Spatial Modelling of Pulmonary TB Distribution in Indonesia Using on Environmental and Socio-economic Variables

Helina Helmy[1,2*], Muhammad Totong Kamaluddin[3], Iskhaq Iskandar[4], Suheryanto[5], Irfannuddin[6], Novrikasari[7]
{helinahelmy@poltekkes-tjk.ac.id}

[1]Graduate School of Environmental Science, Sriwijaya University, Palembang 30139, Indonesia
[2]Department of Environmental Sanitation, Tanjung Karang Health Polytechnic, Lampung Selatan 35145, Indonesia
[3]Department of Pharmacology, Faculty of Medicine, Sriwijaya University, Palembang 30114, Indonesia
[4]Department of Physics, Faculty of Mathematics and Natural Sciences, Sriwijaya University, Indralaya 30662, Indonesia
[5]Department of Chemistry, Faculty of Mathematics and Natural Sciences, Sriwijaya University, Indralaya 30662, Indonesia
[6]Faculty of Medicine, Sriwijaya University, Indralaya 30662, Indonesia
[7]Department of Occupational Health and Safety, Faculty of Public Health, Sriwijaya University, Ogan Ilir 30662, Indonesia

Abstract. Tuberculosis (TB) is a contagious epidemic globally. Based on environmental and socio-economic data, this study aims to develop a spatial model and investigate the factors influencing the spread of pulmonary TB cases. Using the GWR method, this study analyzes influencing aspects and estimates the total of pulmonary TB cases in Indonesia involving seven variables: population density, poverty index, number of health facilities, medical personnel, rainfall, solar radiation, and road network. The results show that the model is accurate with R^2 values of 0.953 and adjusted R^2 values of 0.940. Spatial analysis shows that Indonesia has an average number of pulmonary TB patients with 4,856 cases. Several factors dominate the distribution of pulmonary TB cases in Indonesia: population density, poverty index, number of health facilities, medical personnel, road network, rainfall, and solar radiation. The resulting GWR model can explain the dependent variable by more than 90%. Environmental and socio-economic variables can be adopted to develop spatial models of infectious diseases at an urban level.

Keywords: Spatial Distribution, Pulmonary TB Cases, Environmental, Socio-economic, Geographically Weighted Regression

1. Introduction

The bacillus Mycobacterium tuberculosis causes tuberculosis that affects the lungs (pulmonary TB) [1]. This disease is one of the dominant causes of death globally [2].

SRICOENV 2022, October 05-06, Palembang, Indonesia

DOI 10.4108/eai.5-10-2022.2328785

Geographically, in 2020, there are 30 countries with high TB rates, accounting for 86% of 1.2 million of them all estimated TB cases in the world, and several countries contributing the highest TB cases of the global total: India with 26% of cases, China with 8.5% of cases, Indonesia with 8.4% of cases, Philippines with 6.0% of cases, Pakistan with 5.8% of cases, Nigeria with 4.6% of cases, Bangladesh with 3.6% of cases and South Africa with 3.3% of cases [2].

According to WHO data (2020) [3], 10 million global population endure from tuberculosis, and 1.2 million of them die yearly. Indonesia has an approximated number of people sick with tuberculosis reaching 845,000 with a death rate of 98,000. Several studies mention the factors that cause pulmonary TB, including geographic factors [4,5], environmental factors [6], and socio-economic factors [5–7], population [8,9], which affect the incidence rate of tuberculosis in an area.

Therefore, this study aims to develop a GWR with sufficient accuracy and conduct an in-depth exploration of the distribution aspects of pulmonary TB based on environmental and socio-economic data. The novelty of this research is developing an accurate GWR model based on environmental and socio-economic data and an in-depth exploration of the spatial disparity.

2. Materials and Methods

2.1 Study Area

Indonesia has high number of tuberculosis cases in Southeast Asia. Geographically, in 2020 Indonesia will be the third largest contributor of TB cases globally, namely 8.4% of cases of the estimated 86% of all incident TB cases worldwide [2].

2.2 Data

This study used two classes of data: socio-economic and environmental (Table 1).

Table 1. Data analyzed in this research

No	Data	Classification	Year	References
1	Pulmonary TB Cases	Socio-economic	2020	[10]
2	Population Density	Socio-economic	2020	[10]
3	Poverty Index	Socio-economic	2020	[10]
4	Health Facilities	Socio-economic	2020	[10]
5	Medical Personnel	Socio-economic	2020	[10]
6	Rainfall	Environmental	2020	[11]
7	Solar Radiation	Environmental	2020	[11]
8	Road Network	Environmental	2020	[11]

2.2.1 Socio-economic data

Population density shows the level of distribution of the population of an area. The population density figure shows the average number of people per 1 km^2 [10]. Pulmonary TB cases are data of pulmonary TB sufferers in 2020 [10]. The poverty index is an index that shows the minimum amount of budget needed to meet basic food and beverage needs and non-food basic needs [10]. Health facilities are data on the number of health facilities and health centers in Indonesia [10]. Medical personnel data on medical personnel in health facilities in Indonesia are sourced from the Health HR Information System Indonesia [10]. The road network is road length data by the province in kilometers (km) sourced from the Ministry of Public Works and Public Housing in Indonesian Statistics 2021 [11].

2.2.2 Environmental data

Rainfall is data on the amount of rainfall in millimeters (mm) sourced from the Meteorology and Geophysics Agency in Statistics Indonesia 2020 [11]. Solar radiation is data on the duration of solar radiation sourced from the Meteorology and Geophysics Agency in Statistics Indonesia 2020 [11].

2.3 Methodology

2.3.1 Ordinary Least Square (OLS)

OLS is a spatial regression method used to estimate spatial dependency in regression, test significance, and provide information on the spatial relationship between parameters [12,13]. The ordinary Least Square model is shown in equation (1).

$$Y_{\mathrm{i}} = \beta_0 + \beta_1 X_{1\mathrm{i}} + \beta_2 X_{2\mathrm{i}} + \cdots + \beta_{\mathrm{n}} X_{\mathrm{ni}} + \boldsymbol{\varepsilon}_{\mathrm{i}} \tag{1}$$

Where Y_{i} represents the predicted, $X_{1\mathrm{i}}$, $X_{2\mathrm{i}}$,… X_{ni} represents predictor, $\boldsymbol{\varepsilon}_{\mathrm{i}}$ represents an error, β_0, and β_1 … β_{n} represents intercept and coefficient respectively [14].

2.3.2 Geographically Weighted Regression (GWR)

GWR is a local regression technique used to measure how strong the correlation between a predicted variable and the predictor variable differs from one location to another. GWR collects adjacent data for each location to perform local regression, yielding estimates of local regression coefficients for each predictor variable [16]. Equation (2) shows the Geographically Weighted Regression function.

$$Y_{\mathrm{i}} = \beta_0(u_{\mathrm{i}}, v_{\mathrm{i}}) + \beta_1(u_{\mathrm{i}}, v_{\mathrm{i}})X_{1\mathrm{i}} + \beta_2(u_{\mathrm{i}}, v_{\mathrm{i}})X_{2\mathrm{i}} + \cdots + \beta_{\mathrm{n}}(u_{\mathrm{i}}, v_{\mathrm{i}})X_{\mathrm{ni}} + \boldsymbol{\varepsilon}_{\mathrm{i}}(u_{\mathrm{i}}, v_{\mathrm{i}}) \tag{2}$$

Where $\beta_0(u_{\mathrm{i}}, v_{\mathrm{i}}), \beta_1(u_{\mathrm{i}}, v_{\mathrm{i}}) \ldots \beta_{\mathrm{n}}(u_{\mathrm{i}}, v_{\mathrm{i}})$ Furthermore, $\boldsymbol{\varepsilon}_{\mathrm{i}}(u_{\mathrm{i}}, v_{\mathrm{i}})$ represent intercept, coefficient, and local error estimation, respectively, calculated by the spatial location-weighted OLD model $_{\mathrm{i}}(u_{\mathrm{i}}, v_{\mathrm{i}})$ [14].

3. Result and Discussion

3.1 Correlation and Ordinary Least Square (OLS)

In general, the relationship between cases of pulmonary TB and each variable is illustrated through a scatter plot and correlation coefficient in Figure 1. Based on the seven variables involved in this study, the poverty index variable (0.928), the number of health facilities (0.949), and the number of medical personnel (0.955) have a very significant relationship. In comparison, several other significant variables include population density (0.255) and the number of roads (0.627). In contrast, several variables were considered insignificant in pulmonary TB cases, including rainfall (0.109) and solar radiation (0.060).

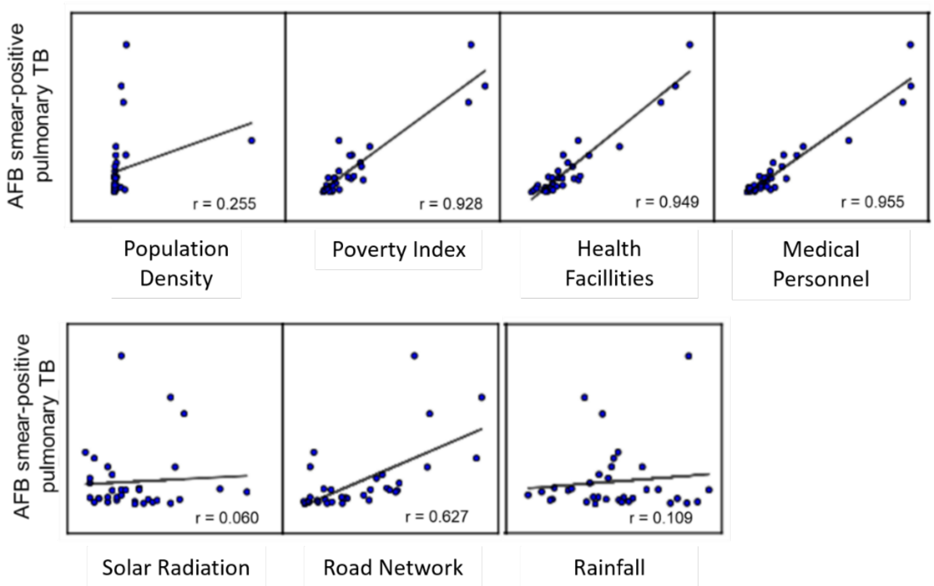

Figure 1. Scatter plot of pulmonary TB cases with each independent variable

Table 2. is the overall statistical test result for the OLS model. Based on the coefficient value, it can be observed that all variables have a positive influence. Several statistically significant and essential variables for the model are the poverty index, health facilities, and medical personnel. In addition, the seven variables used show VIF values ranging from 1.352 to 3.678. Based on Table 2, the multiple R^2, adjusted R^2, and AICc values obtained are 0.953, 0.940, and 611.862, respectively. Meanwhile, statistical probability indicators, including F-Statistic, Koenker (BP) statistic, Joint Wald Statistics, and Jarque-Bera statistic, show values of 75,233, 1532,262, 0.301, and 14,225.

Table 2. Statistical Summary of OLS Result

Variable	Coefficient	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr	VIF
Intercept	1998.676	2389.368	0,836	0.411	1939.205	1.031	0.312	--------
Population Density	-0.034	0.212	-0.162	0.873	0.170	-0.201	0.842	4.670
Poverty Index	0.183	1.449	0.126	0.901	1.409	0.130	0.898	39.256
Health Facilities	15.709	4.920	3.193	0.004*	6.132	2.562	0.017*	35.197
Medical Personnel	0.123	0.094	1.298	0.206	0.082	1.502	0.145	34.484
Rainfall	-0.027	0.331	-0.082	0.935	0.237	-0.115	0.910	1.426
Solar Radiation	-37.448	28.706	-1.304	0.203	24.080	-1.555	0.132	1.304
Road Networks	-0.145	0.059	-2.457	0.021*	0.077	-1.880	0.071	4.866
Diagnostics of OLS Result								
Number of Observations	34			Akaike's Information Criterion (AICc)				611.862
Multiple R-Squared	0.953			Adjusted R-Squared				0.940
Joint F-Statistic	75.233			Probability (>F), (7,26) deg of freedom				0.000*
Joint Wald Statistic	1532.262			Probability (>chi-squared), (7) deg of freedom				0.000*
Koenker (BP) Statistic	14.255			Probability (>chi-squared), (7) deg of freedom				0.047*
Jarque-Bera Statistic	0.336			Probability (>chi-squared), (2) deg of freedom				0.845

Equation (3) presents seven variables used to predict pulmonary TB cases.

$$Y_{\mathrm{i}} = 1998.676 - 0.034X_{1i} + 0.183X_{2i} + 15{,}709X_{3i} + 0{,}123\ X_{4i} - 0.027X_{5i} - 37.448X_{6i} - 0{,}145X_{7i} \quad (3)$$

Seven independent variables are population density (X_1), poverty index (X_2), health facilities (X_3), medical personnel (X_4), rainfall (X_5), solar radiation (X_6), and road network (X_7).

3.2 Pulmonary TB Cases Estimation Using GWR

The estimation of pulmonary TB cases based on GWR processing is divided into five classes, as shown in Figure 2. The very high class has a range of cases from 10577-25861, the high class has a range of cases from 6070-10576, the middle class has a range of cases from 3491-6069, the low class has a range of cases from 1467 to 3490, and very low class has a range of cases from 21 to 1466.

Indonesia has a positive smear-positive pulmonary tuberculosis rate with an average of 4.856 cases. Several provinces have a very high number of pulmonary TB cases, with more than 10,000 cases located in the Provinces, DKI Jakarta, West Java, Central Java, and East Java. The dominant aspects of pulmonary TB cases in Indonesia are the effects of population density, poverty index, number of health facilities, medical personnel, road network, rainfall, and solar radiation. In addition, several provinces were also observed to have a very low number of cases, including Bengkulu, Belitung Islands, Riau Islands, Central Kalimantan, North Kalimantan, West Sulawesi, Gorontalo, North Maluku, and West Papua.

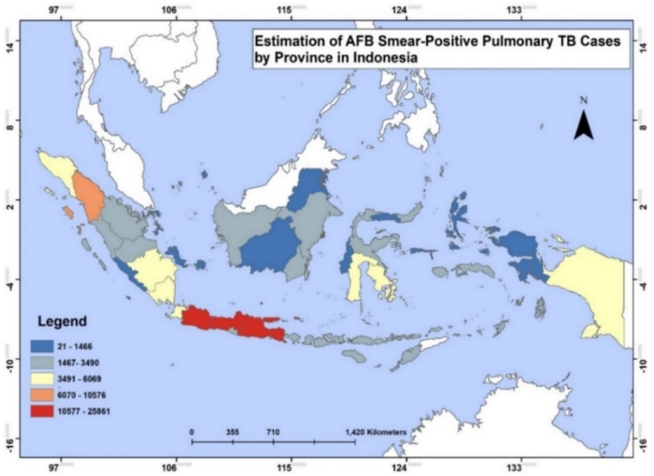

Figure 2. Map of the estimated number of pulmonary TB cases in Indonesia

4. Conclusions

Developing a spatial model of positive smear pulmonary tuberculosis cases using GWR in Indonesia based on environmental and socio-economic variables resulted in an excellent model with multiple R^2 values of 0.953 and adjusted R^2 values of 0.940. The resulting GWR model can explain the dependent variables by more than 90%. Therefore, several independent variables used in this study can be adopted for further model development. Based on the OLS, the resulting model is statistically significant with insignificant heteroscedasticity and non-stationarity, and the residuals are distributed normally and have no bias.

5. Acknowledgments

The author is thankful to appreciate the support from Sriwijaya University, all people and institutions who kindly provided their data for this study.

References

1. Ahmed, A.; Mekonnen, D.; Shiferaw, A.M.; Belayneh, F.; Yenit, M.K. Incidence and determinants of tuberculosis infection among adult patients with HIV attending HIV care in north-east Ethiopia: A retrospective cohort study. *BMJ Open* **2018**, *8*, 1–14, doi:10.1136/bmjopen-2017-016961.
2. World Health Organization *Global Tuberculosis Report 2021*; Geneva, 2021; ISBN 9789240037021.
3. World Health Organization *Global Tuberculosis Report 2020*; Geneva, 2020; ISBN 9789240013131.
4. Li, X.X.; Wang, L.X.; Zhang, H.; Jiang, S.W.; Fang, Q.; Chen, J.X.; Zhou, X.N. Spatial variations of pulmonary tuberculosis prevalence co-impacted by socio-economic and geographic factors in People's Republic of China, 2010. *BMC Public Health* **2014**, *14*, doi:10.1186/1471-2458-14-257.
5. Tanrikulu, A.C.; Acemoglu, H.; Palanci, Y.; Eren Dagli, C. Tuberculosis in Turkey: high altitude and other socio-economic risk factors. *Public Health* **2008**, *122*, 613–619, doi:10.1016/j.puhe.2007.09.005.
6. Li, X.X.; Ren, Z.P.; Wang, L.X.; Zhang, H.; Jiang, S.W.; Chen, J.X.; Wang, J.F.; Zhou, X.N. Co-endemicity of Pulmonary Tuberculosis and Intestinal Helminth Infection in the People's Republic of China. *PLoS Negl. Trop. Dis.* **2016**, *10*, doi:10.1371/journal.pntd.0004580.
7. Masabarakiza, P.; Adel Hassaan, M. Spatial-temporal Analysis of Tuberculosis Incidence in Burundi Using GIS. *Cent. African J. Public Heal.* **2019**, *5*, 280, doi:10.11648/j.cajph.20190506.19.
8. Rosli, N.M.; Shah, S.A.; Mahmood, M.I. Geographical Information System (Gis) Application in tuberculosis spatial clustering studies: A systematic review. *Malaysian J. Public Heal. Med.* **2018**, *18*, 70–80.
9. Tadesse, S.; Enqueselassie, F.; Hagos, S. Spatial and space-time clustering of tuberculosis in Gurage Zone, Southern Ethiopia. *PLoS One* **2018**, *13*, 1–14, doi:10.1371/journal.pone.0198353.
10. Minister of Health Republic of Indonesia *Indonesia Health Profile 2020*; Jakarta, 2020; ISBN 9786024160654.
11. Badan Pusat Statistik Indonesia Statistical Yearbook of Indonesia 2021. *Stat. Indones. 2021* **2021**, 790.
12. Kumar, C.; Singh, P.K.; Rai, R.K. Under-five mortality in high focus states in India: A district level geospatial analysis. *PLoS One* **2012**, *7*, doi:10.1371/journal.pone.0037515.
13. Mahara, G.; Yang, K.; Chen, S.; Wang, W.; Guo, X. Socio-Economic Predictors and Distribution of Tuberculosis Incidence in Beijing, China: A Study Using a Combination of Spatial Statistics and GIS Technology. *Med. Sci.* **2018**, *6*, 26, doi:10.3390/medsci6020026.
14. Li, C.; Li, F.; Wu, Z.; Cheng, J. Exploring spatially varying and scale-dependent relationships between soil contamination and landscape patterns using geographically weighted regression. *Appl. Geogr.* **2017**, *82*, 101–114, doi:10.1016/j.apgeog.2017.03.007.
15. Kirby, R.S.; Delmelle, E.; Eberth, J.M. Advances in spatial epidemiology and geographic information systems. *Ann. Epidemiol.* **2017**, *27*, 1–9, doi:10.1016/j.annepidem.2016.12.001.