Comparison of Spatial Weight Matrices in Spatial Autoregressive Model: Case Study of Intangible Cultural Heritage in Indonesia

Muhamad Sobari[1], Armalia Desiyanti[2], Devi Yanti[3], Putri Monika[4], Atje Setiawan Abdullah[5], Budi Nurani Ruchjana[6*]

[1,2,3]Postgraduate Program in Applied Statistics, Universitas Padjadjaran, Bandung, Indonesia
[1]Central Bureau of Statistics of Tasikmalaya, West Java, Indonesia
[2]Central Bureau of Statistics of Padang, West Sumatra, Indonesia
[3]Central Bureau of Statistics of Rote Ndao, East Nusa Tenggara, Indonesia
[4]Postgraduate Program in Mathematics, Universitas Padjadjaran, Bandung, Indonesia
[5]Department of Computer Science, Universitas Padjadjaran, Bandung, Indonesia
[6]Department of Mathematics, Universitas Padjadjaran, Bandung, Indonesia
budi.nurani@unpad.id[6]

ABSTRACT

Intangible Cultural Heritage (ICH) can effectively contribute to Sustainable Development Goals (SDGs) in all economic, social, and environmental dimensions, along with peace and security. Studying ICH in Indonesia cannot be separated from the spatial aspect of how an area's attributes are related to other areas located close to each other. Spatial regression modeling needs to be done by considering the selection of spatial weight matrix. Using the wrong spatial weight matrix will increase the standard error in parameter estimation. Therefore, this study aims to determine: the best spatial weight matrix to accommodate the spatial autocorrelation in analyzing the description of the spread of ICH in Indonesia; and the variables that are thought to influence the number of ICH determination in Indonesia. The spatial regression modeling used in this study is the Spatial Autoregressive (SAR) model and the spatial weight matrices compared in this study are queen contiguity and inverse distance. The best model is the SAR model used the queen contiguity spatial weight matrix because it has minimum values of AIC, BIC, RMSE and MAPE which are 310.397, 319.555, 18.857 and 57.169 respectively. Simultaneously, involved in performing arts, wearing traditional dress, knowing Indonesian folklore and the spatial lag contribute significantly to number of ICH determination in Indonesia. Partially, only knowing Indonesian folklore have a significant effect on number of ICH determination in Indonesia at significance level α=5%. Each additional 1% of population that knowing Indonesian folklore in an area increases number of ICH determination in that area by 0.6719 units.

—————————————— ◆ ——————————————

A. INTRODUCTION

The sustainable development agenda contained in the Sustainable Development Goals (SDGs) is made to answer the world's demands in addressing sustainable development problems, seen from three dimensions: economy, social, and environment, through 17 goals

or objectives. One of the SDGs, namely Goal 11, aims to create human settlements and inclusive, safe, resilient, and sustainable cities. The cultural aspect is one of the SDGs stated in the Target 11.4, which is a concern in the Sustainable Development Agenda, namely "strengthening efforts to protect and safeguard the world's cultural and natural heritage (Johnston, 2016)".

Protection of cultural heritage is considered very important because cultural heritage contains various ideas, creativity, and values that become the identity of a nation. The collective identity that binds the community will remain strong through the protection of cultural heritage; it then can form security communities that will ensure the community can carry out economic, social, psychological, and cultural activities (Graham & Howard, 2008; Johnston, 2016; Tutchener, 2013).

At first, world cultural heritage was only focused on buildings, monuments, or natural and tangible objects. However, in the 1990s, this concept began to experience a shift regarding the existence of intangible cultural heritage (Patru-Stupariu et al., 2019). UNESCO Convention for The Safeguarding of The Intangible Cultural Heritage of 2003 defines Intangible Cultural Heritage as practices, representations, expressions, knowledge, and skills (including instruments, objects, artefacts, cultural spaces) that communities and groups –in some cases, individuals recognize it as part of their cultural heritage. Individuals who declare themselves as cultural heritage is called living cultural heritage. This intangible cultural heritage is passed down from generation to generation, continuously re-created by communities and groups in response to the human environment.

Cultural heritage, mainly Intangible Cultural Heritage, can effectively contribute to sustainable development in all economic, social, and environmental dimensions, along with peace and security (United Nations, 2011). The development of cultural heritage, mainly Intangible Cultural Heritage, can help improve the life quality of the community and have a positive impact on tourism development.

The Indonesian Ministry of Education, Culture, Research, and Technology [Kemendikbudristek RI] stated that Indonesia has a rich Intangible Cultural Heritage despite there being still many that have not been established. Indonesia has designated 1,528 Intangible Cultural Heritage spread across all provinces in Indonesia from 2013 to 2021 (medcom.id, 2021). The Indonesian Intangible Cultural Heritage determination is one of the government's efforts to maintain the original values, which are the philosophy, source of knowledge, and identity of the Indonesian nation.

Research on intangible cultural heritage (ICH) taking into account spatial autocorrelation has conducted by Meng et al., 2022. In that study, the dependent variable was the number of ICH expressions, while the independent variables were population size, the number of roads and railways, farmland area, GDP, urban-rural construction land area, air temperature, altitude, forest area, and water area. The Geodetector was used to assess the impact of independent variables on the dependent variable.

The distribution of ICH in Shandong Province has a clear positive spatial autocorrelation (Meng et al., 2022). This indicates that the distribution of ICH determination in Indonesia cannot be separated from the spatial aspect of how an area's attributes are related to other areas located close to each other while spatial data is data related to spatial aspects and

presents geographic location information or an accurate picture of an area on the earth's surface (Cressie, 1993).

Unlike Meng's research, this study focuses on cultural product factors rather than social, economic, or natural factors using spatial analysis. Spatial analysis analyses the data that refer to positions, objects, and the relationship among them in the earth's space. The indications of attribute linkages between adjacent regions become the basis for spatial analysis, known as the spatial regression model, which is a model that can define the relationship of variables with spatial linkages (Anselin, 1990).

In addition, unlike Meng's research, which used the Geodetector method, this study used the spatial regression method. Spatial regression models need to be done by considering the relationship between locations, which is described by a weight matrix. A contiguity matrix describes the relationship between regions or the close relationship between regions (Suryowati et al., 2018). Another approach to determine the weight matrix is to use the distance band. (Anselin & Rey, 2014) explain that building a simple weight matrix is based on distance measurements when i and j are considered neighbors if j falls within a predetermined critical distance (distance band) from i.

In examining spatial weight, (Arlinghaus et al., 1996) states that the selection of the weight matrix must be carried out specifically and arranged appropriately. The specification of the weight matrix will represent information on the scope and intensity of spatial effects of a location unit in the geographic system (Trisilia, 2014). In several studies, the use of weight matrix in spatial regression models is still diverse, trial and error, and based on habit. Using the wrong weight will increase the standard error in parameter estimation, while using the correct weight will minimize it. Furthermore, this study aims to determine: (i) the best weight matrix to accommodate the spatial autocorrelation in analysis the description of the spread of Intangible Cultural Heritage in Indonesia, and (ii) the variables that are thought to influence the number of Intangible Cultural Heritage determination in Indonesia.

There is still less research concerning Intangible Cultural Heritage within spatial regression analysis; studies related to Intangible Cultural Heritage are critical to maintaining and protecting cultural heritage; this makes researchers interested in further study, especially in spatial regression model. Spatial Autoregressive Model (SAR) is a spatial regression model that is pretty effective for estimating data with a spatial dependency effect.

B. METHODS

The data used in this study is secondary data sourced from the Central Bureau of Statistics and Ministry of Education and Culture. The data obtained from the cultural heritage page.kemdikbud.go.id is data on the determination of intangible cultural heritage in Indonesia which has been determined until February 2022, which is 1,511 intangible cultural heritages. Meanwhile, the data obtained from BPS is socio-cultural statistical data for 2018. In more detail, it can be seen in Table 1. The units of observation and analysis in this study cover 34 provinces in Indonesia, as shown in Table 1.

Table 1. Definition of research variables

Variable	Variable Name	Description
Y	The Number of ICH determination	The number of Intangible Cultural Heritage (ICH) that has been determined
X_1	Involved in Performing Arts (X_1)	Percentage of population age 5 and over who have been involved in art performances/exhibitions during the last three months
X_2	Wearing Traditional Dress (X_2)	Percentage of households using regional/traditional dress in the last three months
X_3	Knowing Indonesian Folklore (X_3)	Percentage of population aged 5 years and over who know fairy tales/folk stories in Indonesia

Source: Central Bureau of Statistics and Ministry of Education and Culture

The method of analysis in this study is to use the Spatial Autoregressive (SAR) model. Spatial autoregressive (SAR) is a combination of a simple regression model with spatial lag using cross section data (Lesage, 1999). The general models of SAR are:

$$y = \rho \mathbf{W} y + X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n) \tag{1}$$

With y denotes response variable, X denotes predictor variable, ρ denotes spatial autocorrelation coefficient on the response variable, $\mathbf{W}$ denotes spatial weight matrix, β denotes intercept and regression coefficient and ε denotes error.

Spatial autoregressive model is a model whose response variables are spatially correlated. Parameter estimation using Maximum Likelihood (Lesage, 2004) method and the following equation is obtained:

$$\hat{\beta}_{ML} = (X^T X)^{-1} X^T y - \rho (X^T X)^{-1} X^T W_p y = \hat{\beta}_{OLS} - \rho \hat{\beta}_L \tag{2}$$

With $\hat{\beta}_L$ is an estimator of the regression parameters based on the spatial autocorrelation ρ and the weight matrix $\mathbf{W}$. However, this form cannot be solved directly because the value of ρ is unknown. So, the log-likelihood concentrated function is used as follows (Anselin, 2001).

$$\ln L_c(\rho) = C - \frac{n}{2} ln \left[\frac{1}{n} (e_0 - \rho e_L)^T (e_0 - \rho e_L) \right] + ln |I - \rho W_\rho| \tag{3}$$

With C is a constant. Equation (3) is a nonlinear function in one parameter and is maximized using a numerical technique with direct search.

Data processing was carried out using R software version 4.1.2 and thematic map creation using Q.GIS Desktop 3.16.15. Steps of data analysis were carried out as follows:

1. Exploring data based on thematic maps of all provinces in Indonesia so that data distribution patterns can be known;
2. Test the classical assumptions of multiple linear regression. The assumption test carried out is as follows:
    a. Normality test
       Normality test was carried out using the Shapiro Wilk test. This test is used to identify whether a random variable is normally distributed or not, checking the assumption of normality from errors or random errors and is suitable for small samples (Shapiro & Wilk, 1965). The hypothesis testing is:
       H_0: The population follows a normal distribution
       H_1: The population does not follow a normal distribution
       Test Statistics:

$$T_3 = \frac{1}{D}\left[\sum_{i=1}^{n} a_i(x_{n-i+1} - x_i)^2\right] \tag{4}$$

$$D = \sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{5}$$

With a_i denotes Shapiro Wilk's test coefficient, x_{n-i+1} denotes $(n-i+1)$-th data, x_i denotes i-th data, $\bar{x}$ denotes average of data. The significance of the T_3 value test is compared with the table value of Shapiro W, to see the position of the probability value (p). The decision criteria in making conclusion is to reject H_0 if p <α (0.05).

b. Homoscedasticity Test

This test is carried out using the Breusch-Pagan (BP) test, the BP test is used to test whether there is diversity between provinces. However, this test is affected when there is a spatial autocorrelation in the error term. BP test statistics were spatially adjusted for heteroscedasticity (Yamagata et al., 2011). The hypothesis is:

$H_0 : \sigma_i^2 = \alpha_i$

$H_1 : \sigma_i^2 \neq \alpha_i$

Test Statistics:

$$BP = \frac{1}{2}.f^{'}\, \mathbf{Z}(\mathbf{Z}^{'}\,\mathbf{Z})^{-1}\mathbf{Z}^{'}\, f + \frac{1}{T}\left(\frac{e^{'}\,We}{\sigma^2}\right)^2 \sim \chi_{(P+1)}^2 \tag{6}$$

$$f_i = (\sigma^{-1}.e_i)^2 - 1 \tag{7}$$

$$T = tr\left[W^{'}\, W + W^2\right] \tag{8}$$

With e denotes OLS residual vector e_i, σ^2 denotes variance based on OLS residual, $\mathbf{Z}$ denotes matrix $P+1$ of constants and variables that cause heteroscedasticity. The decision criteria in making conclusion is to reject H_0 if BP > $\chi_{(P+1)}^2$.

c. Non-Autocorrelation Test

This test was carried out with the Durbin Watson test. The non-autocorrelation assumption requires freedom between errors for each Y observation value, which means that the error in a certain observation is not affected by the error in other observations (Gujarati, 2003). Hypothesis testing:

$H_0 : \hat{\rho} = 0$ (There is no autocorrelation)

$H_1 : \widehat{\rho} \neq 0$ (There is autocorrelation)

Test Statistics:

$$d = \frac{\sum_{t=2}^{t=N}(e_t - e_{t-1})^2}{\sum_{t=2}^{t=N} e_t^{\,2}} \tag{9}$$

With d denotes Durbin Watson test statistic, e_t denotes error of the regression model at time t, e_{t-1} denotes error of regression model at time (t-1), $e_t^{\,2}$ denotes Square of error of the regression model at time t. The decision criteria are: If $(d < dL)$ or $(d > 4 - dL)$ then H_0 is rejected, meaning that there is an

autocorrelation in the residual; If $(dL < d < dU)$ or $(4 - dU < d < 4 - dL)$ then the Durbin Watson test does not produce definite or inconclusive conclusions. And if $(dU < d < 4 - dU)$ then it fails to reject H_0, meaning that there is no autocorrelation between the residuals.

d. Non-Multicollinearity Test

The detection of multicollinearity according to (Johnson & Wichern, 2007) can be seen through the Pearson correlation coefficient (r_{ij}). If the coefficient between variables > 0.95 then there is a correlation between these variables. In addition, you can also use the Variance Inflation Factors (VIF) which is stated as follows:

$$VIF_j \ = \ \frac{1}{1 - R_j^2} \qquad (10)$$

With R_j^2 is the coefficient of determination between x_j and other predictor. The value of VIF_j which is greater than 10 indicates the existence of collinearity between the predictor variables.

3. Create a spatial weight matrix based on queen contiguity and inverse distance. Spatial weight matrix (**W**) is basically a matrix that describes the relationship between regions and is obtained based on distance or neighbour information. Because the weight matrix shows the relationship between all locations, the dimensions of this matrix are N×N, where N is the number of locations or the number of units across objects (Dubin, 2009). The diagonal of the weighted matrix is zero. Meanwhile, w_{ij} is an element of **W**, namely the element in the i-th row and j-th column, which describes the relationship between the i-th location and the j-th location with $w_{ij} > 0$ if the i-th location is related to the j-th location. Several approaches can be taken to display the spatial relationship between locations, including the concept of contiguity and the concept of distance. The concept of contiguity or the concept of adjacency is based on geographically neighbouring relationships (Dubin, 2009). The types of spatial weight matrices include rook contiguity, bishop contiguity, and queen contiguity. While in the concept of distance, the elements of the spatial weight matrix are represented in the form of a distance function. In principle, the weight of the distance between a location and the surrounding locations is determined by the distance between the two areas. One method that is often used is the inverse distance method. The shorter the distance between the locations, the greater the weight given. The distance matrix that is generally used is the inverse distance as follows:

$$w_{ij} = \frac{1}{d_{ij}} \qquad (11)$$

With d_{ij} represents the Euclidean distance from location i to location j. The general form of the spatial weight matrix is:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}$$

4. Investigate whether there is autocorrelation between provinces by conducting Moran's index (Moran's I). The hypotheses of Moran's I:

H_0 : No spatial autocorrelation under given **W**

H_1 : There is a spatial autocorrelation under the given **W**

Test statistics of Moran's I (Anselin, 2001):

$$Z_I = \frac{I - E(I)}{\sqrt{Var(I)}} \tag{12}$$

$$Var(I) = \frac{N^2 . S_1 - N . S_2 + 3. S_0^2}{(N-1)(N+1)S_0^2} - [E(I)]^2 \tag{13}$$

$$S_2 = \sum_{i=1}^{N} \left(\sum_{j=1}^{N} w_{ij} + \sum_{j=1}^{N} w_{ji} \right)^2 \tag{14}$$

$$S_1 = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (w_{ij} + w_{ji})^2 \tag{15}$$

$$S_0 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \tag{16}$$

With I denotes Moran Index, Z_I denotes Moran index test statistic, $E(I)$ denotes expected value of Moran index and *Var(I)* denotes variance of Moran index. The decision criteria in making conclusion is to reject H_0 if $Z_I > Z_{(\alpha/2)}$. Moran's I is a measure of the correlation (relationship) between observations that are close together. This statistic compares the value of observations in one area with the value of observations in other areas. According to Moran's I can be measured using the equation:

$$I = \frac{N}{S_0} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{17}$$

With N denotes number of observations, $\bar{y}$ denotes average value of $\{y_i\}$ from N locations, y_i denotes value at location i, y_j denotes value at location j, w_{ij} denotes Spatial weight matrix elements. The value of I is the same as the correlation coefficient, which is between -1 to 1. A high value means that the correlation is high, while a value of 0 means that there is no autocorrelation. However, to say whether or not there is autocorrelation, it is necessary to compare the value of the I statistic with the expected value. The expected value of I is:

$$E(I) = -\frac{I}{N-1} \tag{18}$$

According to (Lee & Wong, 2001) the test statistic used is derived from the standard normal distribution in equation (12). An important aspect in determining spatial autocorrelation is determining the relationship between the closest region, the area around the observed area is thought to have an influence on the observed area. According to Tobler's first law, everything is related, but something closer is more

related (Fischer & Wang, 2011). According to (Lembo, 2006) Spatial autocorrelation is the correlation between variables and themselves based on space or can also be interpreted as a measure of the similarity of objects in a space (distance, time and area).

5. Investigate whether there is a spatial dependence on lag by using the Lagrange Multiplier (LM) test. The Lagrange Multiplier (LM) test is used to see whether there is a spatial dependency on lag or error so that it can determine the type of spatial analysis that is suitable for use. The test statistic used to determine spatial dependencies is the LM test (Anselin, 1988). The hypothesis of the LM test is as follows:

$H_0 : \rho = 0$ (no spatial dependence on lag or error)

$H_1 : \rho \neq 0$ (there is a spatial dependence on lag or error)

Test Statistics:

$$LM_{LAG} = \frac{[(e^T W_A y)/(e^T e/N)]^2}{\left[(W_A X \hat{\beta})^2 M (W_A X \hat{\beta})/(e^T e/N)\right] + \left[tr(W_A{}^T W_A + W_A{}^2)\right]} \sim \chi^2_{(1-\alpha);df=1} \quad (19)$$

The decision criterion in making conclusions is if the test statistic value is greater than the chi square value (reject H_0), then the model used is the Spatial Autoregressive (SAR) model. The LM test for error spatial dependencies is as follows:

$$LM_{ERR} = \frac{[(e^T W_A e)/(e^T e/N)]^2}{\left[tr(W_A{}^T W_A + W_A{}^2)\right]} \sim \chi^2_{(1-\alpha);df=1} \quad (20)$$

The decision criterion in making conclusions is if the test statistic value is greater than the chi square value (reject H_0), then the model used is the Spatial Error Model (SEM). If both LM_{LAG} and LM_{ERR} are significant, one of the best models can be chosen by comparing the Akaike Information Criterion (AIC) values. The model with the smallest AIC value is the best model (Zebua & Jaya, 2022).

6. Parameter estimation of two SAR models using the Maximum Likelihood method, one model using the queen contiguity weight matrix and the other model using the inverse distance;

7. Selection of the best model by comparing the values of AIC, BIC, MAPE, and RMSE generated from each model with the spatial weight used (queen contiguity and inverse distance). The criteria for selecting the model used in this study are as follows:

   a. Akaike Info Criterion (AIC)

   To choose the best regression model found by Akaike and Schwarz (Grasa, 1989). The method is based on the Maximum Likelihood Estimation (MLE) method. Denoted by:

   $$AIC = -2L_m + 2m \quad (21)$$

   With L_m denotes Maximum log-likelihood and m denotes number of parameters in the model. Models with small value of AIC are the best (Wei, 2006).

   b. Bayesian Information Criterion (BIC)

   According to (Stoica et al., 2004), the order selection rules (AIC/BIC) have the same form:

   $$-2 \ln p_n(y, \hat{\theta}^n) + \eta(n, N)n \quad (22)$$

   but with different penalty coefficients $\eta(n, N)$. For BIC $\eta(n, N) = \ln N$ while for AIC $\eta(n, N) = 2$. Then the BIC formula becomes:

$$BIC = -2\ln\left(p_n\left(y, \hat{\theta}^n\right)\right) + n \ln N \tag{23}$$

With n denotes number of parameters in the model, $p_n\left(y, \hat{\theta}^n\right)$ denotes the maximum likelihood function of the model, N denotes number of observations. The best model is determined based on the smallest BIC value.

c. Root Mean Square Error (RMSE)

According to (Webster & Oliver, 2007), one of the most frequently used criteria or measures to compare between two or more models in spatial analysis is the Root Mean Square Error (RMSE). The formula for calculating RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|} \tag{24}$$

With n denotes number of observations, $\hat{y}_i$ denotes the value of the prediction, y_i denotes true value of observation. The smaller the RMSE value of a model, the more accurate the model is.

d. Mean Absolute Percentage Error (MAPE)

MAPE is a measure that reflects the mean absolute percentage of prediction error over the true value (Sharma et al., 2012). The smaller value of MAPE value, the more accurate the model is. The formula for calculating MAPE is as follows:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{25}$$

8. Interpretation of the model based on the best model that has been obtained;
9. Diagnostic testing of SAR model.

C. RESULT AND DISCUSSION

1. Descriptive Analysis

Since Indonesia became a State Party to the 2003 Convention on the protection of Intangible Cultural Heritage, according to articles 11 and 12 of the 2003 Convention, Indonesia is required to regulate the identification and inventory of Intangible Cultural Heritage in its territory. The identification and inventory of the Intangible Cultural Heritage are carried out, among others, through the recording and determination of the Intangible Cultural Heritage. The determination of the Intangible Cultural Heritage is proposed by the local government at the national level. 1,528 Intangible Cultural Heritage have been designated and spread across all provinces in Indonesia from 2013 to 2021, as shown in Figure 1.
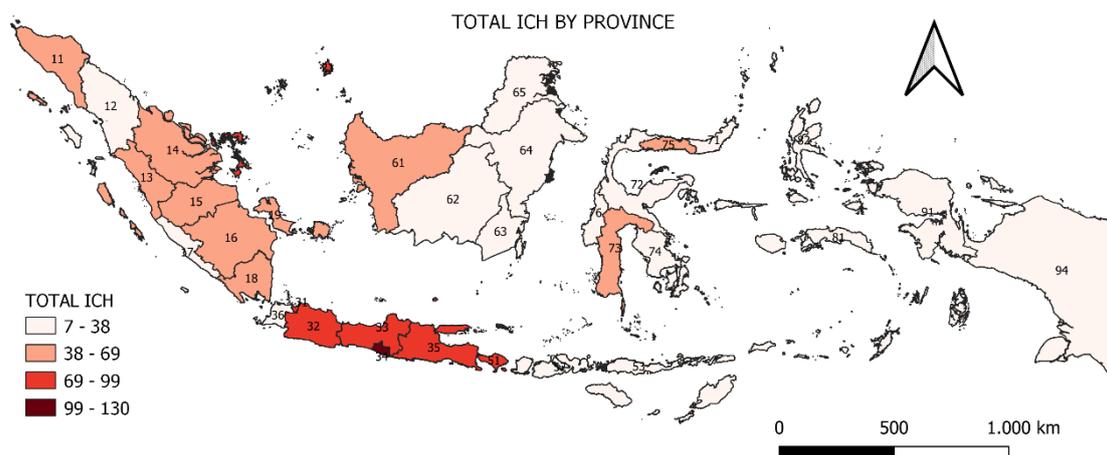
Figure 1. Map of the Distribution of Intangible Cultural Heritage by Provinces in Indonesia

Based on Figure 1, in general the provinces located on the island of Java still dominate the number of Intangible Cultural Heritage determinations in Indonesia, which is 31.41% of the total existing ICH. The province with the highest number of Intangible Cultural Heritage (ICH) determinations in Indonesia is the Special Region of Yogyakarta (130 ICH), followed by Central Java (96 ICH) and West Java (85 ICH). Meanwhile, Central Kalimantan, West Papua, and West Nusa Tenggara occupy the lowest position with the least number of ICH determination. This shows a tendency for spatial autocorrelation, that is, a province with a high number of ICH determinations is close to one another or close to other provinces with a fairly high number of ICH determinations and conversely a province with low ICH determinations is close to one another or close to provinces with fairly low ICH determinations, as shown in Table 2.

Table 2. ID of Province in Indonesia

ID	Province	ID	Province	ID	Province
11	Aceh	32	West Java	64	East Kalimantan
12	North Sumatra	33	Central Java	65	North Kalimantan
13	West Sumatra	34	Yogyakarta	71	North Sulawesi
14	Riau	35	East Java	72	Central Sulawesi
15	Jambi	36	Banten	73	South Sulawesi
16	South Sumatra	51	Bali	74	Southeast Sulawesi
17	Bengkulu	52	West Nusa Tenggara	75	Gorontalo
18	Lampung	53	East Nusa Tenggara	76	West Sulawesi
19	Bangka Belitung Islands	61	West Kalimantan	81	Maluku
21	Riau Islands	62	Central Kalimantan	82	North Maluku
31	Jakarta	63	South Kalimantan	91	West Papua
				94	Papua

One of the efforts made in advancing Indonesian culture according to the Law of the Republic of Indonesia Number 5 of 2017 is through the protection, development, utilization, and fostering of culture. The main targets for promoting cultures are the Cultural

Advancement Object (CAO), which includes performing arts, verbal traditions (fairy tales, folklore), folk games, and traditional knowledge, sports, ceremonies, and products (regional/traditional dress and crafts), as shown in Figure 2.

Figure 2. Distribution Map of the Percentage of Population Involved in Art Performances/Exhibitions in the Last Three Months

One form of appreciating art performances/exhibitions can be seen based on the population percentage (aged five years and over) involved in art performances/exhibitions during the last three months. The profession of an artist is still tiny in demand by the public. This can be seen in Figure 2, which shows the small percentage of the population involved in performing arts/exhibitions in the last three months. Bali and the Special Region of Yogyakarta occupy the top positions for involvement in art performances/exhibitions over the past three months. Here, the involvement concept is deliberately taking the time to perform arts performances/exhibitions or provide direct entertainment to the audience, either as actors or supporters (BPS, 2019).

The next Cultural Advancement Object (CAO) is traditional products, especially regional/traditional dress. Indonesia has a variety of regional/traditional dress. Based on Figure 3, it can be seen that the percentage of households using regional/traditional dress for the last three months varies significantly in each province. The highest percentage of households using regional/traditional dress is in the provinces of Bali (75.09%) and East Nusa Tenggara (60.71%). The high percentage of the use of regional/traditional dress in the province of Bali is inseparable from the religious ritual activities carried out daily, especially for Hindus in Bali, as shown in Figure 3.

Figure 3. Distribution Map of the Percentage of Households Wearing Regional/Traditional Dress for the Last Three Months

In addition to performing arts and traditional products, verbal tradition is one of the cultural advancement objects in Indonesia. Verbal traditions are passed down from generation to generation by the community, including in the form of fairy tales/folklore. The percentage of Indonesia's population (aged five years and over) who knows fairy tales/folklore is relatively high. Almost all provinces are above 50% in terms of knowing fairy tales/folklore in Indonesia, except for the provinces of East Nusa Tenggara (49.89%) and Papua (38.86%). The highest percentages are in the Riau Islands Province (88.63%) and West Sumatra (88.23%). Several fairy tales/folklore in Indonesia have been included in school lessons. In addition to textbooks, people can access Indonesian fairy tales/folklore more freely through other media such as television, radio, or online/streaming, as shown in Figure 4.
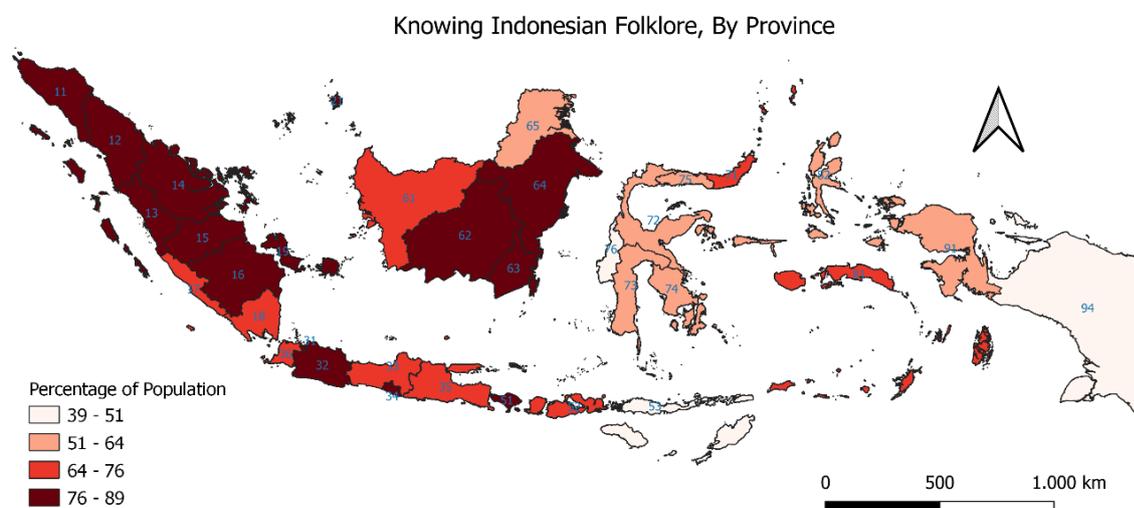
Figure 4. Distribution Map of the Percentage of the Population Knowing Fairy Tales/Folklore in Indonesia

2. Spatial Analysis

Before conducting SAR modeling, it is necessary to test the classical assumptions of multiple linear regression and Moran's I tests. Table 3 shows that almost all of the classic assumptions in multiple linier regression model have been fulfilled, and 33.03% of number of ICH determination in Indonesia can be explained by the three predictor variables. Because d < dL then H_0 is rejected, it means that there is an autocorrelation in the residuals for the multiple regression model, as shown in Table 3.

Table 3. The statistic test result of classic assumptions and R-squared

Statistic test	Result
Normality (Shapiro-Wilk)	P-value = 0.2254
Non-autocorrelation (Durbin-Watson)	1.10142 (dL=1.2707 dU=1.6519)*
Non- multicollinearity (VIF)	1.5163(X_1), 1.5075(X_2), 1.0214(X_3)
Homoscedasticity (Breusch-pagan)	P-value = 0.2161
Adjusted R-squared	0.3303

*) not fulfilled

In this study, two spatial weight matrices were used. The first is a spatial weight matrix with a neighboring basis using a queen contiguity. And the second is a spatial weight matrix based on distance using an inverse distance matrix. The results of the spatial autocorrelation test by conducting the Moran's I Test using two types of spatial weights can briefly be seen in Table 4.

Table 4. Autocorrelation Test with Moran's I Test

Variable	Spatial Weight Matrices	Moran's I Statistic	P-Value (α=5%)
The Number of ICH determination (Y)	Queen Contiguity	0.47461	0.00257
	Inverse Distance	0.16361	8.949×10^{-8}
Involved in Performing Arts (X_1)	Queen Contiguity	0.32599	0.02375
	Inverse Distance	0.04125	0.03628
Wearing Traditional Dress (X_2)	Queen Contiguity	0.32343	0.02694
	Inverse Distance	0.01693	0.17300*
Knowing Indonesian Folklore (X_3)	Queen Contiguity	0.61932	0.00015
	Inverse Distance	0.22377	7.563×10^{-12}

*) not significant

According to the results in Table 4, it can be seen that there is a spatial autocorrelation in the variables Y, X_1, X_2, X_3 and almost all of the results are significant using two types of spatial weight. This can be seen from the p-value which is almost entirely less than α = 5%, there is one that is not significant, which is wearing traditional dress (X2) variable with inverse distance spatial weight. The value of Moran's I statistic is entirely in the range of 0 to 1, which means that the closer an area is, the more similar the variable values will be.

Likelihood Ratio (LR) test is used to determine which model is better, whether the spatial regression model or the multiple regression model. And Lagrange Multiplier (LM) test is used to determine the spatial dependence more specifically whether the dependency on a response

variable (lag), dependency on other variables that are not studied (error), or both (lag and error).

According to the results in Table 5 it can be seen that for both spatial weight matrices LR Test are significant because the p-value is smaller than α=5%, it means there is a very significant difference between the spatial regression model and the multiple regression model. And it can be seen for both spatial weight matrices the spatial dependence in lag are significant because the p-value is smaller than α=5%, and only for queen contiguity spatial weight matrices the spatial dependence in error is significant. So, the SAR model will be applied in this study because for both spatial weight matrices the spatial dependence are significant. The SAR model is a spatial regression model that involves spatial lag in the response variable, as shown in Table 5.

Tables 5. LM test and LR test result

LM-Test	Spatial Weight Matrices	Value	P-Value (α=5%)
LR	Queen Contiguity	5.9471	0.01474
	Inverse Distance	4.0254	0.04482
LM_{LAG}	Queen Contiguity	5.0422	0.02474
	Inverse Distance	4.8019	0.02843
LM_{ERR}	Queen Contiguity	6.5741	0.01035
	Inverse Distance	2.5902	0.10750*

*) not significant

Estimation of SAR model parameters using the Maximum Likelihood method. The estimation results using the two spatial weight matrices are shown in equation (26) and equation (27) respectively for the model with the queen contiguity weight matrix and for the model with the inverse distance weight matrix.

$$\hat{y}_{QC-i} = 0.35542 \sum_{j=1,i\neq j}^{n} w_{QC-ij}y_j - 32.35692 + 4.61519X_{1i} + 0.41397X_{2i} + 0.62838X_{3i} \quad (26)$$

$$\hat{y}_{ID-i} = 0.69575 \sum_{j=1,i\neq j}^{n} w_{ID-ij}y_j - 52.25419 + 4.99668X_{1i} + 0.43694X_{2i} + 0.65052X_{3i} \quad (27)$$

In this study to determine the best model used several statistics such as AIC, BIC, RMSE and MAPE. A good model is a model that has the smallest value for the four statistics. The results of the model selection criteria can be seen in Table 6.

Table 6. Best SAR model selection criteria

Criteria	Spatial Weight Matrices	
	Queen Contiguity	Inverse Distance
AIC	310.3969	312.3186
BIC	319.5550	321.476 7
RMSE	18.85694	19.62264
MAPE	57.16905	57.69117

According to the results in Table 6, information can be obtained that from the two SAR models with two spatial weights used in this study, the model chosen as the best model is the SAR model with queen contiguity spatial weight. This is because the model has minimum AIC, BIC, RMSE and MAPE values. A model with a lower AIC and BIC fits the data better than a model with a higher AIC and BIC, and a model with a lower RMSE and MAPE has lower prediction errors over the true value than a model with a higher RMSE and MAPE. The parameter estimation result of the SAR model with queen contiguity weight can be seen in Table 7.

Table 7. SAR model parameter estimation results

	Estimate	Std Error	Z-Value	Pr (>|z|)
(Intercept)	-32.35692	19.13081	-1.6914	0.09077*
Involved in Performing Arts (X_1)	4.61519	3.82346	1.2071	0.22740
Wearing Traditional Dress (X_2)	0.41397	0.24278	1.7051	0.08817*
Knowing Indonesian Folklore (X_3)	0.62838	0.27377	2.2953	0.02172**

Rho: 0.35542
Asymptotic standard error: 0.12564, z-value: 2.8289, p-value: 0.0046715
Wald statistic: 8.0024, p-value: 0.0046715

Significant codes: **⁾0.05, *⁾ 0.1

According to the results in Table 7, the value of spatial lag variable (Rho) has a positive and significant coefficient in influencing the number of ICH determination in Indonesia. This means that the greater number of ICH is influenced by a large number of ICH in the surrounding area. And the P-value of the Wald test is 0.0046715, indicating that the relationship between number of ICH determination and all predictor variables is spatially dependent. Simultaneously, all predictor variables and the spatial lag all contribute significantly to number of ICH determination in Indonesia. Partially, only knowing Indonesian folklore (X_3) have a significant effect on response variable at significance level $\alpha=5\%$. While involved in performing arts (X_1) and wearing traditional dress (X_2) do not have a significant effect.

In the SAR model, the impact of covariates can be categorized into three types, which is direct impact, indirect impact, and total impact (Zebua & Jaya, 2022). Direct impacts are impacts that occur locally in an area as a result of changes in predictor variables. Indirect impacts are impacts that occur when changes in predictor variables occur in the surrounding area. And total impact is the change that occurs in an area as a result of changes in the area and its surroundings. The size of the direct and indirect impacts of the SAR model in this study can be seen in Table 8.

Table 8. Direct and indirect impact measure of SAR model

Variable	Direct Impact	P-Value Direct Impact	Indirect Impact	P-Value Indirect Impact	Total Impact	P-Value Total Impact
Involved in Performing Arts (X_1)	4.9347	0.1344	2.2253	0.2467	7.1599	0.1490
Wearing Traditional Dress (X_2)	0.4426	0.1084	0.1996	0.1957	0.6422	0.1104
Knowing Indonesian Folklore (X_3)	0.6719	0.0084	0.3030	0.0944	0.9749	0.0106

According to the results in Table 8, one predictor variables have a significant direct impact on number of ICH determination, but no variable have a significant indirect impact. In Indonesia, knowing Indonesian folklore has a direct effect on increasing number of ICH determination. Each additional 1% of population that knowing Indonesian folklore in an area increases number of ICH determination in that area by 0.6719 units, as shown in Table 9.

Table 9. Diagnostic testing of SAR model

Statistic Test	P-Value	Conclusion
Normality (Shapiro-Wilk)	0.9086	Residuals are normally distributed
LM test for residual autocorrelation	0.4864	There is no autocorrelation among residuals
Homogeneity (Breusch-pagan)	0.8693	The assumption of homogeneity is fulfilled

To find out whether the SAR model obtained is good, it is necessary to carry out diagnostic tests, including assumptions of normality, non-autocorrelation, and homogeneity. Table 9 shows the results of diagnostic test, can be concluded the SAR model fulfilled normality, non-autocorrelation and homogeneity assumptions.

D. CONCLUSION AND SUGGESTIONS

In general, provinces located on the Java Island dominate the number of determinations of Intangible Cultural Heritage (ICH) in Indonesia, which is 31.41 percent of the total existing ICH. The number of ICH determinations, the percentage of the population involved in performing arts and the percentage of the population who know Indonesian folklore have spatial dependencies based on the Moran Index test using either the queen contiguity weight matrix and the inverse distance. While the percentage of households wearing traditional dress has a spatial dependence based on the Moran Index test using only the queen contiguity weight matrix.

The best spatial econometrics model was chosen based on the LM test is the Spatial Autoregressive (SAR) model and based on the results of MAPE, AIC, BIC, and RMSE, the model with the queen contiguity weight matrix is better than the model with the inverse distance weight matrix. Percentage of population knowing Indonesian folklore has a direct effect on increasing number of determinations of ICH. Each additional 1% of population that knowing Indonesian folklore in an area increases number of determinations of ICH in that area by 0.6719 units.

Suggestions for the government, especially the Ministry of Education, Culture, Research and Technology in order to preserve Indonesian culture so that every policy taken can support the determination of ICH in Indonesia. For further research, other variables can be added that have a strong correlation with the determination of ICH in Indonesia. Besides that for further research on ICH can use the Spatial Error Model (SEM) method.

ACKNOWLEDGEMENT

The authors would like to thank the Rector of Universitas Padjadjaran, who offered financial help to disseminate studies reports under the Academic Leadership Grant (ALG) with contract number 2023/UN6.3.1/PT.00/2022 . We are also thankful for the discussion on social media analytics through the RISE_SMA project funded through the European Union year 2019-2024 and also studies center of modeling and computation Universitas Padjadjaran.

REFERENCES

Anselin, L. (1988). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*, *20*(1), 1–17. https://doi.org/10.1111/j.1538-4632.1988.tb00159.x

Anselin, L. (1990). Spatial Econometrics: Methods and Models. *Journal of the American Statistical Association*, *85*(411). https://doi.org/10.2307/2290042

Anselin, L. (2001a). *A Companion to Theoretical Econometrics: Spatial Econometrics* (Badi H. Baltagi, Ed.). Blackwell Publishing Ltd.

Anselin, L. (2001b). Spatial Econometrics. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics* (pp. 310–330). Blackwell Publishing Ltd.

Anselin, L., & Rey, S. J. (2014). *Modern spatial econometrics in practice: A guide to GeoDa, GeoDaSpace and PySAL*. GeoDa Press LLC.

Arlinghaus, S. L., Griffith, D. A., Arlinghaus, W. C., Drake, W. D., & Nystuen, J. D. (1996). *Practical handbook of spatial statistics*. CRC Press.

BPS. (2019). Statistik Sosial Budaya 2018. In *Statistik Sosial Budaya 2018*. Badan Pusat Statistik.

Cressie, N. A. C. (1993). *Statistics For Spatial Data* (W. A. Snewhart, S. S. Wllks, V. Barnett, R. A. Bradley, N. I. Fisher, J. S. Hunter, J. B. Kadane, D. G. Kendall, A. F. M. Smith, S. M. Stigler, G. S. Watson, & J. L. Teugels, Eds.). John Wiley & Sons, Inc.

Dubin, R. A. (2009). *The SAGE Handbook of Spatial Analysis* (F. A.S & P. A. Rogerson, Eds.; First). Sage Publications Inc.

Fischer, M. M., & Wang, J. (2011). *Spatial Data Analysis*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21720-3

Graham, B. J., & Howard, P. (2008). *The Ashgate research companion to heritage and identity*. Ashgate Publishing, Ltd..

Grasa, A. A. (1989). *Econometric Model Selection: A New Approach* (Vol. 16). Springer Netherlands. https://doi.org/10.1007/978-94-017-1358-0

Gujarati, D. N. (2003). *Basic econometrics*. McGraw Hill.

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (Sixth). Pearson Education, Inc.

Johnston, R. B. (2016). Arsenic and the 2030 Agenda for sustainable development. *Arsenic Research and Global Sustainability - Proceedings of the 6th International Congress on Arsenic in the Environment, AS 2016*, 12–14. https://doi.org/10.1201/b20466-7

Lee, J., & Wong, D. W. S. (2001). *Statistical Analysis with Arcview Gis*. John Wiley & Sons, Inc.

Lembo, A. J. (2006). *Spatial Autocorrelation*.

Lesage, J. P. (1999). *The Theory and Practice of Spatial Econometrics*. http://www.econ.utoledo.edu

Lesage, J. P. (2004). *Lecture 1: Maximum likelihood estimation of spatial regression models*.

medcom.id. (2021). *2021, Jumlah Warisan Budaya Tak Benda Indonesia Bertambah 289*. https://www.medcom.id/pendidikan/news-pendidikan/nbwXAB6k-2021-jumlah-warisan-budaya-tak-benda-indonesia-bertambah-289

Meng, L., Zhu, C., Pu, J., Wen, B., & Si, W. (2022). Study on the Influence Mechanism of Intangible Cultural Heritage Distribution from Man–Land Relationship Perspective: A Case Study in Shandong Province. *Land*, *11*(8). https://doi.org/10.3390/land11081225

Patru-Stupariu, I., Pascu, M., & Bürgi, M. (2019). Exploring tangible and intangible heritage and its resilience as a basis to understand the cultural landscapes of saxon communities in southern Transylvania (Romania). *Sustainability (Switzerland)*, *11*(11). https://doi.org/10.3390/su11113102

Shapiro, S. S., & Wilk, ; M B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3–4), 591–611. http://links.jstor.org/sici?sici=0006-3444%28196512%2952%3A3%2F4%3C591%3AAAOVTF%3E2.0.CO%3B2-B

Sharma, Suresh, & Sharma, V. (2012). Comparative Analysis of Machine Learning Techniques in Sale Forecasting. In *International Journal of Computer Applications* (Vol. 53, Issue 6, pp. 975–8887).

Stoica, P., Selén, Y., & Li, J. (2004). Multi-model approach to model selection. *Digital Signal Processing: A Review Journal*, *14*(5), 399–412. https://doi.org/10.1016/j.dsp.2004.03.002

Suryowati, K., Bekti, R. D., & Faradila, A. (2018). A Comparison of Weights Matrices on Computation of Dengue Spatial Autocorrelation. *IOP Conference Series: Materials Science and Engineering*, *335*(1). https://doi.org/10.1088/1757-899X/335/1/012052

Trisilia, M. (2014). *Perbandingan Matriks Pembobot Spasial Dalam Model Spatial Autoregressive (SAR) (Studi Kasus Daya Dukung Lahan Pertanian Tanaman Pangan di Kota Batu)* [Pascasarjana]. Universitas Brawijaya.

Tutchener, D. (2013). Collective Identity and Heritage: Ebb and Flow at National and Global Levels. In *Journal of Globalization Studies* (Vol. 4, Issue 2). https://doi.org/10.1.1.1057.3583

United Nations. (2011). *Intangible Cultural Heritage and Sustainable Development.*

Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists Second Edition* (Second). John Wiley & Sons Inc.

Wei, W. W. S. (2006). *Time series analysis: univariate and multivariate methods*. Pearson Addison Wesley.

Yamagata, T., Halunga, A., & Orme, C. D. (2011). *A Heteroskedasticity Robust Breusch-Pagan Test for Contemporaneous Correlation in Dynamic Panel Data Models A Heteroskedasticity Robust Breusch-Pagan Test for Contemporaneous Correlation in Dynamic Panel Data Models*.

Zebua, H. I., & Jaya, I. G. N. M. (2022). Spatial Autoregressive Model of Tuberculosis Cases in Central Java Province 2019. *Cauchy*, *7*(2), 240–248. https://doi.org/10.18860/ca.v7i2.13451