------------|
| Index | Classification    |
|<30    | Ultraoligotrophic |
|30–40  | Oligotrophic     |
|40–50  | Meso-trophic    |
50 – 60  Light Eutrophic
60 – 70  Medium Eutrophic
70 – 80  Heavy Eutrophic
>80 Hyper-eutrophic

TSI score can be calculated using formula following formulas
\[ TSI(\text{SD}) = 60 - 14.41 \ln(\text{SD}) \]
\[ TSI(\text{CHL}) = 30.6 + 9.81 \ln(\text{CHL}) \]
\[ TSI(\text{TP}) = 4.15 + 14.42 \ln(\text{TP}) \]

2.2. Cooks distance

Zakaria et al. [13] stated that the existence of outlier detected by using numerical statistics called Cooks distance. The formula of this measure showed in equation (1).

\[ D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})'X'X(\hat{\beta}_{(i)} - \hat{\beta})}{(k + 1)\hat{s}^2} \]  

where
\( \hat{\beta}_{(i)} \) = vector of parameter estimate when the ith observation is deleted
\( \hat{\beta} \) = vector of parameter estimate using all observations
\( X \) = matrix of predictors and constant
\( k \) = number of parameter
\( \hat{s}^2 \) = variance of the fitted values

Data samples which have Cook’s distance more than four times of mean value may be categorized as outlier or influential data [13].

2.3. Quantile regression

Quantile regression is a method to depict the relationship between variables at different quantile points at dependent variables distribution. Hence, it can be performed non-homogeneous data [14]. The quantile regression model is displayed in equation (1)

\[ y_i = x_i'\beta(\theta) + \varepsilon(\theta)_i \]  

where
\( y_i \) = response variable of the ith observation
\( x_i \) = \((1, x_{i1}, x_{i2}, \ldots, x_{ip})\)
\( \beta(\theta) \) = parameter regression at the \( \theta \)th quantile
\( \varepsilon(\theta)_i \) = error/residual model of the \( \theta \)th quantile
\( i = 1, 2, \ldots, n \)

Estimation of quantile regression parameters can be re-written as the minimization formula as follows in equation

Error! Reference source not found.
\[
\min_{\beta \in \mathbb{R}^p} \left[ \sum_{i \in \{k : x_i \geq \beta\}} \rho_{\theta}(y_i - f(x_i)) \right]
\]

(3)

Where \( \rho_{\theta}(u) = (\theta - 1_{\{u < 0\}})u \) is a check function that presented in Figure 1.

The minimization problem solved by applying linear programming analysis namely simplex method [15].

3. Results and discussion
3.1. Trophic Status Index (TSI) and N/P Ratio in Ranu Pakis
Table 2 shows the trophic status index and N/P Ratio measurement in Ranu Pakis.

| Site | Week | TSI   | Remark          | N/P ratio | Limiting Factor |
|------|------|-------|-----------------|-----------|-----------------|
| 1    | 1    | 73.19 | Heavy eutrophic | 0.75      | N               |
| 2    | 53.20| Light eutrophic | 7           | N         |
| 3    | 69.04| Medium eutrophic | 1.2        | N         |
| 2    | 1    | 60.56 | Medium eutrophic | 3.6      | N               |
| 2    | 69.04| Medium eutrophic | 1.8        | N         |
| 3    | 53.20| Light eutrophic  | 3           | N         |
| 3    | 1    | 57.34 | Light eutrophic | 2.5      | N               |
| 2    | 70.56| Heavy eutrophic | 0.6        | N         |
| 3    | 47.35| Mesotrophic    | 6           | N         |
| 4    | 1    | 53.20 | Light eutrophic | 3.7      | N               |
| 2    | 60.56| Medium eutrophic | 3.4        | N         |
| 3    | 53.20| Light eutrophic | 1.7        | N         |
| 5    | 1    | 65.41 | Medium eutrophic | 13      | N and P         |
| 2    | 60.56| Medium eutrophic | 2.4        | N         |
According to Table 1, the eutrophication status in Ranu Pakis was varied from oligotrophic to heavy eutrophic. The high level category of eutrophication related to the low N/P ratio in the waters where nitrate as the limiting factor. Therefore, the abundance of phytoplankton in Ranu Pakis was mainly dependent to the nitrate concentration. Vrede et al., [16] suggested that low N/P ratio would promote the growth of N-fixing cyanobacteria, and it was proved in this research that more than 40% of phytoplankton biomass in Ranu Pakis come from Cyanophyta division.

3.2. Outlier identification

Data examination to identify outliers in the relationship of N/P ratio and phytoplankton abundance can be seen in Figure 3.

![Influential Observations by Cook's Distance](image)

**Figure 3.** Cooks distance of N/P ratio and phytoplankton abundance.

It can be seen from Figure 2 that outlier existed in the 8th data sample. As the consequence, linear regression was not suitable to model the relationship between N/P ratio and phytoplankton abundance because it was not robust to the outlier. The used of linear regression for data that contain outlier will mislead the results of analysis [17]. Thus, alternative method such as quantile regression must be carried out.

3.3. Outlier identification

The quantile and linear regression analysis between N/P ratio and phytoplankton abundance is presented in Table 3.

| Model         | Distribution | Phyto-abundance (ind/l) | Coefficient of N/P Ratio |
|---------------|--------------|-------------------------|--------------------------|
| Linear regression | Mean         | 13.606                  | 0.184                    |

(810981.1)
Table 3 shows that the coefficients from quantile regression was not constant among quantile, while linear regression only produce one coefficient for the overall data. Therefore, linear regression model cannot well presented the variability of the relationship between phytoplankton abundance N/P ratios. Moreover, using quantile regression method, the N/P ratio threshold to prevent eutrophication can be simply calculated based on the upper quantile (q=0.75). The N/P ratio limit at 2.42, where nitrate as the limiting factor. Furthermore, specific value of nitrate related to eutrophication case then analyze by conducting quantile regression analysis between nitrate and phytoplankton abundance. It revealed that the nitrate level for eutrophication monitoring indicator is accounted for 0.135 mg/l.

4. Conclusion
Linear regression was not suitable to analyse relationship between N/P ratio and phytoplankton abundance in Ranu Pakis since the existence of outlier in the data. Results of quantile regression suggested that N/P ratio threshold corresponding to the eutrophication stood at 2.42, where nitrate as the limiting factor with the level of nitrate accounted for 0.135 mg/l.

References
[1] Yulesty A 2015 *JOM Fekon* 4 1244-54
[2] Suryanto A M 2011 *J. Kelautan* 4 34-39
[3] Donald D B, Bogard M, Finlay K, Bunting L and Leavitt P 2013 *PLoS ONE* 8 1-14
[4] Lv J, Wu H and Chen M 2011 *Linnologica* 41 48-56
[5] Bhagowati B and Ahamad K U 2019 *Ecogy. and Hydrobio*. 19 155-66
[6] Ismaini, Surbakti H and Aryawati R 2014 *Maspari J.* 6 39-45
[7] Paiki K and Kalor J D 2017 *J. of Fisheries and Marine Sci.* 1 65-71
[8] Putri F D M, Widyastuti E and Christiani 2014 a *Scripta Bio.* 1 96-101
[9] Chen C 2005 *An Introduction to Quantile Regression and The Quantreg Procedure* (Colin (Lin) Chen, SAS Institute: NC Statistics and Data Analysis) p 24
[10] Ismail E A-R 2015 *J. of Multidiscipl. Eng. Sci. and Tech.* 2 388-94
[11] Kozumi H and Kobayashi G 2011 *J. of Stat. Comp. and Simu.* 81 1565-78
[12] Carlson R E 1977 *Linn. and Ocean.* 22 361-369
[13] Zakaria A, Howard N K and Nkansah B K 2014 *Ame. J. of Theor. and Appl. Stat.* 3 100-106
[14] Koenker R and Hallock K 2001 *J. of Eco. Persp.* 15 143-56
[15] Koenker R and Machado J A F 1999 *J. of The Ame. Stat. Asso.* 94 1296-310
[16] Vrede T, Ballantyne A, Mille-Lindblom C, Algesten G, Gudasz C, Lindahl S and Brunberg A K 2009 *Freshwater Bio.* 54 331-44
[17] Buhai I S 2005 *Ad. Astra* 4 1-17