

Factors Affecting Dissolved Oxygen at Bengawan Solo River: A Spatial Filtering with Eigenvector Technique

Evellin Dewi Lusiana¹, Arief Darmawan², Sarah Hutahaean³, Muhammad Musa⁴,
Mohammad Mahmudi⁵, Sulastri Arsad⁶

^{1,2,3,4,5,6}Departement of Aquatic Resource Management, Faculty of Fisheries and Marine Science, Universitas Brawijaya, Indonesia

^{1,4,5,6}AquaRES Research Group, Universitas Brawijaya, Indonesia

Article Info

Article history:

Received : 08-20-2021

Revised : 10-11-2021

Accepted : 10-18-2021

Keywords:

Linear Regression;
Morans I;
Residual Diagnostics;
Water Quality.

ABSTRACT

The quality of the river changes according to the development of the surrounding environment which is influenced by various human activities. Analysis of factors affecting Dissolved Oxygen (DO) at Bengawan Solo River is crucial for river management purpose and pollution control. Previous research suggested the use classic multiple linear regression. However, DO measurement were usually took place of sampling sites along the river channel. Therefore, there is a high chance that the measurements results may spatially correlated. As the consequence, the utilization of multiple linear regression technique for the dataset can be inappropriate. In this paper, we applied a modification of multiple linear regression model to incorporate with spatial autocorrelation that exist in the data by adding control variable such vector eigen to the model which known as Spatial Filtering with Eigenvector (SFE). The results showed that nitrate and nitrite were the predictor variables that have a negative and significant effect. However, the model contains spatial autocorrelation. The application of SFE technique by adding three eigenvectors as control variables in the model succeeded in making the residual model free from spatial autocorrelation. However, a new problem arose where there was a violation of the non-heteroscedasticity assumption.



Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: <https://doi.org/10.30812/varian.v5i1.1407>

Corresponding Author:

Evellin Dewi Lusiana,
Departement of Aquatic Resource Management, Universitas Brawijaya.
Email: evellinlusiana@ub.ac.id

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



A. INTRODUCTION

Rivers are water source that have many advantages for human life. The benefits of rivers to humans are for household needs, tourism, fisheries, irrigation and rivers can also be used for transportation facilities (Iloms et al., 2020). Another benefit of the river where the river channel can also be used for human needs (Wurtsbaugh et al., 2019). The river channel can be used to flow into agriculture, because the river water can be used for irrigation, flowing into rice fields and for fishing ponds (Mahmudi et al., 2019). However, the use of river water must meet the water quality criteria in accordance with its utilization (Musa et al., 2019).

The quality of the river changes according to the development of the surrounding environment which is influenced by various human activities. The number of activities along the riverbanks causes the higher the potential for water quality degradation (Arnop et al., 2019). The water quality condition of most rivers in Indonesia is categorized as polluted, mainly due to the density of settlements and the large number of human activities (Hamakonda et al., 2019). This condition can cause disruption of the river ecosystem. As a result, fish die, discoloration in the water that produces odors and causes human health problems (Hertika et al., 2018).

One of the important indicators of river water quality is Dissolved Oxygen (DO). Oxygen is needed for respiration and metabolism of organisms in the water (Lusiana et al., 2021). Dissolved oxygen is used in the process of decomposing waste that discharges to the waters, so the amount of waste in waters causes a decrease in DO levels (Sandi et al., 2017). This can have a negative impact on aquatic organisms where fish are stressed. DO range 0.3-1 mg/l will cause death in fish if it lasts a long time, while DO 1-5 mg/l will cause slow growth of fish (Arend et al., 2011).

Bengawan Solo River is one of the longest rivers on the island of Java (600 km) which crosses two provinces, namely Central Java and East Java with a river flow area of 16,000 km². The area drained by this river passes through 9 districts/municipalities in Central Java and 11 districts/municipalities in East Java (Dani et al., 2015). The condition of the waters in the Bengawan Solo River from upstream to downstream has decreased water quality from year to year (Sutriati, 2012). The upstream and downstream parts of the river physically do not meet the requirements for clean water which are characterized by an unpleasant smell, the color of the river water is yellow-black, and the amount of garbage on the riverbanks (Astuti, 2015). Another factor is that the Bengawan Solo River water is used by the community for fishing, agriculture, industry, domestic activities, and others (Dani et al., 2015). This activity causes a decrease in water quality and there are also development activities along the river so that waste disposal is not controlled (Sutriati, 2012).

DO modelling of Bengawan Solo River is crucial for river management purpose and pollution control. Previous research suggested the use classic multiple linear regression approach to identify factors affecting DO in river (Imran et al., 2014; Ouma et al., 2020; Prasad et al., 2014). However, DO measurement and any other water quality factors were usually took place of sampling sites along the river channel. Therefore, there is a high chance that the measurements results may spatially correlated (Pratiwi et al., 2018). As the consequence, the utilization of multiple linear regression technique for the dataset can be inappropriate (Thayn and Simanis, 2013). This is because multiple linear regression has several underlying assumptions that should be met such as normality, non-heteroscedasticity, non-multicollinearity, and non-autocorrelation (Thayn, 2017). The existence of spatial connectivity among water quality measurements will violent non-autocorrelation assumption and thus making the results of linear regression analysis invalid (Blanchet et al., 2008). In this paper, we applied a modification of multiple linear regression model to incorporate with spatial autocorrelation that exist in the data by adding control variable such vector eigen to the model (Getis and Griffith, 2002). The technique is known as spatial filtering with eigenvector (SFE).

B. LITERATURE REVIEW

1. Linear Regression Model

Linear regression analysis is a statistical method used to determine the effect of one or more independent variables on one dependent variable (Laake and Fagerland, 2015). The general model of linear regression is stated as follows

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i, i = 1, 2, \dots, n \quad (1)$$

where:

- Y_i = i-th observed value of dependent variable
- X_{ij} = i-th observed value of j-th independent variable ($j = 1, 2, \dots, p$)
- p = number of independent variables
- β_0 = intercept
- β_j = regression coefficient of j-th independent variable
- n = sample size
- ε_i = error/residual of i-th observation

The matrix form can be written as follows in equation.

$$Y = X^T \beta + \varepsilon \quad (2)$$

Parameter estimation of linear regression model usually done by using Ordinary Least Square (OLS) with the formula $\hat{\beta} = (X^T X)^{-1} Y$ (Lusiana et al., 2019).

2. Residual Diagnostics Test

Linear regression model is a popular data analysis technique even though it has a simple structure (Altman and Krzywinski, 2016). However, there are certain conditions / assumptions needed where these assumptions can influence the conclusions

generated from the model or the interpretation of the results (Zeileis and Hothorn, 2002). Validation of the fulfillment of the linear regression assumptions is carried out by performing diagnostic testing on the residual models (Cox, 2004). Particularly with regard to linear regression inference, important residual diagnostic tests to do are normality homoscedasticity and non-autocorrelation in residuals for geographic data. In addition, it takes the assumption of non-multicollinearity for the independent variables (Thayn and Simanis, 2013).

3. Spatial Autocorrelation: Global Morans I

Global Morans'I statistic is a measure used to indicate the presence of spatial autocorrelation globally (Getis, 1995; Moran, 1950). This measure is basically a cross product between a variable and its spatial weighting with the variable expressed as the deviation from its average, as shown in equation (3)

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})^2} \quad (3)$$

where:

- I = Global Morans'I statistic
- n = sample size
- w_{ij} = spatial weight between i and j observations
- x_i = i-th observed values
- \bar{x} = mean of observed values

4. Spatial Filtering with Eigenvector (SFE)

Spatial filtering is a method that produces dummy spatial variables which are used as additional independent variables in the regression model. This resolves model specification errors. Thus, spatial data can be appropriately included in the regression model and its diagnostic statistics, making the regression results more robust and easy to interpret (Thayn and Simanis, 2013).

SFE is a spatial filtering technique by creating a set of dummy spatial patterns by searching for eigenvectors associated with independent variables in linear regression models and spatial weighting matrices. (Getis and Griffith, 2002). Eigen vector extraction is obtained from the matrix MCM, where $M = I - X(X^t X)^{-1} X^T$ and C is the spatial weighted connectivity matrix. Furthermore, the selected combination of eigenvectors is determined by a step-wise regression procedure (Thayn, 2017). Finally, a linear regression model with SFE is presented in equation (4) as follows

$$Y = X^T \beta + E\gamma + \eta \quad (4)$$

where $E\gamma$ is misspesification term and $E\gamma + \eta$ is equivalent with ε from equation (2). By substituting $X_{mod}^T \beta_{mod} = X^T \beta + E\gamma$, equation (4) can re-written as equation (5)

$$Y = X_{mod}^T \beta_{mod} + \eta \quad (5)$$

Therefore, the parameter can be estimated using OLS method $\widehat{\beta}_{mod} = (X_{mod}^T X_{mod})^{-1} Y$.

C. RESEARCH METHOD

The study was conducted by collecting secondary data through the Bengawan Solo River Basin Center (BBWS) in the 2016 2020 period. The monitoring site collection in this study was an observation taken by BBWS agency. There were 7 sampling sites as presented by Figure 1. Furthermore, we used several water quality variables which presented in Table 1.

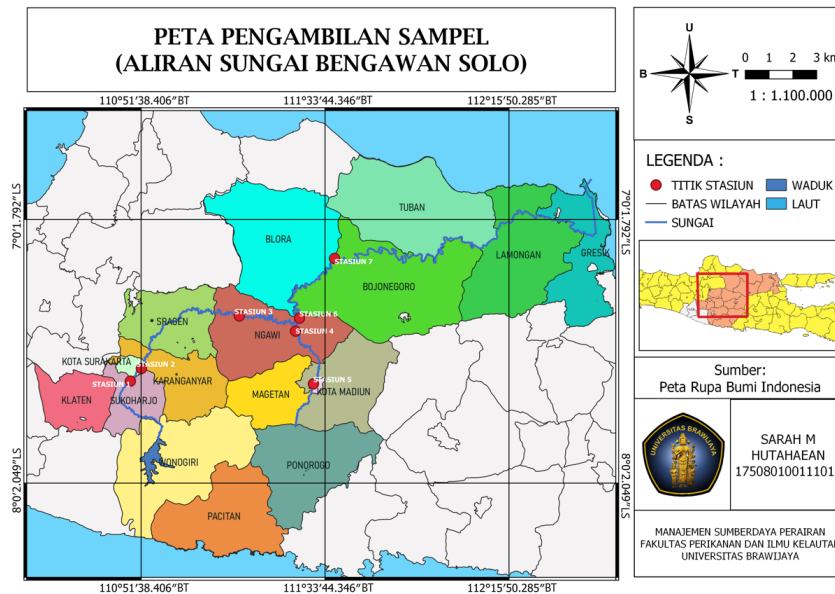


Figure 1. Sampling Site Location in Bengawan Solo River

Table 1. Research Variables Classification

Classification	Variable	Unit
Dependent	Dissolved Oxygen (DO)	mg/L
	Temperature	oC
	pH	-
	Total dissolved solid (TDS)	mg/L
	Total suspended solid (TSS)	mg/L
Independent	Ammonia (NH ₃)	mg/L
	Nitrate (NO ₃)	mg/L
	Nitrite (NO ₂)	mg/L
	Phosphate (PO ₄)	mg/L
	Chemical oxygen dissolved (COD)	mg/L
	Biological oxygen dissolved (BOD)	mg/L

The data analysis steps in this study are as follows

1. Correlation analysis to filter independent variables which has significant association with dependent variable
2. Estimation of linear regression model parameters using the OLS method
3. Linear regression residual diagnostic test, namely spatial autocorrelation (Global Moran's I test), normality (Shapiro-Wilk test), and heteroscedasticity (Breusch-Pagan test)
4. Spatial weighted matrix formation (k-nearest neighbour)
5. Extraction of eigenvectors associated with independent variables and spatial weighting matrix
6. Step-wise selection of eigenvector combinations.
7. Estimation of regression model parameters with spatial filtering (addition of selected eigenvector combinations)
8. Diagnostic test of residual linear regression model of (SFE)

D. RESULT AND DISCUSSION

1. Association Between Variables

The degree of association between independent and dependent variable (DO) in this research is shown in Figure 2. It is clearly seen that DO has significant correlation with pH, NO₃, NO₂, COD and BOD. Therefore, these variables will be utilized in linear regression model.

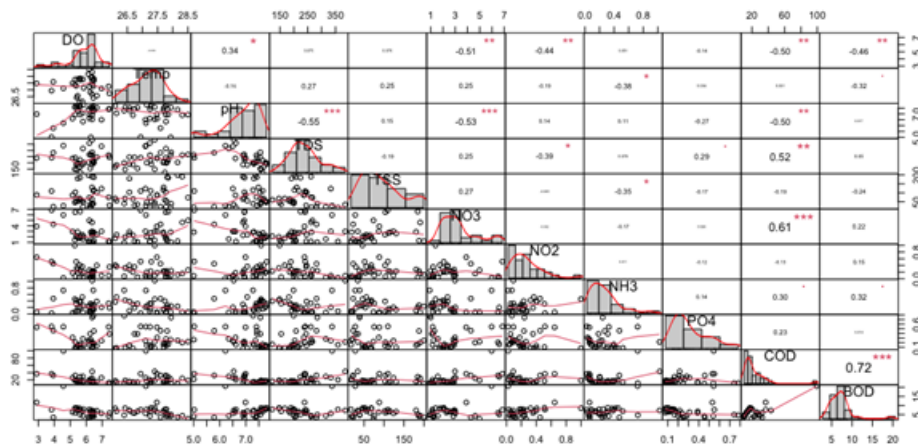


Figure 2. Correlation Chart of Studied Variables

2. OLS Linear Regression Model of DO

The following Table 2 presents linear regression model of DO measured from Bengawan Solo River.

Table 2. Summary of DO Model of Bengawan Solo River by Using Linear Regression

Variable	Coefficient	Standard Error	t	Prob.
Intercept	5.3030	1.7327	3.060	0.00473**
pH	0.3153	0.2429	1.298	0.20450
NO3	-0.1816	0.0955	-1.902	0.06715
NO2	-2.0333	0.6239	-3.259	0.00285**
COD	0.0003	0.0192	0.017	0.9863
BOD	-0.1025	0.0794	-1.291	0.2067

R-square=0.5883; Adj R-square=0.5173
 F=8.289; p-value= 5.843e-05

Note: significant at level 0.05 (), 0.01 (**)

From Table 2 it can be seen that NO2 and NO3 were significantly affect DO at 0.01 and 0.10 level, respectively. Furthermore, both NO2 and NO3 have negative coefficients. It implies that the higher NO3 and NO2, the lower DO value. Further diagnostic on the model residual is presented in Table 3.

Table 3. Diagnostic Residual Test of Linear Regression Model

Diagnostic test	Test Statistics	Prob.	Remark
Global Morans I	0.218	0.020	Spatial autocorrelation
Shapiro-Wilk	0.988	0.958	Normal
Breusch-Pagan	7.924	0.161	Homoscedastic

Table 4. Multicollinearity Test of Linear Regression Model

Variable	VIF	Remark
pH	2.328	Non-multicollinearity
NO3	2.054	
NO2	1.266	
COD	7.340	
BOD	4.640	

The results of the diagnostic residuals in Table 3 and Table 4 shows that the residuals of the linear regression model indicate that they contain a spatial autocorrelation with a positive Moran coefficient. On the other hand, the normality test showed that the model residuals are normally distributed as well as the Breusch-Pagan test show that the residual variance was constant (homoscedastic). In addition, there were no independent variables with VIF value greater than 10. Meaning that the non-multicollinearity assumption was met.

3. SFE Linear Regression Model

Based on the residual diagnostic test of the linear regression model which shows the existence of spatial autocorrelation, then the SFE technique was carried out. This technique produces three eigenvectors which are added as independent variables in the linear regression model, as shown in Table 5. The number of eigen vector that incorporated in the model was based on the eigenvector spatial pattern in Figure 3. It can be seen that on the third eigen vector, the Morans I value was began to stable around zero. Therefore, we add three eigenvectors into the linear regression model.

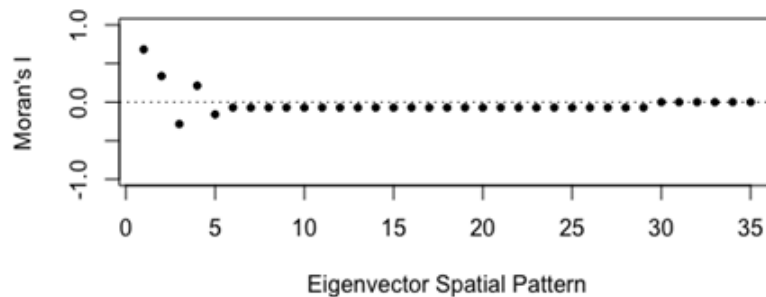


Figure 3. Eigenvector Spatial Pattern in Term of Morans I Statistics

Table 5. Summary of DO Model of Bengawan Solo River by Using Linear Regression

Variable	Coefficient	Standard Error	t	Prob.
Intercept	5.3030	1.4522	3.652	0.00115**
pH	0.3153	0.2036	1.549	0.13351
NO3	-0.1817	0.0800	-2.270	0.03177 *
NO2	-2.0333	0.5229	-3.889	0.00063 ***
COD	0.0003	0.0161	0.021	0.98366
BOD	-0.1025	0.0665	-1.541	0.13540
E.selvec1	1.5838	0.5505	2.877	0.00792 **
E.selvec2	1.0875	0.5505	1.976	0.05891
E.selvec3	-0.9706	0.5505	-1.763	0.08962

R-square=0.7408; Adj R-square=0.661

F=9.287; p-value= 5.862e-06

Note: significant at level 0.05 (), 0.01(*), 0.001 (**), 0.0000 (***)

SFE regression model did not change the coefficient estimation of independent variables compared to classis linear regression. Furthermore, the results of the diagnostic residual test for the SFE linear regression model are presented in Table 6 and Table 7. Based on this table, it appears that the presence of spatial autocorrelation has been successfully resolved. However, the issue of heteroscedasticity eventually emerged in this model where in the standard model this assumption was met.

Table 6. Diagnostic Residual Test of SFE Linear Regression Model

Diagnostic test	Test Statistics	Prob.	Remark
Global Morans I	-0.0733	0.8416	No spatial autocorrelation
Shapiro-Wilk	0.9909	0.9906	Normal
Breusch-Pagan	23.462	0.00282	Heteroscedastic

Table 7. Multicollinearity Test of SFE Linear Regression Model

Variable	VIF	Remark
pH	2.328	Non-multicollinearity
NO3	2.054	
NO2	1.266	
COD	7.340	
BOD	4.640	
E.selvec	1.000	

E. CONCLUSION AND SUGGESTION

Modelling of dissolved oxygen in the Bengawan Solo River with linear regression technique shows that nitrate and nitrite were the predictor variables that have a negative and significant effect. However, based on the results of the diagnostic residual test, it shows that this model contains spatial autocorrelation. The application of SFE technique by adding three eigenvectors as control variables in the model succeeded in making the residual model free from spatial autocorrelation. However, a new problem arose where there was a violation of the non-heteroscedasticity assumption. Therefore, it is recommended that further investigations related to the SFE technique in linear regression models can be carried out which can help to fulfil this assumption.

ACKNOWLEDGEMENT

This research was funded by Hibah Penguatan Kelompok Kajian Universitas Brawijaya 2021 with contract number 1204.9/UN10.C10/PN/2021.

REFERENCES

- Altman, N. and Krzywinski, M. (2016). Points of Significance: Regression Diagnostics. *Nature Methods*, 13(5):385–386.
- Arend, K. K., Beletsky, D., DePinto, J. V., Ludsin, S. A., Roberts, J. J., Rucinski, D. K., SCAVIA, D., SCHWAB, D. J., and HÖÖK, T. O. (2011). Seasonal and Interannual Effects of Hypoxia on Fish Habitat Quality in Central Lake Erie. *Freshwater Biology*, 56(2):366–383.
- Arnop, O., Budiyanto, B., and Saefuddin, R. (2019). Kajian Evaluasi Mutu Sungai Nelas dengan Metode Storet dan Indeks Pencemaran. *Naturalis: Jurnal Penelitian Pengelolaan Sumber Daya Alam dan Lingkungan*, 8(1):15–24.
- Astuti, B. C. (2015). Kualitas Air Sumur Desa Bantaran Sungai Bengawan Solo Berdasarkan Aspek Kemasyarakatan dan Standar Menteri Kesehatan. *Jurnal Matematika, Sains, dan Teknologi*, 16(1):18–25.
- Blanchet, F. G., Legendre, P., and Borcard, D. (2008). Modelling Directional Spatial Processes in Ecological Data. *Ecological Modelling*, 215(4):325–336.
- Cox, N. J. (2004). Speaking Stata: Graphing Model Diagnostics. *The Stata Journal: Promoting communications on statistics and Stata*, 4(4):449–475.
- Dani, T., Suripin, and Sudarno (2015). Analisis Daya Tampung Beban Cemar di DAS Bengawan Solo Segmen Kota Surakarta dan Kabupaten Karanganyar dengan Model Qual2Kw. *Jurnal Ilmu Lingkungan*, 13(2):92–102.
- Getis, A. (1995). Spatial Autocorrelation. *Progress in Human Geography*, 19(2):245–249.

- Getis, A. and Griffith, D. A. (2002). Comparative Spatial Filtering in Regression Analysis. *Geographical Analysis*, 34(2):130–140.
- Hamakonda, U. A., Suharto, B., and Susanawati, L. D. (2019). Analisis Kualitas Air dan Beban Pencemaran Air pada Sub DAS Boentuka Kabupaten Timor Tengah Selatan. *Jurnal Teknologi Pertanian Andalas*, 23(1):56–67.
- Hertika, A. M. S., Kusriani, K., Indrayani, E., Nurdiani, R., and Putra, R. B. (2018). Relationship between Levels of the Heavy Metals Lead, Cadmium and Mercury, and Metallothionein in the Gills and Stomach of *Crassostrea Iredalei* and *Crassostrea Glomerata* [version 1; referees: 2 approved]. *F1000Research*, 7:1–12.
- Iloms, E., Ololade, O. O., and Ogola, H. J. O. (2020). Investigating Industrial Effluent Impact on Municipal Wastewater Treatment Plant in Vaal, South Africa. *International Journal Environmental Research and Public Health*, 17:1–18.
- Imran, M., Sugiharto, E., and Siswanta, D. (2014). Penggunaan Model Regresi Linier untuk Menyatakan Hubungan Fungsional Perubahan Konsentrasi Oksigen Terlarut terhadap Parameter Fisika-Kimia Air Sungai Secang Kulon Progo. *Bimipa*, 24(2):206–218.
- Laake, P. and Fagerland, M. W. (2015). Chapter 11 - Statistical Inference. pages 379–430. Academic Press, Amsterdam.
- Lusiana, E., Arsad, S., Kusriani, Buwono, N., and Putri, I. (2019). Performance of Bayesian Quantile Regression and Its Application to Eutrophication Modelling in Sutami Reservoir, East Java, Indonesia. *Ecological Questions*, 30(2):69–77.
- Lusiana, E., Musa, M., and Ramadhan, S. (2021). Determinants of Nile Tilapia's (*Oreochromis Niloticus*) Growth in Aquaculture Pond in Batu, Indonesia. *Biodiversitas*, 22(2).
- Mahmudi, M., Lusiana, E., Arsad, S., Buwono, N., Darmawan, A., Nisya, T., and Gurinda, G. (2019). A Study on Phosphorus-Based Carrying Capacity and Trophic Status Index of Floating Net Cages Area in Ranu Grati, Indonesia. *AACL Bioflux*, 12(5):1902–1908.
- Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2):17–23.
- Musa, M., Mahmudi, M., Arsad, S., and Buwono, N. R. (2019). Feasibility Study and Potential of Pond as Silvofishery in Coastal Area: Local Case Study in Situbondo Indonesia. *Regional Studies in Marine Science*, page 100971.
- Ouma, Y. O., Okuku, C. O., and Njau, E. N. (2020). Use of Artificial Neural Networks and Multiple Linear Regression Model for The Prediction of Dissolved Oxygen in Rivers: Case Study of Hydrographic Basin of River Nyando, Kenya. *Complexity*, 2020:9570789.
- Prasad, B. S. R. V., Srinivasu, P. D. N., Varma, P. S., Raman, A. V., and Ray, S. (2014). Dynamics of Dissolved Oxygen in Relation to Saturation and Health of An Aquatic Body: A Case for Chilka Lagoon, India. *Journal of Ecosystems*, 2014:526245.
- Pratiwi, L. P., Hanief, S., and Suniantara, I. K. P. (2018). Pemodelan Menggunakan Metode Spasial Durbin Model untuk Data Angka Putus Sekolah Usia Pendidikan Dasar. *Jurnal Varian*, 2(1 SE - Articles).
- Sandi, M. A., Arthana, I. W., and Sari, A. H. W. (2017). Bioassessment dan Kualitas Air Daerah Aliran Sungai Legundi Probolinggo Jawa Timur. *Journal of Marine and Aquatic Sciences*, 3(2):233–241.
- Sutriati, A. (2012). Penilaian Tingkat Pencemaran Air Sungai Bengawan Solo dengan Menggunakan Indeks Kimia-Fisika. *Jurnal Sumber Daya Air*, 8(1):81–94.
- Thayn, J. B. (2017). Eigenvector Spatial Filtering and Spatial Autoregression BT - Encyclopedia of GIS. pages 511–522. Springer International Publishing, Cham.
- Thayn, J. B. and Simanis, J. M. (2013). Accounting for Spatial Autocorrelation in Linear Regression Models Using Spatial Filtering with Eigenvectors. *Annals of the Association of American Geographers*, 103(1):47–66.
- Wurtsbaugh, W. A., Paerl, H. W., and Dodds, W. K. (2019). Nutrients, Eutrophication and Harmful Algal Blooms Along the Freshwater to Marine Continuum. *WIREs Water*, 6(5):e1373.
- Zeileis, A. and Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. *R News*, 2(3):7–10.