The Biotic Integrity of Ciliwung River in West Java Under Multiple Urban Stressors

Dwita Sutjiningsih\textsuperscript{1}, Herr Soeryantono,\textsuperscript{1} Evi Anggraheni\textsuperscript{1} and Toha Saleh\textsuperscript{1}

1 Civil Engineering Department, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia
E-mail: dwita@eng.ui.ac.id

Abstract. Based on its imperviousness, most of the sub-watersheds in the lower part of Ciliwung watershed in West Java are classified as urban drainage. High imperviousness in the urban environment will cause multiple stressors among others on the discharge, water quality and habitat condition, those supposedly support the presence of macroinvertebrates. Macroinvertebrates are commonly used as markers of the biotic integrity of aquatic ecosystem. The aim of the study is to explore the extent of causal relationship between imperviousness, as trigger for the emergence of stressors on biotic integrity, which is represented by Family Biotic Index. The Family Biotic Index is considered as response variables, while imperviousness, discharge, water quality and habitat condition are as explanatory variables. Significance correlation tests among explanatory and response variables from 23 sub-watersheds of Ciliwung using Spearman Rank-order Correlation show that they are correlated. Influence of multiple urban stressors on the biotic integrity was assessed using several alternatives of multiple linear regression model, which is determined based on backward elimination procedure. The standard error of estimate ranges between 0.8563 and 0.9096 are less than standard deviation amounted 1.0876, therefore the models are acceptable for predicting the response variable, the FBI. Although they have not yet fully represented the observed FBI values, especially for the lowest and the highest one, possibly due to very limited data. This can be seen from the low value of the coefficient of determination, $R^2$ ranging from 33.23\% up to 43.86\%, and from the juxtaposed between model results and the observed values.

1. Introduction
The increase in population is always followed by an increase in the construction of facilities and infrastructures supporting the daily activities which will certainly increase the area of impervious land cover in the watershed. A great deal of studies has been carried out throughout the world in order to obtain clarity about the causal relationship between increasing imperviousness in the watershed and the condition of aquatic ecosystems. Review of reports on the influence of imperviousness on stream quality condition resulted in a modified Impervious Cover Model (ICM). Differ from the original model, which was developed based on a study of more than 250 reports on ecoregions around the U.S. and elsewhere, the modified ICM was the result of a review on 35 papers only: 25 from peer-reviewed journals, four from the U.S. Geological Survey, five from peer-reviewed conference proceedings, and one from a state research institute. However, these papers were selected based on four very strict criteria, namely (1) a minimum of 10 individual sub-watersheds must have been sampled; (2) riverine studies that sampled several stations in a progressive downstream direction in the same watershed, were omitted; (3) only studies that directly measured impervious cover or an autocorrelated metric, were considered; (4) the study must have been published in a peer-reviewed, reliable source, such as a scientific journal article or federal report [1].

\textsuperscript{1}Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Stream quality condition as impacted by urbanization in the watershed are generally classified according to four broad categories: changes in hydrologic, physical, water quality or biological indicators. Hundreds of research studies have documented the adverse impact of imperviousness on one or more of those key indicators [2]. According to [3], urbanization that resulted in high imperviousness will generate the effect of multiple stressors syndrome mainly on the discharge, water quality and habitat condition, those supposedly support the presence of macroinvertebrates. Macroinvertebrates are commonly used as marker of the biotic integrity of the aquatic ecosystem [4]. Using simple linear regression equations, the study on Upper Ciliwung subwatershed proved that the relationship between imperviousness and abiotic/biotic elements of its aquatic ecosystems can be used to assess the degree of water-sensitivity of the related regional spatial plan. The abiotic elements were the discharge, water quality and habitat condition, while for representing the biotic element was macroinvertebrate [5].

This study aims to construct a model of the biotic integrity of the aquatic ecosystem by employing the multiple linear regression (MLR) for the whole Ciliwung watershed. The model is used to explore the influence of the imperviousness, the discharge, the water quality and the habitat condition toward the biotic integrity of the aquatic ecosystem. Considering its simplicity in sampling and identification processes, the biotic integrity is represented by Family Biotic Index (FBI) as the response variable, although according to [4] there is a tendency to be less able to describe extreme conditions. The imperviousness (IC) and the three stressors on the biotic integrity, namely instantaneous discharge (Q), water quality index (WQI) and habitat conditions (HAB) are considered as explanatory variables. To diagnose deficiencies in MLR, the classic regression assumption tests are conducted, namely non-multicolinearity, non-autocorrelation, normally distributed residual, and heteroscedasticity tests. Backward elimination procedure is used for selecting MLR with the most significant variable(s) that represents the relation between responsive and explanatory variables [6, 7]. The model can then be used to predict the effect of land use change caused by urbanization, especially changes in imperviousness, on the changes in stream quality based on the predicted FBI value.

Imperviousness IC is obtained from the Digital Elevation Model Indonesia map processing and landuse map from Ciliwung Cisadane River Basin Agency [8], while the three stressors data were obtained from the field survey results. In this case the discharge Q is represented by the instantaneous discharge measurement referring to Indonesian National Standards [9], as well as water sampling [10]. Water quality represented by water quality index WQI calculated using NSF-WQI procedure [11], and habitat condition HAB is assessed using rubrics developed by Ecological Observation and Wetland Conservation, ECOTON [12] that adopt rubrics developed by Mekong River Commission. The habitat condition is a combination of riparian and riverbed substrate condition. Sampling and assessment of macroinvertebrates follows the guidelines prepared also by ECOTON.

The results indicate that all models can be used to predict the FBI values, although they have not yet fully represented the observed FBI values, especially for the lowest and the highest one, possibly due to very limited data. This can be seen from the low value of the coefficient of determination, $R^2$ and from the juxtaposed between model results and the observed values.

2. Material and Method

2.1 The study area
The study was carried out during the period of 2017-2019 at Ciliwung watershed, which springs in Mt. Gede and Pangrango, West Java and debouches into the Java Sea. Twenty-three sub-watershed were selected as samples. Table 1 presents information about the locations of observation/measurement, which are in the administrative areas of Bogor Regency/City, Depok City and DKI Jakarta Province. Figure 1 presents the map of those twenty-three sub-watersheds.
2.2 Conceptual Framework

Human activities as a trigger for changes in the watershed generally will induce the increasing percentage of impervious land cover area on the watershed. Increasing imperviousness will cause multiple stressors in the related aquatic ecosystem. Referring to the concept developed by Center for Watershed Protection [2], which has been modified by [1], the stressors on the urbanized watershed can be in many forms [3]. This study sets three urban stressors induced by high imperviousness as key variables that might affect the condition of aquatic ecosystems in the watershed, namely discharge, water quality index and habitat condition. Furthermore, these three variables are determined to be stressors in the biotic integrity of aquatic ecosystem, which is represented by macroinvertebrates. The relation is presented as conceptual framework on figure 2.

Table 1. Observation/Measurement Locations.

| No  | Sub-watershed | Administrative Region       |
|-----|---------------|------------------------------|
| 1   | Upper Ciliwung 1 | Tugu Selatan, Bogor          |
| 2   | Upper Ciliwung 2 | Tugu Utara, Bogor            |
| 3   | Upper Ciliwung 3 | Tugu Selatan, Bogor          |
| 4   | Upper Ciliwung 4 | Tugu Selatan, Bogor          |
| 5   | Upper Ciliwung 5 | Tugu Selatan, Bogor          |
| 6   | Upper Ciliwung 6 | Cisarua, Bogor               |
| 7   | Upper Ciliwung 7 | Jogjogan, Bogor              |
| 8   | Cisarua        | Kopo, Bogor                  |
| 9   | Cisuren        | Sukakarya, Bogor             |
| 10  | Cisukabirus    | Gadog, Bogor                 |
| 11  | Ciesek         | Pasir Angin, Bogor           |
| 12  | Cibalok        | Pasir Angin, Bogor           |
| 13  | Ciseuseupan    | Sindangsari, Bogor           |
| 14  | Cikumpa        | Jatimulya, Depok             |
| 15  | Cibuluh-Ciluar | Cibinong, Bogor              |
| 16  | Ciparigi       | Pasir Jambu, Bogor           |
| 17  | Sugutamu       | Sukmajaya, Depok             |
| 18  | Salam UI       | Beji, Depok                  |
| 19  | Cijantung 1    | Pasar Rebo, Jakarta Timur    |
| 20  | Cijantung 2    | Pasar Rebo, Jakarta Timur    |
| 21  | Condet 1       | Kramat Jati, Jakarta Timur   |
| 22  | Condet 2       | Kramat Jati, Jakarta Timur   |
| 23  | Lower Ciliwung | Tebet, Jakarta Selatan       |

Figure 1. Sub-watersheds Discretization in Ciliwung.
Based on the above conceptual framework, the study was carried out following the diagram presented on Figure 3.

**Figure 2.** Conceptual framework of the study.

**Figure 3.** Flow diagram of the study.

IC is variable determined based on map processing, while Q, WQI, HAB and FBI are determined based on field survey and laboratory tests. Correlation analysis using Spearman Rank-order Correlation test [13] is conducted to determine the significance of relationships among trigger, stressor and response variables.

Regression is useful tool for prediction or to infer causal relationships among variables [6]. This study focusses on the verification of the causal interpretation among variables. Regression analysis using several alternatives of multiple linear regression models can be used to find out the relation between response variable (FBI) due to multiple stressors (Q, WQI and HAB) triggered by the increasing imperviousness (IC).

2.3 Determination of Variables
Determination of five variables in this study are performed as follows:
IC value is obtained from processing the DEM Indonesia map, which is combined with land use maps from Geospatial Information Agency [8]. As sample, 23 sub-watersheds are set, of which 13 are in Upper Ciliwung and 10 in Lower Ciliwung.

Instantaneous discharge, Q is obtained from field measurements, referring to Indonesian National Standard [9].

Score of Water Quality Index, WQI is obtained from sampling refers also to the Indonesian National Standard [10], which is subsequently tested in laboratory. From nine parameters of physics, chemistry and microbiology, WQI is then calculated according to the National Sanitation Foundation [11]. The scale is 0-100, where 100 is excellent.

Score of habitat condition, HAB is obtained from field observation on riparian and channel condition. The habitat condition score is then calculated based on the rubric developed by Ecoton [12]. The scale is from 1 to 3, where 3 is the best.

Value of FBI is obtained through biomonitoring activities. FBI-value scale ranging from 1 to 10, where 10 is the worst [12].

The value of five variables in 23 sub-watersheds in Ciliwung are presented in Figure 4, while Table 2 presents the statistics.

Figure 4. Variables values of Ciliwung watershed.
Table 2. Observation/Measurement Statistics of Ciliwung.

| Statistics       | IC (%)  | Q (cms) | WQI Score | HAB Score | FBI Value |
|------------------|---------|---------|-----------|-----------|-----------|
| Mean             | 46.4630 | 1.4175  | 63.8478   | 2.0239    | 5.7730    |
| Standard Deviation | 33.9601 | 3.4646  | 15.0849   | 0.4301    | 1.0876    |
| Minimum          | 5.2300  | 0.0030  | 31.0000   | 1.3800    | 3.1300    |
| Maximum          | 97.0700 | 15.7900 | 79.0000   | 2.8800    | 7.7800    |

Figure 4 presents the tendency of all variables values, where IC tends to increase in the downstream direction, while WQI, HAB and FBI have a tendency to become worse in the downstream direction. The variable Q has no clear trend, because it depends on the area and characteristics of sub-watersheds.

Referring to [1], 23 sub-watersheds in Ciliwung have four categories, namely: sensitive (4 sub-watersheds), impacted (5 sub-watersheds), non-supporting (5 sub-watersheds) and urban drainage (9 sub-watersheds). The nine sub-watershed which are classified as urban drainage are in Lower Ciliwung. Details are presented in Figure 5.

Figure 5. Subwatersheds category.

2.4 Correlation and regression analysis

Referring to the conceptual framework, four possible relationships between FBI and IC, Q, WQI and HAB will be considered. Multiple Linear Regression (MLR) models are employed to determine the trend of those relationships based on the results of correlation analysis for those four possible relationships.

Scatterplot. Before conducting correlation and regression analysis, scatterplots of relationships between response and explanatory variables are made. Scatterplot is the most useful graph for displaying the relationship between two quantitative variables. The purpose of a scatterplot is to provide a general illustration of the relationship between variables. Based on the results of the plot it can be seen its relationship trend, whether there is a positive or negative association [6].

Spearman rank-order correlation test. Spearman Rank-order Correlation test is used to find the relationships between response and explanatory variables. This is a statistical procedure that is designed to measure the relationship between two variables on an ordinal scale of measurement for sample size greater than four [14]. The correlation coefficient, Spearman’s rho, \( r_s \), if none of the ranked values are ties,

\[
r_s = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}
\]

Where \( n \) is the number of rank pairs and \( D_i \) is the difference between a ranked pair. If ties are present in the values, the formulas to determine \( r_s \) is the following:
The null hypothesis states that there is no correlation between explanatory and response variable, while the alternative hypothesis is the opposite. The level of risk, $\alpha$ is frequently set at 0.05, meaning that there is a 95% chance that any observed statistical difference will be real and not due to chance. If the critical value is less than or equal to the obtained value, $|r_s|$, must be greater than or equal to the critical value on the table. To be significant, the absolute value of the obtained value, $|r_s|$, must be greater than or equal to the critical value on the table.

Multiple Linear Regression Model. The multiple linear regression (MLR) model is a method to model the relation between response variable (y) with one or more explanatory variables ($x_1,x_2, x_3,\ldots, x_n$), and generally written as follow:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_kx_k + \varepsilon$$  \hspace{1cm} (3)

Where $y$ is the response variable, $\beta_0$ is the intercept, $\beta_i$ is the slope coefficient for the first explanatory variable, $\beta_2$ is the slope coefficient for the second explanatory variable, $\beta_k$ is the slope coefficient for the $k^{th}$ explanatory variable, and $\varepsilon$ is the remaining unexplained noise (the error) in the data [10].

The classic regression assumption tests. In order to diagnose deficiencies in MLR the classic regression assumption tests are conducted, namely: non-multicollinearity, non-autocorrelation and normally distributed residual.

- Assumption of non-multicollinearity. The collinearity test is used to define the presence correlation among explanatory variables. The indication of the presence of collinearity case could be indicated by VIF (Variance Inflation Factors). A good regression equation should not contain multicollinearity among explanatory variables. variable correlation. Value of $R^2$ obtained from regression result each of explanatory variable with another explanatory variables.

$$VIF = \frac{1}{1-R^2_i}$$  \hspace{1cm} (4)

- Assumption of non-autocorrelation. The autocorrelation residual test is applied to determine whether there is correlation among residuals. The approach used in this study was the Durbin-Watson (DW) test. If computed DW greater than upper limit (DU) and less than 4-DU then there is no autocorrelation. In case DW less than lower limit (DL) or greater than 4-DL then there is negative or positive autocorrelation, otherwise there is unclear unless there is autocorrelation or not. DL and DU from Durbin-Watson Significance Table is function of $n, k$ and $\alpha$.

$$DW = \frac{\sum_{t=2}^{n}(e_t-e_{t-1})^2}{\sum_{t=1}^{n}e_t^2}$$  \hspace{1cm} (5)

Where $e_t = y_t - \hat{y}$ are the residuals, $n$ is the number of elements in the sample, $k$ is the number of explanatory variables, and $\alpha$ is significance level (0.05).
Assumption of normally distributed error. The one-sample Kolmogorov-Smirnov test can be used to determine whether a sample comes from a population which is normally distributed. Suppose that the sample comes from a population with cumulative distribution function $F(x)$ then define $D_n$ as follows:

$$D_n = \max_x |F(x) - S_n(x)|$$

Where $x_1, \ldots, x_n$ an ordered residuals (errors) as sample with $x_1 \leq \ldots \leq x_n$ and $S_n(x)$ is defined as follows:

$$S_n(x) = \begin{cases} 0, & x < x_1 \\ k/n, & x_k \leq x_{k+1} \\ 1, & x \geq x_n \end{cases}$$

If $D_n$ less than $D_{n,\alpha}$ then there is no significant difference between the sample and the population which is normally distributed. $D_{n,\alpha}$ is the critical value from Kolmogorov-Smirnov Table. Where $n$ is the number of elements in the sample, $k$ is the number of explanatory variables and $\alpha$ is significance level (0.05).

Assumption of the absence of Heteroscedasticity. A good linear regression model must be met the condition of the absence of heteroscedasticity. The method of determining the Heteroscedasticity test can be done with the Scatterplot. The test determines whether a regression model’s ability to predict a response variable is consistent across all values of that response variable.

Selecting the best MLR model. Selecting the best model can be done by stepwise procedure. It can be forward selection or backward elimination or combination of both which so called stepwise regression. Each procedure has advantages and drawbacks. This study employed the backward elimination procedure, which starts with all explanatory variables in the model and eliminates the one with the lowest partial-F statistic (lowest $|t|$). It stops when all remaining variables are significant. Although the backwards algorithm does not ensure a "best" model because it also cannot consider the combined significance of groups of variables, at least it ensures that the final model has only significant variables.

3. Results and Discussion

3.1 Correlation analysis

Scatterplot. From four scatterplots on Figure 6, can be seen a consistent trend between response variable FBI and the explanatory variables. Positive association of FBI and IC shows an increase in the percentage of impervious cover will cause a decrease in the value of the FBI which indicates deteriorating aquatic ecosystem conditions, while a negative association between FBI and Q, WQI and HAB shows an increase in Q, WQI and HAB indicates the better condition of the aquatic ecosystem.
Figure 6. Scatterplots of response and explanatory variables relations: (a) FBI function of IC, (b) FBI function of Q, (c) FBI function of WQI, and (d) FBI function of HAB.

Spearman rank-order correlation test. The results of Spearman rank-order correlation test for four possible correlation between response variable and the explanation variables are presented on table 3.

Table 3. Results of Spearman rank-order correlation test.

| Correlation | Spearman rho, $r_s$ | P-value | Rho-critical, $r_c$ | Conclusion |
|-------------|---------------------|---------|---------------------|------------|
| FBI = $f$(IC) | 0.656 | 0.0007 | | |
| FBI = $f$(Q) | -0.447 | 0.0324 | 0.415 | $|r_s| > r_c$ |
| FBI = $f$(WQI) | -0.531 | 0.0091 | | $P < 0.05$ |
| FBI = $f$(HAB) | -0.653 | 0.0007 | | |

Since all relations resulted in absolute value of $r_s$ greater than $r_c$, it means that FBI correlated to all four explanatory variables. While P-values which are less than 0.05 (5%) indicated that the correlation are not just by chance.

3.2 Regression analysis

Non-multicollinearity test. On Table 4 are presented the results of non-multicollinearity test.

Table 4. Results of Non-multicollinearity test.

| Explanatory Variables | IC<sup>a</sup> | Q<sup>b</sup> | WQI<sup>c</sup> | HAB<sup>d</sup> |
|-----------------------|----------------|-------------|----------------|----------------|
| $R_i^2$               | 0.7468         | 0.0437      | 0.6168         | 0.6121         |
| VIF                   | 3.9503         | 1.0457      | 2.6094         | 2.5779         |

Conclusion | VIF < 10.0

<sup>a</sup>IC = $f$(Q,WQI,HAB).
<sup>b</sup>Q = $f$(IC,WQI,HAB).
<sup>c</sup>WQI = $f$(IC,Q,HAB).
<sup>d</sup>HAB = $f$(IC,Q,WQI).

VIF values for four possible combination of relation among explanatory variables less than 10.0, these indicated that there is no multicollinearity among explanatory variables.

Non-autocorrelation test. On Table 5 are presented the results of Durbin-Watson tests, which indicate that there is no autocorrelation among the explanatory variables.
Table 5. Results of Non-autocorrelation test using Durbin-Watson.

| MLR Model | M1<sup>a</sup> | M2<sup>b</sup> | M3<sup>c</sup> | M4a<sup>d</sup> | M4b<sup>e</sup> |
|-----------|---------------|---------------|---------------|---------------|---------------|
| DW        | 1.9381        | 1.9326        | 1.9529        | 1.7331        | 1.8086        |
| n         | 23            | 23            | 23            | 23            | 23            |
| k         | 4             | 3             | 2             | 1             | 1             |
| α         | 0.05          | 0.05          | 0.05          | 0.05          | 0.05          |
| DL        | 0.9860        | 1.0780        | 1.1680        | 1.2570        | 1.2570        |
| DU        | 1.7850        | 1.6600        | 1.5430        | 1.4370        | 1.4370        |
| 4-DU      | 2.2150        | 2.3400        | 2.4570        | 2.5630        | 2.5630        |

Conclusion

DW > DU

DW < 4-DU

Normal distributed residual test. Normal residual testing is done by looking at the normal probability plot of the residual pattern following the normal distribution as in figure 7, which is further clarified by the Kolmogorov Smirnov test as presented on the following Table 6.

![Figure 7](image1.png)

(a) (b) (c) (d) (e)

Figure 7. Normally distributed residual variance: (a) Model 1: \(FBI=f(IC,Q,WQI,HAB)\), (b) Model 2: \(FBI=f(IC,WQI,HAB)\), (c) Model 3: \(FBI=f(WQI,HAB)\), (d) Model 4a: \(FBI=f(HAB)\), and (e) Model 4b: \(FBI=f(IC)\).

Table 6. Results of Normally Distributed Residual Test based on One-Sample Kolmogorov-Smirnov Test.

| MLR Model | M1<sup>a</sup> | M2<sup>b</sup> | M3<sup>c</sup> | M4a<sup>d</sup> | M4b<sup>e</sup> |
|-----------|---------------|---------------|---------------|---------------|---------------|
| n         | 23            |               |               |               |               |
| Normal Parameters |
| - Mean    | 0.0000        | 0.0000        | 0.0000        | 0.0000        | 0.0000        |
| - Standard Deviation | 0.8149        | 0.8155        | 0.8164        | 0.8512        | 0.8887        |
| Most Extreme Differences (D<sub>n</sub>) |
| - Absolute | 0.0670        | 0.0868        | 0.0706        | 0.1021        | 0.1607        |
| - Positive | 0.0670        | 0.0868        | 0.0706        | 0.1021        | 0.1607        |
Table 6 shows the significance of the test result that $D_n$ for all studied models are less than $D_{n,0.05}$, so it can be concluded that the residuals are normally distributed.

**Heteroscedasticity test.** Scatterplots of all studied models indicate that heteroscedasticity does not occur. Residuals, which are plotted against the observed values are spread evenly around the zero axis as presented on Figure 8.
Figure 8. Scatterplots of Heteroscedasticity Test: (a) Model 1: \( FBI = f(IC,Q,WQI,HAB) \), (b) Model 2: \( FBI = f(IC,WQI,HAB) \), (c) Model 3: \( FBI = f(WQI,HAB) \) (d) Model 4a: \( FBI = f(HAB) \), and (e) Model 4b: \( FBI = f(IC) \).

Selecting the best MLR model. The first step in the backward elimination procedure, all explanatory variables are used (M1), which resulted in removal of Q (M2). The result of the second step is elimination of IC (M3), then WQI is released in the third step. In the fourth step all that remains is HAB (M4a), means that the most significant explanatory variable is HAB. In order to accommodate IC as the trigger of changes in the aquatic ecosystem, a fifth regression equation is added (M4b), with consideration that if only land use data is available, it can still be predicted the effect of changes in land use on the FBI. Summary of the procedure is presented in Table 7.

### Table 7. Results of Backward Elimination Procedure.

| MLR Model | M1a | M2b | M3c | M4a | M4b |
|-----------|-----|-----|-----|-----|-----|
| IC        | β   | 0.0023 | 0.0022 | -   | -   | 0.0185 |
|           | t   | 0.2012 | 0.2053* | -   | -   | 3.2326 |
| Q         | β   | -0.0092 | -      | -   | -   | -     |
|           | t   | -0.1632* | -      | -   | -   | -     |
| WQI       | β   | -0.0158 | -0.0161 | -0.0188 | -   | -     |
|           | t   | -0.7797 | -0.8232 | -1.3198* | -   | -     |
| HAB       | β   | -1.1389 | -1.1447 | -1.2257 | -1.5740 | -     |
|           | t   | -1.7129 | -1.7699 | -2.4523 | -3.6454 | -     |

*a* \( FBI = f(IC,Q,WQI,HAB) \).
*b* \( FBI = f(IC,WQI,HAB) \).
*c* \( FBI = f(WQI,HAB) \).
*d* \( FBI = f(HAB) \).
*e* \( FBI = f(IC) \).

In the following Table 8, are presented the complete results of Multiple Linear Regression Coefficients of FBI as Response Variable.

### Table 8. Multiple Linear Regression Coefficients of FBI as Response Variable.

| Intercept & Explanatory Variables | M1a | M2b | M3c |
|----------------------------------|-----|-----|-----|
| Intercept                        | 8.9933 | 2.4202 | 3.7159 | 0.0016 |
| IC                               | 0.0023 | 0.0112 | 0.2012 | 0.8428 |
| Q                                | -0.0092 | 0.0567 | -0.1632 | 0.8722 |
| WQI                              | -0.0158 | 0.0202 | -0.7797 | 0.4457 |
| HAB                              | -1.1389 | 0.6649 | -1.7129 | 0.1039 |

*a* \( FBI = f(IC,Q,WQI,HAB) \).
*b* \( FBI = f(IC,WQI,HAB) \).
*c* \( FBI = f(WQI,HAB) \).
From Table 8 can be concluded that the regression equations are as follow:

Model M1: \[ FBI = 8.9933 + 0.0023 IC - 0.0092 Q - 0.0158 WQI - 1.1389 HAB \]

Model M2: \[ FBI = 9.0151 + 0.0022 IC - 0.0161 WQI - 1.1447 HAB \]

Model M3: \[ FBI = 9.4546 - 0.0188 WQI - 1.2257 HAB \]

Model M4a: \[ FBI = 8.9588 - 1.5740 HAB \]

Model M4b: \[ FBI = 4.9153 + 0.0185 IC \]

Based on the above table, the M1 model with \( R^2 = 0.4386 \) means that the percentage of influence contribution of explanatory variables to the response variable is only 43.86%. Or the variation of the independent variables used in the model is able to explain only 43.86% of the explanatory variables variation, while the remaining 56.14% is influenced or explained by other variables which are not included in this study. The Standard Error of the Estimate (SEE) is a measure of the number of errors of...
the regression model in predicting the value of response variable. As a guideline if estimated SEE is less than the standard deviation of observed value of response variable, then the regression model is better at predicting the value of response variable. In this case, value of the standard deviation of observed FBI is 1.0876, thus all models can be used to predict the FBI values.

![Figure 9. Box Plot of FBI from Observation and Models.](image)

Juxtaposed the box plots of FBI resulted from the MLR model with observed FBI value as presented in Figure 9 shows that all models have not yet fully represent the observed FBI values, especially for the lowest and the highest one. Possibly due to very limited data (23 samples), so that despite accomplishment of all classical tests, it turns out that the amount of data describing the highest and lowest FBI values are very limited, so it becomes a kind of outlier. Only one sample with FBI value less than 4.0 and two samples with FBI value greater than 7.0, the remaining 20 samples are in the range between 4.2 and 7.0. The choice of whether to use the model with limited or various explanatory variable(s) is based on a cost-benefit analysis. Considering adding explanatory variable(s) in the model only if they make a significant improvement in the model performance [6].

4. Conclusion

The study is conducted to define the relation among imperviousness, discharge, water quality index, habitat condition as explanatory variables and the family biotic index as response variable, by using the multiple linear regression model. Five variants of models are scrutinized based on 23 samples derived from Ciliwung watershed, which are considered represented the population since all the classical tests had been accomplished. The range of determination coefficient, $R^2$ is $33.23\% - 43.86\%$, while the most significant variable in the model is the habitat condition. The SEE values of all models, which are less than the standard deviation of the observed value, indicate that all models can be use to predict the FBI values. However, the box plot of all models results juxtapose with the observed values, showed that the models have not yet fully represent the observed FBI values, especially for the lowest and the highest one, possibly due to very limited data. Despite accomplishment of all classical tests, it turns out that the amount of data describing the highest and lowest FBI values are very limited, so it becomes a kind of outlier.

References

[1] T R Schueler, L Fraley-McNeal, and K Cappiella 2009 Is impervious cover still important? J. Hydrol. Eng. 14 309

[2] Center for Watershed Protection 2003 Impact of Impervious Cover on Aquatic System. Watershed Protection Research Monograph No. 1. (Ellicot City : Center for Watershed Protection)
Acknowledgments
Authors wishing to acknowledge the support from the Ministry of Research, Technology and Higher Education of the Republic of Indonesia under the scheme of Penelitian Terapan Unggulan Perguruan Tinggi, fiscal year 2019, contract No. NKB-1730/UN2.R3.1/HKP05.00/2019.