Principal Component Analysis (PCA) of Phytoplankton Community Relations Based on Physical-Chemical Structures for Supply Chain Management in the Waters of the Bangka Bay Region of West Bay

Julita Nahar*[1], Elis Hertini[1]

[1]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia.

*julita.nahar@unpad.ac.id

Abstract— **Phytoplankton as one of the most important organisms in the territorial waters, which is the beginning of the food chain in a territorial waters. Phytoplankton have certain tolerance limits to physical-chemical factors so that they will form different community structures. The relationship between chemical physics parameters with phytoplankton communities in a territorial waters, can be used as an indicator of water quality. West Bangka Regency has considerable marine and fisheries resource potential, especially for marine aquaculture. So research on the structure of the phytoplankton community in relation to the physical-chemical parameters of seawater needs to be carried out to see indicators of fertility and availability of natural food in the waters that will be used as a marine aquaculture location. To find out the relationship between physico-chemical parameters and phytoplankton abundance, Principal Component Analysis (PCA) was used. Some physico-chemical parameters observed were temperature, brightness, pH, salinity, coated oxygen (DO), phosphate, nitrate, and ammonia. The results show the eigenvalues value of the main component 1 (PC1) represents about 41.78% of the diversity of data with its main identifying variable, namely temperature with a loading factor value of -0.885, brightness with a loading factor value of -0.8872 and salinity with a loading factor value of 0.824. For the main component 2 (PC2) represents about 28.16% of the diversity of data with its founder variable, pH with a loading factor value of -0.841. As for the main component 3 (PC3), it represents 18.67% of the diversity of data with its originating variable, Nitrate, with a loading factor value of -0,700. So that it can be formed into 3 clusters namely the first cluster is temperature and salinity, the second cluster is pH and nitrate and the third cluster is DO, phosphate, and ammonia.**

***Keywords*—PCA,** *Phytoplankton, community relations, physico-chemical parameters, phytoplankton abundance.*

1. Introduction

Biological resources that have an important role in marine ecosystems are plankton. Plankton is an aquatic organism that plays an important role in maintaining the balance of the ecosystem in the sea [1], [2]. Plankton is divided into two, namely phytoplankton and zooplankton [3], [4]. Phytoplankton act as the main producers that provide organic substances from inorganic substances through the process of photosynthesis and greatly affect the survival of zooplankton and other aquatic consumers living in the waters [5], [6].

Phytoplankton have certain tolerance limits to physical-chemical factors so that they will form different community structures. The relationship between physical-chemical parameters with phytoplankton communities in a territorial waters, can be used as an indicator of water quality [7], [8]. Apart from being the basis of the food chain, phytoplactone is also one of the parameters of the fertility of a waters. Information about this matter is needed to support aquaculture business in a waters because in addition to phytoplankton as the basis of the food chain in the territorial waters, it is also a natural feed needed by aquatic organisms including fish that are kept or cultivated in these waters [9]-[11].

One area that has the potential for the development of marine culture (mariculture) is West Bangka Districts. West Bangka Districts is a coastal district

IJSCM, ISSN: 2050-7399 (Online), 2051-3771 (Print)

on the Main Island, Bangka Island, which has an area of 1,690.28 km^2 of sea waters and small islands scattered around it. This condition makes West Bangka Districts has a large potential of marine and fisheries resources, mainly for marine cultivation or mariculture. Mariculture is the maintenance of marine organisms in seawater, usually in protected coastal waters [12].

Researchers related to the structure of the phytoplankton community have been carried out by previous researchers. For example, research conducted by Tan and Ransangan [13], determined spatial and temporal variations as an effort to increase fishery resources by determining the structure of the phytoplankton community in Marudu Bay. From the results of the study it was found that the structure of the phytoplankton community is more sensitive to the mesotropic environment to changes in zooplankton abundance, nutrient concentrations. In addition, in the research of Zhou et al [14], analysing increased nutritional disorders to the response of marine ecosystems. With the aim of knowing the structure of the phytoplankton community and providing nutritional treatment with nitrates, phosphates, and iron to determine the biodiversity response throughout the tropical ocean. From the results of the study it was found that DIN supply would change the structure of the phytoplankton community and stimulate a regime shift from cyanobacteria to diatoms (relative dominance of R> 0). In a study conducted in Mei et al [15], reconstructed variations in the productivity of phytoplankton and community structure changes in sea level as a result of sedimentation. Based on the results obtained indicate primary productivity and single algae present a consistent trend for the entire core during the middle Pleistocene.

To support the development plan of marine aquaculture in West Bangka Regency, research on the structure of phytoplankton community needs to be done related to the physical-chemical parameters of sea water. This is an indicator of fertility and availability of natural food in territorial waters that will be used as a marine culture location (mariculture). Therefore, this study aims to analyse the structure of the phytoplankton community with respect to the physical-chemical parameters of sea water. This needs to be done to see indicators of fertility and availability of natural food in territorial waters that will be used as a marine aquaculture location. So from the results of the study can find out the relationship between physical-chemical parameters and the abundance of phytoplankton, Principal Component Analysis (PCA) used in the development of marine culture in West Bangka Regency.

2. Literature Review

2.1 Phytoplankton Community Structure

2.1.1 Phytoplankton

Phytoplankton are tiny aquatic plants that live in water, which occupy the position as the first level producers of the food chain in the territorial waters. Phytoplankton have a very important role as primary producers because they are able to absorb sunlight for photosynthesis. Phytoplankton can act as one of the ecological parameters that can describe how the ecological conditions of a territorial waters and is one of the parameters of the fertility of a territorial waters.

2.1.2 Community Structure

Community structure is a collection of various types of microorganisms that interact in a particular zoning. Some factors that are usually used to describe community structure are diversity, uniformity, and dominance. states that the abundance and distribution of phytoplankton in the sea are influenced by many factors both physical, chemical and biological. While according to Parsons et al [16], explained that the biogeographic distribution of plankton is largely determined by environmental factors such as light, temperature, salinity, nutrients and other factors. These factors determine the existence and success of plankton species in an aquatic environment.

2.2 Physical-Chemical Parameters

2.2.1 Physical Parameters

a. Temperature

According to Toral [17], temperature is a measure of the energy of molecular motion. Vital life processes can be affected by temperature and only function in a narrow range, usually between 0-40°C. If it is outside the temperature will be able to damage the function of organs of the organism. Good temperature for plankton in general ranges between 20-30 ºC.

b. Brightness

Brightness is a measure of water transparency that is determined visually using a Secchi disk and expressed in meters. The value of brightness in a waters is influenced by weather conditions, time of measurement, turbidity, suspended solids, and accuracy when measuring. The brightness of a waters will be closely related to the process of phytoplankton photosynthesis in these territoril waters.

2.2.2 Chemical Parameters

a. Acidity (pH)

pH is the amount of hydrogen ions in a solution. The pH of sea water is alkaline. PH values in the sea generally range between 7.5 - 8.4. The highest pH value is found in layers in the surface layer or near the surface layer.

b. Salinity

Salinity is the number of grams of salt dissolved in one kilogram of sea water and expressed in

thousandths. Furthermore it was stated that seawater dissolved in various salts, especially sodium chloride. In addition there are also salts of magnesium, calcium, potassium and so on. The distribution of salinity in the sea is influenced by several factors including circulation patterns, evaporation, rainfall and river flow.

c. Illegal Oxygen (DO)

DO indicates the amount of dissolved oxygen present in water expressed in ppm. The level of dissolved oxygen in natural waters depends on temperature, salinity, water turbulence and atmospheric pressure. Dissolved oxygen levels also fluctuate daily and seasonally depending on mixing, and the movement of water masses, photosynthetic activity, respiration and waste that enters the water [18].

d. Nitrogen element

Nitrogen is present in the forms of protein molecules in dead organisms which are then broken down into inorganic forms by a series of decomposing organisms, especially the formation of nitrates. The next result is nitrate nutrient which is a form that is ready for use by phytoplankton.

2.3 Principle Component Analysis (PCA)

PCA is applied to data that has multicollinearity, because multicollinearity between variables is one violation of the assumptions in cluster analysis [19].

2.3.1. Eigen Value and Eigenvector

If A is a matrix n × n, there is a scalar λ nonzero vector V so that it satisfies the following equation:

$$AV = \lambda V$$

The number λ is the eigenvalue of A and the vector V is called the eigenvector associated with the eigenvalue (λ). To obtain the eigenvalue the equation is rewritten as follows:

$AV = \lambda V$; with $V \neq 0$

$$AV - \lambda V = 0$$

$$AV - \lambda I = 0$$

$$(A - \lambda I)V = 0$$

For λ to be an eigenvalue, there must be a nonzero solution from the Equation. The equation will have zero resolutions if and only if:

$det(A - \lambda I) = 0$ or $|A - \lambda I| = 0$.

2.3.2 Loading Factors

Loading factors are correlations between the original variable and the new variable. The value of loading factors gives an indication of which original variables are most important or affect the formation of new variables. The higher the loading factors, the longer the relevant variable has more influence on the new variable. Loading factors can be calculated using the following formula:

$$l_{ij} = \frac{w_{ij}}{s_j}\sqrt{\lambda_i}$$

with,

l_{ij} = loading of the j-variable for PC i

w_{ij} = the weight of the j-variable to the i-th PC

s_j = standard deviation of the jth variable

λ_i = eigenvalue of the i-th PC

2.3.3 Principle Component Analysis (PCA)

Johnson and Wichern [20] defines the main component as one form of variable transformation which is a linear combination of variables. The process of forming the main component is to determine the matrix which is the observation data of rainfall. From the X matrix, then calculate the covariance matrix, to determine the eigenvalue (λ). Based on the eigen matrix, the main components (PCs) that are formed are:

$$PC_1 = z_j v_{j1} = z_1 v_{11} + z_2 v_{21} + \ldots + z_p v_{j1}$$
$$PC_2 = z_j v_{j2} = z_1 v_{12} + z_2 v_{22} + \ldots + z_p v_{j1}$$
$$\vdots$$
$$PC_p = z_j v_{jp} = z_1 v_{1p} + z_2 v_{2p} + \ldots + z_p v_{jp}$$

with :

PC_1 = main component 1

PC_2 = main component 2

The criteria used to determine the number of components that can be formed are the percentage variance criteria. The number of main components that will be used in cluster analysis is that which has a cumulative percentage of variance of at least 80% (Rencher, 2001). The next step is to calculate the score component (PCj) that will be used as input for cluster analysis. The score component obtained from the main component m (where m <p) will be used in further analysis instead of the initial variable data value. The score component of the results of the analysis of the main components with the original data (raw data) as the input of the analysis is determined by [20]:

$$y_{i1} = e_1 x_i$$
$$y_{i2} = e_2 x_i$$
$$\vdots$$
$$y_{ik} = e_k x_i$$

Results and Discussion

To determine the relationship between physics-chemical parameters and the abundance of phytoplankton in the waters of the West Kelabat Bay used Principal Component Analysis (PCA). Some physico-chemical parameters observed were temperature, brightness, pH, salinity, coated oxygen (DO), phosphate, nitrate, and ammonia. The results of the analysis using PCA will get the root value (eigen value), the cumulative value of variance (loading factors), and the correlation matrix. The analysis results are processed using *Software R 3.4.1.*

3.1 Eigen Value

Eigen value is used to measure the amount of variation maintained by each major component. The eigen value in main component 1 will be greater than the eigen value in the next main component. More than one eigen value is used as a cutoff in selecting the main components that are maintained.

Table 1. Value of the Root Character (Eigen Value) of Physical-Chemical Parameters

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	3.34	2.25	1.49	0.52	0.33	0.05	0.00
Variance (%)	41.79	28.16	18.67	6.54	4.17	0.67	0.00
Cumulative variance (%)	41.79	69.95	88.62	95.16	99.33	100.00	100.00

The results of data analysis in Table 1 show that the main component compilers can be done with 3 main components, namely main component 1, main component 2, and main component 3. The three main components can explain 88.61% of the total variation. While the other main components are not used because they have a fairly low value in explaining the total variation. An alternative method for determining the number of major components is to look at the Scree Plot, which is a plot of the eigen value that is drawn from the largest to the smallest value. The number of components is determined at that point, beyond that the remaining eigen values are all relatively small and of comparable size.

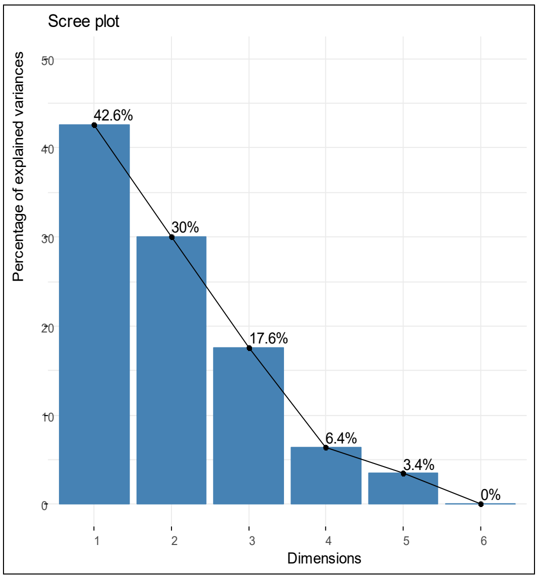

Figure 1. ScreePlot

3.2 Loading Factors

Table 2. Loading Factors on Physical-Chemical Parameters

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Temperature	-0.885	-0.006	-0.410	-0.016	-0.218	-0.040	0.000
Brightness	-0.872	-0.429	-0.234	0.006	0.018	-0.014	0.000
pH	-0.433	-0.841	-0.145	0.272	0.087	-0.061	0.000
Salinity	0.824	0.059	-0.365	0.175	-0.383	-0.083	0.000
Oxygen. Dissolved	-0.632	0.498	0.503	-0.275	-0.082	-0.132	0.000
Phosphate	0.607	-0.632	0.454	-0.005	0.112	-0.121	0.000
Nitrate	0.202	0.590	-0.700	0.002	0.334	-0.095	0.000
Ammonia	0.354	-0.603	-0.409	-0.586	-0.029	0.011	0.000

The results of the data analysis in Table 2 represent the loading factors of each major component. The eigen values table shows that the main component 1 (PC1) represents about 41.78% of the diversity of data with its main identifying variable being temperature with a loading factor value of -0.885,brightness with a loading factor value of -0.887 and salinity with a loading factor value of 0.824. For the main component 2 (PC2) represents about 28.16% of the diversity of data with its founder variable, pH with a loading factor value of -0.841. As for the main component 3 (PC3), it represents 18.67% of the diversity of data with its originating variable, Nitrate, with a loading factor value of -0,700.

3.3 Graph Analysis Based on Parameters and Stations

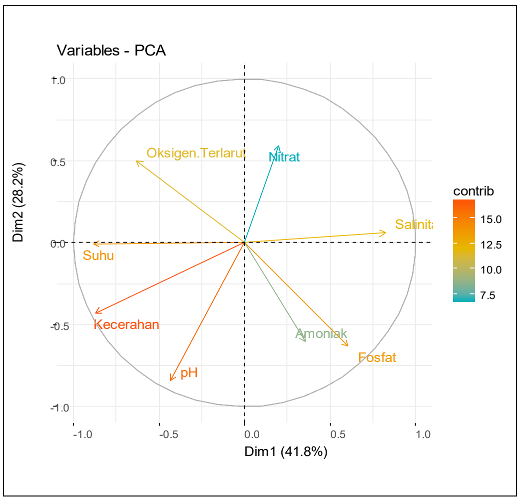

Figure 2. Graph of Results of PCA Analysis of Observed Parameters

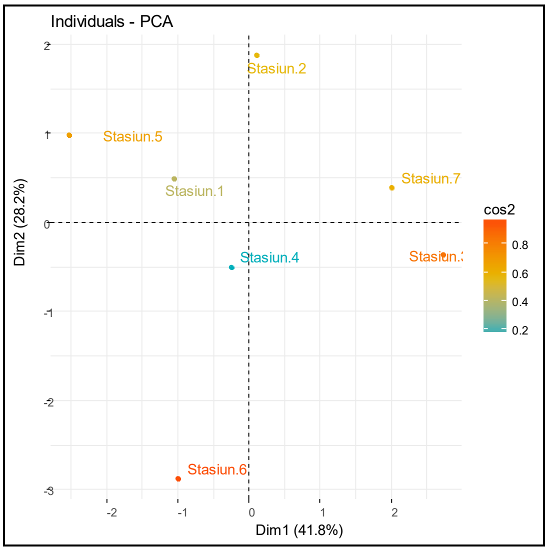

Figure 3. Distribution of Stations Based on Parameters That Affect

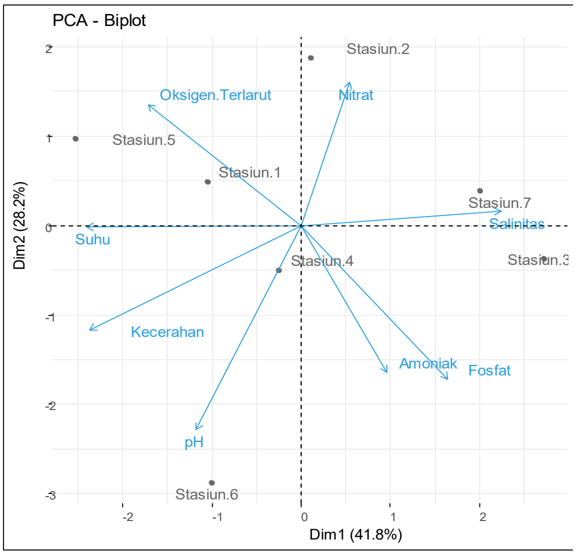

Figure 4. Graph of PCA Results on Parameter and Its Distribution

In Figure 2 is the contribution of variables in calculating the variability in the main component. Variables relating to PC1 are located close to axis 1 (Dim.1). The variable that is most important explains the variability of data. Whereas the variables related to PC 2 are located close to axis 2 (Dim.2). It can be seen from the figure that the one close to axis 1 (Dim.1) is the temperature and salinity parameters, and the one close to the axis 2 (Dim.2) is the pH parameter. Whereas variables that are far from axis 1 and axis 2 have a low contribution in the analysis of the main components. The results of the figure in Figure 3 are the distribution of station groupings based on parameters that affect. Station 1, station 3 and station 7 are adjacent to axis 1 (Dim.1), while station 2, station 4 and station 6 are adjacent to axis 2 (Dim.2).

Merging Figure 2 and Figure 3 can be seen in Figure 4. Figure 4 shows the proximity of a station based on related parameters, where adjacent stations have the same value of these parameters. Station 1, station 3, and station 7 are close to axis 1 (Dim.1) with parameters that characterize them are temperature and salinity. While station 2, station 4, and station 6 are close to axis 2 (Dim.2) with parameters that characterize it are pH and nitrate. However nitrate has a fairly low contribution (see figure 2). Stations that are located far from axis 1 or axis 2 show differences in characteristics from other stations with no parameters that characterize one axis.

3. Conclusions

From the results of the above analysis it can be concluded that the physical-chemical parameters that affect the abundance of phytoplankton can be divided into 2 clusters. This grouping is based on the distribution of stations to the parameters that affect. The first cluster is at station 1, station 3, and station 7 with parameters that characterize it are temperature and salinity, the second cluster is at station 2, station 4, and station 6 with parameters that characterize it are pH and nitrate. Parameters and stations that are located far from axis 1 or axis 2 show the differences in characteristics of other stations in the absence of parameters that characterize one axis and have a low contribution in the analysis of the main components.

Acknowledgments

We would like to thank the Academic Leadership Grant (ALG) led by Prof. Dr. Sudrajat Supian for all support given

References

[1] P. Serret, E. Fernández, C. Robinson, E. M. S.Woodward and V. Pérez. 2006. Local production does not control the balance between plankton photosynthesis and respiration in the open Atlantic Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*. Volume 53, Issues 14–16, pp. 1611-1628.

[2] N. L. Silva, C. R. Marcolin and R. Schwamborn. 2019. Using image analysis to assess the contributions of plankton and particles to tropical coastal ecosystems. *Estuarine, Coastal and Shelf Science.* Volume 219, pp. 252-261.

[3] Z. Chen, Z. Tian, S. Zhang and C. Wei. 2020. The stationary distribution and ergodicity of a stochastic phytoplankton–zooplankton model with toxin-producing phytoplankton under regime switching. *Physica A: Statistical Mechanics and its Applications.* Volume 537, p. 122728.

[4] W. Zheng and J. Sugie. 2019. Global asymptotic stability and equiasymptotic stability for a time-varying phytoplankton–zooplankton–fish system. *Nonlinear Analysis: Real World Applications.* Volume 46, pp. 116-136.

[5] M. Javidi and B. Ahmad. 2015. Dynamic analysis of time fractional order phytoplankton–toxic phytoplankton–zooplankton system. *Ecological Modelling.* Volume 318, pp.8-18.

[6] Y. Wang, H. Wang and W. Jiang. 2014. Hopf-transcritical bifurcation in toxic phytoplankton–zooplankton model with delay. Journal of Mathematical Analysis and Applications. 415(2), pp. 574-594.

[7] J. Goodman, M. A. Brzezinskia, E. R. Halewood and C. A. Carlson. 2012. Sources of phytoplankton to the inner continental shelf in the Santa Barbara Channel inferred from cross-shelf gradients in biological, physical and chemical parameters. *Continental Shelf Research.* Volume 48, pp. 27-39.

[8] Y. Lee, E. J. Yang, J. Park, J. Jung, T. W. Kim and S. H. Lee. 2016. Physical-biological coupling in the Amundsen Sea, Antarctica: Influence of physical factors on phytoplankton community structure and biomass. *Deep Sea Research Part I: Oceanographic Research Papers*. Volume 117, pp. 51-60.

[9] J. Pahissa, J. Catalan, G. Morabito, G. Dörflinger, J. Ferreira, C. Laplace-Treyture, R. Gîrbeag, A. Marchettoc, P. Polykarpouh and C. Hoyos. 2015. Benefits and limitations of an intercalibration of phytoplankton assessment methods based on the Mediterranean GIG reservoir experience. *Science of the Total Environment.* Volume 538, pp. 169-179.

[10] M. Balci and N. Balkis. 2017. Assessment of phytoplankton and environmental variables for water quality and trophic state classification in the Gemlik Gulf, Marmara Sea (Turkey). *Marine Pollution Bulletin*. Volume 115, Issues 1–2, pp. 172-189.

[11] G. Zhao, X. Gao, C. Zhang and G. Sang. 2020. The effects of turbulence on phytoplankton and implications for energy transfer with an integrated water quality-

ecosystem model in a shallow lake. *Journal of Environmental Management*. Volume 256, p. 109954.

[12] J. Yu and Q. Han. 2020. Food security of mariculture in China: Evolution, future potential and policy. *Marine Policy*. Volume 115, p. 103892.

[13] K. S.Tan and J. Ransangan. 2017. Effects of nutrients and zooplankton on the phytoplankton community structure in Marudu Bay. *Estuarine, Coastal and Shelf Science*. Volume 194, pp. 16-29.

[14] Y. Zhou, B. Hu, W. Zhao, D. Cui, L. Tan and J. Wang. 2018. Effects of increasing nutrient disturbances on phytoplankton community structure and biodiversity in two tropical seas. Marine Pollution Bulletin. Volume 135, pp. 239-248.

[15] X. Mei, R. Li, X. Zhang, Z. Wang and Y. Zhang. 2019. Reconstruction of phytoplankton productivity and community structure in the South Yellow Sea. *China Geology*. 2(3), pp. 315-324.

[16] Parsons et al. *Biological Oceanography Process*. Third Edition. Pergamon Press, New York. 1984.

[17] R. Toral. 2003. On the definition of physical temperature and pressure for nonextensive thermostatistics. *Physica A: Statistical Mechanics and its Applications.* Volume 317, Issues 1–2, pp. 209-212.

[18] C. Teodosiu, G. Barjoveanu, and D. Teleman. 2003. Sustainable Water Resources Management 1. River Basin Management and the Ec Water Framework Directive. *Environmental Engineering and Management Journal.* 2(4), pp. 377-394.

[19] Hair, J.F.J., Anderson, R.E., Tatham, R.L. & Black, W.C. *Multivariate Data Analysis*, 5th edition, Upper Saddle River: Prentice Hall. 1998.

[20] Johnson, R. A., & Wichern, D. W. *Applied Multivariate Statistical Analysis.* 3th Edition. New Jersey: Prentice Hall of India Private Limited. 1996.