Semarak International Journal of Entrepreneurship, Economics and Business Development

Journal homepage:
https://semarakilmu.my/index.php/sijeebd/index
ISSN: 3083-8053

The Spatial Analysis for Poverty in Malaysia Using the GWR and the PSDM GWR Model

Nur Edayu Zaini[1], Syerrina Zakaria[1,*], Nor Fatimah Che Sulaiman[2], Nuzlinda Abdul Rahman[3], Wan Saliha Wan Alwi[1]

[1] Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
[2] Faculty of Business, Economics & Social Development, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
[3] School of Mathematical and Sciences, Universiti Sains Malaysia Pulau Pinang, Malaysia

ARTICLE INFO

Article history:
Received 11 December 2024
Received in revised form 26 February 2025
Accepted 21 March 2025
Available online 30 March 2025

Keywords:
Ordinary Least Square (OLS) model; the Parameter-Specific Distance (PSDM) models; Geographical Weighted Regression (GWR); COVID-19

ABSTRACT

This research examines poverty inequalities from every angle, including their origins, effects on people and communities, and possible governmental responses to the problem. The population density, unemployment rate, non-citizen rate, median income, Gini income, mean expenses, crime rate, and COVID-19 incidence rate were among the factors that were included in this study. The parameter-specific distance (PSDM) models outperform the conventional geographical weighted regression (GWR) models using a multi-dimensional spatial methodology. The new PSDM model's findings indicate that poverty was significantly affected by median household income both before and after the pandemic. Poverty was found to be impacted by both the unemployment rate and Gini income after the pandemic. Furthermore, this research yields distinct relevant elements for each district. The spatial inference guidelines would give policymakers improved direction on the spatial analysis process by improving their comprehension of collinearity and type 1 error. Additionally, statistical inference approaches with integrated spatial modifications were used to analyze relevant variables. According to projections for 2024, Sabah districts were found to have a high prevalence of poverty, which calls for the government to take proactive measures by launching programs in the affected districts.

1. Introduction

Spatial poverty studies use geographic information and data analysis to uncover poverty patterns and variables at different spatial levels. Aggregated national or regional data often masks local differences; thus, more detailed geographical data is needed for accurate poverty assessments and targeted policies. In Malaysia, poverty is influenced by various socioeconomic and political factors. Despite significant progress, with the poverty rate dropping from 49.3% in 1970 to 1.7% in 2012, the issue remains complex [1].

* *Corresponding author.*
E-mail address: syerrina@umt.edu.my

https://doi.org/10.37934/sijeebd.3.1.1228a

Spatial poverty study is important because it gives policymakers a thorough grasp of how poverty is distributed and what factors influence it. Spatial analysis techniques using multi-source big data provide a more accurate assessment of urban poverty than traditional methods, which could miss fine-scale changes in poverty [2]. The unique difficulties faced by rural communities, which frequently have greater rates of poverty than urban regions, might not receive enough attention. In Islamic economics, poverty is measured by meeting basic and additional demands, including spiritual needs [3].

Poverty has become a global issue affecting both developed and developing countries due to its increased visibility through the media and its worsened effects brought on by inflation, technological advancement, and growing income disparities [4]. Therefore, poverty is a societal and psychological reality as well as an economic one, necessitating all-encompassing strategies and actions to reduce it. The measurement of poverty has transformed greatly, embracing both one-dimensional and multi-dimensional approaches. Poverty has conventionally been measured through a financial lens, with a focus on the economic means required to uphold a satisfactory standard of living. This approach remains very relevant because of its simplicity and direct connection to economic policies [5]. However, the multi-dimensional approach, which considers aspects like health, education, and other relational resources, has been increasingly acknowledged for its ability to offer a more thorough insight into well-being and lack [5].

Poverty in Malaysia is established by a detailed interplay of macroeconomic, spatial, and socio-demographic factors. The function of macroeconomic components, such as inflation, joblessness, and economic expansion, in moulding poverty levels has been extensively researched, highlighting the intricate nature of the phenomenon [6]. Inflation and unemployment exhibit a positive correlation with poverty levels, underscoring the exacerbating effect of higher inflation and unemployment rates [6,7]. Conversely, a reversed relationship is observed between public debt and poverty, indicating that effective fiscal governance could aid in reducing poverty challenges [7]. Spatial examination indicates that poverty is unevenly distributed throughout Malaysia, particularly concentrated in Kelantan, Terengganu, Penang, Kedah, Perlis, and Perak, where agricultural practices prevail and rural households confront heightened vulnerability to risks [8-10]. Moreover, noticeable regional gaps exist, with rural East Malaysia notably experiencing higher poverty rates compared to urban Peninsular Malaysia, underscoring the necessity for tailored poverty alleviation approaches at a regional level [11].

The approach utilized for spatially modeling poverty in Malaysia encompasses a blend of Geographic Information System (GIS) methodologies, spatial autocorrelation examination, and diverse spatial regression models. GIS is applied for the cartographic representation of poverty rates, demographic burdens, and poverty clusters, offering a visual depiction of poverty's spatial dispersion at sub-district levels [12]. The analysis of hotspots, through approaches like the Zone of Indifference and distance bands, assists in identifying regions with elevated poverty concentrations, notably in the northeastern territories of Kelantan and Terengganu [13]. Geographically Weighted Regression (GWR) stands as a pivotal method used to explore the connections between poverty and various explanatory factors, like expenditure per capita, life expectancy, and access to clean drinking water, which exhibit spatial variability across different areas [13,14]. Furthermore, the utilization of the Spatial Error Model (SEM) is crucial in handling the spatial dependence of errors in observational data, while GWR addresses the heterogeneity of spatial variance.

Spatial analysis is a methodology utilized to examine the spatial data's locations, attributes, and relationships through diverse analytical approaches. The Geographically Weighted Regression (GWR) model represents a sophisticated approach to spatial analysis. Previous studies stated that GWR serves as a localized variant of linear regression employed to depict spatially diverse relationships by enabling distinct associations to occur at different spatial points, thereby tackling spatial non-stationarity issues that conventional global models like Ordinary Least Squares (OLS) struggle with [6,15,16]. Likewise, GWR has been utilized to investigate the influence of lightning strikes on distribution transformers, showcasing enhanced explanatory capabilities over OLS by identifying points with significant explanatory power and regression coefficients [3]. GWR's adaptability enables the accommodation of various models and analyses, demonstrating a commitment to exploring spatial variability in model parameters and results proved by Bagde and Krishnakumar [17,18]. GWR is a potent instrument in spatial analysis, furnishing detailed perspectives on spatial heterogeneity and enhancing the precision of spatial models.

Parameter-specific distance metric (PSDM) for Geographically Weighted Regression (GWR) is a novel approach that allows for the calibration of GWR models using different distance metrics for each parameter, enabling spatial relationships to vary in intensity based on location and direction. This method addresses the limitations of traditional GWR models that use a single distance metric, such as Euclidean or non-Euclidean, by introducing parameter-specific distance metrics that enhance model calibration and prediction accuracy. The PSDM GWR model has been shown to improve goodness of fit and provide insights into how regression relationships may vary across different spatial scales, offering a more nuanced understanding of spatial heterogeneities in data relationships [4-7]. While Lu *et al.,* [5,13] introduced PSDM GWR, its computational burden required attention, and as such, only a relatively limited form of PSDM GWR was demonstrated. Later, Lu *et. al.,* [6] improved GWR calibration with parameter-specific distance metrics and bandwidths by using a Brute-force search for optimal metrics, the Minkowski approach, and empirical validation.

Furthermore, the pandemic's consequences on productivity growth in Malaysia are a cause for concern, given the series of economic setbacks that have impeded productivity advancements in recent times, which were examined by Hassan *et al.,* [20]. The COVID-19 pandemic has influenced shifts in poverty rates in both rural and urban areas. Before the unexpected global health crisis, Malaysia had significantly reduced poverty and income inequality; however, these gains are now in jeopardy, as obtained from Rashid *et al.,* [21]. However, there is still a gap in how robust have Malaysia's poverty alleviation programs been in maintaining poverty reduction achievements during times of crisis. Many households have experienced job loss, income reduction, and financial insecurity due to the disruption of economic operations caused by the imposition of national lockdowns and mobility restrictions to contain the virus [22].

This study seeks to contribute to the existing body of knowledge by providing a comprehensive assessment of the spatial modeling and spatial inference procedure of poverty in Malaysia. There is a need for targeted interventions that address the specific causes of poverty in different regions or communities. National strategies may need to be adapted to the unique local needs of each region, accounting for cultural, economic, and political differences. By understanding the dynamics of how the pandemic has affected the incidence of poverty, this research aims to enlighten targeted policy implications that can address the emerging challenges of poverty disparity and promote a more resilient and equitable economic recovery in the post-pandemic era.

2. Methodology

2.1 The Data and Research Design

The Department of Statistics Malaysia (DOSM) supplied the data for this research. The Household Income and Basic Amenities Survey (HIS&BA), conducted via in-person interviews over

12 months during 2016, 2019, and 2022, yielded household income data. The survey employed probability sampling techniques based on a household list from the Population and Housing Census. Malaysia is composed of two principal regions: West Malaysia and East Malaysia. West Malaysia, or Peninsular Malaysia, encompasses 144 districts across 13 states and three Federal Territories, including Kuala Lumpur, the capital. The Federal Territories of Putrajaya and Labuan are also included. In contrast, East Malaysia consists of two states, Sabah and Sarawak. Each state is subdivided into several administrative districts, as illustrated in the Appendix. The regions of Malaysia are classified as Northern, Central, East Coast, Southern, and East Malaysia. Table 1 presents this study's list of dependent and independent variables. Poverty rate calculation involves all households having monthly gross income below Poverty Line Income. The formula for calculating the Incidence of Poverty (IP) is as Eq. (1).

$$\mathrm{IP} = \frac{\text{Number of households with income below the PLI}}{\text{Total number of households}} \times 100 \tag{1}$$

Table 1
The list of dependent and independent variables applied in this study

Indicators	Variables
POV	Incidence of poverty (as dependent)
POPDEN	Population Density
NONCITIRATE	Noncitizen Rate
UNEMPRATE	Unemployment Rate
MFRATIO	Male Female Ratio
GININC	Gini Income
MEXP	Mean Expenses
MEDINC	Median household Income
CRIME	Number of crimes
COVID19	The incidence of COVID-19

The unemployment rate is defined as those who did not work during reference work but are interested in working and seeking a job. They are classified into two groups: actively and inactively unemployed. This can be calculated by Eq. (2). Meanwhile, the noncitizen rate is computed as Eq. (3), where the number of noncitizens is divided by number of populations.

$$\text{Unemployment Rate} = \frac{\text{Number of unemployed person}}{\text{Number of person in labour force}} \times 100 \tag{2}$$

$$\text{Noncitizen Rate} = \frac{\text{Number of noncitizen}}{\text{Number of population}} \times 100 \tag{3}$$

The mean expenses can be obtained by Eq. (4), where the total expenditure in each district will be divided by the number of households.

$$\text{Mean Expenses} = \frac{\text{Total Expenditure}}{\text{Number of households}} \tag{4}$$

Next, the number of COVID-19 cases was transformed into the incidence rate per 100,000 people. The objective of the transformation was to portray the actual frequency of COVID-19 cases over a specific period as a proportion of the people at risk for the disease (population). This rate, obtained from an earlier study by Tosepu *et al.,* [23], is a good reflection of a community's state of the pandemic. The incidence rate can be calculated as Eq. (5).

$$\text{Incidence Rate (IR)} = \frac{\text{The number of cases}}{\text{Total population}} \times 100{,}000 \tag{5}$$

Figure 1 shows a flowchart of this study. First, the spatial data were entered for dependent and independent variables with latitude and longitude. Second, univariate and spatial modeling was conducted to obtain the estimated poverty incidence coefficient. The GWR and PSDM models were compared during the model diagnostic phase in the third phase. The best model was selected for collinearity diagnostic testing to proceed or to repeat the process. After several independent variables were removable, the new model was constructed in this study. In the last phase, the model selected will go through model evaluation and prediction for the incidence of poverty.

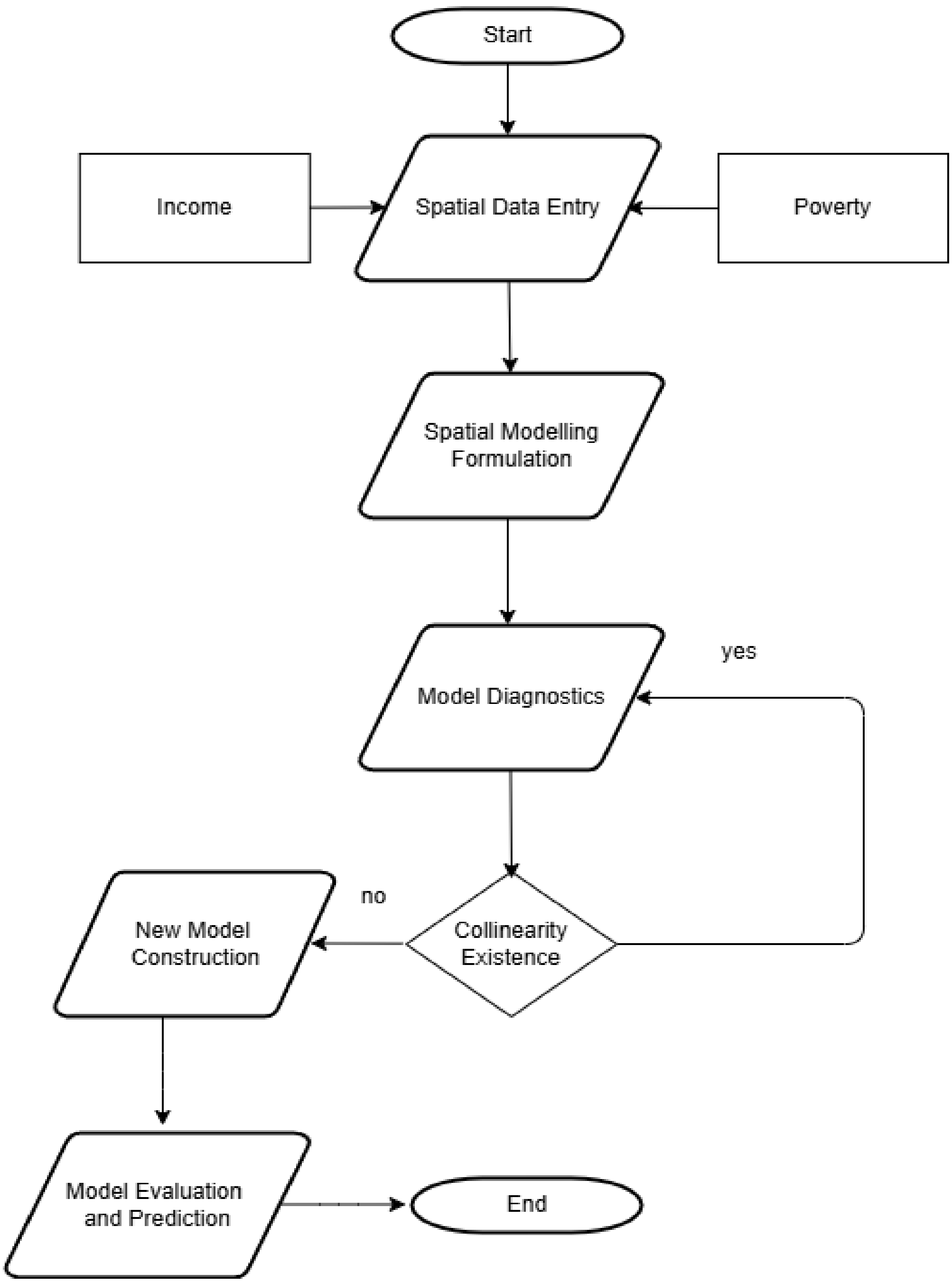

Fig. 1. The research flow chart

2.2 The GWR Model

The Geographical Weighted Regression, as introduced by Fotheringham *et al.,* [24] represents the initial approach. It extends the multiple linear regression models, incorporating considerations for local variations. The coefficients within the model deviate from global estimates specific to location i. The primary aim of GWR is to generate predictions and parameter estimates across a predetermined set of locations, with each of these locations corresponding to regression points. Within this model, the integration of local spatial relationships between dependent and independent variables is a crucial feature of the regression framework.

$$y_i = \beta_0(\mu_i, v_i) + \sum_{j=1}^{m} \beta_j (\mu_i, v_i) x_{ij} + \varepsilon_i \tag{6}$$

The Eq. (10) indicates $\beta_0(\mu_i, v_i)$ is a constant function and $\beta_j(\mu_i, v_i)$ is a continuous function, representing a GWR coefficient in subdistrict i. The location point of subdistrict i is defined by latitude and longitude coordinates (μ_i, v_i) and ε is a random error term at the location. In the GWR modelling, the initial step involves defining the latitude and longitude coordinates (μ_i, v_i) for each subdistrict, followed by determining the appropriate distance to be utilised. In the GWR model, these geographical coordinates are employed to calculate the Euclidean distance between the observed data in subdistrict i and subdistrict j within the village, as illustrated in Table 2. Meanwhile, the distances used in this study for the PSDM models were Euclidean and Minkowski.

Table 2
The different types of distance function

Distance types	Functions	
Euclidean	$d = \sqrt{(\mu_i - \mu_j)^2 + (v_i - v_j)^2}$	(7)
Minkowski	$d = (\sum_{i=1}^{m} \lvert \mu_i - v_j \rvert^p)^{1/p}$ *M-dimension* Euclidean space and *p* is a positive real number.	(8)

The distance specified in Eq. (7) and Eq. (8) is a fundamental metric for weighting the data in the parameter estimation process for GWR. Perhaps a Minkowski distance is another option for spatial analysis—a shorter distance between subdistricts results in a higher weighting of the data during parameter estimation. The process of weighting was carried out by utilising an adaptive bi-square of function, which is detailed in Eq. (9)

$$\psi_{ij} = \begin{cases} \left(1 - (d_{ij}/b_{i(k)})^2\right)^2 & d_{ij} < b_{i(k)} \\ 0 & d_{ij} > b_{i(k)} \end{cases} \tag{9}$$

The $b > 0$ is a constant bandwidth which is determined through a cross-validation method, and $\psi_{ij}$is the weight assigned to an observation at the location j is used to estimate the coefficient at that location i. Meanwhile, d_{ij} is the Euclidean distance between two points is denoted as i and j. An adaptive bandwidth size $b_{i(k)}$ is defined as the kth nearest neighbourhood distance was mentioned by Waller *et al.,* [28] All data observed within a specific subdistrict demonstrate a consistent estimated parameter. This phenomenon can be attributed to the data weighting exclusively considering the distances between the subdistricts represented in Eq. (10).

$$\hat{\beta}(\mu_i, v_i) = [X^T W(\mu_i, v_i) X]^{-1} X^T W(\mu_i, v_i) Y \quad (10)$$

An approach to estimate the kernel bandwidth not based on a prediction of the response variable is the corrected AIC, adopted in form from locally weighted regression to GWR. Instead, it is based on minimising the estimation error of the response variable. It is a compromise between the model's goodness-of-fit and model complexity in that there is a penalty in the criterion for the effective number of parameters in the model. The corrected AIC for GWR is presented as Eq. (11).

$$AIC_C = 2n\log(\hat{\sigma}) + n\log(2\pi) + n\left(\frac{n+\text{trace}(H)}{n-2-\text{trace}\ (H)}\right) \quad (11)$$

2.3 The PSDM model

In standard GWR, the weighting matrix is calculated using an ED metric and a unique bandwidth. This assumes that 'as the crow flies' distances are appropriate throughout and that any dependent/independent variable relationship varies at the same spatial scale. However, the spatially varying scale or intensity of the different dependent/independent variable relationships may differ, and as such, each should have its distinct weighting scheme within the same model [27,28]. Such situations are catered for with PSDM GWR, which is implemented via an adjusted back-fitting algorithm as used in mixed GWR and FBGWR, each a particular case of PSDM GWR. The PSDM GWR model can be expressed as Eq. (12),

$$y_i = \beta_{0i}^{(DM_0, bw_0)}(u_i, v_i) + \sum_{k=1}^{M} \beta_{ki}^{(DM_k, bw_k)}(u_i, v_i)\, x_{ik} + \varepsilon_i \quad (12)$$

where DM_k and bw_k (k =0,1, ⋯, m) represent the specific distance metric and bandwidth for each independent variable (and intercept) parameter estimate. A complete account of PSDM GWR is provided in research by Lu *et. al.,* [27]. To choose an optimum bandwidth for each parameter of PSDM GWR, an optimisation can be conducted by minimising the CV score or the corrected AIC (AICc) within the back-fitting iterations was conducted by several authors [27,28]. Note that AICc takes into account the adequate sample size. However, as spatial autocorrelation is likely, the effective sample size is expected to be much smaller than the nominal sample size.

At the latest, PSDM GWR is calibrated with ED and TT candidate metrics to a London house price data set, following the brute-force procedure described. In addition, the Minkowski approach is applied, where seven different Minkowski distances are viewed as candidate metrics was also conducted by Lu *et al.,* [29]. The Minkowski approach can help approximate the underlying 'optimum' metric when no prior or practical knowledge of the geographical process is known. Furthermore, it uses simulated data to verify the empirical results. Table 3 shows a difference and similarity for the GWR and PSDM models used in this study.

3. Results

3.1 The Preliminary Analysis

Figure 4 illustrates the decreasing trend of poverty incidence from 1995 to 2020, which reflects the effectiveness of the government's initiative to eradicate poverty. From 19.5 per cent in 1995 to 17.4 per cent in 2007 and 16.2 per cent in 2020, the poverty rate has been steadily declining. The 2005 Poverty Line Income (PLI) technique is applied to determine this poverty incidence.

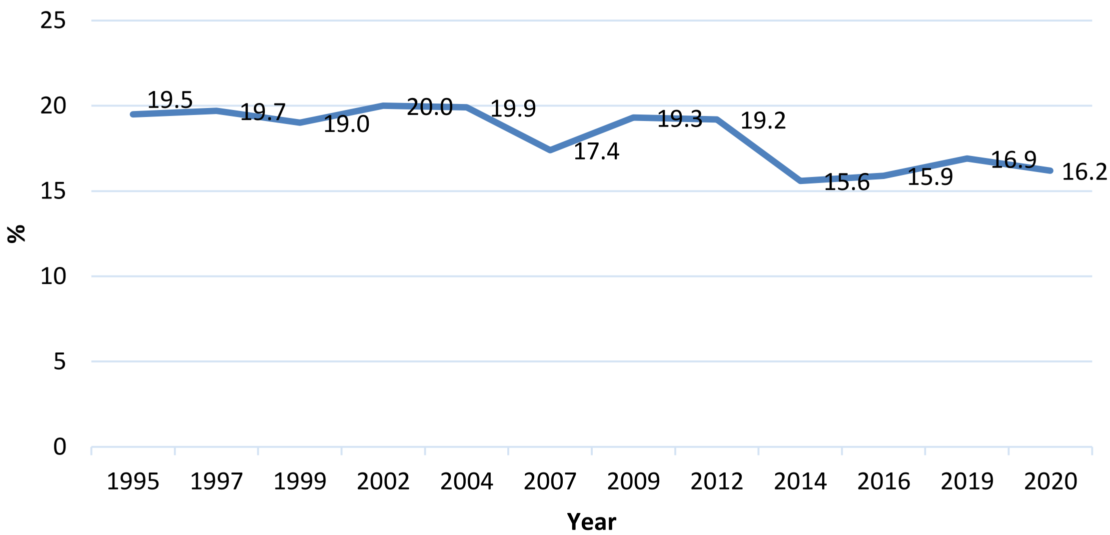

Fig. 4. The incidence of poverty in Malaysia, 1995 -2020

The idea of assessing relative poverty differs from that of absolute poverty. The relative poverty standard is established according to the median income of entire households. In the data represented in Figure 5, Sarawak (16.2), Johor (15.9), and Pulau Pinang (15.3) manifested the highest levels of relative poverty in the year 2022. Post the emergence of the COVID-19 pandemic, most states encountered a growth in poverty rates, except for Terengganu, Selangor, Perlis, and Melaka. Terengganu reported the lowest poverty rate at 6.9 and is the sole state displaying a positive trend towards poverty reduction compared to others.

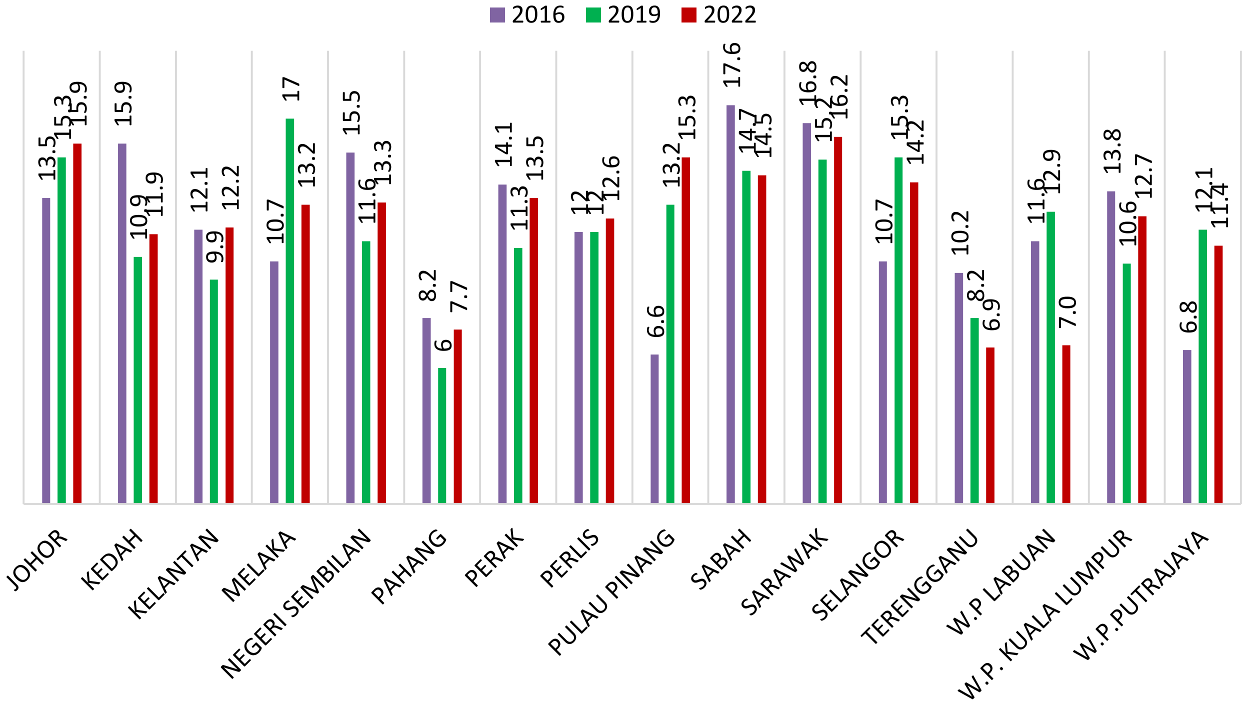

Fig. 5. The incidence of poverty by state in Malaysia

3.2 The GWR and PSDM Modelling

The GWR and PSDM models' bandwidth selections were optimized using a cross-validation approach under an adaptive bi-square weighting kernel. The GWR model used a Euclidean distance function for three years, while the PSDM with GWR for the incidence of poverty used a Minkowski distance for 2019, and 2022. The estimated coefficients of GWR and PSDM were summarised in Tables 3, 4 and 5. Moreover, different bandwidths were available for each independent variable.

Table 3 shows that the GWR coefficient estimates show a higher variation than the PSDM ones—as indicated by the interquartile range (IQR)—except for the intercept. The low variation of the intercept could be caused by the single average bandwidth of the GWR model, which is narrower than the bespoke bandwidth for the individual predictor but broader than the bandwidth for the intercept from the PSDM.

Table 3
The coefficient estimates arising from the GWR and PSDM models (1Q = first quartile, Med = median, 3Q = third quartile, IQR = interquartile range), 2016

Coefficient	GWR 1Q	Bandwidth: 142 Med	3Q	IQR	PSDM Bandwidth (km)	1Q	Med	3Q	IQR
Intercept	5.955	7.156	11.426	17.322	29	-0.735	7.206	13.625	28.601
POPDEN16	0	0	0.001	0.003	143	0.001	0.001	0.001	0.001
NONCITIRATE16	-7.745	1.977	-2.305	2.508	106	-3.507	3.76	5.511	15.767
UNEMPRATE16	-0.031	-0.031	-0.001	0.037	143	-0.035	-0.034	-0.031	-0.032
MFRATIO16	0.151	0.668	0.456	0.674	86	0.126	0.13	0.135	0.136
MEDINC16	-0.004	-0.004	-0.004	-0.004	29	-0.006	-0.006	-0.005	-0.002
GININC16	53.98	55.528	76.582	105.513	92	44.207	44.733	50.957	52.035
MEXP16	-0.005	-0.001	-0.002	-0.001	143	-0.001	-0.001	-0.001	-0.001

Table 4
The coefficient estimates arising from the GWR and PSDM models (1Q = first quartile, Med = median, 3Q = third quartile, IQR = interquartile range), 2019

Coefficient	GWR 1Q	Bandwidth: 142 Med	3Q	IQR	PSDM Bandwidth (km)	1Q	Med	3Q	IQR
Intercept	2.611	21.371	20.031	25.222	113	13.365	28.421	31.189	39.356
POPDEN19	0.001	0.002	0.003	0.006	114	0.001	0.001	0.002	0.003
MFRATIO19	-9.733	1.697	0.479	5.731	104	-3.582	-3.452	-1.786	0.494
NONCITIRATE19	0.011	0.04	0.072	0.12	132	0.013	0.014	0.041	0.079
UNEMPRATE19	-1.406	0.191	-0.307	0.317	143	-0.214	-0.164	0.152	0.644
MEDINC19	-0.006	-0.002	-0.003	-0.001	14	-0.006	-0.003	-0.004	-0.001
GININC19	7.424	29.019	36.614	66.029	92	9.578	24.604	16.809	26.611
MEXP19	-0.006	-0.001	-0.003	0	124	-0.004	-0.002	-0.003	-0.002
CRIME19	-0.001	0	0.001	0.001	92	-0.001	-0.001	0.001	0.003
COVID19	-0.003	0	-0.001	0	107	-0.001	0	-0.001	0

Table 5 shows that the GWR coefficient estimates show almost the same variation as the PSDM ones—as indicated by the interquartile range (IQR)—except for the intercept. The GWR coefficients MFRATIO22 and GININC22 have high variation compared to PSDM. Meanwhile, the PSDM coefficient UNEMPRATE22 recorded a higher variation than the GWR model.

Table 5
The coefficient estimates arising from the GWR and PSDM models (1Q = first quartile, Med = median, 3Q = third quartile, IQR = interquartile range), 2022

Coefficient	GWR 1Q	Bandwidth: 142 Med	3Q	IQR	PSDM Bandwidth (km)	1Q	Med	3Q	IQR
Intercept	-11.333	-2.524	-2.82	4.357	49	-8.744	-4.637	0.781	2.935
POPDEN22	-0.001	-0.001	0	0	141	0	0	0	0.001
MFRATIO22	2.96	7.478	8.344	16.953	137	-0.084	-0.081	-0.019	0.08
NONCITIRATE22	-0.127	0.179	0.056	0.218	119	-0.41	-0.237	-0.179	-0.043
UNEMPRATE22	-0.732	-0.567	-0.378	-0.069	143	9.846	9.908	12.327	15.157
MEDINC22	-0.005	-0.004	-0.004	-0.001	19	-0.006	-0.003	-0.004	-0.001
GININC22	41.221	56.097	52.246	67.519	68	38.159	42.041	44.64	49.176
MEXP22	-0.001	0.002	0.002	0.005	101	0	0	0	0
CRIME22	-0.003	0	-0.002	0	143	-0.001	-0.001	0.001	0.004

Table 6 illustrates a model diagnostic for the occurrences of poverty utilising four distinct models. The progression from the global model to the spatial model is demonstrated in the adjusted values R^2 shown. The outcomes of poverty incidence over three years reveal that the PSDM model exhibits the lowest AICc and the highest adjusted R^2 value. Moreover, there is an enhancement in the adjusted R-squared values for the PSDM model, with percentages of 0.853 in 2016, 0.8959 in 2019, and 0.8749 in 2022. The AICc values for the PSDM model are 875.134 in 2016, 811.027 in 2019, and 810.03 in 2022. In conclusion, despite the GWR model yielding a comparable output value, the PSDM model demonstrates superior performance.

Table 6
The model diagnostic analysis for the incidence of poverty for all models

Year	Adjusted R^2			AICc		
	GWR	PSDM	New PSDM	GWR	PSDM	New PSDM

2016	0.7645629	0.8531161	0.8399091	922.6104	875.1346	885.1961
2019	0.8202552	0.8959085	0.8828845	910.5249	811.0277	808.6063
2022	0.8077986	0.8749844	0.876566	896.4817	810.0319	803.5507

3.3 The Collinearity Testing

As collinearity between variables may degrade coefficient estimate precision in GWR and potentially invalidate interpretation about the coefficient, locally weighted VIFs are checked here as a diagnostic of collinearity. A local VIF is calculated for each independent variable concerning each district's corresponding geographically weighted scheme. While VIFs that exceed ten are usually regarded as problematic as conducted by Gollini *et. al.,* [30], local VIFs greater than a stricter value of 3 are counted as performed by previous by Yang [31]. Moreover, it was reported before to be more cautious and consistent with the global multicollinearity check. Table 7 lists the count of districts that have local VIFs greater than 3 for each variable in the context of various year models. Meanwhile, the local variance decomposition proportions (VDPs) at districts greater than 0.5 are stated here to be more sensitive to the global multicollinearity problem. These diagnostics are considered an integral part of an analytical toolkit that should always be employed in any GW regression analysis.

Table 7 shows a collinearity result for incidences of poverty in 2016, 2019 and 2022. For 2019, the VIFs value was higher for POPDEN, MEXP and COVID-19, with 37 districts, 44 districts and 44 districts, respectively. For 2022, the VIF value was higher for MFRATIO, NONCITIRATE, and MEXP, with eight districts, 78 districts, and 144 districts, respectively. Meanwhile, Table 7 shows 46 districts registered as VDP greater than 0.5 for MFRATIO for 2016, 17 for 2019 and 144 for 2022. The NONCITIRATE obtained 66 districts registered as VDP greater than 0.5 for 2022. The GININC register as VDP was more significant than 0.5 by 138 districts in 2016 and 87 districts in 2019. Lastly, the number of crimes obtains a VDP greater than 0.5 by ten districts for 2016 and 15 districts for 2022. As the collinearity problem exists after performing the VIFs and VDPs, the simple removal of one variable from the analysis may go some way in alleviating this problem before proceeding to a more locally focused analysis.

Table 7
The count of districts of collinearity test for incidence of poverty for 2019

Variable	(VIFs)			(VDPs)		
	2016	2019	2022	2016	2019	2022
POPDEN	0	37	0	0	0	0
MFRATIO	0	0	8	46	17	144
UNEMPRATE	0	0	0	0	0	0
NONCITIRATE	0	0	78	0	0	66
MEDINC	0	0	0	0	0	0
GININC	0	0	0	138	87	0
CRIME	0	0	0	10	0	15
MEXP	0	44	144	0	0	0
COVID19	-	44	-	-	0	-

3.4 The New PSDM Model

The new PSDM GWR for the incidence of poverty uses an adaptive bi-square kernel for 2016, 2019, and 2022, as shown in Table 8. It applies a Minkowski distance for three years. Moreover, different bandwidths were available for each independent variable.

Table 8
The selected bandwidth of the new PSDM GWR for the poverty model

Year	Intercept	POPDEN	NONCITIRATE	UNEMPRATE	MFRATIO	MEDINC	GININC	MEXP	CRIME	COVID19
2016	29	143	89	143	X	20	X	143	X	-
2019	49	X	143	52	X	19	X	X	108	X
2022	49	143	X	49	X	18	69	X	X	-

Remark: 'X' for Deleted variables and '- 'for unavailability

Table 9 represents the estimated coefficient using the new PSDM model of poverty for the year 2016. The coefficient's value consists of a minimum, first quartile, median, mean, third quartile, and maximum. Table 9 represents the estimated coefficient values such as POPDEN16, NONCITIRATE16, UNEMPRATE16, MEDINC16, and MEXP16. The estimated coefficient of MEDINC16 on the incidence of poverty ranged between -0.015 and 0.002, and there is a negative coefficient in the study area. The NONCITIRATE16 rate gives a negative coefficient with the incidence of poverty. Moreover, a pessimistic estimate coefficient in the local model contradicts the OLS model for the NONCITIRATE16.

Table 9
The summarised estimated coefficient new PSDM model for the incidence of poverty in Malaysia, 2016

	Intercept	POPDEN16	NONCITIRATE16	UNEMPRATE16	MEDINC16	MEXP16	Intercept
Min.	15.492	0.000	-0.097	0.121	-0.015	-0.002	15.492
1st Qu.	22.97	0.001	-0.04	0.146	-0.008	-0.001	22.97
Median	39.991	0.001	-0.011	0.179	-0.006	0.000	39.991
Mean	40.115	0.001	-0.027	0.191	-0.006	-0.001	40.115
3rd Qu.	46.705	0.001	-0.008	0.24	-0.003	0.000	46.705
Max.	91.699	0.001	-0.004	0.261	-0.002	0.000	91.699

Table 10 shows an example of selected districts for the new PSDM model for poverty in 2016. Eq. (13) represents the new PSDM model at Kluang. Meanwhile, Eq. (14) represents the new PSDM model for Limbang, and Eq. (15) represents the new PSDM model for Marang.

Table 10
The example of the new PSDM model of poverty 2016 for selected districts

ID	Districts	PSDM equations	Eq.
3	Kluang	$21.73996 + 0.0005POPDEN16 - 0.00545NONCITI16 + 0.1988UNEMPRATE16 - 0.00286MEDINC16 - 0.00027MEXP16$	(13)
110	Limbang	$49.84519 + 0.00343POPDEN16 - 0.0435NONCITI16 + 0.2333UNEMPRATE16 - 0.00641MEDINC16 - 0.00151MEXP16$	(14)
143	Marang	$35.69909 + 0.000562PODEN16 - 0.01162NONCITIRATE16 + 0.122516UNEMPRATE16 - 0.00593MEDINC16 - 0.00027MEXP16$	(15)

Table 11 represents the estimated coefficient using the new PSDM poverty for 2019. The coefficient's value consists of a minimum, first quartile, median, mean, third quartile, and maximum. The coefficient shows a positive and negative value in different districts. Table 11 represents the estimated coefficient values of NONCITIRATE19, UNEMPRATE19, MEDINC19, and CRIME19. The estimated coefficient of UNEMPRATE19 on the incidence of poverty ranged between -1.377 and 0.965, and there is a positive and negative relationship in the study area.

Table 11
The summarised estimated coefficient of the new PSDM model for the incidence of poverty in Malaysia, 2019

	Intercept	NONCITIRATE19	UNEMPRATE19	MEDINC19	CRIME19
Min.	8.353	0.025	-1.377	-0.018	0.00
1st Qu.	11.633	0.025	-0.847	-0.006	0.00
Median	30.451	0.026	0.136	-0.005	0.00
Mean	31.59	0.066	-0.063	-0.005	0.001
3rd Qu.	38.845	0.128	0.297	-0.001	0.003
Max.	86.968	0.136	0.965	-0.001	0.003

Table 12 shows an example of selected districts for the new PSDM model for poverty in 2019. Eq. (16) represents the new PSDM model at Kota Tinggi. Meanwhile, Eq. (17) represents the new PSDM model for Padang Terap, and Eq. (18) represents the new PSDM model for WP Putrajaya.

Table 12
The example of the new PSDM model of poverty 2019 for selected districts

ID	Districts	PSDM equations	Eq.
4	Kota Tinggi	$10.90661 - 0.026395NONCITI - 0.13875UNEMPRATE - 0.00108MEDINC + 0.0001CRIME$	(16)
18	Padang Terap	$42.47983 + 0.024862NONCITI - 1.2913UNEMPRATE - 0.00693MEDINC + 0.000131CRIME$	(17)
72	WP Putrajaya	$8.747227 + 0.05NONCITI + 0.1905UNEMPRATE - 0.001MEDINC + 0.000119CRIME$	(18)

Table 13 represents the estimated coefficient using the new PSDM model of poverty for 2022; the coefficient value consists of a minimum, first quartile, median, mean, third quartile, and maximum. The value coefficient shows a positive and negative value in different districts. In addition, Table 13 represents the estimated coefficient values such as POPDEN22, NONCITIRATE22, UNEMPRATE22, MEDINC22, and GININC22. The estimated coefficient of UNEMPRATE22 ranged between -1.682 and 0.466, with a positive and negative relationship in the study area.

Table 13
The summarised estimated coefficient for the new PSDM model of incidence of poverty in Malaysia, 2022

	Intercept	POPDEN22	UNEMPRATE22	MEDINC22	GININC22
Min.	-2.389	-0.001	-1.682	-0.017	22.042
1st Qu.	1.318	-0.001	-0.489	-0.005	33.049
Median	7.555	-0.001	-0.203	-0.003	51.631
Mean	14.144	0.000	-0.223	-0.004	46.193
3rd Qu.	17.817	0.0000	0.213	-0.002	55.369
Max.	79.304	0.000	0.466	-0.001	62.91
Global	2.315	0	0.255	-0.001	69.286

Table 14 shows an example of selected districts for the new PSDM model for poverty in 2022. Eq. (19) represents the new PSDM model at Kuala Muda. Meanwhile, Eq. (20) represents the new PSDM model for Putatan, and Eq. (21) represents the new PSDM model for Petaling.

Table 14
The example of the new PSDM model of poverty 2022 for selected districts

ID	Districts	The PSDM equations	Eq.
14	Kuala Muda	$2.017493 - 0.0005POPDEN - 0.48943UNEMPRATE - 0.00278MEDINC + 59.47194GININC$	(19)
88	Putatan	$28.10287 + 0.000457POPDEN - 0.22195UNEMPRATE + 51.70338MEDINC + 10.37805GININC$	(20)
135	Petaling	$-1.52375 - 0.0005POPDEN + 0.248359UNEMPRATE - 0.0074MEDINC + 30.70121GININC$	(21)

In conclusion, the PSDM model provides an improved output from the GWR output. Generally, these analyses of the PSDM coefficient give a picture of their positive or negative influence on the dependent variable. The details on the percentage of districts that influenced the dependent variable were obtained through the hypothesis procedure of spatial data. The following section describes this procedure.

3.5 The Model Evaluation and Prediction

For the evaluation part, Table 15 summarises the measurement error for the income and poverty models. The results show that the new PSDM model provides the lowest values of MSE and RMSE. For income, the new PSDM for poverty shows 18.12 for MSE and 4.257 for RMSE. Overall, the new PSDM model gives more accurate predictions than traditional GWR.

Table 15
The measurement error of forecasting

	MSE	RMSE
GWR	26.09951573	5.108768514
New PSDM	18.12427017	4.257260876

Figure 6 predicts the incidence of poverty in Malaysia for the year 2024. The results of high value indicate that districts were severely impoverished. About six districts, located in several Sabah districts, obtained a high incidence of poverty above 33.7. About 50 districts obtained about 12.37 to 33.77 incidence of poverty, generally located in East Malaysia, Perak, and Kelantan. Around 88 districts have an incidence of poverty of less than 12.377, primarily located in Peninsular Malaysia. In terms of the incidence of poverty, peninsular Malaysia has successfully reduced the amount of poverty, except for some districts in Kelantan. Meanwhile, the government should focus more on East Malaysia in terms of improving infrastructure, providing the chance for education, increasing job employment for citizens, and other potential solutions.

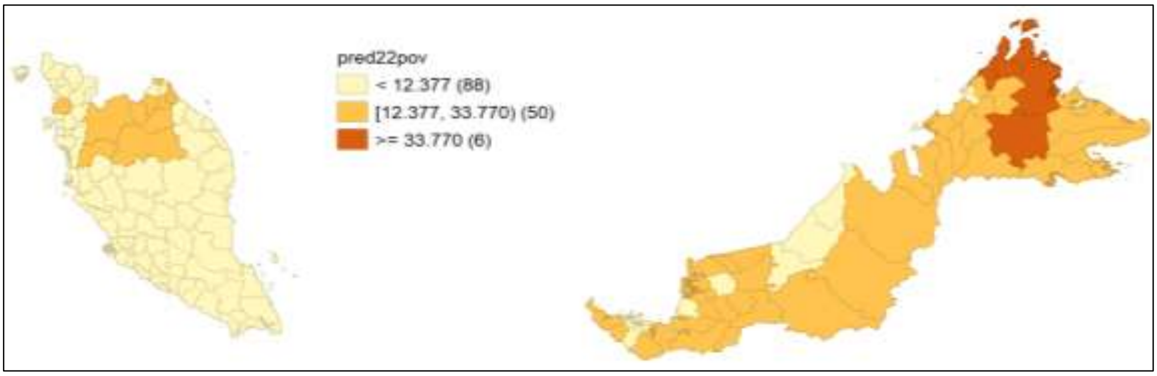

Fig. 6. The prediction for the incidence of poverty in 2024

4. Conclusion

To gauge the district's success in terms of poverty, the incidence of poverty was chosen. In Malaysia, the income gap between districts is still rather large and is predicted to get worse over the next several years. Furthermore, poverty rates show a strong resemblance to those in Sabah's neighbouring districts. Sabah's poverty rate is unchanged, although it has somewhat changed after the COVID-19 pandemic affected Pahang and Kedah, Malaysia.

This paper provides a thorough analysis of both global and spatial models, highlighting how spatial modelling may improve outcomes. When it comes to analyzing socioeconomic variables like the prevalence of poverty, the PSDM performs better. In terms of analysis, the GWR outperforms global analysis as the best model for individual districts. AICc and adjusted R square are important metrics for choosing the best model to solve the problem. Additionally, districts that show strong positive or negative associations with independent variables can be identified using the GWR mapping coefficient. In terms of median poverty, the findings show a wide range of differences between before (2016 and 2019) and after (2022) the COVID-19 pandemic in terms of population density, gender ratios, unemployment rates, non-citizen percentages, and income disparity. According to projections for 2024, Sabah districts were found to have a high prevalence of poverty, which calls for the government to take proactive measures by launching programs in the affected districts.

In a similar vein, the PSDM model provides the best findings in terms of poverty incidence and provides a thorough spatial representation via coefficient mapping. The mapping shows the extent to which districts have a positive or negative influence on poverty. The PSDM model's hypothesis procedure has a major impact on the proportion of districts that agree with the relationship's effects. Population density, unemployment rate, non-citizen proportion, crime rates, median household income, and Gini income were all factors that affected poverty before the pandemic, according to the PSDM model. The prevalence of poverty is significantly influenced by unemployment and unemployment rates both during and after pandemic events. In conclusion, factors such as population density, unemployment, non-citizen rates, and crime levels affect variations in the prevalence of poverty in Malaysian districts.

Acknowledgement
This research was not funded by any grant.

References

[1] Abd Manap, Nurliyana, Z. Zalmiyah, and H. Rohayanti. "Investigation of poverty indicators for designing case representation to determine urban poverty." *Int J Advance Soft Comput Appl* 9, no. 2 (2017): 90-106.
[2] Wang, Zhicheng, and Kun Tang. "Combating COVID-19: health equity matters." *Nature medicine* 26, no. 4 (2020): 458-458. https://doi.org/10.1038/s41591-020-0823-6
[3] Younsi, Moheddine, and Marwa Bechtini. "Economic growth, financial development, and income inequality in BRICS countries: does Kuznets' inverted U-shaped curve exist?." *Journal of the Knowledge Economy* 11, no. 2 (2020): 721-742. https://doi.org/10.1007/s13132-018-0569-2
[4] Ozdil, Tuncer, Ilham Yilmaz, and Gunce Demir. "Sociological and Economic Importance of the Concept of Poverty." *Journal of Eurasian Economies*, January. 2024. https://doi.org/10.36880/J03.1.0132
[5] Wagle, Udaya R. "Poverty measurement: evolving concepts and measurement frameworks." In *Research Handbook on Poverty and Inequality*, pp. 15-35. Edward Elgar Publishing, 2023. https://doi.org/10.4337/9781800882300.00008
[6] Norasikin, N. N. K., & Azhar, N. A. Z. A. M. The Factors Affecting Poverty in Malaysia: Macroeconomic Perspectives. *Advances in Social Sciences Research Journal 11,* no. 2.2 (2024): 457–468. https://doi.org/10.14738/assrj.112.2.16411

[7] Sulong, Amri, and Raja Nor Azuwah Raja Yusof. "Unravelling the Interplay of Public Debt, Inflation and Unemployment on Poverty in Malaysia: A Comprehensive Examination." *Journal of Social Transformation and Regional Development* 5, no. 2 (2023): 22-29. https://doi.org/10.30880/jstard.2023.05.02.003

[8] Majid, M. Rafee, Abdul Razak Jaffar, Noordini Che Man, Mehrdad Vaziri, and Mohamed Sulemana. "Mapping poverty hot spots in Peninsular Malaysia using spatial autocorrelation analysis." *Planning Malaysia* 4 (2016). https://doi.org/10.21837/pm.v14i4.144

[9] Senadjki, Abdelhak, Saidatulakmal Mohd, Zakaria Bahari, and Abdul Fatah Che Hamat. "Assets, risks and vulnerability to poverty traps: A study of Northern region of Malaysia." *The Journal of Asian Finance, Economics and Business* 4, no. 4 (2017): 5-15. https://doi.org/10.13106/jafeb.2017.vol4.no4.5

[10] Zakaria, Syerrina, Nuzlinda Abdul Rahman, and Nuredayu Zaini. "Spatial analysis on distribution patterns of poverty incidences in Peninsular Malaysia." In *AIP Conference proceedings*, vol. 2184, no. 1. AIP Publishing, 2019. https://doi.org/10.1063/1.5136415

[11] Rongen, Gerton, Zainab Ali Ahmad, Peter Lanjouw, and Kenneth Simler. *The Interplay of Regional and Ethnic Inequalities in Malaysian Poverty Dynamics*. World Bank, 2022. https://doi.org/10.1596/1813-9450-9898

[12] Samat, Narimah, S. M. R. Rashid, and Yasin Abdalla Elhadary. "Analyzing spatial distribution of poverty incidence in northern region of peninsular Malaysia." *Asian Social Science* 14, no. 12 (2018): 86-96. https://doi.org/10.5539/ass.v14n12p86

[13] Vaziri, Mehrdad. "Spatial Distribution of Poverty Incidence in Peninsular Malaysia." PhD diss., Universiti Teknologi Malaysia, 2012.

[14] Sunusi, Nurtiti, and Aan Subarkah. "Geographically Weighted Regression with Different Kernels: Application to Model Poverty." *Indonesian Journal of Applied Research (IJAR)* 4, no. 1 (2023): 27-41. https://doi.org/10.30997/ijar.v4i1.283

[15] Perchinunno, Paola, Antonella Massari, Samuela L'Abbate, and Monica Carbonara. "A Spatial Statistical Approach for the Analysis of Urban Poverty." In *International Conference on Computational Science and Its Applications*, pp. 389-404. Cham: Springer Nature Switzerland, 2023. https://doi.org/10.1007/978-3-031-37111-0_28

[16] Vista, Brandon Manalo, and Yuji Murayama. "Spatial determinants of poverty using GIS-based mapping." *Spatial analysis and modeling in geographical transformation process: GIS-based applications* (2011): 275-296. https://doi.org/10.1007/978-94-007-0671-2_16

[17] Bagde, Rakshit Madan. "भारतातील दारिद्याची मीमांसा." (2023). https://doi.org/10.31219/osf.io/g69bu

[18] Wang, Xiaolin. "The Evolution of Thinking about the Concept of Poverty." In *Multidimensional Poverty Measurement: Theory and Methodology*, pp. 1-22. Singapore: Springer Nature Singapore, 2022. https://doi.org/10.1007/978-981-19-1189-7

[19] Drago, Carlo. "Measuring Poverty: An Interval Based Composite Indicator Approach." *Available at SSRN 3306685* (2018). https://doi.org/10.2139/ssrn.3306685

[20] Hassan, N. N. N. M., Mohd Hakimi Harman, and Nur Farihah Mohammad Azhar. "The Determinants of Labour Productivity Growth in Malaysia." *International Journal of Academic Research in Business and Social Sciences* 13, no. 5 (2023): 2063-2077. https://doi.org/10.6007/IJARBSS/v13-i5/16893

[21] Rashid, M. F., N. S. Yusoff, and K. H. Kamarudin. "The impact of COVID-19 pandemic towards the resilience of small medium enterprises (SMEs) in Malaysia." In *IOP Conference Series: Earth and Environmental Science*, vol. 1082, no. 1, p. 012001. IOP Publishing, 2022. https://doi.org/10.1088/1755-1315/1082/1/012001

[22] Shah, Ain Umaira Md, Syafiqah Nur Azrie Safri, Rathedevi Thevadas, Nor Kamariah Noordin, Azmawani Abd Rahman, Zamberi Sekawi, Aini Ideris, and Mohamed Thariq Hameed Sultan. "COVID-19 outbreak in Malaysia: Actions taken by the Malaysian government." *International journal of infectious diseases* 97 (2020): 108-116. https://doi.org/10.1016/j.ijid.2020.05.093

[23] Tosepu, Ramadhan, Joko Gunawan, Devi Savitri Effendy, La Ode Ali Imran Ahmad, Hariati Lestari, Hartati Bahar, and Pitrah Asfian. "Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia." *Science of the total environment* 725 (2020): 138436. https://doi.org/10.1016/j.scitotenv.2020.138436

[24] Brunsdon, Chris, Stewart Fotheringham, and Martin Charlton. "Geographically weighted regression." *Journal of the Royal Statistical Society: Series D (The Statistician)* 47, no. 3 (1998): 431-443. https://doi.org/10.1111/1467-9884.00145

[25] Yang, W., A. S. Fotheringham, and P. Harris. "Model selection in GWR: the development of a flexible bandwidth GWR." In *The 11th international conference on geocomputation*, pp. 20-22. 2011.

[26] CheonKyu Park, Chongeun Cho, & Do-In Kim (2022). GWR Analysis on the Effect of Lightning Strikes and Damages in Distribution Transformer. *The transactions of The Korean Institute of Electrical Engineers*, 71(12), 1874-1880. https://doi.org/10.5370/KIEE.2022.71.12.1874

[27] Lu, Binbin, Chris Brunsdon, Martin Charlton, and Paul Harris. "Geographically weighted regression with parameter-specific distance metrics." *International Journal of Geographical Information Science* 31, no. 5 (2017): 982-998. https://doi.org/10.1080/13658816.2016.1263731
[28] Yang, W., A. S. Fotheringham, and P. Harris. "Model selection in GWR: the development of a flexible bandwidth GWR." In *The 11th international conference on geocomputation*, pp. 20-22. 2011.
[29] Lu, Binbin, Martin Charlton, Chris Brunsdon, and Paul Harris. "The Minkowski approach for choosing the distance metric in geographically weighted regression." *International Journal of Geographical Information Science* 30, no. 2 (2016): 351-368. https://doi.org/10.1080/13658816.2015.1087001
[30] Gollini, Isabella, Binbin Lu, Martin Charlton, Christopher Brunsdon, and Paul Harris. "GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models." *Journal of Statistical Software* 63 (2015): 1-50. https://doi.org/10.18637/jss.v063.i17
[31] Yang, Wenbai. "An extension of geographically weighted regression with flexible bandwidths." PhD diss., University of St Andrews, 2014.