The modelling of ground water quality in urban area based on demographics factor and building coverage ratio by using geographically weighted regression approach (case study in Jakarta, Indonesia)

Fitria Fitria^{1*}, Dwita Sutjningsih¹, and Titin Siswantining²

¹Civil Engineering Department, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia

²Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok, Indonesia

Abstract. The population density as consequences of urbanisation in certain area may cause the occurrence of land utility shifting from natural site to impervious. Thus, this impervious cover can be used as indicator of environment quality, particularly the water quality, such being the case of Jakarta, that 80% of ground water sample in Jakarta has been contaminated by E.Coli. Due to this condition, there are indicators which might be correlated with the water quality, namely building coverage ratio (BCR) and demographic factors such as population and building density. This study argues the influence of those three factors toward the ground water quality. The aim of this study is to formulate a spatial regression-correlation model between those three factors with the main indicator of water quality, namely E.Coli. The sampling data were obtained from a densely populated areas in Duren Sawit sub-district, in Jakarta. In term of statistical process, Geographically Weighted Regression (GWR) was used to analyse the spatial data. The result suggested that the value of R^2 at 68.9% and the largest influence of E.Coli value in this models was more affected by the BCR rather than by demographic factors. This outcome would be an early recommendation for Jakarta.

1 Introduction

The land transfer function in urban areas tends to increase the surface area with the closure of cement, asphalt, and waterproof area (*impervious*) [1,2]. In the last few decades, a large number of big cities in Indonesia are faced by a similar problem, that is a high population growth rate. This condition is triggering an increasing demand of residential areas and city infrastructures. In the further development, it would impact on the shifting of land usage in urban area. The rise of impervious cover could lead to eco-hydrologist problems such as an increase of surface runoff, a declining water quality, the damage on habitat of water aquatic animal [3,4,5]. Nevertheless, the land covers (impervious cover) is the characteristic of urban area [2,6]. The impervious cover could be used as indicator of environmental quality mainly related to the water resources quality [2]. In addition, the previous studies discuss majority on how much the influence of population density toward the water quality in a certain watershed [4,5,6,7]. The data mapping has been conducted by employing GIS and statistical method includes anova, correlation, regression as well as spatial regression (Geographical Weighted Regression) [8,9]. Returning briefly to the environmental issues, DKI Jakarta become one of the largest metropolitan cities

which experienced with the biggest impact of development. This condition was illustrated by its rate of city growth and development which is higher than other cities [10]. Regarding to the environmental quality monitoring report in DKI Jakarta in 2015, there has been a decreasing quality of river water, ground water and air, thereby the water pollution in Jakarta has reached a harmful level. This water quality monitoring was conducted toward well water in 197 urban village. The data sample was obtained twice during period September – October and November – December by employing few parameters such as physical, chemistry, and biology (microbiology). In term of quantity and quality aspects, the ground water notes better performance than the surface water. Although, 80% of ground water samples from 75 wells in 75 urban villages have contaminated with the *Escherichia Coli* (E.Coli) [11]. The spread of E. Coli bacteria in the ground is heavily influenced by the ground porosity. The horizontal movement is hard to be ascertained because it depends on several factors such as a type of soil, the height of ground water surface, the flow of ground water, the construction of hand pumps well as well the number of people who use it, also the number of people who throw feces [12]. Simultaneously, other impacts from the fast development in Jakarta is a quality reduction of space utilisation and the imbalance of space function and structure. According to the result

* Corresponding author: fitria65@eng.ui.ac.id

of satellite images data processing lansat 8 and GIS technology in 2017, DKI Jakarta showed a high building density. Consequently, the building density and building coverage ratio (BCR) increased, which is caused by a higher population density in the middle of the constant variable of land areas. For those purpose, this study attempts to obtain a spatial regression correlation model among the variables of population density, the building density, and the building coverage ratio, toward one main indicator of water quality, that is E.Coli bacteria in urban area. The outcome of this study is expected would be able to deliver a positive contribution and the policy direction related to the law enforcement.

In the meantime, the research would be conducted in East Jakarta as one of municipalities in DKI Jakarta with densely residential. The location being precisely taken for research is Duren Sawit, as one of the most densely populated sub-districts in East Jakarta and considered to be housing zone based on the Local Regulation DKI Jakarta No.1 year 2014.

Subsequently, the modelling research of ground water quality was implemented through the spatial regression approach or Geographical Weighted Regression (GWR), in which the geographical location was applied as predictor variable that could influence the response variable[13].

2 Materials and Methods

2.1 Study Site

The location of this study was determined in urban area in DKI Jakarta, in Duren Sawit sub-district, East Jakarta municipality to be precisely. This sub-district consist of 7 urban villages.

2.2 Sampling Methods

The data sample was collected from Duren Sawit sub-district in 30 site of location based on the result of population density mapping. The ground water sampling was represented by the resident well water and the measurement of building coverage ratio in 30 houses. The proportional probability sampling (pps) method was applied by using the population density as guidance, thereby each urban village (study location) hold a different number of sample.

2.3 Data Collection

2.1.1 Population density (PD)

The population density data was obtained from Indonesia Central Bureau Statistic 2016, which includes the landmass of 7 urban villages, as well as the population number and density. Based on the population

density as a guidance, the data sample was determined by using the proportional probability technique, namely the stratified random sampling. Thence, the number of sample was defined around 30 from total number of population density in entire urban villages.

2.1.2 Building density (NDBI)

The building density was estimated through the approach of NDBI (Normalized Difference Building Index). This index data was derived from the processing of satellite images data of landsat 8. Specially NDBI is a building density index in which the type of data shaped raster with pixel value between -1 until 1. The value of -1 determined the area with low building density, whereas the value of 1 for the high density.

2.1.3 Building Coverage Ratio (BCR)

The value of BCR is calculated by formula:

$$BCR = \frac{\text{building area}}{\text{site area}} \quad (1)$$

The value of BCR is presented in percentage (%).

2.1.4 Water Quality

The ground water quality sample was tested by using E Coli bacteria. The early interview conducted in 30 site was concluded that the ground water was mostly used as drinking water and sanitation. In term of microbiology, the sampling of ground water quality procedure has followed the SNI [14].

2.4 Modeling Methods

The Geographically weighted regression (GWR) method is a regression model for each site of location. The GWR method would generate a unique models because each site of location is different. This model involves a spatial factor (latitude and longitude coordinate) from each sample. The GWR model is developed by using the software of R 3.2.1 version. The model is presented by formula (2):

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \quad (2)$$

The weighted scheme in GWR could apply different method, such as Kernel Gaussian as the most commonly used techniques and the weighted function of bisquare [13] While, the Gaussian function to calculate the closest point -n is:

$$W_{ij} = \exp \left(-\frac{d_{ij}/b}{2} \right)^2 \quad (3)$$

Where:

j = the point - n, as the closest site to point - i

b = the distance of point - n as the closest (bandwidth)
 d_{ij} = the Euclidean distance between point - i to point - j

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad (4)$$

Bandwidth is the radius of a circle where the point in the radius of circle is still considered influential in forming parameter of location model. One method to determine the value of optimum bandwidth is cross-validation (CV) method.

$$CV = \sum_{i=1}^n (y_i - \hat{y}_{-i}(b))^2 \quad (5)$$

The estimation of regression coefficient in Geographically Weighted Regression (GWR) was conducted by using the Weighted Least Square (WLS). This WLS method provides different weight to each location of observation. Nevertheless, the presence of spatial heterogeneity is one of the requirement of GWR could be applied. The spatial effect test (spatial heterogeneity) referse to the variation in each location. The heterogeneity data could be spatially tested by using statistical method of Breusch-Pagan (BP) with value:

$$BP = \left(\frac{1}{2}\right) f^T X (X^T X)^{-1} X^T f \sim \chi^2_k \quad (6)$$

While the vector element f is

$$f_i = \left(\frac{\varepsilon_i^2}{\sigma^2} - 1\right) \quad (7)$$

Where:

ε_i : least squares residual for observation - i

X : the matrix with size $n \times (k+1)$, which is consist of standardised vector (z) for every observation.

The decision making: H_0 is rejected if $BP > \chi^2_k$, which means there is spatial heterogeneity.

3 Result and Discussion

3.1 Data and research variable

The response variable (Y) is E.Coli bacteria, whereas the predictor variable (X) is the Population density (X_1). While the building density is represented by NDBI (X_2) and Building Coverage Ratio (X_3).

Fig. 1. Sampling Location

3.2 Regression Model

Prior to the analysis of GWR model, the OLS regression model analysis need to be employed by equation:

$\hat{y} = -87.3594 + 0.1088X_1 + 152.9582X_2 + 162.6955X_3$ (8)
GWR is frequently used in data analysis related with the spatial heterogeneity. In this study, the value of BP was obtained equal to $=34.217 > 7.815 = \chi^2_{(0.05;3)}$ (H_0 is rejected). It seem that there was a spatial heterogeneity in the model.

3.3 Model GWR

The result of GWR model by using software R 3.2.1 version is as follow:

Table 1. GWR Diagnostic information

Parameter	Values
bandwidth	0.05709193
CV score	22215.88
Number of data point	30
SS GWR residual	15582.71
SS OLS residual	15582.71
Quasi-global R^2	68.53%

Table 1. described the value of Quasi R^2 at 68.53 %, which means that the predictor variable in the model only able to explain the response variable at 68.53%, while 31.47% is explained by other variable outside the model. The GWR model generated the value of SS residual smaller than OLS model. This condition indicated that the GWR model in more appropriate to use than other. Also, the GWR model provided an equation in each sampling location.

Table 2. Estimation of local parameter variable GWR in every point

Location no	β_0	β_1	β_2	β_3	R^2
1	-80.12	0.0073	229.69	155.23	0.684
2	-80.30	0.0072	229.54	155.44	0.684
3	-80.78	0.0064	227.49	156.49	0.681
4	-79.34	0.0049	223.44	156.28	0.679
5	-78.53	0.0057	224.86	154.92	0.682
6	-78.48	0.0056	224.59	154.94	0.682
7	-82.06	0.0068	228.64	157.45	0.681
8	-81.97	0.0067	228.52	157.39	0.681
9	-81.94	0.0068	228.59	157.35	0.681
10	-81.90	0.0067	228.31	157.38	0.680
11	-81.70	0.0067	228.49	157.14	0.681
12	-81.36	0.0070	229.17	156.61	0.682
13	-81.23	0.0070	229.16	156.48	0.682
14	-81.28	0.0069	228.99	156.58	0.682
15	-80.86	0.0070	229.03	156.14	0.683
16	-80.81	0.0070	229.19	156.06	0.683
17	-80.82	0.0070	229.15	156.08	0.683
18	-83.98	0.0047	223.55	160.98	0.675
19	-83.76	0.0047	223.69	160.71	0.675

Location no	β_0	β_1	β_2	β_3	R^2
20	-83.77	0.0046	223.43	160.83	0.675
21	-83.52	0.0047	223.61	160.51	0.675
22	-83.62	0.0046	223.44	160.68	0.675
23	-83.46	0.0049	223.97	160.31	0.675
24	-83.27	0.0050	224.32	160.00	0.676
25	-81.91	0.0059	226.35	157.97	0.679
26	-81.25	0.0056	225.48	157.58	0.679
27	-80.78	0.0055	225.18	157.19	0.679
28	-83.71	0.0041	222.64	161.11	0.674
29	-83.31	0.0042	222.79	160.66	0.675
30	-82.94	0.0040	222.45	160.45	0.675

Table 2. showed the the GWR model introduced β for every point which eventually formed the model in every point, thereby it created 30 models. Based on the data above, the water quality variable and the parameter of E. Coli bacteria in entire site is affected by the building coverage ratio variable. Also, from the Table 2 , the GWR model in site-1 in Pondok Kopi village is:

$$\hat{y} = -80.12 + 0.0073X_1 + 229.69X_2 + 155.23X_3 \quad (9)$$

From t test stated that the rising of 100% building coverage ratio would lead to the increase of E.Coli value.

4 CONCLUSION

This study was designed to determine the influence of building coverage ratio, population density and building density toward the ground water quality. The result of this research showed that the GWR approach is more appropriate to build a model of ground water quality in urban area, because this approach has covered the spatial element required in the research (sampling of geographic area). The GWR approach generated a local regression model in every site (in every neighbourhood association) with building coverage ratio as influential variable. In addition, a high level of building coverage ratio approaching 100% become one of the reasons why the septic tank in certain house is built nearby to the other house. This condition increase the possibility that the ground water may be polluted by E.Coli bacteria. Beside, a high level of building coverage ratio impact on the decreasing of rain-water infiltration to the ground, thereby it would not be able to dilute the pollutant E.Coli in the soil. In particular, the local model obtained from each neighbourhood association would provide a recommendation in enforcing the DKI Jakarta Local Regulation No.1 year 2014 related to the arrangement of building coverage ratio in each house in entire areas.

Acknowledgements

The author would like to thank the financial support provided by Universitas Indonesia through the PITTA 2018 funding scheme under grant No. 2422/UN.2.R3.1/HKP.05.00/2018,

managed by the Directorate for Research and Public Service and Water resources Management Civil Engineering Universitas Indonesia. Special thanks to Duren Sawit Sub-District, and One-door integrated investment and public services agency Government DKI Jakarta for their permit to survey.

References

1. S. F. Indung, *System design of green open space for green city*. Dissertation IPB (2012)
2. C.L.Arnold and C.J.Gibbons, Impervious surface coverage: the emergence of a key environmental indicator, J. of the American planning Association, **62**, no. 2, 243-258 (1996)
3. C. Belluci, Stormwater and aquatic life: Making the connection between impervious cover and aquatic life impairments for TMDL development in Connecticut streams, in *Proceedings of the Water Environment Federation*(2007), **5**, 1003-1018 (2007)
4. E. Brabec, S. Schulte and P. L. Richards, Impervious surfaces and water quality: a review of current literature and its implications for watershed planning. CPL bibliography, **16**, no. 4, 499-514 (2002)
5. D. Sutjiningsih. Water quality index for determining the development threshold of urbanized catchment area in indonesia, Int. J. of Technology, **8**, no. 1, 143-157 (2017)
6. D. Lu and Q. Weng. Use of impervious surface in urban land-use classification, Remote Sensing of Environment, **102**, no. 1, 146-160 (2006)
7. T. Schueler, Impacts of impervious cover on aquatic systems, Watershed protection research Monograph, **1**, 142 (2003)
8. Q. Chen, K. Mei, R. A. Dahlgren, T. Wang, J. Gong et M. Zhang, Sc. Tot. Env, **572**, , 450-466 (2016)
9. J. Tu and Z.-G. Xia, Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation, Sc. Tot. Env, **407**, no. 1, 358-378 (2008)
10. B.P.L.H.D. Jakarta, *Laporan Status Lingkungan Hidup Daerah (SLHD) DKI Jakarta*, DKI Jakarta, (2015)
11. B. R. Kosasih, S. and N. I. Astuty, Kualitas Air Tanah di Kecamatan Tebet Jakarta Selatan Ditinjau dari Sebaran Escherichia Coli, J.Teknik Lingkungan Universitas Trisakti, **5**, No. 1, 1-7 (2009)
12. Marsono, *Faktor-faktor yang berhubungan dengan kualitas bakteriologis air sumur gali di permukiman (Studi di Desa Karanganom, Kecamatan Klaten utara, Kabupaten Klaten)*, (2009)
13. A. S. Fotheringham, C. Brunson and M. Charlton, *Geographically weighted regression: the analysis of spatially varying relationships*, John Wiley & Sons Ltd (2002)
14. SNI No.6989.58:2008, National Standardization Agency Indonesia (2008)